Proceedings of the

1992 IEEE Frequency Control Symposium

Sponsored by the



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. 92CH3083-3

Library of Congress No. 87-654207

27-29 May 1992

The Hershey Lodge and Convention Center
Hershey, Pennsylvania, 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 © 1992 by the Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 92CH3083-3
ISBN 0-7803-0476-4 (Softbound)
ISBN 0-7803-0477-2 (Casebound)
ISBN 0-7803-0478-0 (Microfiche)
Library of Congress Number: 87-645207

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

SYMPOSIUM CHAIRMEN

General Chairman.....Raymond L. Filler

Technical Program Chairman.....Jack Kusters

Local Arrangements Chairman.....Charles J. Jensik

Publicity/Editorial Chairman.....John R. Vig

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Materials & Resonators:	Arthur Ballato, Army Research Laboratory
Oscillators & Circuits:	Fred Walls, National Institute of Standards & Technology
Atomic Standards & Timing:	Helmut Hellwig, Air Force Office of Scientific Research

TECHNICAL PROGRAM COMMITTEE

David Allan, NIST
Ivan Avramov, Inst. Solid State Physics, Bulgaria
Arthur Ballato, Army Research Laboratory
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, NRC, Canada
Jan Brown, Fischer Controls Int'l., Inc.
Leonard Cutler, Hewlett-Packard Company
Michael Driscoll, Westinghouse Systems Center
Errol EerNisse, Quartztronics, Inc.
Raymond Filler, Army Research Laboratory
Marvin Frerking, Rockwell International
Satoru Fujishima, Murata Mfg. Co., Japan
Jean-Jacques Gagnepain, CNRS-LPMO, France
Edward Garber, TRW
Michael Garvey, Frequency & Time Systems, Inc.
Helmut Hellwig, Air Force Office of Scientific Research
James Pon-Wei Hou, Quartz Freq. Tech. Ltd., Taiwan
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
Jack Kusters, Hewlett-Packard Company
Theodore Lukaszek, Army Research Laboratory
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Thrygve Meeker, Consultant
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
Dan Stevens, AT&T Bell Laboratories
Joseph Suter, Johns Hopkins University
Richard Sydnor, Jet Propulsion Laboratory
John Vetelino, University of Maine
John Vig, Army Research Laboratory
Fred Walls, NIST
S. Clark Wardrip, Bendix Field Engineering Corp.
Rolf Weglein, Consultant
Werner Weidemann, Ball Efratom Division
Joseph White, Naval Research Laboratory
Roger Williamson, STC Technology, United Kingdom
Gernot Winkler, U.S. Naval Observatory
Nicholas Yannoni, U.S. Air Force/RADC
Lidia Zhourkina, Soviet Committee IEC, Russia

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
Jack Kusters, Hewlett-Packard Company

GROWTH & SWEEPING OF CRYSTALLINE MATERIAL
Gary Johnson, Sawyer Research Products

CONFINED ATOMS AND ION STANDARDS
Leonard S. Cutler, Hewlett-Packard Labs

FREQUENCY SYNTHESIZERS
Victor S. Reinhardt, Hughes Aircraft Company

GUIDELINES FOR THE MEASUREMENT OF ENVIRONMENTAL SENSITIVITIES OF PRECISION OSCILLATORS
Helmut Hellwig, Air Force Office of Scientific Research

CRYSTAL OSCILLATORS
Raymond L. Filler, Army Research Laboratory

VIBRATIONS OF BOUNDED RESONATORS
Daniel Stevens, AT&T Bell Labs

STRATEGIES FOR THE NEXT GENERATION OF HIGH STABILITY FREQUENCY STANDARDS
Fred L. Walls, National Institute of Standards & Technology (NIST)

ACCELERATION EFFECTS & MODE SHAPES IN QUARTZ RESONATORS
Errol P. EerNisse, Quartztronics, Inc.

PIEZOELECTRIC MATERIALS OTHER THAN QUARTZ
Arthur Ballato, Army Research Laboratory

POSTER SESSION
Cryogenic Oscillators and Resonators: John R. Vig, Army Research Laboratory
Devices & Packaging: Jan Brown, Fisher Control International, Inc.
SAW Devices and Techniques: Donald C. Malocha, University of Central Florida
Oscillators: Raymond L. Filler, Army Research Laboratory
Measurement Techniques, Results & Applications: Leonard S. Cutler, Hewlett-Packard Laboratories
Strategies for the Next Generation of High Stabiltiy Frequency Standards: Fred L. Walls, NIST

NOISE AND OSCILLATORS
Thomas E. Parker, Raytheon Research Division

MEASUREMENTS AND DISTRIBUTION
Richard Sydnor, Jet Propulsion Laboratory

RESONATORS
Robert Kinsman, Motorola, Inc.

RUBIDIUM AND CESIUM STANDARDS
John Prestage, Jet Propulsion Laboratory

TABLE OF CONTENTS

PAGE

AWARD PRESENTATIONS 1

ATOMIC AND OTHER HIGH STABILITY FREQUENCY STANDARDS

Local Oscillator Requirements and Strategies for the Next Generation of High-Stability Frequency Standards 2
F.L. Walls, National Institute of Standards and Technology

Local Oscillator Requirements for Timekeeping in the $10^{-14}\tau^{-1/2}$ Era 6
R.J. Douglas & J.-S. Boulanger, National Research Council of Canada, Canada

Optical Frequency Synthesis by Interval Division *(INVITED)* 27
H.R. Telle, Physikalisch-Technische Bundesanstalt, Germany

Optical Parametric Division *(INVITED)* 32
N.C. Wong & D. Lee, Massachusetts Institute of Technology

Coherent Frequency Synthesis in the Infrared *(INVITED)* 39
G. Kramer, B. Lipphardt & C.O. Weiss, Physikalisch-Technische Bundesanstalt, Germany

Short and Long Term Stability of Optical Oscillators *(INVITED)* 44
M. Zhu & J.L. Hall, University of Colorado and National Institute of Standards and Technology

Laser-Cooled Neutral-Atom Frequency Standards *(INVITED)* 56
W.D. Phillips & S.L. Rolston, National Institute of Standards and Technology

Hg^+ Trapped Ion Standard Performance with the Superconducting Cavity Maser Oscillator as L.O. 58
J.D. Prestage, R.L. Tjoelker, R.T. Wang, G.J. Dick & L. Maleki, Jet Propulsion Laboratory

Progress on Prototype Synthesizer Electronics for $^{199}Hg^+$ at 40.5 GHz 64
C.W. Nelson, F.L. Walls, F.G. Ascarrunz & P.A. Pond, National Institute of Standards and Technology

The Development of a Ytterbium Ion Frequency Standard 70
D.J. Seidel, A. Williams, R.W. Berends, & L. Maleki, Jet Propulsion Laboratory

Frequency Measurement of a Single Barium Ion Transition with an Ammonia Laser 76
K.J.Siemsen, A.A. Madej, J.D. Sankey, B.G. Whitford & R.F. Clark, National Research Council of Canada, Canada & G. Magerl, Technische Universität Wien, Austria

Doppler Effects in a Multi-Wavelength Rabi Cavity for the Excitation of Atomic Beams in Microwave Frequency Standards 81
E. Bava, Politecnico di Milano, P. Tavella, IEN "Galileo Ferraris" & A. DeMarchi, Politecnico di Torino, Italy

Frequency, Phase and Amplitude Changes of the Hydrogen Maser Oscillation 86
C. Audoin, Laboratoire de l'Horloge Atomique, France & W.A. Diener, Jet Propulsion Laboratory

Experimental Frequency and Phase Stability of the Hydrogen Maser Standard Output as Affected by Cavity Auto-Tuning 92
H.B. Owings, P.A. Koppang, C.C. MacMillan & H.E. Peters, Sigma Tau Standards Corporation

Diode-Laser-Pumped, Gas-Cell Atomic Clocks 104
R.E. Drullinger & C. Szekely, National Institute of Standards and Technology, & J.C. Camparo, The Aerospace Corporation

Rubidium Frequency Standard with a High Resolution Digital Synthesizer 108
A. Stern, B. Levy & M. Bootnik, Time and Frequency Ltd., Israel, & E. Detoma & G. Pedrotto, SEPA, Italy

Towards a Miniature Laser-Pumped Cesium Cell Frequency Standard 114
P.J. Chantry, B.R. McAvoy, J.M. Zomp, and I. Liberman, Westinghouse Science & Technology Center

Ways of Improvement of Long-Term Frequency Stability of Onboard Atomic Beam Frequency Standard for GLONASS System 123
A.G. Gevorkyan, V.S. Zholnerov & A.B. Basevich, Russian Institute of Radio-navigation and Time, Russia

Architecture and Algorithms for New Cesium Beam Frequency Standard Electronics 127
L.S. Cutler & R.P. Giffard, Hewlett-Packard Laboratories

A New RF Architecture for Cesium Frequency Standards 134
R.K. Karlquist, Hewlett-Packard Company

A New Cesium Beam Frequency Standard Performance Data 143
J.A. Kusters & J.L. Johnson, Hewlett-Packard Company

A Study of Atomic Time Scale Stability 151
W. Guo, Shaanxi Astronomical Observatory, China

Low Noise, Microwave Signal Generation Using Cryogenic, Sapphire Dielectric Resonators: An Update 157
M.M. Driscoll, Westinghouse Electronics Systems Group & R.W. Weinert, Westinghouse Science & Technology Center

Higher Order Modes, High Q Dielectric Resonators for Oscillator Applications 163
K.A. Zaki & X-P. Liang, University of Maryland

Ultra-Stable Cryogenic Sapphire Dielectric Microwave Resonators 167
A.G. Mann, A.N. Luiten, D.G. Blair & M.J. Buckingham, University of Western Australia

Environmental Sensitivity of the Superconducting Cavity Maser Oscillator 172
R.T. Wang & G.J. Dick, Jet Propulsion Laboratory

Microwave Frequency Discriminator with a Cooled Sapphire Resonator for Ultra-Low Phase Noise 176
D.G. Santiago & G.J. Dick, Jet Propulsion Laboratory

FREQUENCY SYNTHESIZERS, NOISE, MEASUREMENT TECHNIQUES AND DISTRIBUTION

Ultra-High Stability Synthesizer for Diode Laser Pumped Rubidium 183
J.P. Lowe, F.L. Walls & R.E. Drullinger, National Institute of Standards and Technology

A Low-Frequency, High Resolution Digital Synthesizer *(INVITED)* 188
R.P. Giffard and L.S. Cutler, Hewlett-Packard Laboratories

Spectral Performance of Frequency Multipliers and Dividers *(INVITED)* 193
M.M. Driscoll & T.D. Merrell, Westinghouse Electronic Systems

Frequency Instabilities in Phase-Locked Synthesizers Induced by Time Delays 201
W. Wischert, M. Olivier & J. Groslambert, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, France

Stability Analysis of an 18 GHz Microwave Frequency Synthesizer 207
S.D. Turner, MAXTECH, Inc. & A. Behagi, Pennsylvania State University

Sensitivity of Fast Settling PLLs to Differential Loop Filter Component Variations 213
T.J. Endres & J.B. Kirkpatrick, Hughes Aircraft Company

Improving Noise in Measurement Systems 225
R.L. Hamell, P.F. Kuhnle, R.L. Sydnor, G.J. Dick, A. Kirk & W.A. Diener, Jet Propulsion Laboratory

Application of Kalman Filtering Techniques to the Precision Clock With Non-Constant Aging 231
W. Su, Geo-Centers, Inc., & R.L. Filler, Army Research Laboratory

Performance of a PID Phase Lock Loop With Kalman Filtered Input Data 238
R. Hardin & M. Yankowski, Ball Efratom Division

Cross-Correlation Phase Noise Measurements 257
W.F. Walls, Femtosecond Systems

A Shortcut for Computing the Modified Allan Variance 262
C.A. Greenhall, Jet Propulsion Laboratory

Noise Induced Time Interval Measurement Biases 265
E. Rubiola, A. Del Casale & A. De Marchi, Politecnico di Torino, Italy

The Greatest Common Factor Frequency and its Application in the Accurate Measurement of Periodic Signals 270
W. Zhou, Xidian University, P.R. China

Discrete Simulation of Power Law Noise 274
N.J. Kasdin & T. Walter, Stanford University

A New Multi-Variance Method for the Oscillator Noise Analysis 284
F. Vernotte, Observatoire de Besançon, E. Lantz, Laboratoire d'Optique CNRS, J. Groslambert & J.J. Gagnepain, Laboratoire de Physique et Metrologie des Oscillateurs du CNRS, France

Secondary Standard for PM and AM Noise at 5, 10, and 100 MHz 290
F.L. Walls, National Institute of Standards and Technology

Confidence on the Second Difference Estimation of Frequency Drift 300
M.A. Weiss, D.W. Allan & D.A. Howe, National Institute of Standards and Technology

Accuracy Estimation of GLONASS Satellite Oscillators 306
Y.G. Gouzhva, A.G. Gevorkyan & P.P. Bogdanov, Russian Institute of Radio-navigation and Time, Russia

High Stability Microwave Fiber Optic Systems: Demonstrations and Applications 310
R.T. Logan, Jr. & G.F. Lutes, Jet Propulsion Laboratory

NRL/USNO Two Way Time Transfer Modem Design and Test Results 317
G.P. Landis & I. Galysh, Naval Research Laboratory

A New Model of 1/F Noise in BAW Quartz Resonators 327
F.L. Walls, National Institute of Standards and Technology, P.H. Handel, University of Missouri, R. Besson, ENSMM, France & J.J. Gagnepain, CNRS, France

Reduction of Quartz Crystal Oscillator Flicker-of-Frequency and White Phase Noise (Floor) Levels and Acceleration Sensitivity Via Use of Multiple Resonators 334
M.M. Driscoll, Westinghouse Electronic Systems Group

BAW, SAW AND OTHER ACOUSTIC OSCILLATORS AND FILTERS; AND FREQUENCY CONTROL DEVICE COMPONENTS

Spectral Purity of Acoustic Resonator Oscillators *(INVITED)* 340
T.E. Parker & G.K. Montress, Raytheon Research Division

Microwave Oscillators for Superior Short Term Stability & Ultra-Low Phase Noise *(INVITED)* 349
G.J. Dick, Jet Propulsion Laboratory

Design and Performance of a Low Noise, Wide Tuning Range AQP SAW Delay Line VCO 356
G.K. Montress, T.E. Parker, Raytheon Research Division, D.W. Kress, Naval Air Warfare Center, & J.A. Kosinski, Army Research Laboratory

Optimal and Suboptimal Synthesis of Surface Acoustic Wave Filters 371
A.S. Rukhlenko, Minsk Radioengineering Institute, Republic of Belarus, CIS

Analysis of Leaky Surface Acoustic Waves on $LiTaO_3$ Substrate..... 378
H. Meier & P. Russer, Technischen Universität München, Germany

Parameter Extraction of SAW Resonator Equivalent Circuit Parameters and Package Parasitics 384
M. Cavin, N. Eisenhauer & D.C. Malocha, University of Central Florida

Microwave Oscillators Stabilized with Surface Transverse Wave Resonant Devices..... 391
I.D. Avramov, Institute of Solid State Physics, Republic of Bulgaria

An X-Band, 2.5 Watt Continuous Wave Dielectric Resonator Oscillator for Future Military Systems 409
M. Mizan, Army Research Laboratory

Piece-Wise Linear Modeling of Solid-State Varactor-Tuned Microwave Oscillators..... 415
A. Behagi, Penn State University & S. Turner MAXTECH, Inc.

Ultra-Stable Microwave and Millimeter Wave Photonic Oscillators..... 420
R.T. Logan, Jr. & L. Maleki, Jet Propulsion Laboratory

Temperature Dynamic Model and Temperature Dynamic Compensation of Crystal Oscillators..... 425
A.V. Kosykh, B.P. Ionov, Omsk Polytechnical Institute, & A.M. Vasiliev, Microsistem, Inc., Russia

Frequency Temperature Behavior and Its Hysteresis of S-TCXOs Using New Shape GT Cut Quartz Crystal Resonators 434
H. Kawashima and K. Sunaga, Seiko Electronic Components Co. Ltd., Japan

Two-Mode Quartz Resonators for Digital Temperature Compensated Quartz Oscillators..... 442
I.V. Abramson, Omsk Research Institute of Device-Making, Russia

Analysis of Start-Up Characteristics of CMOS Crystal Oscillators..... 448
M. Toki & Y. Tsuzuki, Yokohama National Univiersity, Japan

Universal, Computer Facilitated, Steady State Oscillator, Closed Loop Analysis Theory 453
B. Parzen, Consulting Engineer

Long-Term Performance of Precision Crystal Oscillators in a Near-Earth Orbital Environment..... 465
L.J. Rueger, Bendix Field Engineering Corporation, J.R. Norton, Johns Hopkins University, Applied Physics Laboratory, & P.T. Lasewicz, Naval Satellite Operations Ctr.

Long Term Aging of Oscillators..... 470
R.L. Filler & J.R. Vig, US Army LABCOM

Analysis and Results of Compensated Quartz Crystal Oscillator Ovens 485
R. Brendel, G. Marianneau, & F. Djian, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, & E. Robert, Centre National d'Etudes Spatiales, France

A Pulsed Carrier Recovery Using a Discrete Conception of the PLL 492
A. Aguasca, J.M. Plana & A. Broquetas, Polytechnic University of Catalonia, Spain

QUARTZ AND OTHER PIEZOELECTRIC/DIELECTRIC RESONATORS, SENSORS AND OTHER DEVICES

Improvement of Characteristics of Quartz Resonator-Thermostate With Direct Heating Piezoelement 499
I.V. Abramson & A.N. Dikidzhi, Omsk Research Institute of Device Making, Russia

The Further Development of Internally Heated Quartz Resonators and the Oscillators on Their Basis for Use in Up-To-Date Search-and-Rescue and Communication Systems 505
Y.L. Vorohkovsky, I.G. Petrosian, B.G. Drakhlis, A.A. Volkov, Morion Mfg. Co., Russia

Piezoelectric Sensors and Actuators: Smart Materials *(INVITED PLENARY SESSION PAPER)* 513
R.E. Newnham, The Pennsylvania State University

New Cuts for Width-Extensional Mode Quartz Crystal Resonators 525
H. Kawashima, Seiko Electronic Components, Ltd., Japan

Forced Vibrations of Thickness-Flexure, Face-Shear and Face-Flexure in Rectangular AT-Cut Quartz Plates 532
H. Sekimoto & Y. Watanabe, Tokyo Metropolitan University & M. Nakazawa, Shinsyu University, Japan

Vibrations of AT-Cut Quartz Strips of Narrow Width and Finite Length 537
P.C.Y. Lee & J. Wang, Princeton University

Lateral Field Excitation and Coupled Modes in AT Quartz: Synchrotron Radiation X-Ray Topography 553
B. Capelle, Y. Zheng, A. Zarka, L.M.C.P. Université P&M Curie, & J. Detaint & J. Schwartzel, CNET, France

Non-Uniform Distribution of Motion Influence on the Effective Mass-Loading in AT-Cut Quartz Resonators 561
I. Mateescu & E. Candet, Institute of Physics and Technology of Materials, Romania

On Straight Crested Waves in a Third Overtone SC-Cut Quartz Resonator 567
Y-K. Yong & Z. Zhang, Rutgers University

Four-Point SC-Cut Crystal Unit Design Calculations for Reduced Acceleration Sensitivity 582
L.D. Clayton & E.P. Eernisse, Quartztronics, Inc.

Low Temperature Limitation on the Quality Factor of Quartz Resonators 597
A. El Habti & F. Bastien, Laboratoire de Physique et Métrologie des Oscillateurs du CNRS, France

The Effect of the Temperature on the Vibration Frequencies of Plane-Convex Plate of Quartz Resonators 603
S.M. Bauer, S.B. Filippov, B.N. Semenov, P.E. Tovstik & Y.L. Vorokhovsky, Sankt-Petersburg State University, "Morion" Mfg. Co., Russia

Investigation of Electrode-Quartz Interface 607
P.I. Asseev & V.B. Grouzinenko, Scientific-Research Institute "Fonon", Russia

A Theoretical Study of Quartz Crystal Resonators for Chemical Sensors, Vibrating in C Mode 610
M. Nakazawa & M. Takeuchi, Shinshu University, T. Moriizumi & H. Niiyama, Tokyo Institute of Technology, Japan & A. Ballato & T. Lukaszek, Army Research Laboratory

An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips 614
Y.S. Zhou & H.F. Tiersten, Rensselaer Polytechnic Institute

Vibrations of Circular Disk Dielectric Resonators 626
P.C.Y. Lee & J.S. Yang, Princeton University

Energy Trapping Resonators and Integrated Filters Using the AT Cut of Quartz Like Materials 639
J. Detaint, J. Schwartzel & C. Joly, CNET, A. Zarka, B. Capelle, & Y. Zheng, L.M.C.P. Université Paris, Y. Toudic, CNET, & E. Philippot, L.P.C.M.S. Université du Languedoc, France

New Processing Techniques: Sweeping of Quartz Wafers and a Practical Method for Processing Quartz Resonators Under Controlled Conditions 648
C. Bowen & G. Houck, Precision Quartz Products

Sweeping Investigations on Grown-In Al-Li^+ and Al-OH^- Centers in Natural Crystalline Quartz 657
H. Bahadur, National Physical Laboratory, India

Ageing of Crystal Elements Manufactured of Swept Synthetic Quartz 667
L.N. Romanov & S.V. Kolodieva, All-Union Research Institute for the Synthesis of Materials, Russia

Radiation Resistance of Synthetic Quartz Crystals 673
I.V. Kabanovich, Scientific Research Institute "Fonon", Russia

The Development and Production of High Quality Quartz Material for Precision Crystals 679
A.P. Pogrebnyak & S.N. Abdrafikov, The Plant "Kristall", Russia

Numerical Calculation of Temperature Field for the Case of Simple Convection Model in Autoclave 683
W. Hofman, Institute of Electronic Materials Technology, Poland

Disregarded Sources of Random Errors in the Nonlinear Constants of Quartz 690
C.K. Hruska, York University, Canada

Lattice Distortion and Parameter Variation in Quartz and Berlinite Crystals: Methods for Improving the Quality 696
A. Zarka, B. Capelle & Y. Zheng, LMCP, J. Schwartzel & J. Detaint, CNET, X. Buisson, SICN, & Y. Toudic, CNET, France

Piezoelectric Resonators from $La_3Ga_5SiO_{14}$ (Langasite) - Single Crystals 707
V.B. Grouzinenko & V.V. Bezdelkin, Scientific-Research Institute "Fonon", Russia

Application of Langasite Crystals in Monolothic Filters Operating in Shear Modes *(INVITED)* 713
S.A. Ssakharov, I.M. Larionov & A.V. Medvedev, Scientific-Research Institute "Fonon", Russia

Dilithium Tetraborate ($Li_2B_4O_7$) Fabrication Technology 724
J.G. Gualtieri, J.A. Kosinski & W.D. Wilber, US Army LABCOM, Y. Lu, Rutgers University, S.T. Lin, M. Murray & W. Ruderman, INRAD, Inc.

Piezoelectric, Elastic, and Ferroelectric Properties of $KTiOPO_4$ and Its Isomorphs 732
D.K.T. Chu & J.D. Bierlein, DuPont Experimental Station & R.G. Hunsperger, University of Delaware

New Approach of Crystal Growth and Characterization of a Quartz and Berlinite Isomorph: $GaPO_4$ 744
E. Philippot, A. Ibanez & A. Goiffon, CNRS, A. Zarka, Univ. Paris, J. Schwartzel & J. Detaint, CNET, France

Robust Solder Joint Attachment of Coaxial Cable Leads to Piezoelectric Ceramic Electrodes 753
P.T. Vianco, Sandia National Laboratories

ENVIRONMENTAL SENSITIVITIES OF PRECISION OSCILLATORS AND STANDARDIZATION

Introduction to Guidelines for the Measurement of Environmental Sensitivities of Precision Oscillators 762
H. Hellwig, Air Force Office of Scientific Research

Acceleration, Vibration and Shock Effects - IEEE Standards Project P1193 763
J.R. Vig, Army Research Laboratory, C. Audoin, Université Paris-Sud, L.S. Cutler, Hewlett-Packard Co., M.M. Driscoll, Westinghouse Defense and Electronic Systems Center, E.P. EerNisse, Quartztronics, Inc., R.L. Filler, Army Research Laboratory, R.M. Garvey, Frequency and Time Systems, Inc., W.J. Riley, EG&G Frequency Products, R.C. Smythe, Piezo Technology, Inc., & R.D. Weglein, Consultant

Precision Oscillators: Dependence of Frequency on Temperature, Humidity and Pressure - Working Group 3 of the IEEE SCC27 Committee 782
D.W. Allan, NIST, J. Barnes, Austron, F. Cordara, IEN, Italy, R.M. Garvey, Frequency and Time Systems, Inc., W. Hanson, Piezo Crystal Co., R. Kinsman, Motorola, Inc., J. Kusters, Hewlett-Packard Co., R. Smythe, Piezo Technology, Inc., & F.L. Walls, NIST

IEEE Standards Project P1193 Electric and Magnetic Fields and Effects 794
R.L. Sydnor, Jet Propulsion Laboratory, R. Brendel, CNRS, France, & W. Wiedemann, Efratom/Ball Corporation

The Effects of Ionizing and Particle Radiation on Precision Frequency Sources - (Proposal for IEEE Standards Project Pll93) 798
J.J. Suter, R.H. Maurer & J.D. Kinnison, Johns Hopkins University, R. Besson, Ecole Nationale Supérieure de Méchanique et des Microtechniques, France, J.R. Vig, Army Research Laboratory, & A. Koehler, Harvard University

Aging, Warm-Up Time and Retrace; Important Characteristics of Standard Frequency Generators - (Proposal for IEEE Standards, Project P1193) 807
J. Vanier, National Research Council, Canada, J.J. Gagnepain, Centre National de la Recherche Scientifique, France, W.J. Riley, EG&G, F.L. Walls, National Institute of Standards and Technology, & M. Granveaud, LPTF, France

General Considerations in the Metrology of the Environmental Sensitivities of Standard Frequency Generators - IEEE Standards Committee P1193 Working Group 6 Report 816
L.A. Breakiron & G.M.R. Winkler, US Naval Observatory, L.S. Cutler, Hewlett-Packard Co., H.W. Hellwig, US Air Force Office of Scientific Research, & J.R. Vig, Army Research Laboratory

Environmental Effects in Mixers and Frequency Distribution Systems 831
L.M. Nelson & F.L. Walls, National Institute of Standards and Technology

Advances in Acceleration Sensitivity Measurement and Modeling 838
J.A. Kosinski & A. Ballato, Army Research Laboratory

Market Survey of Acceleration-Insensitive SAW Oscillators 849
J. Himmel, R. McGowan, J. Kosinski & T. Lukaszek, Army Research Laboratory

International Standardization in the Global Marketplace 861
D.C. Bradley, EG&G Mound Applied Technologies, Inc.

AUTHOR INDEX 866

PROCEEDINGS AVAILABILITY INFORMATION 868

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL 869

1992 AWARD WINNERS

The Cady Award

The Cady Award was presented to Raymond J. Besson, ENSMM, Besancon, France "for fundamental contributions to both quartz resonator fabrication technology and understanding of nonlinear effects leading to devices of superior performance." The award was presented by Errol P. EerNisse, Quartztronics, Inc.

The Rabi Award

The Rabi Award was presented to James A. Barnes, Austron, Inc., "for contributions and leadership in the development of the statistical theory, simulation and practical understanding of clock noise and the application of this understanding to the characterization of precision oscillators and atomic clocks." The award was presented by David W. Allan, NIST.

The Sawyer Award

The Sawyer Award was presented to Martin B. Bloch, Frequency Electronics, Inc., "for that rare combination of scientific and entrepreneurial excellence dedicated to the frequency control industry." The award was presented by John R. Vig, U.S. Army LABCOM.



Raymond J. Besson, Cady Award winner; James A. Barnes, Rabi Award winner; and Martin B. Bloch , Sawyer Award winner; after the award presentations.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LOCAL OSCILLATOR REQUIREMENTS AND STRATEGIES FOR THE NEXT GENERATION OF HIGH-STABILITY FREQUENCY STANDARDS

F. L. Walls
Time and frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

This paper provides a brief introduction to the special session on high-stability frequency synthesis from rf to the visible. The special session is motivated by a number of proposals that have been made for developing frequency standards with frequency stabilities in the 10^{-16} to 10^{-18} range. Currently available local oscillators, frequency synthesis techniques, and the measurement methods are not, however, sufficient to support this level of stability. Some general approaches to solve these problems, such as the effect of interrogation type on local oscillator induced white frequency noise, are briefly introduced. The details are to be found in the individual presentations.

Introduction

The purpose of this paper is to set the stage for the special session on high-stability frequency synthesis from the rf to the visible. A number of groups around the world are working on new frequency standards that hold promise of attaining fractional frequency stabilities in the 10^{-16} to 10^{-18} range. Frequency standards based on Ba^+, Be^+, Hg^+, Mg^+, Yb^+, have been proposed in the microwave region. Frequency standards based on Ba^+, Hg^+, Sr^+, Yb^+, and Ca have been proposed in the region from 10^{12} to 10^{15} Hz [1-7]. To achieve the projected frequency stabilities of 10^{-16} to 10^{-18} and to translate this stability to other frequencies will require substantial improvements in local oscillators, frequency synthesis techniques, and measurement procedures. Although the frequencies of the clocks range over 5 orders of magnitude, there are many common elements which must be addressed and that is the focus of this special session.

Contribution of the U. S .Government, not subject to copyright.

Projections for Local Oscillators Requirements

The requirements for local oscillators depend on the frequency stability, the line width of the atomic resonance, and the type of modulation. Most of the new ultra-high precision atomic clocks under development use some type of confinement scheme to reduce the linewidth of the resonance. Digital modulation schemes are usually indicated for frequency standards with linewidths under a few hertz. Several common approaches are shown in Fig. 1. In these approaches, the phase accumulation in the clock during portions of the cycle when the atomic reference is not sensitive to the clock frequency contributed to a pseudo white frequency level that limits the short-term frequency stability of the standard. This has been investigated in considerable detail by Dick et al [8]. Their results summarized in Fig. 2 show that for square-wave frequency modulation and a single atomic reference sample, the local oscillator induces white frequency noise at approximately 50% of the open-loop stability of the local oscillator, at the period of the interrogation cycle.

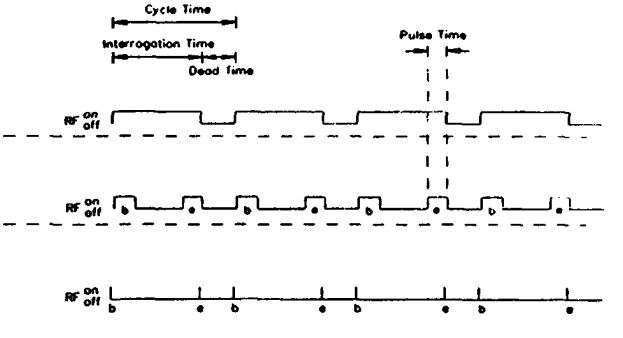


Figure 1. Timing diagram for (a) Rabi type square-wave frequency modulation with dead time, (b) Ramsey type time-domain interrogation with dead time, and (c) Ramsey type time-domain interrogation with very short pulses. Adapted from [8]

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

If a time-domain Ramsey interrogation cycle is used, the local oscillator induced white frequency can be reduced as much as a factor of 30 for small fractional dead times. If, however, two or more reference samples are available, it is possible to shape the interrogation process such that the white frequency contribution due to noise in the local oscillator is reduced even more. The amount of reduction depends on the details of the interrogation cycle and the type of noise in the local oscillator. Flicker frequency was assumed for the calculations of Fig. 2.

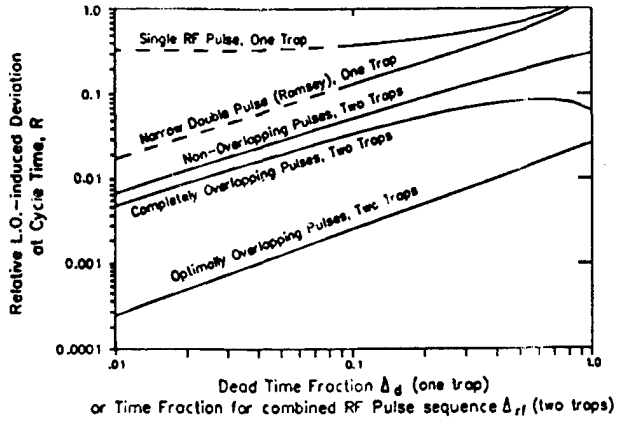


Figure 2a. Relative level of white frequency induced by the local oscillator as a function of dead time for both one and two atomic samples as a function of interrogation type and fractional dead time.

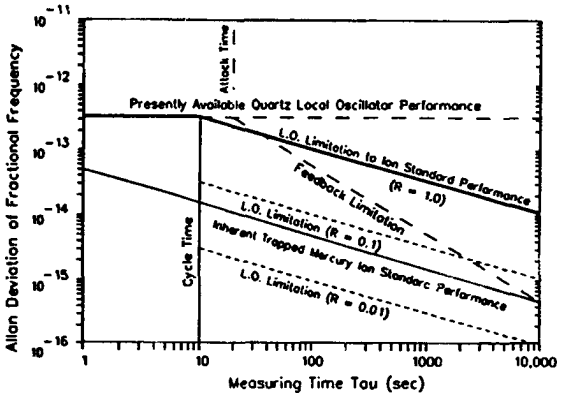


Figure 2b. Limitation on the frequency stability due to local oscillator induced noise as a function of measuring time for selected values of R. Adapted from [8]. Some of the oscillators from Norton exhibit flicker levels a factor of 5 better than assumed in this example [11].

Frequency standards having linewidths larger than a few hertz are potential candidates for sine-wave modulation. Sine-wave modulation also induces a pseudo contribution to the white frequency noise which has been examined in detail by Barillet et al [9]. They show that the phase noise in the local oscillator at the second harmonic and even higher harmonics of the modulation frequency, f_m, limits the frequency stability to

$$\sigma_y(f) = f_m/\nu_o\,[0.723\,S^{1/2}_{\phi}(2\,f_m) + CS_{\phi}(4f_m)] \qquad (1)$$

where $S_{\phi}(2f_m)$ and $S_{\phi}(4f_m)$ are the power spectral densities of phase noise in the local oscillator at 2 and 4 times the modulation frequency, and C is typically of order 0.1. For sine-wave modulation the frequency stability of the source at long times is of little concern since the induced white frequency noise level is due primarily to noise at the second harmonic of the modulation. Lowe et al examine the practical limitation and potential improvements that might be obtained by using a notch filter at $\pm$ 2 f_m from the carrier [10]. They show that, by careful selection of the configuration, it is possible to realize substantial reductions in the induce white frequency.

Using these results as a guide, we can now estimate the effect of various local oscillators on the overall performance of the proposed frequency standards as a function of modulation type. In general the stability of presently available local oscillators does not meet the needs of the potential super clocks in either the rf or optical regions.

Local Oscillators in the rf, Microwave, and Optical Regions

A large number of the talks in the special session and throughout the conference examine the present stability and prospects for future improvements in oscillators based on quartz acoustic resonators, microwave oscillators at cryogenic and room temperature, composite rf and microwave oscillators, and various optical sources. In most cases the best local oscillator is distributed in nature. The frequency stability at different times (or Fourier frequencies) is derived from a combination of several oscillators or oscillators and filters. The distributed approach of necessity involves high-precision frequency synthesis.

Frequency Synthesis Requirements

There are generally two drivers for improving techniques for frequency synthesis. The first is the need to fabricate the local oscillator from several

sources to optimize the frequency stability at a variety of measurement times or Fourier frequencies. The second is the need to translate the precision output frequency to another frequency or to keep time.

In the rf region frequency multiplication, division, and translation are well established techniques. The very high precision and accuracy of the new clocks will, however, require a factor of approximately 100 improvement in the phase stability of these techniques at frequencies of 5 to 10 MHz and a factor of 10 improvement at 100 MHz. At 10 GHz present techniques are sufficient.

In the optical region, frequency multiplication and translation have been used for a long time, but frequency division is still in its infancy. In this region the phase stability requirements are often just that there be no cycle slips. Reliability of the sources and the nonlinear devices are still important issues in this region. We have a number of exciting talks in the special session which focus on improving frequency division, translation, multiplication, and the reliability of sources in the optical region.

Measurements and Distribution Strategies

Measurement strategies for determining the frequency stability of local oscillators in the short term and long term differ greatly from the rf to the optical region.

In the region below 10 GHz, many of the problems of frequency synthesis and metrology are due to environmentally induced phase shifts in signal transmission and detection. The magnitude of the problem at low frequencies can be visualized by noting that at 5 MHz only 3 x 10^{12} rad accumulate in a day. To achieve a fractional frequency stability of 10^{-17} requires a phase stability of 30 μrad/d! A number of the talks and posters address this problem in frequency synthesis and/or in measurement systems. Several of the talks in the special session focus on developing better phase detectors for both frequency synthesis and time-domain measurements.

Conclusion

The development of new clocks with frequency stabilities in the 10^{-16} to 10^{-18} range will require major improvements in the stability and reliability of reference sources, frequency synthesis, and measurement techniques. Potential solutions to many of these problems appear in the talks of this conference. Although many of these techniques have yet to be fully implemented and refined, we have hope that several types of frequency standards and clocks exhibiting frequency stabilities in this range will be realized within the next few years.

Acknowledgements

I am grateful to all the presenters, the technical program committee, and especially Jack Kusters for making this special session on frequency synthesis a reality.

References

[1] D.J. Wineland, J.C. Bergquist, J.J. Bollinger, W.M. Itano, D.J. Heinzen, S.L. Gilbert, C.H. Manney, and M.G. Raizen, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, vol. 37, No. 6, pp. 515-523, Nov. 1990.

[2] J.D. Prestage, G.J. Dick, and L. Maleki, "JPL Trapped Ion Frequency Standard Development," IEEE 41st Annual Frequency Control Symposium, pp. 20-24, 1987.

[3] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, D.J. Wineland, "Ionic Crystals in a Linear Paul Trap," Physical Review, vol. 45, Number 9, pp. 6493-6501, May 1, 1992.

[4] A. de Marchi."The Optically Pumped Caesium Fountain: 10^{-15} Frequency Accuracy?," Metrologia, vol. 18, pp. 103-116, 1982.

[5] A. Clairon, C. Salomon, S. Guellati, and W.D. Phillips, "Ramsey Resonance in a Zacharias Fountain," Europhysics Letters vol. 16(2), pp. 165-170, September, 1991.

[6] S. Chu "Laser Manipulation of Atoms and Particals," Science, vol. 253, pp. 861-866, 1991.

[7] J.L. Hall, M. Zhu, and P. Buch, "Prospects for Using Laser-Prepared Atomic Fountains for Optical Frequency Standards

Applications," J. Opt. Soc. Am. B, vol 6, pp. 2194-2205, 1989.

[8] G.J. Dick, J.D. Prestage, C.A. Greenhall, and L. Maleki, "Local Oscillator Induced Degradation of Medium-Term Stability in Passive Atomic Frequency Standards," 22nd PTTI, NASA CP 3116, pp. 487-508, 1990.

[9] R. Barillet, V. Giordano, J. Viennet, C. Audoin, "Microwave Interrogation Frequency Noise and Clock Frequency Stability: Experimental Results," Proc. 6th EFTF, 1992.

[10] J. P. Lowe, F. L. Walls, and R. E. Drullinger, "Ultra-high stability synthesizer for diode laser pumped rubidium," 1992 IEEE Frequency Control Symposium, these proceedings.

[11] J. Norton, "BVA-Type Quartz Oscillators for Spacecraft," Proc. of 45th Annual Frequency Control Symposium, 1991, pp. 426-430.

LOCAL OSCILLATOR REQUIREMENTS FOR TIMEKEEPING IN THE $10^{-14}\,\tau^{-1/2}$ ERA

R.J. Douglas and J.-S. Boulanger
Institute for National Measurement Standards
National Research Council of Canada, Ottawa, Canada, K1A 0R6

Abstract: *To exploit the timekeeping capabilities of new passive atomic frequency standards, improved local oscillators or new ways of processing the response of the atomic resonance are needed. We present and analyze a new method of using current local oscillators for timekeeping standards that use Ramsey interrogation of large ensembles of ions or atoms, such as a linear ion trap or a pulsed atomic fountain. We use the time domain to analyze these systems. The atomic ensemble is regarded as the reference oscillator to measure the local oscillator's accumulation of the phase error during the time interval between the two Ramsey interactions. Thus the local oscillator's degradation of timekeeping can be largely limited to the accumulation of local oscillator phase error during the dead time between Ramsey interrogation cycles. The degradation is studied and presented for different local oscillators in the context of a pulsed atomic fountain. For timekeeping, current quartz local oscillators could be used in a pulsed cesium fountain frequency standard to approach the $\sigma_y(\tau)$ regime of $10^{-14}\,\tau^{-1/2}$, and with improved local oscillators that are under development, could be spectacularly better.*

Introduction

Traditionally, atomic frequency standards have been operated continuously with a local oscillator under servo control of the atomic resonance, aimed at optimizing the stability of the standard's real-time output. Pulsed atomic (and ionic) frequency standards with a large number of atoms per pulse[1],[2] are under development which challenge the capabilities of traditional local oscillators in two ways: firstly, their intrinsic stability could be very good ($10^{-14}\,\tau^{-1/2}$); and secondly, with traditional frequency control servos their cycle times are long - in the range of 1-100 s and demand really good local oscillator performance, beyond the capacity of the current state-of-the-art of quartz oscillators, in the 10-1000 s range ($\sigma_y(\tau)$ of 3×10^{-14} to 3×10^{-15}).

Many atomic frequency standards are actually used for timekeeping which can incorporate post-processing, rather than for the stability of their raw real-time output. Applications which demand the best performance from a frequency standard (or an ensemble of frequency standards) are normally converted to this timekeeping mode. Post-processing for time scales (e.g. TAI [3], to which most frequency standard work ultimately refers) is very widely practised, and can include retrospective processing which a traditional servo designer would characterize as a violation of causality since it uses a knowledge of the future. Frequency standards used in this timekeeping mode do not rely exclusively on the traditional servo of the unit to establish its frequency.

We propose here a new mode of operation of pulsed frequency standards for the purposes of timekeeping, where the requirements on the servo and local oscillator can be greatly relaxed at the cost of a modest amount of additional post-processing that would only delay "final" results by a few cycle times (1-100 s). This mode works for a reasonably large number (>100) of atoms detected per cycle, and appears to offer greatest benefits for pulsed atomic frequency standards with small dead time fractions - the fraction of time between the completion of one Ramsey cycle and the commencement of the next cycle. For some standards, a dead time fraction of 0.01 might be obtained relatively easily, even without incorporating the elegant solutions proposed for linear ion traps [4]. In this proposed mode the local oscillator is used to transfer frequency stability from the cycle time interval back to the dead time interval, rather than the traditional transfer of frequency stability from the cycle time interval out to the integrator knee of a conventional frequency standard's servo.

To analyze the relaxed requirements for the performance needed from the local oscillator, we choose the time domain, since we expect the timing uncertainty to depend on the phasing of the start and stop times of a

0-7803-0476-4/92 $3.00 © 1992 IEEE

user's interval with respect to the Ramsey interrogations of the atomic ensemble.

The time domain simplifies the analysis of the errors contributed by the local oscillator and also greatly simplifies presentation and use of the results. For example, consider evaluating the importance of synchronization in phase measurement methods to be employed with a pulsed frequency standard. We would relate phase measurements of the external unit to the phase of the local oscillator, and have to choose an optimal averaging strategy, regarding the phase error of the local oscillator as a quantity to be (imperfectly) measured by the atomic polarization of the (two-level) atomic ensemble, and the local oscillator frequency through the dead time to be a quantity to be (imperfectly) predicted from its frequency over adjacent measured intervals. We feel that this requires significantly more effort in the frequency domain, and appeals less to an intuitive appreciation of the situation.

To quantitatively analyze the contributions of the local oscillator we use the conventional power-law spectral density of phase noise, but we examine the root-mean-square timing residuals across the Ramsey cycle (this timing residual is the root-mean-square deviation from a fitted functional form of local oscillator phase vs time, averaged over many cycles but not averaged across the cycle). We restrict ourselves to analyzing residuals from linear fits of phase versus time, but consider three cases: (1) using only the frequency calibration from the previous Ramsey cycle, (2) using the frequency calibration from the current Ramsey cycle (perhaps incomplete at the real time under consideration) and (3) using the frequency calibrations from both the previous and subsequent Ramsey cycles. The residuals give exactly what we want: a prediction of the time uncertainty due to the combination of the local oscillator and any given processing scheme. We present our methods of calculating residuals from the commonest types of phase noise power laws in good local oscillators: our answers reproduce the results which can be derived from conventional N-sample variances where the two methods can legitimately be compared.

We conclude that current local oscillators can give spectacularly good performance when used with pulsed atomic-ensemble frequency standards, and that superior short-term ($\tau < 1$ s) frequency stability of the local oscillator can be particularly beneficial for Zacharias fountain frequency standards that are under development [2],[5],[6].

Pulsed Frequency Standards

A traditional Ramsey atomic frequency standard is "pulsed" as far as an individual atom (or ion) is concerned, in that each atom experiences two "pulses" of the electromagnetic radiation that drives the transition [7], but generally the beam of atoms controlling the servo is continuous, perhaps with dead times or inconstant sensitivities added by the servo. The local oscillator contribution to the stability of this class of frequency standard has been studied.[8] Longer dead times are added for evaluation of C-fields or for longer evaluation procedures such as beam reversal in laboratory standards, and frequency stability through the dead time must be carried by another reference frequency standard (either the unit's local oscillator or other members of an ensemble of frequency standards). With proper processing, a frequency standard with such dead times has a time scale which accurately reflects the frequency of the standard alone, although it includes an admixture of the stability of the reference frequency standards.

A different class of pulsed atomic frequency standard has also been developed (Be^+[8], Hg^+[1] ...., Cs fountain[2]), where an atomic (or ionic) ensemble is prepared simultaneously (into one atomic state) as a pulse (or burst), and the whole ensemble experiences the first Ramsey "pulse", the interpulse evolution (C field), the second Ramsey "pulse", and ensemble interrogation (measuring the fraction of the atoms in the different atomic states). This class of pulsed standard is different from the more continuous type in that a traditional servo has to cope with a pulsed error signal. However there is a second and fundamental difference which we propose to exploit: the ensemble can be used to measure the phase error accumulated by the local oscillator between two Ramsey pulses (i.e. the frequency error) *without averaging over multiple Ramsey cycles.*

Zacharias Fountain

Where we feel that clarity may be improved by a specific example of a pulsed atomic-ensemble frequency standard, we will choose to consider the Zacharias fountain [2],[5],[6] type of primary (i.e. cesium hyperfine interval)

frequency standards. In a Zacharias fountain, slow atoms travelling vertically upwards at 3-10 m/s rise, stop, and fall back as in a "fountain" of atoms in a vacuum. The largest apparatus to date was Zacharias' original fountain (5m tall, 2 s flight time giving a 0.25 Hz Ramsey linewidth) which could detect no slow atoms in the thermal beams he used in the mid-50's. Recent experiments on laser cooling, trapping, and launching cesium atoms show that with modest power (50-100 mW) diode lasers (852 nm) it is possible to launch more than 10^8 atoms per pulse and flight times of 0.25 s have been achieved[2], and that with more powerful lasers $3x10^{10}$ atoms per pulse has been achieved[6]. With simple shutters, one might prepare and launch ensembles of atoms that cross the Ramsey microwave region so that there is only a short dead time between the downwards passage of one ensemble and the upwards passage of the next ensemble, as shown in Figure 1. Relative speeds within the ensemble of a few cm/s (rms), the distance of the trap below the microwave cavity, the launch velocity of the ensemble through the cavity (up to 7 m/s for our example) the microwave cavity dimension (1 cm) and the time between the two Ramsey passages all dictate that it should be relatively easy to achieve a dead time of 0.01 s for a 1 s flight time. We choose this size largely for reasons of potential accuracy (the taller the better) and convenient size (1.2 m fountain height above the cavity, but considerable additional space may be desirable for additional state preparation (introducing non-velocity perturbing changes of m_F to measure the magnetic field on the same atomic trajectory, or magnetically focusing the dispersing ensemble). However, the basic conclusions of our work do not depend critically on the choice of fountain height.

The details of the sensitivity to local oscillator phase fluctuations during a Ramsey passage has been worked out and clearly presented[4],[9]. In our estimates presented below, we take a worst-case approach, first using familiar statistical tools, and then we adapt another familiar tool that we employ for the purposes of frequency standard design. Modest gains in performance over these estimates could be properly calculated for a particular local oscillator given the details of a proposed standard's dimensions, the time dependence of its Ramsey fields, and the longi-

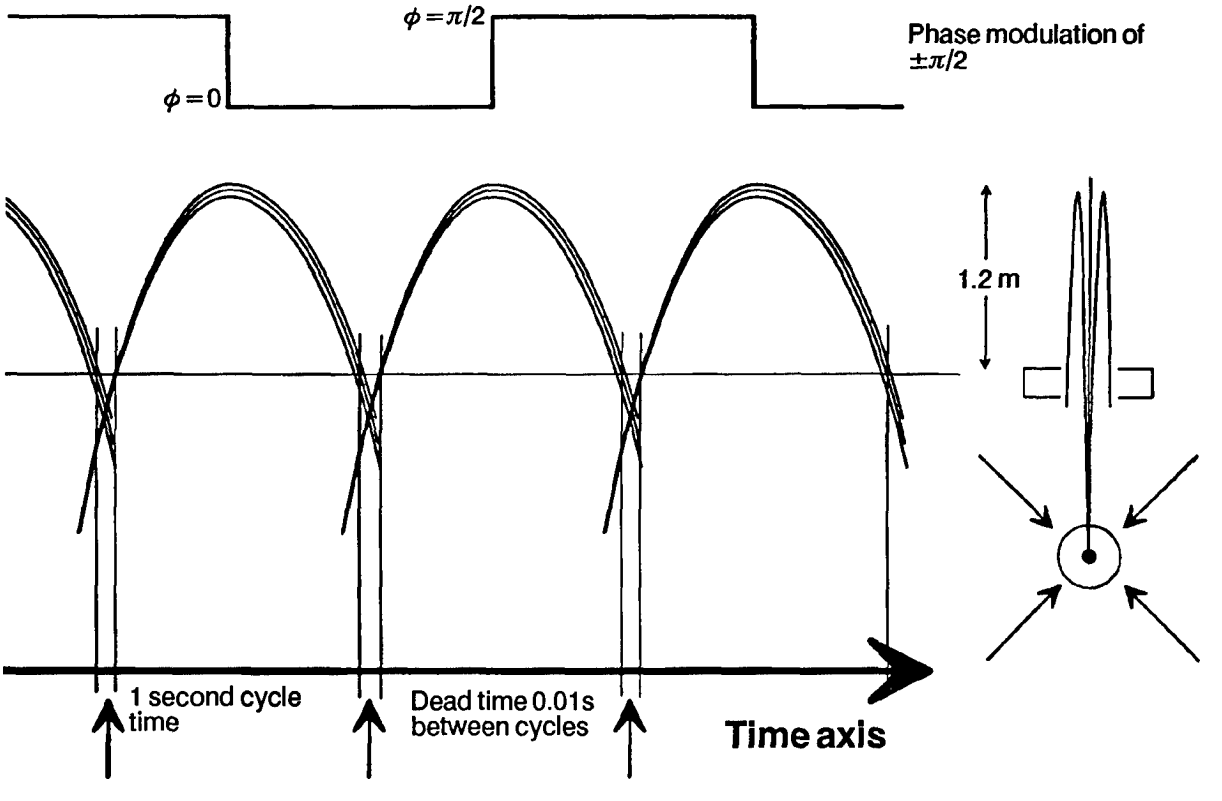


Figure 1. Example of a pulsed atomic-ensemble frequency standard, a cesium Zacharias fountain with a 1 second cycle time.

tudinal spreading of its atomic ensemble with time. These details are not necessary for appreciating the strikingly large improvement offered by a "timekeeping servo" applied to a pulsed atomic-ensemble frequency standard.

Pulsed Atomic Ensemble as a Ramsey Phase Discriminator

We consider a large number (>100) of atoms (or ions) that are simultaneously prepared in one atomic state A, and then (almost) simultaneously experience their first Ramsey interaction at almost exactly the resonant frequency ($f_0 = |E_A - E_B|/h$, from state A with energy E_A to a second atomic state B with energy E_B) and at optimum power. Atoms form a mixed state of A and B which evolves with the characteristic frequency f_0, and some time t_a later they experience (again almost) simultaneously the second Ramsey interaction with an electromagnetic field which has a phase difference $\delta\phi$ with that of the evolving atomic ensemble. Subsequently, the number of atoms detected in state A, N_A, would depend on $\delta\phi$ as $\cos(\delta\phi)$, while N_B - the number of atoms detected in state B - would depend on $\delta\phi$ as $\{1-\cos(\delta\phi)\}$.

During t_a, we can deliberately impose a square-wave phase modulation of $\pm\pi/2$ on the electromagnetic radiation, as shown in Figure 1, to obtain phase discrimination as shown in Figure 2, with a $\pm\pi/2$ range for measuring the unintended variation of the phase of the electromagnetic radiation. That is, we propose to measure the phase error accumulated by the local oscillator (relative to the atomic reference) during the time interval t_a, by measuring the atomic polarization of the ensemble. Post-processing these measurements, a corrected time scale can be constructed relative to the time scale of the local oscillator. We wish to evaluate the residual local oscillator noise that is left in this corrected time scale for different types of processing. The use of two phase offsets of $+\pi/2$ and $-\pi/2$ allows cancellation of some systematics. The modulation frequency is $1/(2t_c)$, where $t_c = t_a + t_d$.

We require that the local oscillator only rarely allow the confusion of cycle slips by exceeding the $\pm\pi/2$ unambiguous range of the Ramsey phase detector, a relatively easy to satisfy requirement for a cesium Zacharias fountain. If the detector noise is uniform with $\delta\phi$, this is also the region of greatest phase stability. Since the number of atoms $N_A + N_B$ is expected to vary widely from burst to burst in a Zacharias fountain, both N_A and N_B must be measured.

If N_A and N_B can be measured without background, the independent atomic shot noise of the two measurements can be shown to combine to give a polarization uncertainty with the same form as the slope of the Ramsey phase discriminator curve. Thus the shot noise limit for the phase uncertainty is independent of $\delta\phi$, and the system can behave as an ideal phase discriminator with an expected stability limit of

$$SHOT\ \sigma_y^2(\tau) = \frac{1}{(2\pi f_0)^2 t_c (N_A + N_B)\tau} \qquad (1)$$

Outside the shot-noise regime, an additional local oscil-

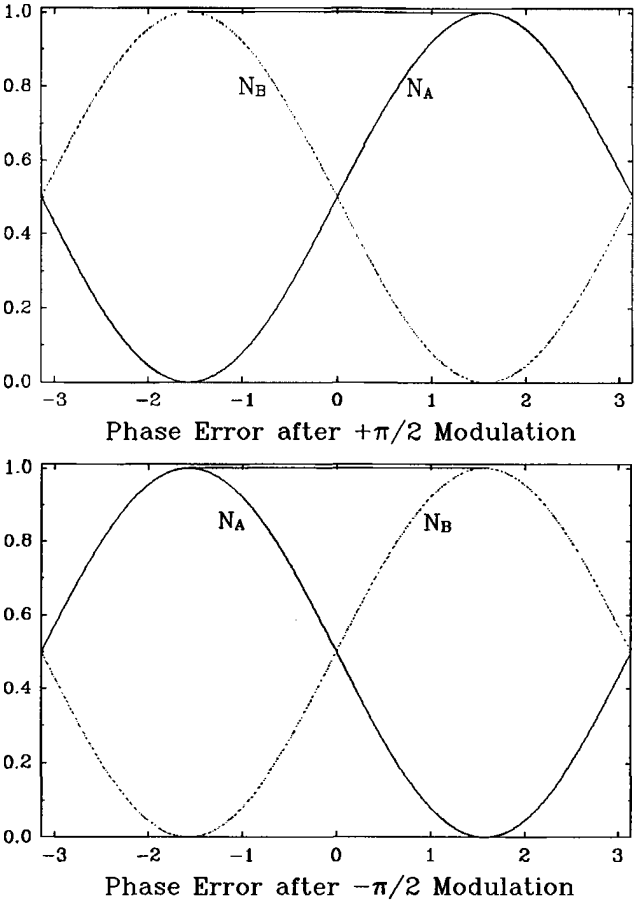


Figure 2. Ramsey phase discriminator for $\pm\pi/2$ phase modulation at the resonance frequency and at optimum power, showing the expected fraction of atoms detected in the two energy levels A and B versus the local oscillator phase error (in radians) relative to the atomic resonance. For the cesium hyperfine transition the unambiguous locking range is ± 27 ps, and over this range a measurement of the fraction can give the local oscillator phase error developed over the Ramsey interval.

lator requirements needs to be specified: to further reduce the allowable phase variation during t_a, so that the nonlinearity of the phase discriminator slope does not significantly degrade the predicted performance of these frequency standards.

Estimating Local Oscillator Requirements with Traditional Stability Tools

The N-sample frequency deviation, (from which the familiar Allan deviation is derived by setting N = 2) provides an established basis for estimating the average frequency deviation of the local oscillator average frequency $\bar{y}_n$ over a subinterval (of length τ/N), relative to the average frequency over the entire interval of N equal subintervals, $<\bar{y}>_N$. The N sample frequency deviation with no dead time, $<\sigma^2_y(N,\tau,\tau)>^{1/2}$, is defined by[10]

$$<\sigma^2_{\bar{y}}(N,\tau,\tau)> = \left<\frac{1}{N-1}\sum_{n=1}^{N}[\bar{y}_n - <\bar{y}>_N]^2\right> \quad (2)$$

Interpolation: for evaluating the average frequency deviation for a time t from the beginning of an interval of length t_a, we should evaluate $\sigma_y(N = t_a/t,t_a,t_a)$ for the specific power law noise processes of the local oscillator. In normal form, an abrupt high frequency cutoff, f_h is used to approximate the frequency sensitivity of the measurement process - here the square of the Fourier transform of the rise-time of the phase sensitivity function g(t) described by Dick[4],[9]. For our example of a Zacharias fountain with t_a = 1s, and with individual atoms taking 1.4 ms to traverse the microwave cavity, we choose a value of f_h = 1000 Hz that, when coupled to a dead time of 0.01 s, should overestimate the effects of local oscillator noise. For N large, the discrete nature of N and the effects of correlation are less important and the variation averaged over the interval t_a will be almost the same as the variation averaged over the first interval t of many cycles t_a (this will be least true for noise power laws which emphasize low-frequency noise). Thus after processing to take account of the measured frequency over the interval t_a, the time uncertainty developed by a local oscillator at a time t into the same interval t_a, for $t << t_a$, is approximately normally distributed with a standard deviation

$$^p\sigma_x(t) = [t^2 <\sigma_y^2(N = t_a/t,t_a,t_a)>]^{1/2} \quad (3)$$

The condition that $t << t_a$ is partly a condition that N be large: the discrete nature of N, and the normalization to the number of degrees of freedom (N-1) are then less troubling here. Yet the main difficulty is that unless $t << t_a$, the effects of correlation are not being correctly accounted for here: the covariance of the frequency over the interval t with the frequency over the interval t_a is important.

For example, if we are interpolating for a time t_l near the end of the interval t_a, it is possible for timekeeping systems to take the time error by interpolating backwards in time from the end of t_a. Since the variance of the frequency near the end of the interval t_a, at t_a-t_l, will be the same as its variance near the beginning of the interval t_a, at t_l, the time uncertainty developed by the local oscillator can be much smaller if the time is interpolated backwards in time from the end of the interval t_a (an allowed "violation of causality" in a post-processed timekeeping system), with an uncertainty approximately normally distributed with a standard deviation of

$$^p\sigma_x(t_l) = [(t_a-t_l)^2\sigma^2_y(N = t_a/(t_a-t_l),t_a,t_a) + t_a^2\sigma_a^2]^{1/2} \quad (4)$$

where σ_a is the frequency uncertainty over t_a due to the measurement uncertainty in the calibration of the average frequency of the local oscillator over the time interval t_a. In the shot noise limit σ_a^2 is given by Equation 1.

Near the middle of the interval t_a, the time error interpolated forward and backward in time are partially independent estimates of the same quantity, and the average might approach $\sqrt{2}$ times smaller than either. Furthermore since this is intended as an uncertainty prediction rather than a statistical tool seeking N-convergence, the division by the number of degrees of freedom (1), rather than by the number of intervals averaged (2), in this context should be questioned as well. These concerns are all resolved with the development and use of residuals, below.

Extrapolation: If t is very small compared to t_a, then there is a scant difference whether the time interval t resides within t_a (interpolation, discussed above) or adjacent to t_a (extrapolation). The time error developed by the local oscillator during the dead time t_d between Ramsey cycles is determined by extrapolation from the previous t_a. (Using both previous and subsequent calibration intervals t_a only gives an appreciable advantage only where the frequency over the first t_a is largely uncorrelated with the frequency over the second t_a. This is presented rigorously below, using residuals.) Since the dead time t_d is much less than t_a, the time uncertainty developed over each dead

time is approximately normally distributed with a normal distribution having a standard deviation of

$$^{p}\sigma_x(t_d) = [t_d^2 <\sigma_y^2(N=t_a/t_d,t_a,t_a)>]^{1/2} \quad (5)$$

Here the major question that is difficult to resolve with traditional frequency stability statistical tools is the degree to which an improvement is expected using the subsequent as well as the previous frequency calibration intervals t_a. In the post-processing environment, both calibrations are available, and if the two calibrations are uncorrelated there might be a $\sqrt{2}$ improvement in the time uncertainty, if proper use could be made of both the forward and backwards interpolations in a post-processing environment. Again, the question is resolved with the development of residuals, below.

Results with Traditional Stability Tools

We consider two quartz oscillator models: one is derived from the phase noise specification of the 5 MHz BVA Oscilloquartz 8600-3, and the second one is derived from a high-level 5 Mhz oscillator that would be regarded as three times worse as a local oscillator in a traditional frequency standard with a classical servo. Our analysis shows that the "worse" local oscillator *can* be the better choice, by a significant margin, for a pulsed atomic-ensemble frequency standard.

For a continuous Zacharias fountain with a flight time of 1 s (Ramsey linewidth $\Delta\nu$ of 0.5 Hz), operated with a conventional servo with square-wave frequency modulation with modulation frequency f_m just at the quasi-static limit ($f_m < \Delta\nu/2$ or 0.25 Hz). Choosing $f_m = 0.2$ Hz, the traditional stability limit of Equation 11 of reference [7] would be an asymptote of $1.52\times10^{-13}\ \tau^{-1/2}$ for the Oscilloquartz 8600-3, and $4.77\times10^{-13}\ \tau^{-1/2}$ for the high level 5 Mhz quartz oscillator. The simple analysis in Table I shows that the pulsed-mode asymptote is a more than a factor of 2 improvement for the 8600-3, and that a further 3-fold improvement is possible by using a "worse" 5 Mhz local oscillator. For the high-level oscillator it is worth recalling here that low-frequency noise from the strong 1/f character to $S_y(f)$ may have an N-sample variance where correlations over 1 s make it underestimate the variance from one second to the next.

Table I shows the individual terms of the noise models for the two quartz oscillators (expressed as phase noise as

Table I. N-sample Variance Analysis of Local Oscillator Degradation of Pulsed Atomic-ensemble Standards for two 5 MHz quartz oscillators, with dead time fraction $t_d/t_c = .01$, so $N = 100$.

Noise type:	Flicker frequency	Flicker Phase		White phase
Sy(f)	$h_{-1}\,f^{-1}$	$h_{+1}\,f$		$h_{+2}\,f^{2}$
$S\phi(f)$	$(2.5\times10^{13})h_{-1}f^{-3}$	$(2.5\times10^{13})h_{+1}\,f^{-1}$		$(2.5\times10^{13})h_{+2}$
$<\sigma_y^2(N,\tau,\tau)>$	$h_{-1}\,N\ln(N)/(N-1)$	$h_{+1}\,2(N+1)[\gamma+\ln(2\pi f_h\tau)-\ln(N)/(N^2-1)]/(N(2\pi\tau)^2)$		$h_{+2}2f_h(N+1)/(N(2\pi\tau)^2)$
Oscilloquartz 8600-3 5 MHz quartz oscillator specifications				
Sy(f)	$8\times10^{-26}\,f^{-1}$ +	$8\times10^{-27}\,f$ +		$5.6\times10^{-29}\,f^{2}$
$\sigma_y^2(N=100,1s,1s)$	3.71×10^{-25} +	1.93×10^{-23} +		2.87×10^{-23}
Eq. 5 $\sigma_x^2(t_d)$	3.71×10^{-27} +	1.93×10^{-25} +		$2.87\times10^{-25}\ s^2$
$\sigma_x(t_c)$ in s	6.95×10^{-14} @$t_c=1s$	5.95×10^{-14} @$t_c=0.5s$	8.10×10^{-14} @$t_c=2s$	6.37×10^{-13} $t_c=100s/f_h=10Hz$
$\sigma_y(\tau)$ asymptote	$6.95\times10^{-14}\tau^{-1/2}$	$1.19\times10^{-13}\tau^{-1/2}$	$4.05\times10^{-14}\tau^{-1/2}$	$6.37\times10^{-15}\tau^{-1/2}$
High Level 5 MHz quartz oscillator.				
Sy(f)	$8\times10^{-25}\,f^{-1}$ +	$8\times10^{-29}\,f$ +		$5.6\times10^{-32}\,f^{2}$
$\sigma_y^2(N=100,1s,1s)$	3.71×10^{-24} +	1.93×10^{-25} +		2.87×10^{-26}
Eq. 5 $\sigma_x^2(t_d)$	3.71×10^{-26} +	1.93×10^{-27} +		$2.87\times10^{-28}\ s^2$
$\sigma_x(t_c)$ in s	1.99×10^{-14}	1.05×10^{-14} @$t_c=0.5s$	3.89×10^{-14} @$t_c=2s$	1.93×10^{-12} @$t_c=100s$
$\sigma_y(\tau)$ asymptote	$1.99\times10^{-14}\tau^{-1/2}$	$2.10\times10^{-14}\tau^{-1/2}$	$1.94\times10^{-14}\tau^{-1/2}$	$1.93\times10^{-14}\tau^{-1/2}$

well as frequency noise), and the expected mean-square time uncertainty from each term. Also shown is the expected root-mean square time error per cycle, and the frequency-uncertainty asymptote: this asymptotic curve should be added in quadrature to the asymptotic frequency deviation of the random noise from interrogating the atomic reference.

Also shown in Table I are the asymptotic limits for Zacharias fountains 0.3 m tall ($t_c = 0.5$ s) and 4.9 m tall ($t_c = 2.0$ s) having the same dead time fraction of 0.01 and with f_h scaling with the atomic speed through the Ramsey cavity.

Also shown in Table I is the asymptotic limit for a system such as a linear ion trap with a cycle time of 100 s and a dead time fraction of 0.01. Here, where the local oscillator is being asked to carry the frequency averaged over 100 s back to 1 s, the Oscilloquartz 8600-3 may be seen to be the better choice, as we would expect.

The asymptotic limit for the local oscillator's contribution is particularly simple for local oscillators with predominantly flicker frequency noise in and around the frequency range from $1/t_c$ to $1/t_d$ (equivalent to a constant Allan deviation in the range t_d to t_c). The asymptotic form for large time intervals is

$$ {}^{p}\sigma_y^2(\tau) = \frac{\ln N}{N(N-1)2\ln 2}\,\sigma_y^2(t_c)\,t_c\,\tau^{-1} \qquad (6) $$

where $N = t_c/t_d$ and $\sigma_y^2(t_c)$ is the Allan variance of the local oscillator at the cycle time.

When this asymptotic limit for a pulsed atomic-ensemble timekeeper is compared to the continuous frequency standard case[] with square-wave frequency modulation at the quasi-static modulation frequency limit ($f_m < 1/(4t_c)$), for a local oscillator with predominantly flicker-frequency noise it can be shown that the advantage in the rms frequency deviation asymptote is

$$ N\left[\frac{N-1}{N\ln N}\right]^{1/2}, \qquad (7) $$

in favour of the pulsed system over the continuous system.

The above forms for a local oscillator with predominantly flicker frequency noise, and Table I's estimates of asymptotic limits of the contributions of noise from realistic local oscillators, are to say the least, very encouraging. These local oscillators could be used unmodified to make timekeeping systems for the $10^{-14}\,\tau^{-1/2}$ era.

Within the limitations discussed at length above, estimates can be made for the degraded performance generated by imperfect synchronization of the measurement periods with the Ramsey cycles, adding time error due to the local oscillator due to interpolating both near the beginning and near the end of the measurement period. We prefer to present these effects more rigorously, in terms of the rms residuals analysis developed below.

Estimating Local Oscillator Requirements Using Residuals

The above analysis convinced us that pulsed systems merited further, more rigorous, study. A better treatment of correlation effects across the cycle time was highly desirable.

Because of the multiple difficulties outlined above in trying to use the traditional N-sample variance of frequency to estimate the time errors in a pulsed atomic-ensemble frequency standard, we found it is necessary to develop statistical tools that unambiguously answer the question that we are asking: *for a pulsed atomic-ensemble frequency standard's time scale (constructed for a specific algorithm), at any particular moment in time what is the time uncertainty that is added by the local oscillator's phase noise?* At a given time in the cycle, the answer is found in the root-mean-square residual of the local oscillator's difference of phase relative to the best estimate of the phase formed from some time scale algorithm. The mean is to be taken over many cycles, and is not averaged across the Ramsey cycle. Like the stability-measure statistical tools such as the Allan deviation, this one does not necessarily have to be measured directly. A conventional measurement of the phase noise of the local oscillator will prove sufficient to allow calculation of the expected residual.

The rms residual serves as an aid to pulsed atomic-ensemble frequency standard design: to optimise the choice of local oscillator from those available, to understand the effects of varying the cycle time and the dead time, to permit optimal combination of the time scales of multiple pulsed atomic-ensemble frequency standards and to permit quantitative design of phase comparators that minimize the effects of the nonuniform uncertainty of the time

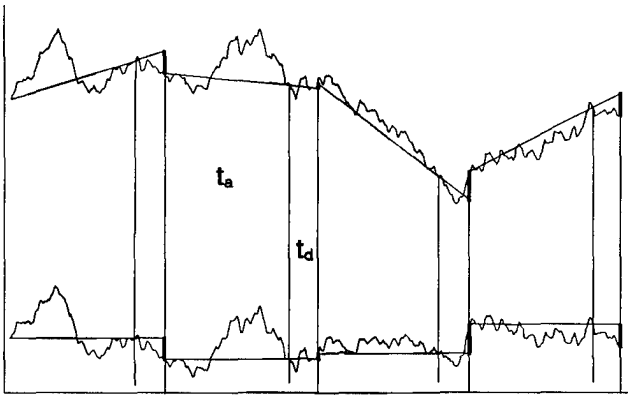


Figure 3. Local Oscillator Phase Noise Measured over Intervals t_a and Extrapolated over Intervals t_d. Top curve: raw oscillator phase and measured slopes (over t_a). Bottom curve: post-processed time scale. The heavy vertical lines are the residual phase errors of the local oscillator over the preceeding dead time, having an rms of $RESID_1\ \sigma_y(t_a,t_d)$.

scale due to the uncorrected local oscillator phase noise. The time or phase residual is a particularly convenient tool for the practical design of pulsed atomic frequency standards and ancillary equipment.

Residuals from Two Calibration Intervals

The residual is to be determined for a specific time τ after the completion of a calibration interval - the active Ramsey cycle (time interval t_a). The ith time scale algorithm is used to construct the estimate $x_{Ei}(\tau)$ of the local oscillator phase at time τ, using local oscillator frequency calibrations from the nearby Ramsey intervals.

$$RESID_i\ \sigma_x^2(t_a,\tau) = < \left[x(\tau) - x_{Ei}(\tau)\right]^2 > \qquad (8)$$

To evaluate the noise added by the local oscillator, we consider the mean square time residual (from a linear extrapolation of x) at a time τ after the (one) most recent Ramsey cycle (period t_c, dead time t_d, active time $t_a = t_c - t_d$)

$$RESID_1\ \sigma_x^2(t_a,\tau) =$$

$$< \left[x(\tau) - \left\{x(0) + \frac{(x(0) - x(-t_a))}{t_a}\tau\right\}\right]^2 > . \qquad (9.a)$$

The utility of this residual for our problem is apparent if we consider Figure 3. For each dead time, a linear extrapolation is made of the phase from the last active time t_a which calibrated the frequency of the local oscillator. However, during each dead time the local oscillator departs from this extrapolated prediction by an amount indicated by the vertical step at the end of each dead time, as shown on the lower curve of Figure 3. The rms size of these steps is given by $RESID_1\ \sigma_y^2(t_a,t_d)$. After post-processing to remove the (measured) slope at each interval, these same vertical steps are still present, as shown in the upper curve of Figure 3, and still have an rms size of $RESID_1\ \sigma_y(t_a,t_d)$.

Equation 9.a would become equation 9.c below, if the "sampling time" for determining the phase is infinitely short. More generally, where g(t) is the frequency sensitivity function as defined by Dick et al[4], this would generalize to two different frequency sensitivity functions $g_1(t)$ and $g_2(t)$ for these two intervals, and to

$$RESID_1\ \sigma_x^2(t_a,\tau) =$$

$$= \tau^2 < \left[\frac{\int\{1-g_1(t')\}y(t')dt'}{\int\{1-g_1(t')\}dt'} - \frac{\int g_2(t)y(t)dt}{\int g_2(t)dt}\right]^2 > \qquad (9.b)$$

where for the first term the integrations are taken over a bit more than the interval $[0,\tau]$, and represent the accumulation of the local oscillator error unmeasured by the atomic reference - starting with the fall of $g_1(t)$ near the end of a cycle ($t = 0$) - and extending towards τ where the rise time of $g_1(t)$ will be defined by the particulars of the phase comparator being considered for the measurement of $x(\tau)$. A useful simplification results if $x(\tau)$ is thought of as being measured with the same sensitivity slope $g_1'(\tau+\delta)$ as $-g_1'(0+\delta)$, and there will be then only one function g'(t) determining the rise and fall characteristics of the g_1 weights applied to y(t). This will naturally be the case when similar standards, or multiple pulse trains through the same standard are being compared, and will overestimate (if only slightly) the effects of local oscillator noise on systems that average the phase over longer intervals than the assumed risetime. For the second term the integration is over the calibration interval: a bit more than $[-t_a,0]$. With a single interaction time t_a for all atoms of the ensemble, the rise and fall times of g_2 will be equal (and will be equal to those of g_1). If the interaction time cannot be kept the same for all atoms, an integration over the distribution of interaction times could be done.

Where the four rise and fall times discussed above can

Table II. RMS Residual Analysis of Local Oscillator Degradation of a Pulsed Cesium Fountain's Stability using $RESID_1$ with cycle time $t_c = 1s$, and dead time $t_d = .01t_c$.

Noise type:	Flicker frequency	Flicker Phase		White phase
Oscilloquartz 8600-3 5 MHz quartz oscillator specifications				
Sy(f)	$8x10^{-26}\ f^{-1}$ +	$8x10^{-27}\ f$ +		$5.6x10^{-29}\ f^2$
Eq. A.8 $\sigma_x^2(t_d)$	$4.53x10^{-29}$ +	$1.93x10^{-27}$ +		$2.87x10^{-27}\ s^2$
$\sigma_x(t_c)$in s	$6.96x10^{-14}@t_c=1s$	$5.97x10^{-14}@t_c=0.5s$	$8.10x10^{-14}@t_c=2s$	$6.74x10^{-13}@t_c=100s$
$\sigma_y(\tau)$ asymptote	$6.96x10^{-14}\tau^{-1/2}$	$1.19x10^{-13}\tau^{-1/2}$	$4.05x10^{-14}\tau^{-1/2}$	$6.74x10^{-15}\tau^{-1/2}$
High Level 5 MHz quartz oscillator.				
Sy(f)	$8x10^{-25}\ f^{-1}$ +	$8x10^{-29}\ f$ +		$5.6x10^{-32}\ f^2$
Eq. A.8 $\sigma_x^2(t_d)$	$4.53x10^{-28}$ +	$1.93x10^{-29}$ +		$2.87x10^{-30}\ s^2$
$\sigma_x(t_c)$ in s	$2.18x10^{-14}@t_c=1s$	$1.14x10^{-14}@t_c=0.5s$	$4.29x10^{-14}@t_c=2s$	$2.12x10^{-12}@t_c=100s$
$\sigma_y(\tau)$ asymptote	$2.18x10^{-14}\tau^{-1/2}$	$2.28x10^{-14}\tau^{-1/2}$	$2.15x10^{-14}\tau^{-1/2}$	$2.12x10^{-14}\tau^{-1/2}$

be considered equal, a significant simplification results. The slower rise and fall times effectively smooth y(t), preferentially removing high-frequency noise in y(t). An approximation of this is that used in the standard form derived for the N-sample variance (Barnes et al.) where this rolloff is approximated by a high-frequency cutoff f_h.

$$RESID_1\ \sigma_x^2(t_a,\tau) =$$

$$= \tau^2 < \left[\frac{1}{\tau}\int_0^{\tau} y(t')dt' - \frac{1}{t_a}\int_{-t_a}^{0} y(t)\,dt \right]^2 > \qquad (9.c)$$

which can be expressed in terms of the local oscillator's spectral density of frequency fluctuations Sy(f), as given in equation A.5 of appendix A, derived in a manner that is highly reminiscent of the derivation for the N-sample variance [10].

We note that the extrapolated ($\tau > 0$) and interpolated ($\tau < 0$) residuals are related by

$$RESID_1\ \sigma_x^2(t_a,\tau) = \left\{ \frac{(t_a+\tau)^2}{t_a^2} \right\} RESID_1\ \sigma_x^2(t_a+\tau,-\tau) \qquad (10)$$

and that interpolation is symmetric about the interval's midpoint

$$RESID_1\ \sigma_x^2(t_a,-\frac{t_a}{2}+t) = RESID_1\ \sigma_x^2(t_a,-\frac{t_a}{2}-t) \qquad (11)$$

and that, at the interval's midpoint ($t_a/2$), this residual has a particularly familiar form

$$RESID_1\ \sigma_x^2(t_a,\frac{t_a}{2}) = \frac{1}{2}\left(\frac{t_a}{2}\right)^2 \sigma_y^2(\frac{t_a}{2}) \qquad (12)$$

where $\sigma_y^2(t_a/2)$ is the Allan variance at the half-interval time. This expression may be compared with our earlier estimate of equation 3, with the deviation reduced by $\sqrt{2}$ to account for the two largely independent extrapolations possible (forward or backwards in time). The residual estimated in this way with equation 3 may be seen to be wrong by the ratio $<\sigma_y^2(2,t_a,t_a)>/\sigma_y^2(t_a/2)$, or $\sigma_y^2(t_a)/\sigma_y^2(t_a/2)$. For many local oscillators this ratio will be 1 for the value chosen for t_a.

In Appendices A, B and C, we develop and present fully the method for calculating the residuals from a single-interval calibration, $RESID_1$, for the usual power law spectral density of frequency fluctuations $S_y(f)$. The explicit expressions for $RESID_1\ \sigma_y^2(t_a,\tau)$ for the five common power-laws that contribute to $S_y(f)$ are given in Appendix A as equations A.8a through A.8e. Appendix B derives a generalized two-interval covariance, and applies it to the cases at hand. Appendix C gives the integrals required for the different power laws, as well as numerical approximations and transformations that are useful in simplifying the calculation of these integrals.

The results of the residuals calculations are presented in Table II. As expected, when dominated by flicker frequency noise, a local oscillator is seen not to perform as well as the overly optimistic N-sample variance would predict, due to the correlations within a 1 s interval. The

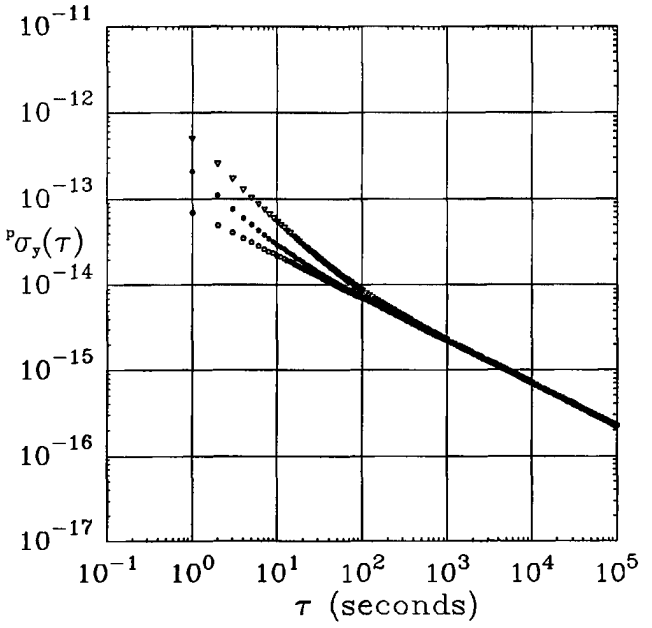


Figure 4. The predicted deviation of the average frequency over an interval τ expected due to an Oscilloquartz 8600-3 5 MHz local oscillator in a Cs fountain with a cycle time of 1 s, a dead time of 0.01 s per cycle and $3x10^6$ atoms per cycle. The three curves are described in the text: a well-designed system can access the lowest curve.

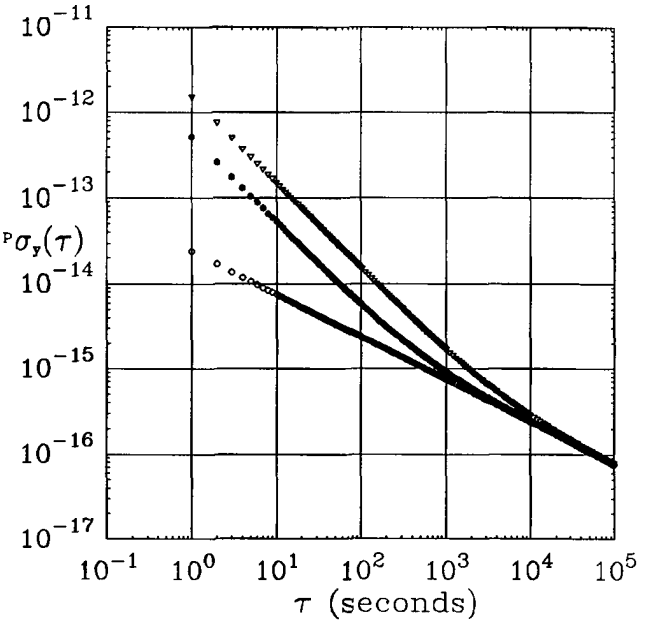


Figure 5. The predicted frequency deviation over an interval τ due to a high-level quartz 5 MHz local oscillator in a Cs fountain like that for Fig. 4. Although this oscillator would commonly be considered a worse local oscillator than that of Fig. 4, it is in fact better for our proposed mode of operation.

N-sample variance is seen to have predicted the variances correctly except for the flicker frequency noise. The size of the underestimate with the N-sample variance is less than 20% (<10% for deviations) for flicker frequency noise. The rms residual is handling correlations correctly within the 1 s cycle, and it is only in the cycle-to-cycle variation that correlations could still trouble us. The low frequencies of Sy(f) are effectively filtered by the linear fitting process, so the correlations they induce are smaller still, and here we do not consider them further.

The mean frequency deviation due to the uncompensated fluctuations of the local oscillator, over a time interval τ is

$$^{p}\sigma_y(\tau) = \frac{1}{\tau}\left[\begin{array}{l}\left\{A^2+RESID_i\sigma_x^2(t_a,t_d)\right\}n \; + \\ RESID_i\sigma_x^2(t_a,-t_1)+RESID_i\sigma_x^2(t_a,-t_2)\end{array}\right]^{1/2} \quad (13)$$

where $\tau = t_1 + n\,t_c + t_2$: n fountain cycles, plus time intervals t_1 ($<t_c$) at the beginning of τ, and t_2 (also $< t_c$) at the end of τ, the minus signs for t_1 and t_2 indicate, in the notation developed in the Appendices, that the residual is being calculated by interpolation: the interpolative interval lies within the interval t_a used for the frequency calibration. A is the limit from interrogation of the atomic resonance, arbitrarily modeled for most of our presentation below as $A = 10^{-14}t_c^{1/2}$, (i.e. a $\sigma(\tau)$ of $10^{-14}\tau^{-1/2}$ or $3x10^6$ atoms detected per cycle in a cesium Zacharias fountain operated at the shot noise limit with a 1 second cycle time, and a 0.01 second dead time per cycle). Clearly, to make best use of a pulsed frequency standard for short times intervals τ, we must consider the role of the position of the interval's start and stop position within the cycle. For example, the design of a phase comparator to calibrate a bank of H masers must consider when the phase comparisons should be made. The best and worst performance on a $^{p}\sigma_y(\tau)$ curve, with this scheme of post-processing, will occur for time intervals that are multiples of the cycle time.

In Figures 4 and 5, the best behaviour of the two quartz oscillators may be compared, as calculated with $RESID_1$ $\sigma_y(t_a,t_d)$. This would be for a post-processed system being measured just after the dead time. The "worst" behaviour,

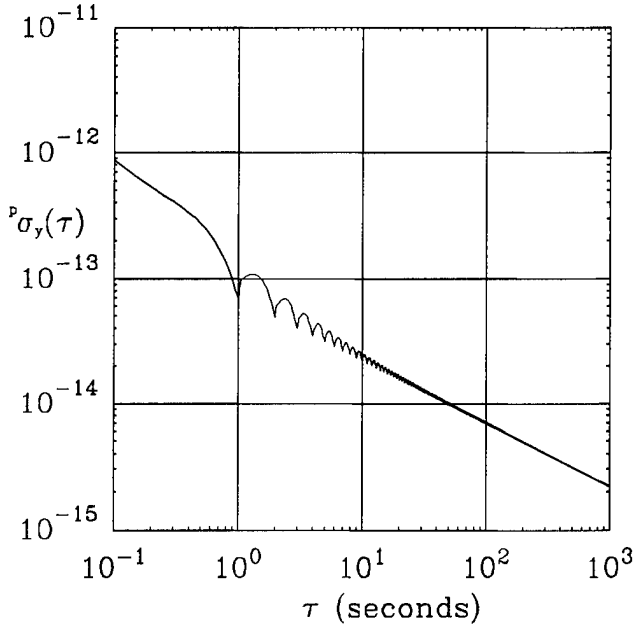


Figure 6. The predicted deviation of the average frequency over an interval τ expected due to an Oscilloquartz 8600-3 5 MHz local oscillator in a Cs fountain like that for Fig. 4. The interval is synchronised to start or stop at the end of a dead time.

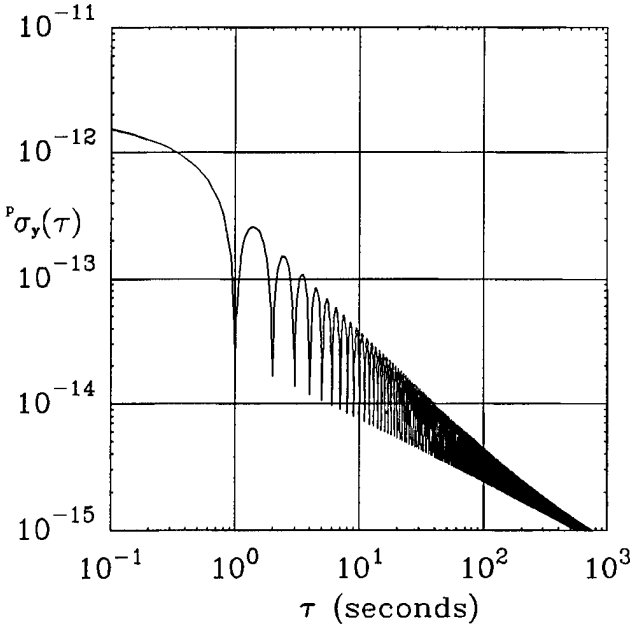


Figure 7. The predicted frequency deviation over an interval τ due to a high-level quartz 5 MHz local oscillator in a Cs fountain like that for Fig. 4. The interval is synchronised to start or stop at the end of a dead time.

shown as the top curve in each graph, has effectively no post-processing. They model a full cycle of extrapolation both at the start and at the end of the interval. With a microstepper-like phase step correction after each cycle, it should be possible to obtain these curves from a pulsed atomic-ensemble system operated as a traditional frequency standard with an only slightly unconventional servo. We see no point in doing so, since the intermediate curve shows the worst case performance for the simplest form of post processing (delaying by 1 second a knowledge of the post-processed phase of the standard). This postprocessing worst case is for time intervals equal to a multiple of the cycle time, but phased to start and stop half-way through the cycle. Post-processing is used to interpolate over the full calibration cycle t_a which is still incomplete when the start or stop phase measurement for the interval τ is called for. The advantage of post-processing is clear for shorter times, but these examples of one-calibration-interval processing have the same asymptote.

In Figures 6 and 7 the variation of the predicted deviation with the length of the time interval not restricted to be a multiple of the cycle time is shown for single-calibration interval post-processing of the two quartz local oscillators. The start (or, equivalently the stop) of the interval is synchronized with the end of a dead time. The curves just fit under the envelope of the middle curves of Figures 4 and 5 and above the asymptote.

Residuals from Two Calibration Intervals

If the average frequency from the two nearest frequency calibration intervals (each of length t_a) is to be used for establishing the local oscillator's frequency over the dead time td between the two calibration intervals, then the equal-weight average of the two calibrations will be the optimal average. If the frequencies of the two calibrations are highly correlated, then the two-sample frequency average will give only a small improvement in the time residual over the dead time. The two-sample frequency average will reduce the residuals due to uncorrelated frequency fluctuations of the local oscillator, and so could give up to a $\sqrt{2}$ improvement in the performance of some local oscillators. Thus we consider the residuals from the equal-weight two-sample frequency average.

Table III. RMS Residual Analysis of Local Oscillator Degradation of a Pulsed Cesium Fountain's Stability using $RESID_2$ with cycle time $t_c = 1s$, and dead time $t_d = .01t_c$.

Noise type:	Flicker frequency	Flicker Phase		White phase
Oscilloquartz 8600-3 5 MHz quartz oscillator specifications				
Sy(f)	$8x10^{-26} f^{-1}$ +	$8x10^{-27} f$ +		$5.6x10^{-29} f^2$
Eq. A.8 $\sigma_x^2(t_d)$	$3.97x10^{-29}$ +	$1.93x10^{-27}$ +		$2.87x10^{-27}$
$\sigma_x(t_c)$in s	$6.95x10^{-14}$ @$t_c = 1s$	$5.97x10^{-14}$ @$t_c = 0.5s$	$8.09x10^{-14}$ @$t_c = 2s$	$6.31x10^{-13}$ @$t_c = 100s$
$\sigma_y(\tau)$ asymptote	$6.95x10^{-14}\tau^{-1/2}$	$1.19x10^{-14}\tau^{-1/2}$	$4.04x10^{-14}\tau^{-1/2}$	$6.31x10^{-15}\tau^{-1/2}$
High Level 5 MHz quartz oscillator.				
Sy(f)	$8x10^{-25} f^{-1}$ +	$8x10^{-29} f$ +		$5.6x10^{-32} f^2$
Eq. A.8 $\sigma_x^2(t_d)$	$3.97x10^{-28}$ +	$1.93x10^{-29}$ +		$2.87x10^{-30}$
$\sigma_x(t_c)$ in s	$2.05x10^{-14}$ @$t_c = 1s$	$1.08x10^{-14}$ @$t_c = 0.5s$	$4.02x10^{-13}$ @$t_c = 2s$	$1.99x10^{-12}$ @$t_c = 100s$
$\sigma_y(\tau)$ asymptote	$2.05x10^{-14}\tau^{-1/2}$	$2.04x10^{-14}\tau^{-1/2}$	$2.01x10^{-14}\tau^{-1/2}$	$1.99x10^{-14}\tau^{-1/2}$

$$RESID_2\ \sigma_x^2(t_a,\tau) =$$

$$< \left[x(\tau) - \left\{x(0) + \left(\frac{x(0)-x(-t_a)}{t_a} + \frac{x(t_d+t_a)-x(t_d)}{t_a}\right)\frac{\tau}{2}\right\}\right]^2 >$$

$$= \tau^2 < \left[\frac{1}{\tau}\int_0^\tau y(t')dt' - \frac{1}{2t_a}\left\{\int_{-t_a}^{0} y(t)\,dt + \int_{t_d}^{t_d+t_a} y(t)\,dt\right\}\right]^2 > \quad (14)$$

which can also be expressed in terms of the local oscillator's spectral density of frequency fluctuations $S_y(f)$, as developed in Appendix D, and given in equation D.6.

Table III shows the expected rms residuals using both adjacent calibration intervals t_a, for our example local oscillators. $RESID_2$ shows an improvement in asymptotic performance of the deviation of up to 7% over $RESID_1$, for a dead time fraction of 0.01. The improvement increases to about 30% as the dead time fraction increases towards 0.90. By comparing Table III with Table II, it may be seen that of the three noise terms, only the flicker frequency noise has a reduced contribution to the residual when two active intervals are used.

More Improvements

Going beyond a linear fit for the residuals will introduce higher-order autocorrelation functions into a treatment using $S_y(f)$, which we wish to avoid. Even so, this is not the end of the possible reduction of local oscillator effects. If another similar pulsed-ensemble atomic frequency standard is operating from the same local oscillator, but delayed by $t_c/2$, then the frequency calibration interval can be reduced by a factor of two to $\pm\ t_c/2$. Shorter values still of t_c might be used, with a pulsed frequency standard optimized to assist the local oscillator calibration over the dead time. At the present time we do not see the need to do this type of work on atomic resonance assisted local oscillators for pulsed atomic-ensemble frequency standards, currently the 80 K sapphire frequency discriminator[11] seems to answer many Zacharias-fountain local-oscillator dream requirements, with a $RESID_2$ asymptote of $7x10^{-16}\ \tau^{-1/2}$, an almost 30-fold improvement over the high-level 5 MHz quartz oscillator considered in Table III.

Perhaps more important to discuss briefly is the case of multiple identical pulsed-ensemble frequency standards. Since a pulsed Zacharias fountain might be able to support a large number m ($m < 100$) of simultaneous atomic-ensemble pulses without overlap, it is interesting to consider a formalism for optimally creating a time scale from a large number of quasi-independent overlapping frequency measurements of the local oscillator. If the characteristics of the local oscillator are known, the $RESID_1\ \sigma_x^2(t_a,\tau)$ or $RESID_2\ \sigma_x^2(t_a,\tau)$ for each of the quasi-independent pulse sequences could be calculated to obtain the optimal statistical weights at each time through the cycle for an initial weighted average of the quasi-independent time scales. The improved estimate of the local oscillator frequency through each dead time

could then be incorporated into the next iteration of the average time scale. To the extent that the remaining frequency errors in each dead time are correlated, they should be removable in such an iterative process, and to the extent that they are not correlated they should average out: therefore from a fountain tossing m atomic ensembles per cycle time, we might hope to achieve a further $m^{1/2}$-fold improvement in the short-term frequency performance *without requiring a higher performance from the local oscillator.*

Difficulties of Implementation

The pulsed atomic-ensemble timekeeper, with post-processing largely replacing the servo, appears very attractive compared to traditional frequency standard systems. Nonetheless, it does not provide a complete solution for all local oscillator requirements. For evaluation of most systematic effects and for stability measurements either a highly superior reference frequency standard or a second timekeeper will be required.

Although the requirements for measuring the Ramsey phase differences have been discussed, we have not yet discussed the requirements on establishing t_a and t_d accurately. If the offset of the local oscillator is relatively large (10^{-11}, say), then if there is an error of $-\varepsilon$ in the establishment of t_a (and a corresponding error of ε in t_d) there will be an error in the frequency calibration of $10^{-11}\ 1/(1-\varepsilon/t_a)$, and a time error over each cycle of $10^{-11}\ \varepsilon\ [1+t_d/t_a]$, or a frequency error of $10^{-11}\ \varepsilon/t_c\ [1+t_d/t_a]$. So, to achieve a 10^{-16} accuracy in frequency in the face of a systematic local oscillator offset of 10^{-11}, one has to establish t_a to a fractional accuracy of 10^{-5}, or equivalently to establish t_d to an accuracy of 10^{-3} in our example fountain: this corresponds to knowing the centre of the microwave Ramsey interaction to an accuracy of 0.5%. The accurate definitions of t_a and t_d are given by the g(t) functions used in Equation 7.b, but clearly at this level spatial assymetries due to minor differences in construction of the microwave cavity or temporal power fluctuations could be significant. All of these concerns can be most easily managed merely by steering the local oscillator so that its long-term frequency offset is more nearly zero (10^{-13} should not be difficult).

Similarly, the microwave power in the cavity must be maintained near the optimum level if the frequency offset of the local oscillator is large, since otherwise the Ramsey phase discriminator's background and slope will be altered and the phase calibration will be wrong. While means can (and probably would) be invoked to measure the microwave power on-line, the easiest defence is once again to keep the local oscillator's offset small.

In any measurement system designed to compare a pulsed atomic-ensemble frequency standard with another source, the question of the limited duty cycle for such high precision phase comparison must be considered. A synchronized, windowed phase comparator or other measurement system is likely needed, which could have a duty cycle in the 0.2% to 1% range. Based on our experience with 5 MHz phase comparators (using a 500 Hz offset oscillator, we can obtain phase noise of 0.2 ps by averaging for 1 second), we would anticipate that a 100 MHz phase comparator could give 10^{-14} s noise averaging for "1 second" (spread over .01 s intervals in 100 successive seconds), and that a similar X-band phase comparator could give this level of performance in a time interval of 0.01 s.

Coherent noise and bright lines can give rise to noise which our analysis above will not be able to predict. In our example, where $\pm\pi/2$ phase modulation is proposed with 1 second toss time, the modulation cycle time is 2 seconds, and some of the local oscillator noise (a fraction t_d/t_c of it) at that frequency (0.5 Hz and its harmonics) will get through to perturb the output frequency produced by such a timekeeping system. The analysis correctly accounts for all uncorrelated local oscillator noise, but any bright lines at a harmonic of the modulation frequency can cause trouble. This will likely mean that the example of a system with exactly a 1 s cycle time could be highly unsuitable for use near a time lab which has so many signals modulated at just these harmonics!

Conclusions

Requirements for local oscillator stability are reduced for pulsed systems over continuous systems by a factor $N\ [\ (N-1)/(N \ln N)\ ]^{1/2}$, where N is the ratio of cycle time to dead time, and the local oscillator noise is predominantly flicker frequency. This result is also found to apply to the case of postprocessing the two adjacent calibration intervals.

Provided that the dead time can be kept to 1%, current quartz oscillators are good enough to give asymptotic performance in or below the low $10^{-14}\,\tau^{-1/2}$ range for a cesium Zacharias fountain, and for linear ion traps. Two avenues for further improvements (80 K sapphire X-band frequency discriminator[11], and a large number of overlapping cycles) have been identified which might be considered in the quest to reach below the level of $10^{-15}\,\tau^{-1/2}$.

The Ramsey fringes can be used as a calibrated phase measuring device, and its phase sensitivity in the shot noise limit has been presented: by measuring the number of atoms in both states a $\sqrt{2}$ reduction in phase noise is possible and the phase noise can be made uniform across the range of $\pm\pi/2$.

We have shown that the potentially difficult problem of the frequency control servo could be sidestepped by adopting a timekeeping paradigm rather than a strict frequency control one.

The postprocessing done by timekeeping has been shown to have advantages that we have quantified for the short-term worst-case uncertainty and a more modest improvement in asymptote. The required phase modulation and processing seem straightforward, and even the need for calibrating the Ramsey phase detector can be simplified by a not-too-stringent requirement on the local oscillator's medium-term average frequency offset (10^{-11} to 10^{-13} for a Cs fountain). The frequency averaged over one cycle (1 s) must also have a small probability of a cycle slip, which can be met easily for a Cs fountain with a 1s accuracy of 10^{-11}.

The concept of a local oscillator's rms residual from a linear fit to adjacent frequency calibration cycles has been developed and the results presented for the common power-law spectral distributions of local oscillator frequency noise. Forms that permit straightforward computation of the residuals exist. Low-frequency divergences are avoided in this problem. The results have been used to check the values derived with the N-sample variance where the results should be comparable. The details of these useful forms have been presented. We have touched on the use of the residual in designing phase comparators, and its role in planning to compare with other classical frequency standards, or with a second pulsed atomic-ensemble system.

Acknowledgement

It is a pleasure to acknowledge helpful conversations with D. Morris and J. Vanier.

References

[1] "Linear Ion Trap based Atomic Frequency Standard", J.D. Prestage, G.J. Dick and L. Maleki, *Proc. 44th Ann. Symp. on Frequency Control*, 82-88, 1990.

[2] "Ramsey Resonance in a Zacharias Fountain", A. Clairon, C. Salomon, S. Guellati and W.D. Phillips, Europhys. Lett. **16**, 165-170, 1991.

[3] "Comparative Study of Time Scale Algorithms", P Tavella and C. Thomas, *Metrologia* **28**, 57-63, 1991.

[4] "Local Oscillator Induced Degradation of Medium-term Stability in Passive Atomic Frequency Standards", G.J. Dick, J.D. Prestage, C.A. Greenhall and L. Maleki, *Proc. 22nd Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 487-508 (1990).

[5] "Observation of the Cesium Clock Transition Using Laser-cooled Atoms in a Vapor Cell", C. Monroe, H. Robinson and C. Wieman, *Optics Lett.* **16**, 50-52, 1991.

[6] "Improved Magneto-optic Trapping in a Vapor Cell", K.E. Gibble, S. Kasapi and S. Chu, *Optics Lett.* **17**, 526-528, 1992.

[7] "A Limit to the Frequency Stability of Passive Frequency Standards", C. Audoin, V. Candelier and N. Dimarcq, *IEEE Trans. Instrum. Meas.* **IM-40**, 121-125, 1991.

[8] "A 303-MHz Frequency Standard Based on Trapped Be+ Ions", J.J. Bollinger, D.J. Heinzen, W.M. Itano, S.L. Gilbert and D.J. Wineland, *IEEE Trans. Instrum. Meas.* **IM-40**, 126-128, 1991.

[9] "Calculation of Trapped ion Local Oscillator Requirements", G.J. Dick, *Proc. 19th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 133-147 (1988).

[10] "Characterization of Frequency Stability", 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, *IEEE Trans. Instrum. Meas.* **IM-20**, 105-120, 1971.

[11] "Microwave Frequency Discriminator With a Cooled Sapphire Resonator for Ultra-low Phase Noise", D.G. Santiago and G.J. Dick, *Proc. 46th Ann. Symp. on Frequency Control*, 1992.

[12] Handbook of Mathematical Functions, M. Abramowitz and I. Stegun, 9th printing by Dover Publications, New York

Appendix A

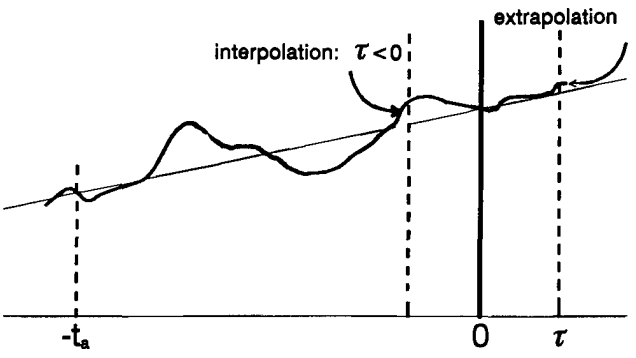


Figure A.1. Calculating the RMS Residual Relative to a Linear Estimate

Referring to Figure A.1 , the residual error at time τ, due to extrapolation (for τ greater than zero) or interpolation (for τ smaller than zero) from a calibrating phase difference measurement of an oscillator made over the time interval [-t_a to 0] is called $RESID\ 1\ \sigma_x^2(t_a,\tau)$, and is defined as

$$< \left[x(\tau) - \left\{ x(0) + \frac{\left(x(0) - x(t_a)\right)}{t_a} \right\} \tau \right]^2 > \quad \text{(A.1)}$$

where $x(\tau)$ is the value of the phase at time τ, $x(t_a)$ is the value of the phase at $-t_a$, time at which the active cycle begins, ending at time 0 , where the phase is $x(0)$.

This expression can be expanded :

$$< \left(x(\tau) - x(0)\right)^2 + \frac{\tau^2}{t_a^2}\left(x(0) - x(t_a)\right)^2 - \frac{2\tau}{t_a}\left(x(\tau) - x(0)\right)\left(x(\tau) - x(t_a)\right) > \quad \text{(A.2)}$$

Since $x(t_b) - x(t_a) = (t_b - t_a)\left\{ \frac{1}{t_b - t_a} \int_{t_a}^{t_b} y(t)\,dt \right\}$ where the term in brackets is the average frequency over the time interval under consideration, Eq. A.2 can be rewritten as

$$< \left(\int_0^\tau y(t)\,dt \right)^2 + \frac{\tau^2}{t_a^2}\left(\int_{-t_a}^{\tau} y(t)\,dt \right)^2 - \frac{2\tau}{t_a}\left(\int_0^\tau y(t)\,dt \int_{-t_a}^{\tau} y(t)\,dt \right) > \quad \text{(A.3)}$$

or

$$\begin{aligned} RESID\ 1\ \sigma_x^2(t_a,\tau) = {} & \tau^2\left\{ \frac{1}{\tau^2} \int_0^\tau \int_0^\tau < y(t')\,y(t)\,dt'\,dt > \right\} \\ & + \tau^2\left\{ \frac{1}{t_a^2} \int_{-t_a}^{\tau} \int_{-t_a}^{\tau} < y(t')\,y(t)\,dt'\,dt > \right\} \\ & - 2\tau^2\left\{ \frac{1}{\tau t_a} \int_0^\tau \int_{-t_a}^{\tau} < y(t')\,y(t)\,dt'\,dt > \right\} \end{aligned} \quad \text{(A.4)}$$

We can see that the residual is zero if we are at one end of the active, or calibrating measurement period. With the help of the transformations calculated in Appendix B, the previous equation can be rewritten in terms of the spectral density of the noise.

$$\begin{aligned} \frac{2}{2\pi} \int_0^\infty \frac{S_y(\omega)}{\omega^2} \Big[& \left\{ 1 - \cos(\omega\tau) \right\} + \frac{\tau^2}{t_a^2}\left\{ 1 - \cos(\omega t_a) \right\} \\ & + \frac{2\tau}{t_a}\left\{ 1 - \cos(\omega t_a) - \cos(\omega\tau) + \cos(\omega(t_a + \tau)) \right\} \Big]\, d\omega \end{aligned} \quad \text{(A.5)}$$

In order to evaluate this equation, the terms in brackets are reorganised in the following way:

$$\frac{2}{2\pi}\int_0^{\infty}\frac{S_y(\omega)}{\omega^2}\Big[\,(1+\tfrac{\tau}{t_a})\,\{1-\cos(\omega\tau)\}$$
$$+\tfrac{\tau}{t_a}(1+\tfrac{\tau}{t_a})\,\{1-\cos(\omega t_a)\} \qquad (A.6)$$
$$-\tfrac{\tau}{t_a}\{1-\cos(\omega(t_a+\tau)\}\Big]\,d\omega$$

Using the conventional definition[10] for the coefficients h_α of spectral density of frequency noise, and changing variable from f to $\omega=2\pi f$, we get

$$S_y(\omega)=\sum_{\alpha=-2}^{\alpha=2}\frac{h_\alpha\omega^\alpha}{(2\pi)^\alpha}\ ,$$

We can rewrite Eq. A.6, reordering the summation and integral:

$$\sum_{\alpha=-2}^{\alpha=2}\frac{2h_\alpha}{(2\pi)^{\alpha+1}}\int_0^{\infty}\omega^{\alpha-2}\Big[\,(1+\tfrac{t}{t_a})\,\{1-\cos(\omega t)\}$$
$$+\tfrac{t}{t_a}(1+\tfrac{t}{t_a})\,\{1-\cos(\omega t_a)\} \qquad (A.7)$$
$$-\tfrac{t}{t_a}\{1-\cos(\omega(t_a+t)\}\Big]\,d\omega$$

Using the results of Appendix C, where we evaluate the integral part of the previous expression:

$$I(\alpha,\omega_h,t)\ =\ \frac{2h_\alpha}{(2\pi)^{\alpha+1}}\int_0^{\omega_h}\omega^{\alpha-2}\{\,(1-\cos(\omega t))\,\}\,d\omega$$

Eq. A.7 can then be expressed in terms of this general function.

$$RESID\,1\,\sigma_x^2(t_a,\tau)\ =\ \sum_{\alpha=-2}^{\alpha=2}\Big[\,(1+\tfrac{\tau}{t_a})\,I(\alpha,\omega,\tau)$$
$$+\tfrac{\tau}{t_a}(1+\tfrac{\tau}{t_a})\,I(\alpha,\omega,t_a) \qquad (A.8)$$
$$-\tfrac{\tau}{t_a}I(\alpha,\omega,t_a+\tau)\,\Big]$$

Using the results of Appendix C, and assuming a sharp high frequency cutoff f_h such that $\omega = 2\pi f_h$ and a low frequency cutoff of $\varepsilon = 0$ we can write the expression for each α value from -2 to 2.

For white phase noise ($\alpha = 2$), using Eq. C.3a and after some reordering of the terms, Eq. A.8 reduces to

$$RESID\,1\,\sigma_x^2(t_a,\tau)_{\alpha=2}\ =\ \frac{2h_2}{(2\pi)^3}\omega\Big[\,(1+\tfrac{\tau}{t_a})\,\{1-\tfrac{\sin(\omega\tau)}{(\omega\tau)}\}$$
$$+\tfrac{\tau}{t_a}(1+\tfrac{\tau}{t_a})\,\{1-\tfrac{\sin(\omega t_a)}{(\omega t_a)}\} \qquad (A.8a)$$
$$-\tfrac{\tau}{t_a}\{1-\tfrac{\sin(\omega(t_a+\tau))}{(\omega(t_a+\tau))}\}\,\Big]$$

For flicker phase noise ($\alpha = 1$), using Eq. C.3b, Eq. A.8 reduces to

$$RESID\ 1\ \sigma_x^2(t_a,\tau)_{\alpha=1} = \frac{2h_1}{(2\pi)^2}\left[(1+\frac{\tau}{t_a})\left\{\gamma+\ln(\omega\tau)-C_i(\omega\tau)\right\}\right.$$
$$+\frac{\tau}{t_a}(1+\frac{\tau}{t_a})\left\{\gamma+\ln(\omega t_a)-C_i(\omega t_a)\right\} \qquad \text{(A.8b)}$$
$$\left.-\frac{\tau}{t_a}\left\{\gamma+\ln(\omega(t_a+\tau))-C_i(\omega(t_a+\tau))\right\}\right]$$

To calculate the cosine integral $C_i(x)$, see the end of Appendix C.

For white frequency noise ($\alpha = 0$), using Eq. C.3c and after some reordering of the terms, Eq. A.8 reduces to

$$RESID\ 1\ \sigma_x^2(t_a,\tau)_{\alpha=0} = \frac{2h_0}{2\pi}\tau\,(1+\frac{\tau}{t_a})\left[S_i(\omega\tau)-\frac{1-\cos(\omega\tau)}{\omega\tau}\right.$$
$$+S_i(\omega t_a)-\frac{1-\cos(\omega t_a)}{\omega t_a} \qquad \text{(A.8c)}$$
$$\left.-S_i(\omega(t_a+\tau))+\frac{1-\cos(\omega(t_a+\tau))}{\omega(t_a+\tau)}\right]$$

To calculate the sine integral $S_i(x)$, see the end of Appendix C.

For flicker frequency noise ($\alpha = -1$), using Eq. C.3d and after some reordering of the terms, Eq. A.8 expands to

$$RESID\ 1\ \sigma_x^2(t_a,\tau)_{\alpha=-1} =$$

$$h_{-1}\left[(1+\frac{\tau}{t_a})\,\tau^2\left\{C_i(\omega\tau)-\frac{1-\cos(\omega\tau)}{(\omega\tau)^2}-\frac{\sin(\omega\tau)}{\omega\tau}\right\}\right.$$
$$+\frac{\tau}{t_a}(1+\frac{\tau}{t_a})\,t_a^2\left\{C_i(\omega t_a)-\frac{1-\cos(\omega t_a)}{(\omega t_a)^2}-\frac{\sin(\omega t_a)}{\omega t_a}\right\} \qquad \text{(A.8d)}$$
$$\left.-\frac{\tau}{t_a}(t_a+\tau)^2\left\{C_i(\omega(t_a+\tau))-\frac{1-\cos(\omega(t_a+\tau))}{(\omega(t_a+\tau))^2}-\frac{\sin(\omega(t_a+\tau))}{\omega(t_a+\tau)}\right\}\right]$$

For random-walk frequency noise ($\alpha = -2$) we have to be careful when using Eq. C.3e. With reordering of the terms in $3/\varepsilon$ the low frequency cutoff terms disappear. Eq. A.8 expands to

$$RESID\ 1\ \sigma_x^2(t_a,\tau)_{\alpha=-2} =$$

$$h_{-2}\frac{2\pi}{3}\left[\frac{\tau}{t_a}(t_a+\tau)^3\left\{2\frac{1-\cos(\omega(t_a+\tau))}{(\omega(t_a+\tau))^3}+\frac{\sin(\omega(t_a+\tau))}{(\omega(t_a+\tau))^2}+\frac{\cos(\omega(t_a+\tau))}{\omega(t_a+\tau)}+S_i(\omega(t_a+\tau))\right\}\right.$$
$$-\frac{\tau}{t_a}(1+\frac{\tau}{t_a})\,t_a^3\left\{2\frac{1-\cos(\omega t_a)}{(\omega t_a)^3}+\frac{\sin(\omega t_a)}{(\omega t_a)^2}+\frac{\cos(\omega t_a)}{\omega t_a}+S_i(\omega t_a)\right\} \qquad \text{(A.8e}$$
$$\left.-(1+\frac{\tau}{t_a})\,\tau^3\left\{2\frac{1-\cos(\omega\tau)}{(\omega\tau)^3}+\frac{\sin(\omega\tau)}{(\omega\tau)^2}+\frac{\cos(\omega\tau)}{\omega\tau}+S_i(\omega\tau)\right\}\right]$$

Appendix B

We will consider the general case of the estimate of the covariance of the frequency over two independent time intervals $[t_1, t_2]$ and $[t_3, t_4]$, which may overlap:

$$\frac{1}{(t_2-t_1)(t_4-t_3)}\int_{t_1}^{t_2}\int_{t_3}^{t_4} \langle y(\tau')y(\tau'')\rangle \, d\tau'' d\tau' \tag{B.1}$$

Here $\langle y(\tau')y(\tau'')\rangle$ is equivalent to the autocorrelation function $R_y(\tau'-\tau'')$. Replacing this expression by the autocorrelation function definition:

$$R_y(\tau) = \int_0^{\infty} S_y(f) \cos(2\pi f \tau)\, df \tag{B.2}$$

and changing the order of the integrals:

$$\frac{1}{(t_2-t_1)(t_4-t_3)}\int_0^{\infty} S_y(f) \int_{t_1}^{t_2}\int_{t_3}^{t_4} \cos(2\pi f(\tau'-\tau''))\, d\tau'' d\tau'\, df \tag{B.3}$$

Substituting ω for $(2\pi f)$ and doing the integrals over τ' and τ'':

$$\frac{1}{(2\pi)(t_2-t_1)(t_4-t_3)}\int_0^{\infty} \frac{S_y(\omega)}{\omega^2} \Big[\cos(\omega(t_2-t_4)) + \cos(\omega(t_1-t_3)) \tag{B.4}$$

$$-\{\cos(\omega(t_1-t_4)) + \cos(\omega(t_2-t_3))\}\Big]\, d\omega$$

There are two special cases of interest. If the two time intervals are the same, i.e. $t_1 = t_3$ and $t_2 = t_4$, then substituting $T = t_4 - t_2 = t_3 - t_1$:

$$\frac{2}{2\pi T^2}\int_0^{\infty} \frac{S_y(\omega)}{\omega^2} \{1-\cos(\omega T)\}\, d\omega \tag{B.5}$$

The second interesting case is when the two intervals are contiguous, ($t_2 = t_3 = 0$, $t_1 = -T$, $t_4 = \tau$):

$$\frac{1}{2\pi T\tau}\int_0^{\infty} \frac{S_y(\omega)}{\omega^2} \{\cos(\omega T) + \cos(\omega\tau) - [1+\cos(\omega(T+\tau))]\}\, d\omega \tag{B.6}$$

Appendix C

Evaluation of the following integral will allow us to use one function to solve all the possible cases of residuals

$$I(\alpha, \omega_h, \tau) = \frac{2h_\alpha}{(2\pi)^{\alpha+1}} \int_0^{\omega_h} \omega^{\alpha-2} \left\{ 1-\cos(\omega\tau) \right\} d\omega \qquad (C.1)$$

Let us assume a high frequency cutoff of ω. It is then possible to write Eq. C.1 as follows substituting $\omega\tau$ by u :

$$I(\alpha, \omega, \tau) = 2\frac{h_\alpha t^{1-\alpha}}{(2\pi)^{\alpha+1}} \int_0^{\omega\tau} u^{\alpha-2} (1-\cos(u)) du \qquad (C.2)$$

With the help of mathematical tables, this equation can be solved for our values of α .

For $\alpha = 2$ (white phase noise):

$$I(2, \omega, \tau) = 2\frac{h_2}{(2\pi)^3} \omega \left\{ 1-\frac{\sin(\omega\tau)}{\omega\tau} \right\} \qquad (C.3a)$$

We can readily see that terms involving the white phase noise will have a contribution depending linearly on the frequency cutoff, as in the Allan deviation.

For $\alpha = 1$ (flicker phase noise) :

$$I(1, \omega, \tau) = 2\frac{h_1}{(2\pi)^2} \left\{ \gamma+\ln(\omega\tau)-C_i(\omega\tau) \right\} \qquad (C.3b)$$

$C_i(x)$ is the Cosine Integral function, which can be easily approximated using numerical approximation as in Abramovitz[12] for arguments greater than 1, or direct numerical integration of the expression between brackets (see end of Appendix C). The dependency on the frequency cutoff is tempered by the time interval τ.

For $\alpha = 0$ (white frequency noise) :

$$I(0, \omega, \tau) = 2\frac{h_0}{2\pi} t \left\{ \frac{1-\cos(\omega\tau)}{\omega\tau} + S_i(\omega\tau) \right\} \qquad (C.3c)$$

$S_i(x)$ is the Sine Integral function, which can be treated the same way as the Cosine integral for arguments greater than 1, or direct numerical integration of the definition of $S_i(x)$ for small arguments. For high frequency cutoff, this term depends only on the time interval t.

For $\alpha = -1$ (flicker frequency noise):

$$I(-1, \omega, \tau) = 2h_{-1}\frac{\tau^2}{2} \left[1.5-\gamma-\ln(\varepsilon)-\left\{ \frac{1-\cos(\omega\tau)}{(\omega\tau)^2} + \frac{\sin(\omega\tau)}{\omega\tau} - C_i(\omega\tau) \right\} \right] \qquad (C.3d)$$

This term looks as if it will diverge to infinity for a low frequency cutoff $\varepsilon = 0$, but the combination of terms in the equation leading to the solution for flicker frequency noise is such that all constants arising from $I(-1, \omega, \tau)$ will cancel out, leaving a term depending only on the square of the time interval, as expected.

For $\alpha = -2$ (random walk frequency noise) :

$$I(-2, \omega, \tau) = 2h_{-2}\frac{\tau^3}{6} \left(\frac{3}{\varepsilon\tau} - \left[2\frac{1-\cos(\omega\tau)}{(\omega\tau)^3} + \frac{\sin(\omega\tau)}{(\omega\tau)^2} + \frac{\cos(\omega\tau)}{\omega\tau} + S_i(\omega\tau) \right] \right) \qquad (C.3e)$$

As expected, this expression will diverge to infinity for a low frequency cutoff $\varepsilon = 0$, but the combination of terms will cancel that in the problems under study here. In practice, the low frequency cutoff will depend on the observation (active) time and is always some value greater than zero.

Using formula 5.2.8 and formula 5.2.9 from Abramovitz & Stegun[12], pages 232 and 233, the rational approximations 5.2.38 and 5.2.39, it is easy to compute with enough accuracy the values needed for sine and cosine integrals of arguments greater than 1. Here are the formulas involved:

$$S_i(x) = \int_0^x \frac{\sin(x)}{x} dx = \frac{\pi}{2} - f(x)\cos(x) - g(x)\sin(x).$$

$$C_i(x) = \gamma + \ln(x) - \int_0^x \frac{1-\cos(x)}{x} dx = f(x)\sin(x) - g(x)\cos(x).$$

where f(x) and g(x) are computed using:

$$f(x) = \frac{1}{x}\left(\frac{x^8+a_1x^6+a_2x^4+a_3x^2+a_4}{x^8+b_1x^6+b_2x^4+b_3x^2+b_4}\right)+\varepsilon(x) \qquad g(x) = \frac{1}{x^2}\left(\frac{x^8+a_1x^6+a_2x^4+a_3x^2+a_4}{x^8+b_1x^6+b_2x^4+b_3x^2+b_4}\right)+\varepsilon(x)$$

$1 \le x < \infty$ $\qquad$ $1 \le x < \infty$

$|\varepsilon(x)| < 5\times10^{-7}$ $\qquad$ $|\varepsilon(x)| < 3\times10^{-7}$

$a_1 = 38.027264$	$b_1 = 40.021433$	$a_1 = 42.242855$	$b_1 = 48.196927$
$a_2 = 265.187033$	$b_2 = 322.624911$	$a_2 = 302.757865$	$b_2 = 482.485984$
$a_3 = 335.677320$	$b_3 = 570.236280$	$a_3 = 352.018498$	$b_3 = 1114.978885$
$a_4 = 38.102495$	$b_4 = 157.105423$	$a_4 = 21.821899$	$b_4 = 449.690326$

For arguments smaller than one, numerical integration can be done easily for $S_i(x)$. For Ci(x), the following transformation is very helpful:

$$C_i(x) = \gamma + \ln(x) - \int_0^x \frac{1-\cos(x)}{x} dx = \gamma + \ln(x) - \int_0^x \frac{\sin^2(\frac{x}{2})}{x} dx.$$

The second integral is easy to compute.

Appendix D

The residual error at time τ due to the extrapolation from the previous and next active times is

$$RESID\ 2\ \sigma_x^2(t_a,\tau) = < \left[x(\tau) - \left\{x(0) + \left(\frac{x(0)-x(-t_a)}{t_a} + \frac{x(t_d+t_a)-x(t_d)}{t_a}\right)\frac{\tau}{2}\right\}\right]^2 > \qquad (D.1)$$

This can be expanded into:

$$< \left[\left(x(\tau)-x(0)\right) - \frac{\tau}{2t_a}\left(x(0)-x(-t_a)\right) - \frac{\tau}{2t_a}\left(x(t_d+t_a)-x(t_d)\right)\right]^2 > \qquad (D.2)$$

Note that we use equal weighting for the calibrations from the two active periods. In the case of the fountain this is a good choice, since we know only the frequency over each period. In the more general case where we want to estimate the deviation between two fully characterised periods, we could introduce a weighting factor in front of the two last terms in Eq. D.2.

After expansion Eq. D.2 becomes:

$$
\begin{aligned}
< \quad & \big(x(\tau)-x(0)\big)^2+\frac{\tau^2}{4t_a{}^2}\big(x(0)-x(-t_a)\big)^2+\frac{\tau^2}{4t_a{}^2}\big(x(t_d+t_a)-x(t_d)\big)^2 \\
& -\frac{\tau}{t_a}\big(x(\tau)-x(0)\big)\big(x(0)-x(-t_a)\big)-\frac{\tau}{t_a}\big(x(\tau)-x(0)\big)\big(x(t_d+t_a)-x(t_d)\big) \\
& +\frac{\tau^2}{2t_a{}^2}\big(x(0)-x(-t_a)\big)\big(x(t_d+t_a)-x(t_d)\big) >
\end{aligned}
\tag{D.3}
$$

Using the results of Appendix B:

$$
\begin{aligned}
\frac{2}{2\pi}\int_0^\infty \frac{S_y(\omega)}{\omega^2}\Big[& \big\{1-\cos(\omega\tau)\big\}+\frac{\tau^2}{2t_a{}^2}\big\{1-\cos(\omega t_a)\big\}+\frac{\tau}{2t_a}\big\{1-\cos(\omega t_a)-\cos(\omega\tau)+\cos(\omega(t_a+\tau))\big\} \\
& +\frac{\tau}{2t_a}\big\{\cos(\omega(t_d-\tau))-\cos(\omega(t_a+t_d-\tau))-\cos(\omega t_d)+\cos(\omega(t_a+t_d))\big\} \\
& +\frac{\tau^2}{4t_a{}^2}\big\{2\cos(\omega(t_a+t_d))-\cos(\omega t_d)-\cos(\omega(2t_a+t_d))\big\}\Big]\,d\omega
\end{aligned}
\tag{D.4}
$$

Reorganising the terms to express the residuals as functions of integral of {1-cos(ωt)}:

$$
\begin{aligned}
\frac{2}{2\pi}\int_0^\infty \frac{S_y(\omega)}{\omega^2}\Big[& \Big(1+\frac{\tau}{2t_a}\Big)\big\{1-\cos(\omega\tau)\big\}+\frac{\tau}{2t_a}\Big(1+\frac{\tau}{t_a}\Big)\big\{1-\cos(\omega t_a)\big\} \\
& -\frac{\tau}{2t_a}\big[1-\cos\{\omega(t_a+\tau)\}\big]-\frac{\tau}{2t_a}\Big(1+\frac{\tau}{t_a}\Big)\big[1-\cos\{\omega(t_a+t_d)\}\big] \\
& +\frac{\tau^2}{4t_a{}^2}\big[1-\cos\{\omega(2t_a+t_d)\}\big]+\frac{\tau}{2t_a}\Big(1+\frac{\tau}{2t_a}\Big)\big[1-\cos\{\omega(t_d)\}\big] \\
& -\frac{\tau}{2t_a}\big[1-\cos\{\omega(t_d-\tau)\}\big]+\frac{\tau}{2t_a}\big[1-\cos\{\omega(t_a+t_d-\tau)\}\big]\Big]\,d\omega
\end{aligned}
\tag{D.5}
$$

Using Appendix C,

$$
\begin{aligned}
RESID_{\,1}\,\sigma_x^2(t_a,\tau) = \sum_{\alpha=-2}^{\alpha=2}\Big[& \Big(1+\frac{\tau}{2t_a}\Big)\mathrm{I}(\alpha,\omega,\tau)+\frac{\tau}{2t_a}\Big(1+\frac{\tau}{t_a}\Big)\mathrm{I}(\alpha,\omega,t_a) \\
& -\frac{\tau}{2t_a}\mathrm{I}(\alpha,\omega,t_a+\tau)-\frac{\tau}{2t_a}\Big(1+\frac{\tau}{t_a}\Big)\mathrm{I}(\alpha,\omega,t_a+t_d) \\
& +\frac{\tau^2}{4t_a{}^2}\mathrm{I}(\alpha,\omega,2t_a+t_d)+\frac{\tau}{2t_a}\Big(1+\frac{\tau}{2t_a}\Big)\mathrm{I}(\alpha,\omega,t_d) \\
& -\frac{\tau}{2t_a}\mathrm{I}(\alpha,\omega,t_d-\tau)+\frac{\tau}{2t_a}\mathrm{I}(\alpha,\omega,t_a+t_d-\tau)\Big]
\end{aligned}
\tag{D.6}
$$

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

OPTICAL FREQUENCY SYNTHESIS BY INTERVAL DIVISION

H. R. Telle

Lab. 4.42, Physikalisch-Technische Bundesanstalt
PF 3345, D W-3300 Braunschweig, Germany

Abstract

The concept of an interval-division optical frequency chain is presented, which relies entirely on laser diodes as transfer oscillators. Experimental data are given for frequency-stabilized lasers and mixing elements involved. The application of multiple quantum well laser diodes as 4-wave-mixing elements in the low frequency part of the chain is demonstrated.

Direct frequency measurement of visible or infrared radiation instead of conventional wavelength comparison has been of long standing interest since the development of stable vis/ir frequency standards and low-noise laser transfer oscillators.
Traditional vis/ir frequency measurement chains [1] rely on the synthesis-by-harmonic-generation approach, which is well known from rf techniques. In this scheme a rf-standard frequency is subsequently multiplied in several stages in order to reach the vis/ir region, i. e. each stage operates at an absolute frequency being integer-related to the rf master frequency (offset-frequencies introduced for technical reasons are neglected). In contrast to this absolute frequency scheme, difference-frequency synthesis systems (Fig.1.) accomplish the multiplication/division process in the difference frequency domain. [2,3]. However a absolute-to-difference frequency conversion is required in this case at the high- and low-frequency ends of the chain. The converter at the low-frequency end could be a fast photodiode, which transforms a beat signal, i. e. a difference frequency into an ac photo current. The conversion process at the higher end can be carried out, for instance, by second harmonic generation, which connects the absolute fundamental frequency with the width of the resulting octave interval [2]. Relating higher harmonics of a difference frequency to an absolute frequency, an alternative approach, has been demonstrated in the ir by using MIM diodes as nonlinear elements [3].
Two possible schemes for the individual difference frequency divider stages are depicted in Fig. 2. One may employ one mixer with strong $\chi^{(3)}$-nonlinearity to generate various inter-modulation products

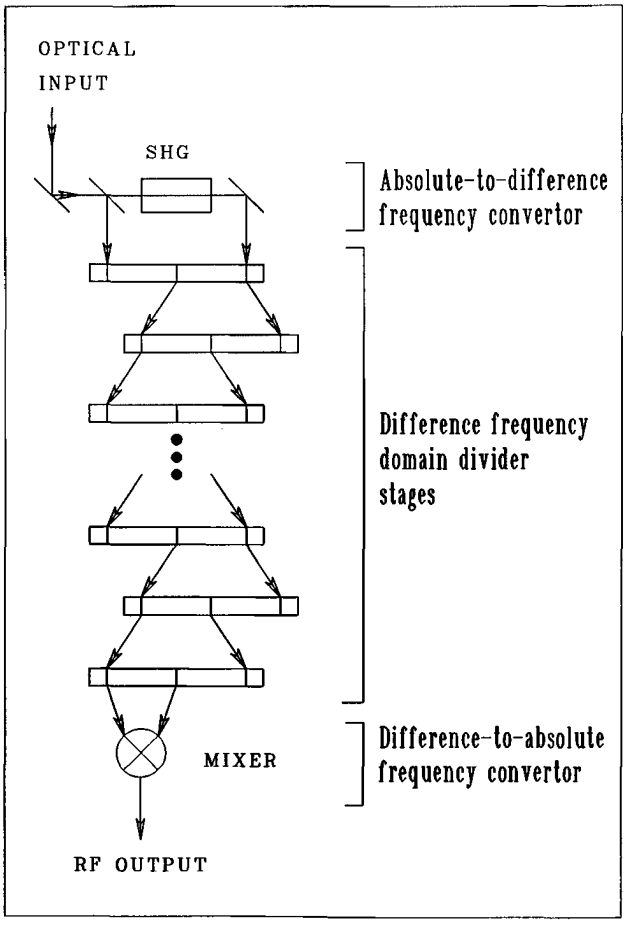


Fig. 1. Scheme of a difference-frequency synthesis system.

0-7803-0476-4/92 $3.00 © 1992 IEEE

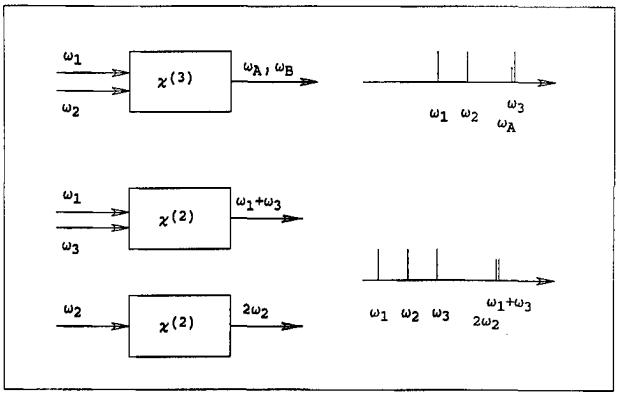


Fig. 2. Possibilities for dividing a frequency interval.

ω_A, ω_B...of the input signals at ω_1 and ω_2, see Fig. 2. Phase-locking a third laser at ω_3 to the four-wave mixing product at ω_A results in bisection of interval ω_1-ω_3. Similarly, two $\chi^{(2)}$-nonlinearities can be used to compare the frequency and phase of the sum-frequency-signal at $\omega_1+\omega_3$ and the second harmonic at $2\omega_2$. Interval bisection can than be accomplished by stabilizing the relative phase between these two three-wave mixing products.

The scheme and frequency plan of the PTB interval division chain, which is currently under construction, is briefly discussed in the following.

The scheme is depicted in Fig. 3. The starting point on this 'phase-information flowchart' is a frequency-stabilized InGaAlAs laser diode emitting at ν_1=c/1.315 μm. A red-emitting InGaAlP (ν_2=c/.657 μm) is phase-locked to the second harmonic of this field, which is generated in a nonlinear LBO-crystal.

These two frequencies form the octave starting interval, which is bisected in the following stages.

The first divider stage consists of a nonlinear crystal ($KNbO_3$), which generate the sum of ν_1 and ν_2, a GaAlAs LD (ν_3=c/.87 μm) together with a frequency doubler ($KNbO_3$-crystal) and a photomultiplier (PM), which compares the phases of the fields oscillating at $\nu_1+\nu_2$ and $2\nu_3$, respectively. The PM provides the error signal for a PLL which controls ν_3.

This twin 3-wave-mixing-approach will be employed for the following 5 stages, i. e. down to interval widths of 1.3 THz whereas 4-wave-mixing in laser diodes will be used for the narrower interval widths. Some experimental results for such devices are given below.

After 14 bisection steps, the starting octave-interval is divided down to ~28 GHz,

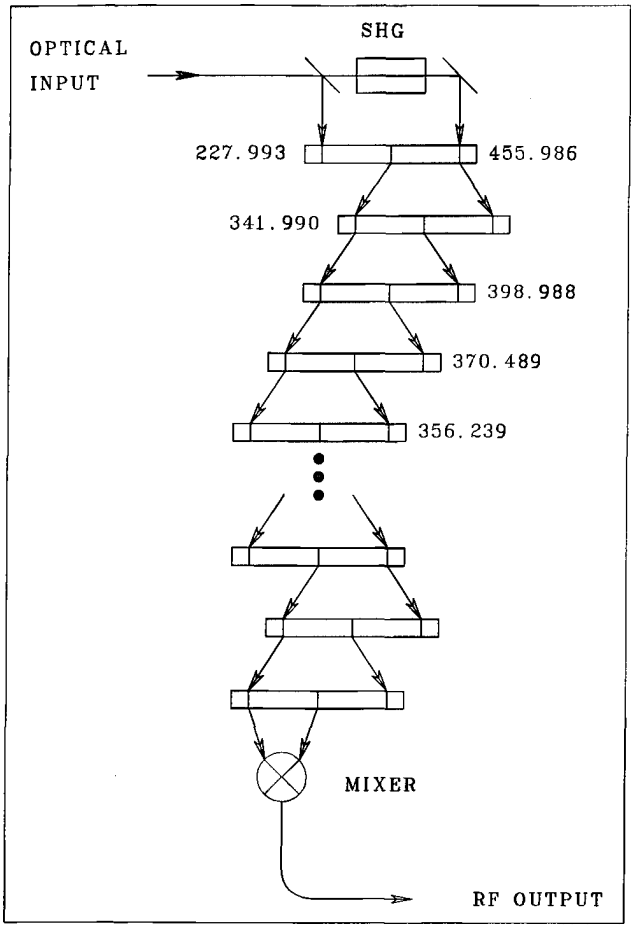


Fig. 3. The planned interval division scheme of PTB.

which can be detected with a fast photodiode and compared with a Cs-referenced standard frequency by means of conventional rf electronics. The width of the starting interval, and thus the value of ν_1, can be easily calculated from this frequency (multiplication with 2^n, where n is the number of stages). The other frequencies involved can be deduced analogously.

All frequencies fall within the starting interval, i. e. very similar transfer oscillators can be used. In our case, all lasers with the exception of those at ν_1 and ν_2 are extended-cavity GaAlAs laser diodes.

The chain directly allows the frequency measurement of the Ca intercombination line at ν_2 = 455.986241 THz, which has the accuracy potential for an optical clock [4]. However, many other optical frequencies of interest can be synthesized with only moderate additional effort by employing combinations of the known frequencies of the chain.

The convergence frequency ν_c is chosen to be $\nu_c \approx$ c/.85 μm, where both highly efficient

mixers (non-critically phase matched $KNbO_3$-crystals near room temperature) and highly coherent laser oscillators (GaAlAs LDs) are available.
In the following, experimental results will be presented for four-wave mixing in GaAlAs LDs, which may substantially simplify the low frequency part of the synthesis chain.
The experimental set up is depicted in Fig. 4. Two narrow band signals, pump (p) and

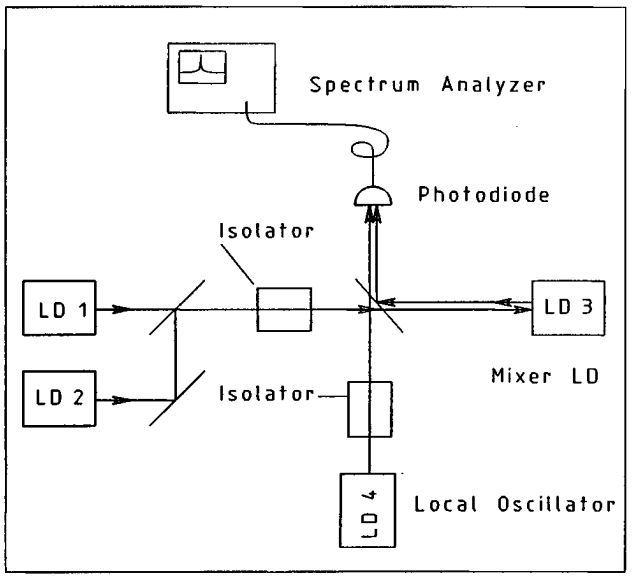


Fig. 4. Experimental setup for demonstration of four-wave mixing in LD.

co-pump(cp), generated by frequency-stabilized LDs (LD1, LD2) were injected into a solitary LD (LD3) acting as four-wave mixing element. The mixer LD was operated with injection currents of about 30 mA above threshold. However, its free-running oscillation was completely suppressed by proper setting of the injection powers (~200 μW per laser). The pump and co-pump frequencies were chosen to coincide within a few GHz with longitudinal modes of the mixer LD. The optical spectrum of the mixer LD output in the vicinity of the mixing product frequency $\nu_m=2\nu_p-\nu_{cp}$ was heterodyne-detected with the help of an additional stabilized LD (LD4), acting as local oscillator (LO), a fast Si photodiode and an rf spectrum analyzer. The LO field was coaxially combined with the mixer output (~50 mW) by means of a beam splitter.
All four lasers are MQW GaAlAs LDs (STC LT-50) with a free running emission wavelength of ~850 nm and a longitudinal mode spacing of ~90 GHz. Extended cavity set-ups are used for the two pump lasers and for the LO in order to achieve both, broad tuning range (~10 THz) and efficient frequency noise suppression (>20 dB). These identical lasers are housed in small temperature-stabilized aluminum boxes together with cavity extensions, each consisting of a holographic 1000 l/mm transmission grating and a highly reflecting mirror. The transmitted zeroth diffraction order is used for output coupling, while the first order (~50% diffraction efficiency) is employed for optical feedback and mode selection.
A beat note spectrum between the pump and co-pump laser is shown in Fig. 5. Both lasers were operated without additional electronic frequency control. Thus, acoustic and 1/f-noise components at low Fourier frequencies, seen at the slopes of the line, were still present. However, one estimates the fast emission line width from the wings of the line to be of the order of 10^4 Hz as compared to ~10^7 Hz for the same laser without optical feedback.
Most remarkably, the phase noise floor is

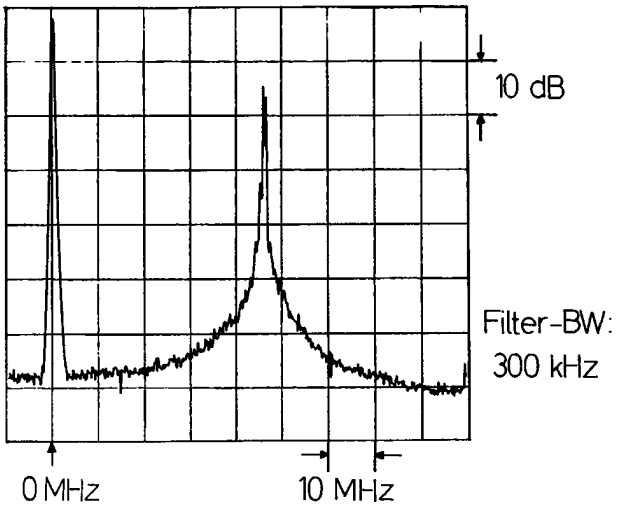


Fig. 5. Rf spectrum of beat note between two stabilized LD.

below -110 dBc (1 Hz BW) which underlines the outstanding noise suppression capability of the extended cavity scheme.
A typical heterodyne detected spectrum of the mixing product is shown in Fig. 6 for a fundamental interval width $\delta\nu=|\nu_p-\nu_{cp}| \simeq$ 270 GHz which corresponds to 3 longitudinal mode spacings of the mixer LD. Note that the optical line-width of the mixing product is slightly narrower than in Fig. 6 since the rf spectrum represents a convolution with the LO emission spectrum.
The four-wave mixing origin of the narrow line in Fig. 6 was verified by locking the phase of the beat note between LO and mixing product to an rf reference signal. The phase servo loop (PLL) employed consists of an rf preamplifier, a double balanced mixer which generates the phase error signal and a shunt J-FET for control-

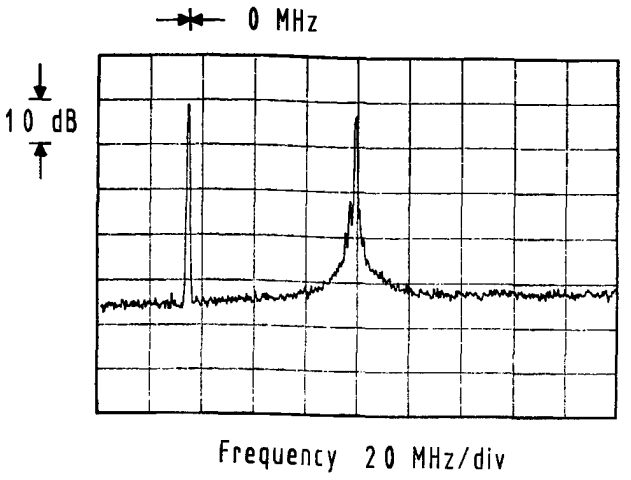


Fig. 6. Rf spectrum of heterodyne-detected mixing signal.

ling the pump injection current. The obtained beat note spectrum seen in Fig. 7 shows the characteristic δ-spike of a PLL at the rf oscillator frequency and servo loop-induced noise side bands. On estimates that the δ-spike contains more than 80% of the total mixing product power.
A phase lock in any synthesis chains can be understood as an time-transfer device which adds a certain amount of timing noise to the overall time delay, owing to noise sources being present in every practical PLL. Assuming the noise floors to be almost frequency independent, e. g. shot noise (passive mixers) or broad band amplified spontaneous emission (mixer LDs), the equivalent phase noise floors $S_{\varphi n}$ are expected to be white, too, where the index n denotes the n-th bisection stage. The correspondent timing noise floors are $S_{xn}=S_{\varphi n}/(2\pi\Delta\nu_n)^2$, where $\Delta\nu_n$ is the width of the n-th interval.

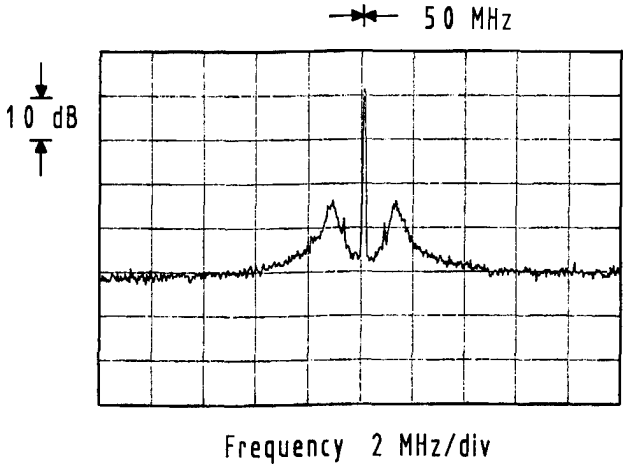


Fig. 7. Rf spectrum of phaselocked mixing product.

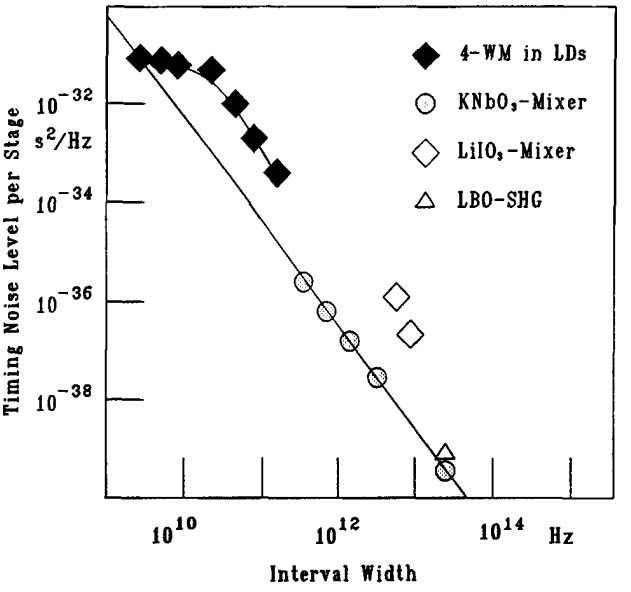


Fig. 8. Timing noise level of the various stages of the frequency division chain.

The measured and/or estimated timing noise contributions of the n=14 individual bisection stages of a 228 THz-to-28 GHz divider chain are depicted in Fig. 8. If almost independent noise processes are assumed, the total timing noise level is simply given by the sum of all timing noise spectral densities:

$$S_{\varphi sum} = 1.4\cdot10^{-32}\ s^2/Hz$$

The corresponding Allan variance is of the order of

$$\sigma_y(\tau) \simeq (10^{-14}\ s)/\tau$$

for PLL control bandwidths of the order of 10^5 Hz, which indicates that laser diode based interval division synthesis chains can meet the performance level of state-of-the-art microwave and optical frequency standards.

This work was partly supported by the Deutsche Forschungsgemeinschaft.
The author is indebted to Ch. Koch for performing the four-wave mixing experiments.

References

[1] G. Kramer, B. Lipphardt, C. O. Weiss, "Coherent frequency synthesis in the infrared", presented at this conference.

[2] H. R. Telle, D. Meschede, T. W. Hänsch, "Realization of a new concept for visible frequency division: Phase locking of harmonic and sum frequencies", Opt. Lett. 15, pp. 532-534, 1990.

[3] B. G. Withford, "Simultaneous phase-lock of five CO_2 Lasers to a primary Cs frequency standard", Appl. Phys. B 35, pp. 119-122, 1984.

[4] A. Morinaga, F. Riehle, J. Ishikawa, J. Helmcke, "A Ca optical frequency standard: Frequency stabilization by means of nonlinear Ramsey resonances", Appl. Phys. B 48, pp. 165-171, 1989.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

OPTICAL PARAMETRIC DIVISION

N. C. Wong and D. Lee

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract: A novel concept of frequency division in the optical domain based on an efficient, one-step parametric downconversion process is described. In this method an input optical frequency is halved by phase locking the beat frequency of the two subharmonic outputs of an optical parametric oscillator to a microwave frequency source. This approach to optical frequency division was demonstrated in a tunable KTP optical parametric oscillator whose phase-locked beat-note linewidth was at the Hz level. A wideband phase modulator that matched the optical and microwave phase velocities was fabricated to facilitate difference frequency measurements. We obtained a modulation index of 0.36 with a microwave power of 0.5 W at 12.4 GHz. When the modulator was enclosed in an optical cavity to enhance the modulation, it generated a fifth order sideband at 62 GHz with 0.1% sideband power. Finally, a new method for measuring optical frequencies from the UV to the near-IR relative to the cesium clock is described. In this scheme the frequency difference of two known ratios (1/2 and 2/3) of an optical frequency f relative to the cesium clock is measured. By employing parallel-connected parametric oscillators and wideband modulators to link the $(1/2)f$ and $(2/3)f$ frequencies, a precise and accurate optical frequency comb can be provided in the ~1–2 μm wavelength region, with an extended measurement range of 0.4–7 μm.

1. Introduction

In recent years there is rapid progress in high-stability lasers with Hz linewidths that are suitable local oscillators for interrogating high Q optical resonances in ion trap and atomic fountain experiments. The tantalizing possibility of an optical clock with a Q of 10^{15} or more promises a new level of precision and accuracy in frequency metrology and precision measurements. It also serves to remind us of the urgent need of a practical method for linking optical and microwave frequencies. Here the goal is develop the capability for measuring and synthesizing optical frequencies from a microwave frequency standard, and vice versa.

The traditional technique of a frequency chain[1] involving harmonics of laser and klystron sources is particularly useful for a specific optical-microwave measurement link, as in the comparison of the I_2-stabilized 633-nm He-Ne laser to the cesium clock. However, it is not a practical method for measuring a broad range of frequencies owing to the lack of tuning capability in the transfer laser oscillators.

Here we describe a new method of optical frequency division that takes advantage of the inherent frequency dividing property of an optical parametric oscillator (OPO).[2] In section 2, we briefly review the concept of optical parametric division (OPD). This approach to frequency division has been demonstrated in an experiment that utilized a KTP OPO is described in section 3. In section 4 we report preliminary experimental results of a wideband phase modulator that is a key element for achieving a significant OPO tuning capability. We will then describe in section 5 an OPO-based method for measuring optical frequencies from the UV to the near-IR relative to a microwave frequency standard.

2. OPO frequency division

An OPO converts with high efficiency an input pump wave of frequency f_p into two intense, coherent subharmonic outputs, a signal (f_1) and an idler (f_2), which obey the energy conservation relation

$$f_1 + f_2 = f_p \ , \qquad (1)$$

and whose linewidths are essentially limited by the input pump linewidth. OPO's have been used mainly for generating tunable radiation over a considerable wavelength range using a single pump source. The actual output frequencies are determined by the usual phase

0-7803-0476-4/92 $3.00 © 1992 IEEE

matching condition. For example, for a biaxial crystal such as KTP, angle tuning is used to achieve the proper phase matching for the interacting signal, idler, and pump waves.

The availability of high-quality crystals with low-loss coatings and advances in stable pump sources have made the current generation of OPO's easier to use and more stable to operate. For frequency division applications, we use a doubly resonant oscillator configuration in which both the signal and idler waves are resonant inside the OPO cavity. This increases the cavity storage time for the signal and idler waves leading to significantly lower pump threshold, narrower linewidths, but also reduced stability.

Figure 1 shows a schematic of an OPO frequency divider.[3] A measurement of the output difference frequency

$$\delta = f_1 - f_2 \,, \tag{2}$$

in conjunction with the sum-frequency constraint (1), precisely determines the signal and idler frequencies. As a frequency divider, the difference frequency δ is phase locked to a microwave reference frequency source, such that the known input frequency f_p is divided into two precisely known frequencies

$$f_{1,2} = f_p/2 \pm \delta/2 \,. \tag{3}$$

Since δ is tunable and small compared with f_p, the OPO serves as a tunable 2:1 frequency divider.

The phases of the OPO-divider output beams have constraints similar to their corresponding frequencies. The sum of the output phases follow the phase of the input pump: $\phi_1 + \phi_2 = \phi_p$. Because the phase difference $\phi_1 - \phi_2$ is locked to the phase of the microwave frequency source, the output phases are precisely determined. For an OPO-based optical-to-microwave frequency chain, as will be described in section 5, the above phase constraints imply that the frequency link between optical and microwave frequencies can be phase coherent.

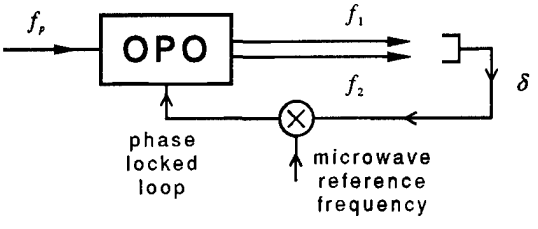


Fig. 1. Schematic configuration of an OPO-divider.

3. Phase locked OPO-divider

Optical parametric division has recently been demonstrated[4] in a highly efficient, low-threshold KTP OPO. The nonlinear crystal KTP was chosen for its good optical quality, high damage threshold, large temperature and angular tolerances, and type-II phase matching. Our two-element OPO consisted of a 25-mm-radius mirror and a 8-mm-long flux grown KTP crystal. One end of the crystal, coated for maximum reflection at 1.06 μm and maximum transmission at 532 nm, had a 40-mm radius of curvature and served as the input port for the pump beam which was derived from a krypton ion laser at 531 nm. The other end of the crystal was antireflection coated at both 1.06 μm and 532 nm. The output mirror was coated for maximum reflection at 532 nm and 0.5% transmission at 1.06 μm, thus forming a single-ended cavity configuration for both the pump and the IR beams. Besides facilitating the separation of the IR output beams from the pump input, this configuration effectively doubled the interaction length when compared with a single-pass configuration. The KTP OPO had a minimum threshold of ~22 mW, and an operating threshold of ~40 mW near the frequency degenerate point $f_1 \approx f_2$. The total conversion efficiency was ~30% over a large range of pumping levels of as much as 5 times above threshold.

Four different elements were used to tune the output frequencies: incident pump angle, cavity length, crystal temperature, and electro-optic effect. By changing the incident angle of the pump beam and thereby angle tuning the KTP crystal, the output frequency separation could be varied over a δ range of ~1 THz. A piezoelectric transducer (PZT) attached to the output mirror controlled the cavity length and different mode-pairs of signal and idler waves could be selected by scanning the PZT, with a δ range of ~100 GHz. The PZT-mirror was used as part of an intensity servo that stabilized the cw IR output powers of a particular mode-pair.

Continuous tuning of the output frequencies was made possible by the slow temperature tuning of the crystal using a thermoelectric (TE) cooler and by electric-field (E-field) tuning via the electro-optic effect. The tuning range of the TE cooler was ~500 MHz while that of the E-field tuning was ~40 MHz.

The signal-idler beat frequency δ was detected with a 25-GHz InGaAs/InP photodetector with a typical dc photocurrent of ~300 μA. The ac photocurrent was

demodulated in a microwave double-balanced mixer whose local oscillator was derived from a synthesized microwave signal. The mixer output was amplified, phase detected, and fed back to the E-field tuning circuit to phase lock the beat frequency to the microwave reference signal. Figure 2 shows a demodulated signal-idler beatnote spectrum under phase-locked conditions. The microwave reference frequency was 9.36 GHz and the rf reference frequency for the phase detector was 50 MHz. The spectrum in Fig. 2 was taken with a span of 2 kHz, a resolution bandwidth of 30 Hz, and a total sweep time of 50 s. The 3-dB linewidth (FWHM) of the peak in Fig. 2 was ~30 Hz, suggesting that the beatnote linewidth was substantially less than the 30-Hz resolution bandwidth of the rf spectrum analyzer. A period counter would allow us to measure the Allan variance of the phase-locked beatnote with the possibility of observing the quantum limited phase diffusion noise of the OPO.[3, 5] The beatnote signal-to-noise ratio of ~40 dB was significantly less than the expected value when the known phtodetection efficiency and other conversion losses were taken into account. The likely source of the additional phase noise was the phase locked loop.

The observed linewidth of a few Hz of the signal-idler beatnote implies that the intrinsic noise of the OPO frequency divider is in general small compared to the pump linewidth. Equation (2) suggests that the accuracy of the output frequencies, f_1 and f_2, is mainly limited by the accuracy of the input pump laser with little contribution from the OPO apparatus. Indeed this is one of the advantages of this one-step parametric division process: the non-resonant parametric interaction does not generate much noise as compared with resonant processes in laser oscillators.

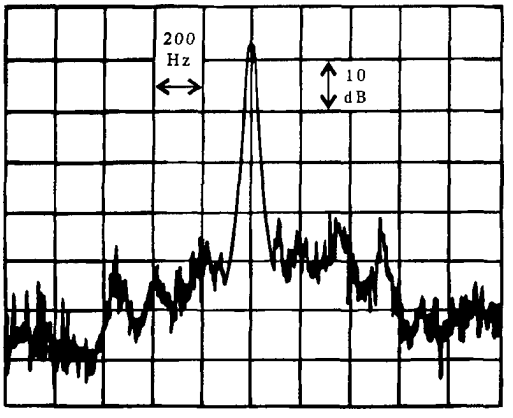


Fig. 2. Trace of demodulated signal-idler beatnote spectrum showing a 3-dB width that was limited by the 30-Hz resolution bandwidth of the rf spectrum analyzer (center frequency, 50 MHz; sweep time, 50 s).

The demonstration of frequency division using a high performance OPO with precision tuning capability suggests new opportunities in areas of laser spectroscopy and frequency metrology. The high quality KTP crystal and our two-element cavity design allow us to stabilize, tune, and phase lock the OPO for an extended period of time. We expect that further improvement in crystal quality, crystal waveguide fabrication, and cavity design will result in an OPO that is very stable and easy to use for various applications.

4. High-frequency modulator

The tuning range of δ is normally limited by the bandwidth of the high-speed photodetector, ~100 GHz for a commercially available product. It can be significantly increased to THz range by applying an external phase modulation to the OPO outputs. Figure 3 is a frequency plot of the two modulation combs after f_1 and f_2 are modulated at the a frequency f_m. By measuring the beat frequency f_b between the two closest sidebands, one from each modulation comb, the frequency separation δ can be obtained. In Fig. 3 where we assume each modulation comb has 16 sidebands, the frequency separation is given by $\delta = 16f_m + f_b = 340$ GHz, for $f_m = f_b = 20$ GHz. By dithering the modulation frequency and observing the change in the beat frequency which equals the dither deviation times the sideband order, the exact order of the sideband and therefore δ can be determined.

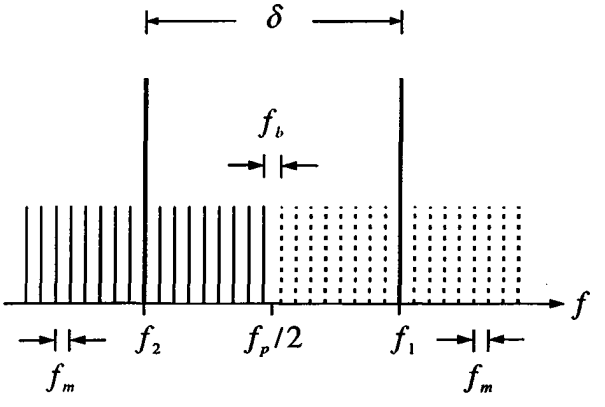


Fig. 3. Measurement of signal-idler difference frequency δ by modulating OPO outputs f_1 and f_2 at the modulation frequency f_m: $\delta = 16f_m + f_b$.

In order to generate many sidebands at a microwave frequency of ~20 GHz, a new type of phase modulator is required. We have fabricated a 25-mm-long $LiNbO_3$ phase modulator that contained a microwave waveguide structure such that the phase velocities of the microwave and optical waves could be matched to optimize the modulation,[6] as shown in Fig. 4. The modulator was 0.5 mm thick and 4 mm wide, with a gold-plated ground at the bottom and a 1.94-mm-wide stripline on the top. The microwave field was nominally confined to a volume of the crystal defined by the width of the stripline, which also determined the microwave frequency at which the optical and microwave phase velocities were matched. This velocity-matched frequency was 12.44 GHz for our modulator. Our microwave coupling was a SMA connector pin that rested on of the crystal near the center of the modulator in a position that was perpendicular to the stripline. The coupling efficiency depended critically on the distance between the pin and the stripline. For an initial trial we used a thin strip of aluminum foil that connected the pin and the stripline and obtained adequate but less than optimal coupling.

A 1.06 μm beam from a diode-pumped YAG laser was propagated through the crystal with the beam aligned along the edge of the stripline where the electric field strength was maximum. To reduce optical losses, the crystal was antireflection coated at 1.06 μm. At a microwave input power of ~0.5 W, we observed ~3% power in the first order sideband in a single-pass configuration, which implied a modulation index of 0.36. The observed modulation index was ~30% of the expected value because of low coupling efficiency, estimated at ~30%. Improving the microwave coupling efficiency should allow us to obtain at least an index of ~1 for a microwave power of 1 W.

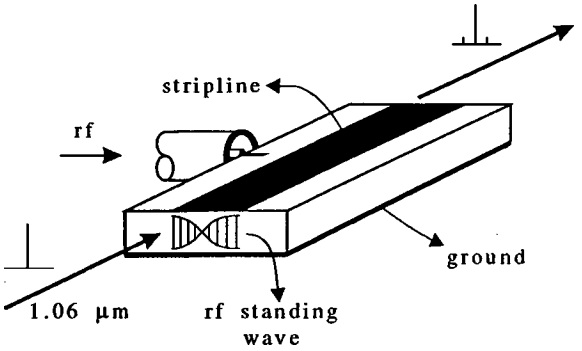


Fig. 4. Sketch of phase-velocity-matched electro-optic modulator (height, 0.5 mm; length, 25 mm, stripline width, 1.94 mm).

We have placed the velocity-matched modulator inside an optical cavity to enhance the modulation effect. For a cavity with a finesse of 23, an enhancement of the modulation index of ~10 was observed. Using an optical spectrum analyzer as a detector we have observed the fifth order sideband at ~62 GHz with a sideband power of 0.1%. Future improvements with a higher finesse cavity and a higher modulation frequency should make it possible to obtain a modulation comb that spans over a THz or more.

5. Optical frequency counter

One of the goals in the study of OPD is to develop a practical method for measuring the frequency of light. A useful optical frequency measuring device should be capable of measuring a broad range of optical frequencies with a single apparatus. Also, the ease of use, cost, and availability of components are important considerations of a practical optical frequency counter.

An obvious extension of OPO frequency division for this purpose is to connect several OPO's in a serial configuration so that an input signal frequency can be divided until it is low enough to be directly measured. However, this serial method requires efficient OPO's in the IR and far-IR regions. Furthermore, the optical power requirements for operating OPO's in a chain are severe. Such disadvantages make the serial method less attractive as a practical optical frequency counter.

A different scheme that uses a parallel-connected OPO configuration is superior to the serial method in many ways. Instead of repeatedly dividing an optical frequency f until it reaches the countable microwave region, the scheme seeks to measure the frequency difference between two of its integral subharmonics relative to a microwave frequency. A set of similar OPO's that operate in parallel are used to generate the subharmonics and to facilitate the difference-frequency measurement, as will be described. The system produces an ultra-wideband, precision optical frequency comb that makes it possible to measure accurately any optical frequency from the UV to the near-IR.[7] Here the pump frequency f serves as a "calibration" frequency for the comb and therefore it can be chosen to optimize the OPO system.

The scheme uses a stable laser at frequency f to pump a set of two master OPO's and N auxiliary OPO's that are in a parallel phase-locked configuration, as sketched in Fig. 5. The first master OPO

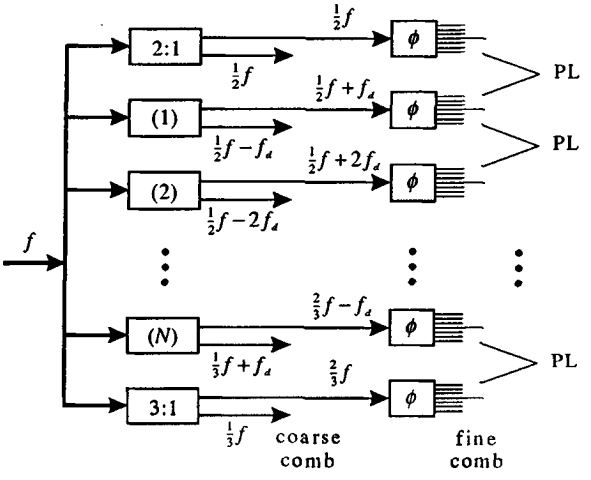


Fig. 5. Schematic of a parallel configuration of phase-locked OPO's as a frequency counter. PL: phase locked; ϕ: modulator.

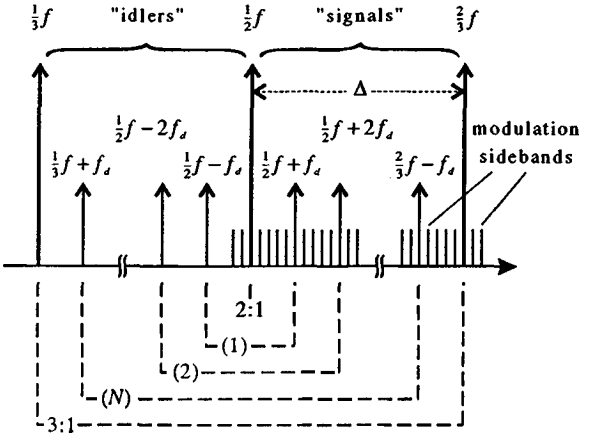


Fig. 6. Frequency plot of the parallel-OPO configuration of Fig. 5.

is an exact 2:1 divider which generates the frequency $(1/2)f$; the other is an exact 3:1 divider with output frequencies $(1/3)f$ and $(2/3)f$. Exact divisions by 2 and 3 are chosen because they can be easily generated. A 2:1 frequency divider has been successfully demonstrated as described in section 3. For a 3:1 division the signal frequency is exactly twice that of the idler. Therefore, by doubling the idler wave and phase locking it to the signal wave, an exact 3:1 divider can be obtained.

The N auxiliary OPO's form a very wideband frequency comb, spanning the spectrum from $(1/3)f$ to $(2/3)f$ with a comb interval f_d. The OPO's are controlled by phase-locked loops that assure that the offset frequencies f_d's are precisely known. Figure 6 shows how the OPO output frequencies evenly fill the frequency space between $(1/3)f$ and $(2/3)f$. (We have assumed equal f_d's, but this is not a required condition.) In Figs. 5 and 6, the higher-frequency signals lie in the range $(1/2)f$ to $(2/3)f$ while the lower-frequency idlers lie in the range $(1/3)f$ to $(1/2)f$. Either range can be used for phase locking the OPO's.

Figure 5 shows that the OPO signals are modulated by a microwave signal that is phase locked to a microwave frequency standard to create secondary combs of modulation sidebands, the "fine combs". These fine combs can be obtained using the wideband modulator-cavities described in section 4 for a precise determination of the spacing f_d of the coarse OPO frequency comb. The measurements of f_d between adjacent OPO signals provide an absolute determination of the frequency difference $\Delta = (2/3)f - (1/2)f$ between the upper frequency marker $(2/3)f$ and the degenerate frequency marker $(1/2)f$ of the coarse comb. Once Δ is measured as a function of a microwave frequency reference, the pump frequency $f = 6\Delta$ is accurately determined. f can then be phase-locked to and controlled by the microwave signal. Consequently, both the coarse and fine combs of frequency markers are fully characterized. In this way, conventional counting techniques can be used for measuring Δ and the entire comb is calibrated.

To illustrate the method, we describe a system that directly measures frequencies in the 1.2–2.4 μm range, and which can be extended to 0.4–7 μm. Consider a pump laser with frequency $f = 380$ THz ($\lambda = 790$ nm). This is chosen to be near the peak of the Ti:Al_2O_3 laser which has high power and can be easily stabilized to sub-kHz linewidths. The outputs of the two master OPO's are $(1/2)f = 190$ THz (1.58 μm), $(1/3)f = 127$ THz (2.37 μm), and $(2/3)f = 253$ THz (1.18 μm), which give $\Delta \approx 63$ THz. Assuming a difference-frequency measurement capability $f_d \sim 10$ THz, six auxiliary OPO's are required to span Δ. To facilitate the f_d measurements, the OPO signals are strongly phase modulated at a frequency ~20 GHz with a modulation index of ~250, thus imposing on each signal a fine comb of modulation sidebands with a span of ~10 THz.

The proposed system forms a versatile optical frequency counter. Consider the measurement of an unknown optical frequency that is within the range $(1/2)f$ to $(2/3)f$ (1.18–1.58 μm). The closest OPO

signal is identified by a conventional wavelength measurement. Then a simple beat measurement between the unknown frequency and the nearest sideband of the OPO signal yields the desired frequency information. Frequencies between $(1/3)f$ and $(1/2)f$ (1.58–2.37 μm) can also be measured accurately, in a similar way, by applying a strong modulation to either the unknown frequency or the nearest OPO idler. Thus the proposed optical frequency counter covers a range from $(1/3)f$ to $(2/3)f$ using only a single pump laser that can be conveniently chosen.

The method can readily be extended to longer wavelengths without changing the pump frequency f. Consider an unknown frequency in the range $(1/9)f$ to $(1/3)f$: its second or third harmonic is within the measurement range $(1/3)f$ to $(2/3)f$. It can also be extended to shorter wavelengths. For example, if a signal in the range $(2/3)f$ to $2f$ has sufficient power (~40 mW) to drive an OPO, a 2:1 or 3:1 division brings it within the measurement capability of the counter. Therefore the measurement capability can be extended to the range $(1/9)f$ to $2f$ (~0.4–7 μm). Furthermore, it is apparent that the proposed system is capable of not only measuring but also synthesizing optical frequencies throughout this range.

The parallel-OPO scheme has a number of operational advantages. First, it provides flexibility in the choice of the cailbration pump laser. Examples include a Ti:Al_2O_3 laser for its ease of operation and high power, and a phase-locked diode laser array for its compactness and portability. Second, the fundamental measurement noise is due to the intrinsic phase diffusion noise of the OPO's: no intermediate reference sources contribute. The phase noises of the input pump and the microwave reference source are considered as extrinsic to the division process. Third, the parallel scheme is easier to operate and requires less pump power than the serial method. Finally, the optical frequency comb is situated in the spectral region of ~1–2 μ in which excellent components are available, such as nonlinear crystals, modulators, and detectors.

KTP is transparent in the 0.5–3.5 μm wavelength region and is an attractive nonlinear material for constructing OPO's for the proposed counter. Initial results of an efficient, tunable KTP OPO 2:1 divider are reported in section 3. A high conversion efficiency is paramount in our frequency counting scheme because of the need to generate a large number of modulation sidebands. For a total of 500 sidebands with 50 μW per sideband, each OPO must provide at least a signal carrier power of ~25 mW which requires a minimum pump power of ~170 mW. The pump power can be reduced by using a higher modulation frequency and a lower modulation index. For our proposed system using a total of 8 OPO's, the required pump power is ~2 W, which is readily available from a Ti:Al_2O_3 laser.

6. Conclusion

In summary, we have described a new approach to frequency division in the optical domain and a practical method for measuring optical frequencies relative to the cesium clock. The parallel-OPO frequency counter can be realized with currently available technology in the 1–2 μm spectral region and with 20-GHz microwave electronics. Two key elements in this scheme have been demonstrated with good results. A phase-locked OPO-divider was found to contribute little noise to the division process and a velocity-matched phase modulator was operated to yield the fifth order sideband at 62 GHz with the possibility of efficient modulation that spans several THz. Future advances in modulators, detectors, and nonlinear materials and in n:1 parametric division techniques for $n > 3$ will further simplify and extend the proposed optical frequency counting system.

7. Acknowledgement

The authors gratefully acknowledge the contribution of L. R. Brothers to the fabrication and testing of the wideband phase modulator. This research was supported in part by the U.S. Army Research Office under Grant No. DAAL03-90-G0128, and the National Institute of Standards and Technology under a Precision Measurement Grant 60NANBOD1052.

8. References

1. D.A. Jennings, C.R. Pollock, F.R. Petersen, R.E. Drullinger, K.M. Evenson, J.S. Wells, J.L. Hall, and H.P. Layer, "Direct frequency measurement of the I_2-stabilized He-Ne 473-THz (633-nm) laser," Opt. Lett., vol. 8, pp.136–138, March 1983.

2. R. L. Byer, "Optical parametric oscillator," in *Treatise in Quantum Electronics*, H. Rabin and C. L. Tang, Eds. New York, NY: Academic Press, 1973, pp. 587–702.

3. N.C. Wong, "Optical frequency division using an optical parametric oscillator," Opt. Lett., vol. 15, pp. 1129–1131, Oct. 1990.

4. D. Lee and N.C. Wong, "Tunable optical frequency division using a phase-locked optical parametric oscillator," Opt. Lett., vol. 17, pp. 13–15, Jan. 1992.

5. R. Graham and H. Haken, "The quantum-fluctuations of the optical parametric oscillator. I," Z. Physik, vol. 210, pp. 276–302, (1968).

6. B.Y. Lee, T. Kobayashi, A. Morimoto, and T. Sueta, "Picosecond electro-optic modulator/deflector with velocity matching," in *Technical Digest, Conference on Laers and Electro-optics 1991*, Baltimore, MD, p. 146, May 1991.

7. N.C. Wong, "Optical frequency counting from the UV to the near-IR," to appear in Opt. Lett., 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

COHERENT FREQUENCY SYNTHESIS IN THE INFRARED

G. Kramer, B. Lipphardt, C.O. Weiss

Physikalisch-Technische Bundesanstalt, Bundesallee 100,
D-3300 Braunschweig, Germany

Abstract: The paper describes a conventional frequency multiplication chain that has been set up at PTB to compare the 88 THz IR-frequency of the methane stabilized He-Ne laser to the primary frequency standard [1], [2]. Phase locks and coherent measurement techniques reduce all errors contributed by the chain to a negligible level. Use of a hydrogen maser as the reference allows very precise measurements in a short time.

Introduction: Optical frequency synthesis or coherent frequency measurement through the infrared and to the visible radiation is meeting growing interest. One reason is the fact that the primary length standard can be applied only by calculating the wavelength of some near infrared or visible radiation from its measured frequency. Another even more important reason is the fact that a number of new and superior time/length standards have been conceived and are being developed in various laboratories, that all use optical quantum transitions. There are good physical arguments that stored ions and manipulated atoms as well as Ramsey excited molecular beams will reach their full potential only at the highest frequencies. Furthermore some fundamental constants and theories can be checked to very high accuracy using optical frequency measurements.

At PTB in Braunschweig we have therefore developed - and installed as a permanent facility - a frequency multiplication chain linking our methane beam optical frequency standard to the cesium primary clocks with practically no additional uncertainty. Use of a hydrogen maser as a flywheel and continuous phase time recording with sub-ps resolution gives us observed frequency fluctuations below a part in 10^{13} from 5 s averaging on. Fig. 1) shows the general scheme of the chain. In the following we will shortly describe its various components.

Oscillators

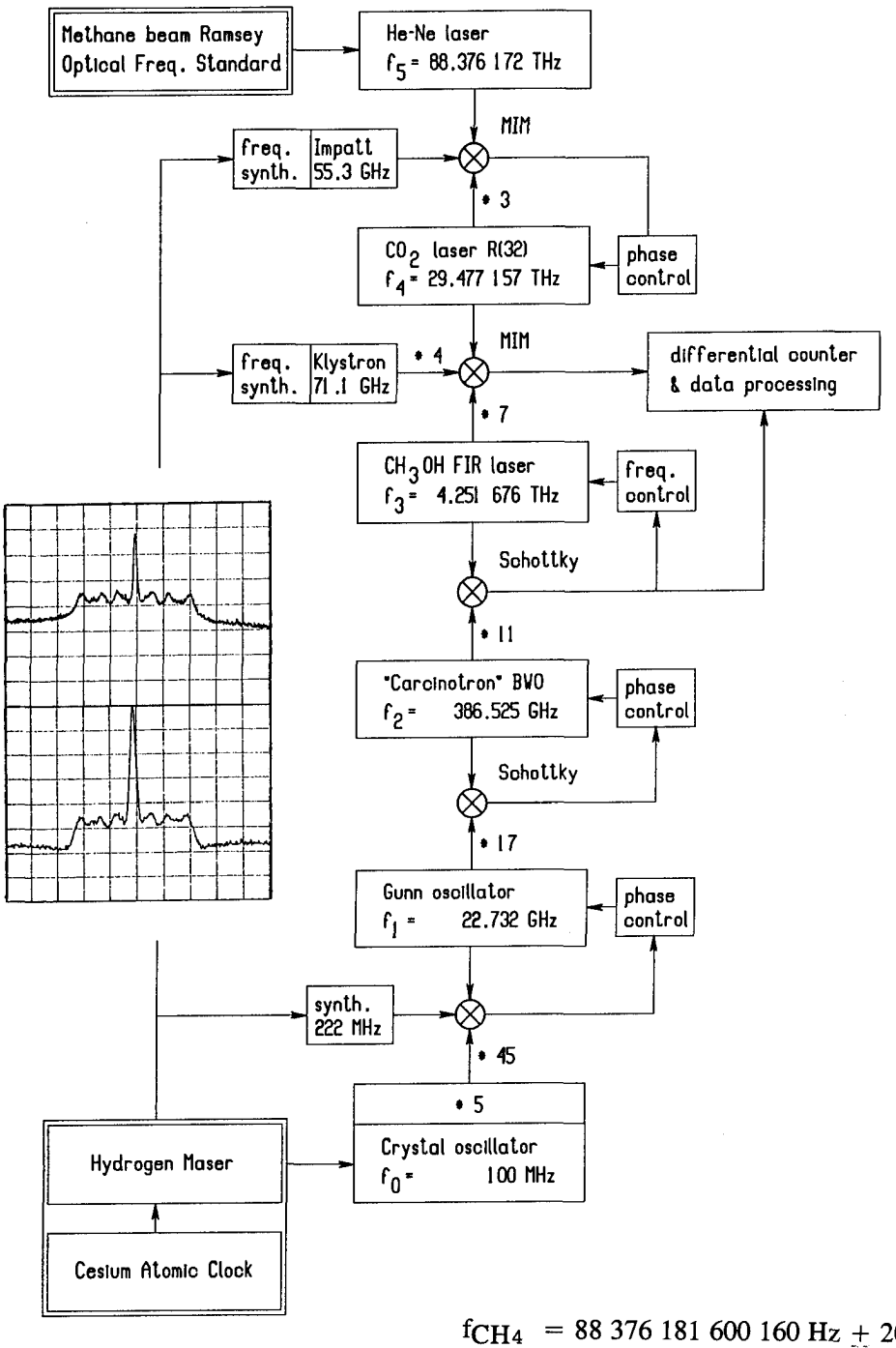
The three main specifications of the oscillators are: frequency, output power, and spectral purity. Starting at the top, there are three gas lasers. The He-Ne laser at 3.39 μm, which can be phase-locked to a CH_4 optical frequency standard, consists of a short laser followed by an amplifier tube, 1.6 m long, to produce the 10 mW of optical power required for the highest mixing step. The CO_2 laser R(32) line is the closest subharmonic, but still a 55 GHz microwave frequency has to be added in order to produce a low frequency ($\lesssim$100 MHz) IF, that can be easily extracted from the MIM diode. This mixing step was for the first time demonstrated by K.M. Evenson 20 years ago.

The methanol FIR laser at 4.2 THz is optically pumped by another CO_2 laser and produces just enough power (35 mW) for the MIM diode working at a relatively high mixing order due to the necessity of adding the 4th harmonic of a 72 GHz klystron. A problem of the FIR laser is that reflexions back into the pump laser can cause pulsing or chaotic behaviour; properly aligned, however, this laser produces reliable and stable beat signals of sufficient S/N ratio (20 dB in 100 kHz BW).

Gas lasers, due to their high Q-factor resonators, have practically no wide-band phase noise, and frequency fluctuations are mainly caused by acoustic vibrations at low Fourier frequencies. With proper processing of the IF signals and application of digital (event-counting) phase lock loops (PLL) [3], low frequency FM noise does not cause any problems to the coherence of the chain. The situation is quite different with the electronic oscillators used in the lower part of the chain. The 386 GHz Backward Wave Oscillator ("Carcinotron") has a phase noise level too high to allow generation of a coherent harmonic beat at 4.2 THz. The Carcinotron was therefore tightly (4 MHz servo BW) locked to the 17th harmonic of a cavity stabilized 22 GHz Gunn oscillator serving as a fly-wheel at the higher Fourier frequencies. The insert in Fig. 1 shows the residual phase noise of the Carcinotron with the characteristic "servo bumps" as observed on the two beat signals. Please note the 21 dB reduction in carrier-to-phase-noise ratio caused by the frequency multiplication by the factor 11.

A low phase noise level of the 22 GHz source is achieved by means of what we call a "self injection locking" scheme (Fig. 2). The output signal of a Gunn oscillator is passed through a high-Q resonator and reinjected into the

0-7803-0476-4/92 $3.00 © 1992 IEEE



f_{CH_4} = 88 376 181 600 160 Hz ± 200 Hz

(unperturbed transition, hfs resolved)

Fig. 1: Schematic of optical frequency chain. The spectra show the two beat signals involving the "Carcinotron" . (Scale: 20 dB; 2 MHz; 100 kHZ.

oscillator. Use of a circulator ensures that there is only a single oscillation mode possible. The stabilization factor of this set-up is 36 dB. Dielectric resonator oscillators (DRO) have now become available, which could replace the system described.

At Fourier frequencies below 100 kHz, as shown in Fig. 3, the lowest timing noise level is that of a 100 MHz quartz crystal oscillator, to which the Gunn is therefore phase-locked with about that servo band-width. The crystal oscillator is then slowly (1 Hz) locked to an atomic reference.

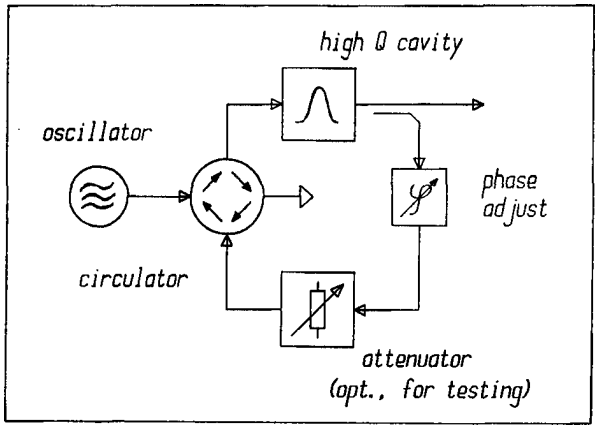


Fig. 2: "Self Injection Locking". This scheme is used to reduce the phase-noise level of the 22 GHz Gunn oscillator by 36 dB.

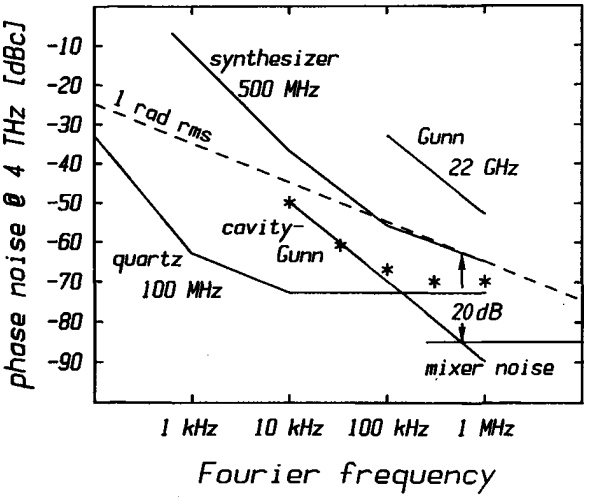


Fig. 3: Phase-noise spectral densities of various oscillators, as multiplied to 4 THz.

Mixers

In the chain, MIM diodes are employed for the two highest frequency steps, semiconductor (Schottky) diodes for high order harmonic mixing up to 4 THz.

Processing of the mixing products:

The frequencies in the chain are chosen so as to produce beat frequencies in the 30 to 100 MHz range. These are used for phase-locking the various oscillators in the chain and for taking the final frequency measurement. The chain is effectively cut in the middle: the upper part is down-locked from the methane optical frequency standard, the lower part is locked to the 100 MHz standard frequency output of a hydrogen maser. In the middle, at 4 THz/ 28 THz the data are taken by means of a digital counter.

The FIR laser was difficult to phase-lock because of its slow tuning response and large vibration induced frequency fluctuations. This laser is therefore being used quasi free-running, i.e. with a slow frequency servo control only. Combining the two beats of this laser, with the Carcinotron and with the CO_2 laser, yields a signal frequency independent of the fluctuating FIR frequency (Fig. 4).

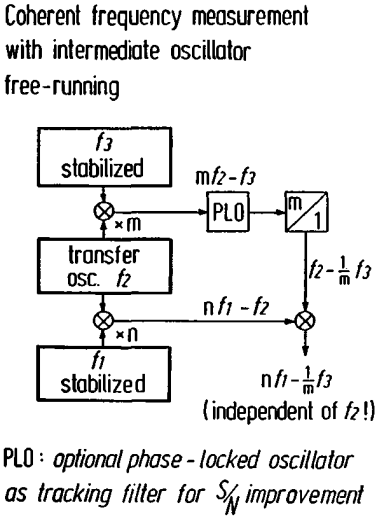


Fig. 4

Instead of taking frequency measurements we could, by comparing to a synthesized nominal frequency, also have generated an error signal at this stage and used it for controlling either the He-Ne laser ("Optical Frequency Synthesis") or the crystal oscillator ("Optical Clock").

Tracking Oscillators

Coherent measurement of a frequency means counting the zero-crossings of a signal, as opposed to the determination of the center of a spectral line. For the generation of unambiguously defined zero-crossings, we make use of (phase-locked) electronic tracking oscillators (Fig. 5), which process the beat signals. They are not only useful for the final data taking by means of a frequency counter, but also work as signal conditioners for the prescalers and digital phase-lock loops [3], which reliably and coherently lock the various members of the chain. While piezo driven mirrors severely limit the loop band width in phase-controlling a laser, the electronic tracking oscillator can track the signal with an optimized band width which may exceed several MHz, over a tuning range of more than 100 MHz.

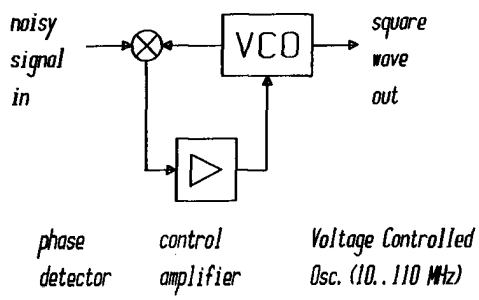


Fig. 5: Tracking oscillator schematic. The control loop band-width can be matched to the noise characteristics of the signal.

Coherence Margin

Counting errors (cycle slipping) may occur, if the phase excursions at the phase detector amount to more than about a radian. These phase excursions are caused by the frequency fluctuations of the oscillators involved, and by additive noise from the harmonic mixers. The phase detector cannot distinguish the noise from the frequency fluctuations (which constitute the signal to be tracked!), and the control loop tries to balance the noise by additional fluctuations introduced into the VCO frequency.

If the noise spectra are known, the total rms phase excursion can be calculated as the sum of the additive noise S_n picked up by the control loop and the portion of the oscillator phase noise $S_{e\varphi}$ not being tracked out, i.e. falling outside the control band width f_c:

$$\langle \varphi^2 \rangle \approx \int_0^{f_c} S_n \, df + \int_{f_c}^{\infty} S_\varphi \, df$$

With the loop transfer functions and for white frequency noise:

$$\langle \varphi^2 \rangle = \frac{\pi}{2} \cdot f_c \cdot \left[S_n + S_\varphi(f_c) \right]$$

The phase noise diagram in Fig. 3 illustrates the situation for measurement of the 4 THz laser in the chain. If the frequency axis is interpreted as the selected control band width, then the dashed line indicates the 1 radian reference. The intercept point of the cavity Gunn phase noise with the mixer noise level is below this line by a "coherence margin" of 20 dB. That such a margin is sufficient, was tested by locking two tracking oscillators with widely different band widths (1 MHz and 10 kHz) and comparing their counts. For this purpose their outputs were each frequency divided (by 10) and the phase difference of the divider outputs was recorded. Cycle slipping rates below 1 per 600 s were observed, which, at 4 THz, corresponds to a frequency measurement error of less than a part in 10^{15}.

Continuous Counting

As noted before, the actual measurement and recording of the optical frequency is carried out by counting the beat frequency generated at 4.2 THz. The $\pm$ 1 count uncertainty then roughly corresponds to $\pm\, 3 \cdot 10^{-13}$ for a single 1 s gate time measurement. Averaging a number n of successive measurements would bring the statistical error down at a rate of $1\sqrt{n}$, but leave any systematic error unchanged. Long gate times have the disadvantage that short disturbances may go unnoticed. We have therefore replaced the gated counter with a totalizing counter, which is never reset but whose contents is periodically (at 1 pps) strobed and transferred to a latch register. It is clear that some synchronization is necessary and that the counter has to be given time to stabilize. Details of this will be given elsewhere. A shift register is used for the latch, and the data is subsequently transmitted to a PC in standard asynchronous format. Several counters of this kind can easily be combined and all calculations, including the one shown in Fig. 4, can be carried out in the PC. As yet the method has only been applied to the two beat frequencies involving the free-running FIR laser, but it is clear that it can replace the phase locks of at least all the gas lasers in the chain. This means that instead of phase-locking all the members or links of the chain, we would take synchronized phase/count measurements of the beat signals.

Fig. 6 shows the Allan variance of a selection of frequency standards and of three optical frequency measurements. The resolution of the counting scheme indicated is for the most straightforward evaluation.

Outlook: While its capability of measuring the methane line is satisfactory, the chain will have to be extended to other interesting optical frequency standards. At Max-

Planck-Institut in Garching (Prof. T.W. Hänsch) work is underway to a high-precision measurement of the 1S-2S uv line of hydrogen, using a transportable methane stabilized laser calibrated by our chain. In our laboratory we are preparing connection of the P(12) and P(14) lines of the CO_2 laser, which are about 1 THz away from the R(32) used in the chain. P(12) is close to the 16th subharmonic of the Calcium frequency standard. Measurement of this optical frequency is pursued in collaboration with the PTB length metrology laboratory. Both, P(12) and P(14) can be tuned to strong OsO_4 molecular resonances, which make very promising frequency references in the mid-infrared.

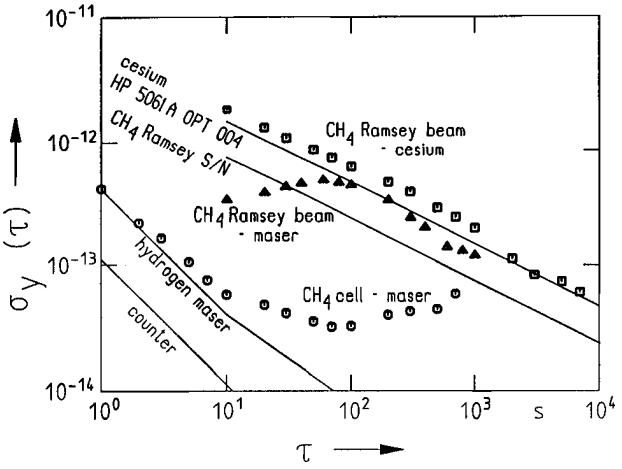


Fig. 6: Allan variances of two optical frequency measurements.

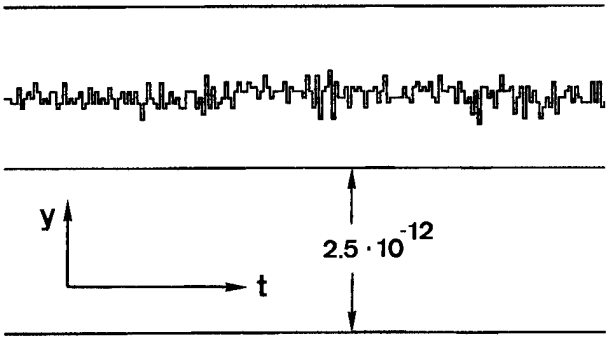


Fig. 7: Sample recording of optical frequency, using continuous counting. Averaging interval: 3 s, time resolution: 240 fs, duration 10 minutes

References

[1] C.O. Weiss, G. Kramer, B. Lipphardt, E. Garcia, "Frequency Measurement of a CH_4 Hyperfine Line at 88 THz/"Optical Clock", IEEE Journal of Quantum Electronics, Vol. 24, pp.1970-1972, Oct. 1988.

[2] G. Kramer, C.O. Weiss, B. Lipphardt, "Coherent Frequency Measurement of the Hfs-Resolved Methane Line", in Proceedings of the Fourth Symposium, Ancona, Italy, September 5-9,1988, "Frequency Standards and Metrology", Springer-Verlag, Berlin, 1989, pp. 181-186.

[3] J.L. Hall, M. Long-Sheng, G. Kramer, "Principles of Optical Phase-Locking: Application to Internal Mirror He-Ne Lasers Phase-Locked via Fast Control of the Discharge Current", IEEE Journal of Quantum Electronics, Vol. QE-23, pp. 427-437, Apr. 1987.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

Short and Long Term Stability of Optical Oscillators

Miao Zhu and John L. Hall *

Joint Institute for Laboratory Astrophysics
University of Colorado and National Institute of Standards and Technology
Boulder, Colorado 80309-0440

Abstract

We discuss the requirements for laser systems to be used as the local oscillator in an optical atomic frequency standard. We also present our experimental results for short, intermediate, and long term stability of the laser systems.

I. Introduction

The goal of producing a better time/frequency standard is of great interest. Since higher frequency transitions can have potentially higher Q-values, it is worthwhile to consider in detail the challenges associated with making an atomic frequency standard in the optical frequency domain.

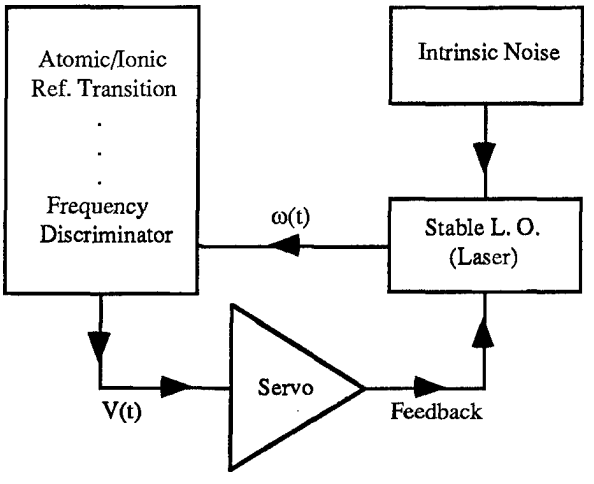


Fig. 1. Block diagram for optical atomic frequency standard.

The principles of an optical atomic frequency standard are the same as those for the microwave atomic frequency standard. As shown in Fig. 1, the frequency error between a local oscillator (a laser) and a pre-chosen atomic (or ionic) transition is fed back so that the frequency of the local oscillator is locked to the center of the reference transition. However, the optical atomic frequency standard presents two new challenges: firstly one must find an atomic (or ionic) reference transition with properties suitable for high resolution optical spectroscopy, and secondly one must develop a suitable coherent optical source (a laser system) to serve as the local oscillator.

The use of narrow transitions in ions as the reference transitions for an optical atomic frequency standard has been studied in detail.[1] The ion trapping technique uses an electric monopole of the ion to provide a long interaction time which is required by the uncertainty principle in order to obtain a narrow transition linewidth. Electrostatic interactions between the ions, however, limit the number of ions that can be usefully stored in a trap, thus limiting the attainable signal-to-noise ratio. On the other hand, the mutual interaction among neutral atoms is much smaller, so that a large number of atoms could be used in a single measurement to improve the signal-to-noise ratio. Laser cooling/trapping techniques [2, 3] now allow samples containing more than 10^{10} cold atoms to be prepared. Furthermore, optical transitions with narrow natural linewidths are known to exist in atoms. For example, the intercombination transitions in Mg and Ca have sub-kHz linewidths, while atoms such as Ag or Ba have optical transitions with sub-Hz linewidths.[4]

The use of the appropriate laser, atom-cooling/atom-trapping technique can provide a group of atoms with sub-mK kinetic energy which gives the second order Doppler shift $\frac{\Delta\nu}{\nu} < 10^{-20}$. However the interaction time between the laser-cooled atoms and the interrogating field is limited by the free-falling time of the atoms in the Earth's gravity field. A good method that circumvents this limitation requires that the laser-cooled atoms be launched vertically to form an atomic fountain.[5, 6] Optical Ramsey fringes can be generated after the atoms pass the interaction zone twice. In Fig. 2 we show the atomic fountain signal using a frequency-shift method with laser-cooled Na atoms. The fraction of atoms that is re-caught in the three dimensional cooling region (often called "optical molasses") is limited by the laser beam configuration (12 degrees off vertical, limited by the existing apparatus).

0-7803-0476-4/92 $3.00 © 1992 IEEE

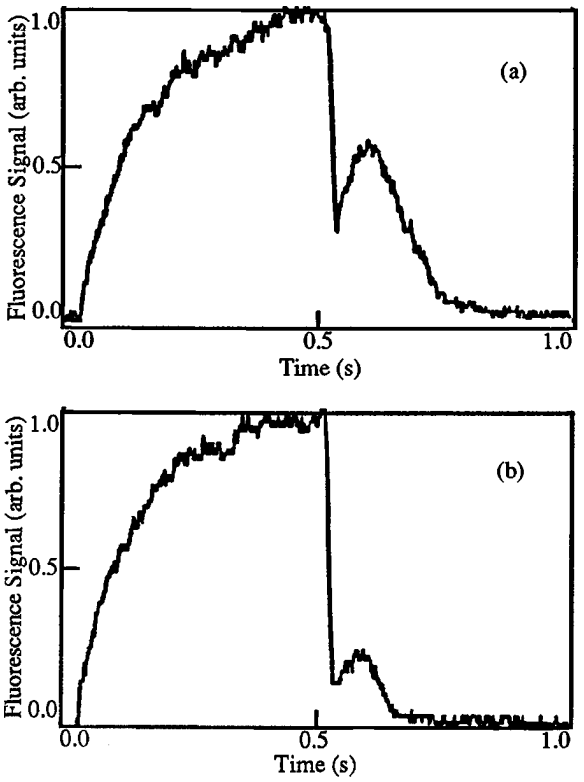


Fig. 2. Fluorescence signal of atomic fountain using laser-cooled Na atoms. The optical molasses is loaded with slow Na atoms up to 0.5 s. Then the atoms are launched upwards in a direction 12 degrees off vertical for 20 ms by shifting the frequency of the downward laser by (a) 1 MHz, and (b) 2 MHz corresponding to making optical molasses in a moving frame with 0.3 m/s and 0.6 m/s respectively along the launching direction. After 20 ms launching time the frequency is shifted back so that the optical molasses is again in the laboratory frame. Part of the launched atoms fall back to the optical molasses region and get re-caught. The fraction of the atoms which is re-caught is limited by the 12 degree off vertical launching direction. See text.

The second challenge requires a local oscillator (the laser system) with the following characteristics: its short term linewidth is much narrower than the atomic reference transition, it has precision tunability such that its frequency can be tuned across the atomic reference transition or its frequency can be locked to the center of the atomic reference transition frequency, and in the long term its frequency is predictable even when it is not locked on the atomic reference transition. In this paper we concentrate on this challenge, i. e., to make a laser system with good short, intermediate, and long-term stability.

Figure 3 illustrates a laser system that can be used as the local oscillator in an optical atomic frequency standard. Since the linewidth of a free-running laser is often broader than that of the atomic reference transition, the laser linewidth has to be reduced, e. g., by locking to a passive reference with good short term stability. Typically such a device is a high finesse Fabry-Pérot cavity. With contemporary cavity mirrors, the laser linewidth can be reduced to the sub-Hz or milliHz level. A large number of environmental disturbances, e. g., acoustic noise, building floor vibrations, etc., can change the resonance frequency of the cavity, thus modulating the frequency of the laser which is locked to the cavity. Therefore the cavity needs to be isolated from environmental disturbances. So as to have long term stability, the cavity usually does not have any tuning elements. To fulfill the tunability requirement, an optical phase-locked-loop (PLL) can be used to lock the optical phase of another laser to the cavity-locked laser such that most of the power of the second laser is within the bandwidth of the first laser. When used with an rf frequency synthesizer, the additional heterodyne stage required by this optical PLL provides the necessary programmable precision tunability without degrading the optical linewidth of the system. Hence this second laser is ready to be used to interrogate the atomic reference transition, and its frequency can be locked to the center of the reference transition.

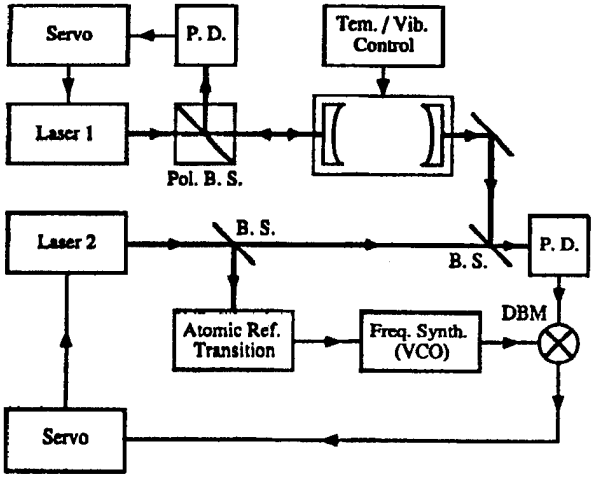


Fig. 3. Laser system to be used for optical atomic frequency standard.

In the following sections, we describe in detail the experimental investigations of the short, intermediate, and long term stability of a laser system performed in our laboratory.

II. Short term stability

To reduce the linewidth of a free-running laser, we can stabilize its optical phase/frequency relative to a passive reference with good short term stability. Although the pre-selected atomic reference transition in the optical atomic frequency standard, in principle, could serve as this

passive reference, the nonlinear nature of the field-atom interaction and the available number of (slow) atoms usually limit the signal-to-noise ratio within the required servo bandwidth. Instead we use either a high finesse Fabry-Pérot cavity (or other kinds of optical interferometers) or another stable laser as this optical phase/frequency reference. As shown in Fig. 3, a cavity is used as the reference to reduce the linewidth of Laser 1. Then stabilized Laser 1 (the transmission beam of the cavity) serves as the reference for Laser 2.

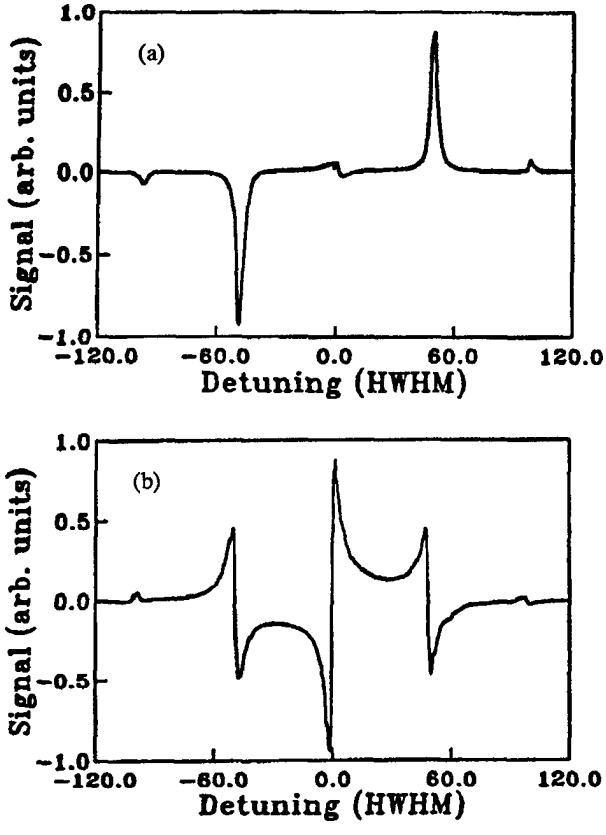


Fig. 4. The detected signal versus detuning in the reflection scheme for (a) absorption, (b) dispersion. Modulation frequency is about 48 times the cavity linewidth (HWHM). The detuning shown is expressed in terms of cavity linewidth (HWHM).

II A. Fabry-Pérot cavity locking

First, let us look at the case of stabilizing the laser frequency to a Fabry-Pérot cavity. When the laser frequency is in resonance, a Fabry-Pérot cavity (HWHM linewidth γ) stores the laser's electromagnetic field. The build-up time (or decay time) of this field is $1/\gamma$. The transmitted beam is the stored field which leaks out through the output mirror. The reflected beam consists of the coherent superposition of the stored field leaking out of the input mirror, and the current instantaneous field applied to, and reflected from, the cavity input mirror. Thus the signal detected in the reflection beam contains optical *phase* information on a short time scale ($t < 1/\gamma$), and the optical *frequency* information on a long time scale ($t > 1/\gamma$). Either the transmitted beam [7, 8] or the reflected beam [9, 10] from the cavity can be used to stabilize the laser frequency. Moreover, the signal can be derived at either dc [10] or rf frequencies,[8, 9] from a few tens of kHz to several hundred MHz. Because most lasers have large amplitude fluctuations at low frequencies, it can be difficult to reach the full signal-to-noise ratio as limited by shot noise in the detection process. This, and the absence of offsets from stray light or detector dark current, are the reasons one employs rf detection schemes.

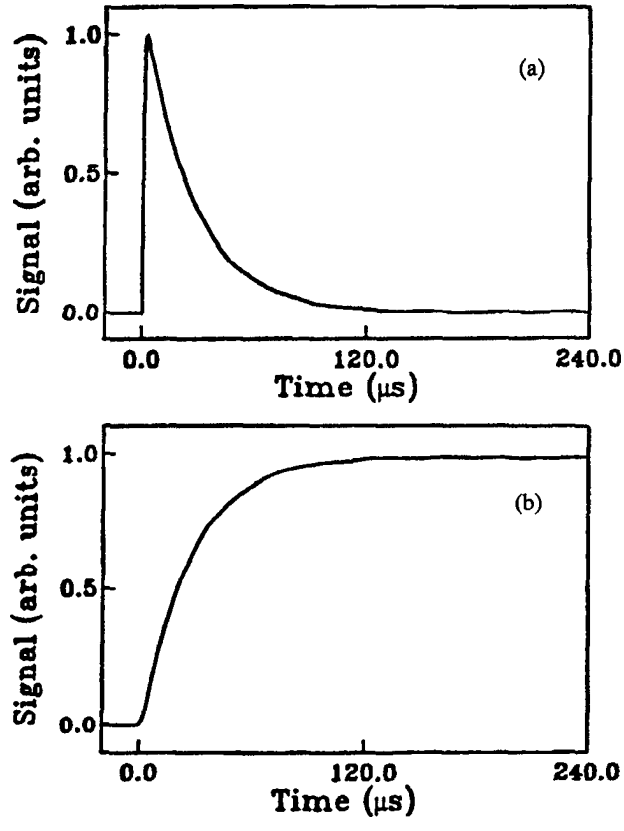


Fig. 5. Transient responses of a Fabry-Pérot cavity using the Pound-Drever scheme in the dispersion mode with zero detuning: (a) to a step phase disturbance, (b) to a step frequency disturbance. Cavity linewidth (HWHM) is 6 kHz. The modulation frequency is 12.5 MHz with modulation index ~ 0.9.

To take full advantage of the information available, the error signal can be detected at an rf frequency using the reflection mode topology (Pound-Drever scheme).[9] Figure 4 shows the steady response, in both the absorption mode and the dispersion mode, of the signal versus detuning, obtained using the Pound-Drever scheme. The odd symmetry makes the dispersion mode ideal for

improving the laser's short stability. The transient responses to step phase and step frequency disturbances are shown in Fig. 5 for the case where the detuning is zero in the dispersion mode. Figure 6 shows the steady state response (Bode plot) when the detuning is zero in the dispersion mode. In general, a modulation frequency higher than γ should be chosen in order to recover the full optical phase information.

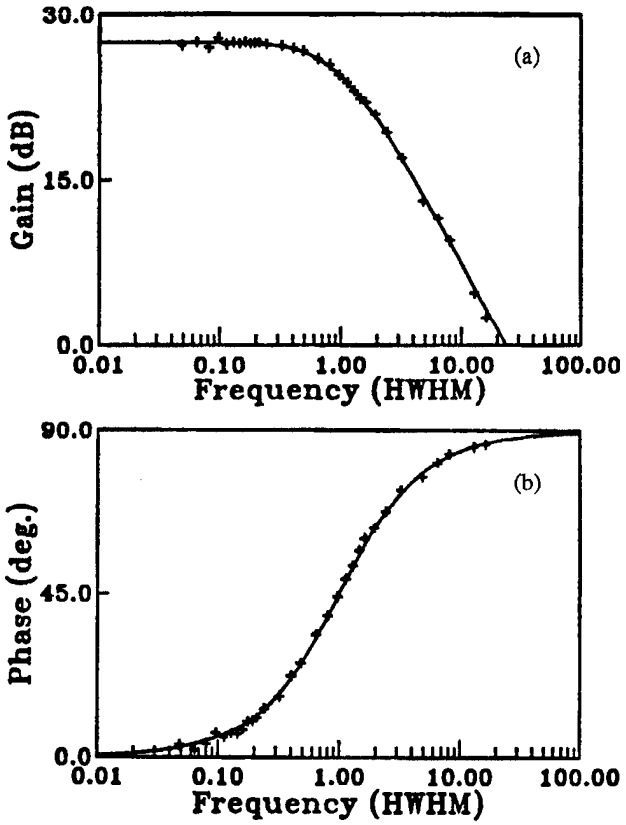


Fig. 6. Steady responses of a Fabry-Pérot cavity to FM noise using the Pound-Drever scheme in the dispersion mode with zero detuning: (a) relative gain, (b) phase delay. Cavity linewidth (HWHM) is 6 kHz. The modulation frequency is 12.5 MHz with modulation index ~ 0.9. The crosses are the measured data while the solid lines are theoretical curves.

To see what can be expected when one locks a laser's frequency to a Fabry-Pérot cavity, let us take one of our cavities as an example. The cavity linewidth is 12 kHz (FWHM) with a contrast of 1/3 for the fringes. The optical power on the detector is 100 μW, and we use a servo bandwidth of 10 kHz. Within this bandwidth the shot noise limited signal-to-noise ratio is 3.5×10^4 which gives the frequency noise spectral density 3.4 mHz/$\sqrt{\text{Hz}}$. The optical power spectral density $P_E(\omega)$ in this case is given by [11, 12]

$$P_E(\omega) = \frac{E_0^2}{\pi} \frac{\frac{\pi^2\langle f_N^2\rangle}{B}}{(\omega-\omega_0)^2 + \left(\frac{\pi^2\langle f_N^2\rangle}{B}\right)^2} \qquad (1)$$

where E_0 is the nominal amplitude of the laser's electric field, and $\sqrt{\langle f_N^2\rangle/B}$ is frequency noise spectral density in units of Hz/$\sqrt{\text{Hz}}$. The optical line shape is Lorentzian with a full linewidth of $\Delta f_{FWHM} = \pi\langle f_N^2\rangle/B = 36$ μHz. Note that the servo bandwidth is less than the cavity linewidth, so we are utilizing only optical frequency information. Thus we do not expect to put the power in the carrier itself because of the measurement noise.

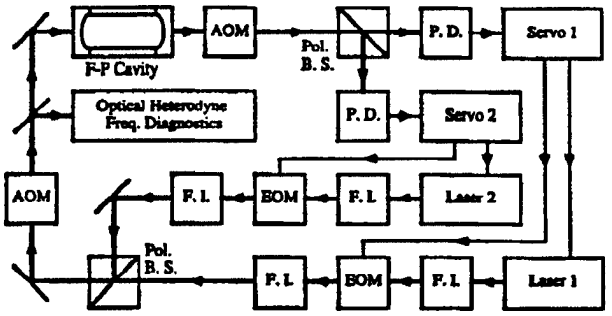


Fig. 7. Schematic for He-Ne laser cavity locking. See text for details.

To check the reality of such remarkable laser frequency stabilization, it is useful to lock a pair of lasers to a single cavity, and so to reveal the laser frequency stability with only a small influence of the cavity resonance frequency fluctuation. Figure 7 is a schematic of the setup for stabilizing the laser frequency to a Fabry-Pérot cavity.[13] This cavity, with a Zerodur [14, 15] spacer, has a free spectral range (FSR) of 250 MHz and a linewidth of 70 kHz (FWHM). There is a three-level hierarchy of frequency control stages for each of these two identical He-Ne lasers: a thin heating tape bonded on the laser tube, a coaxial tubular piezoelectric transducer (PZT) cemented to the laser tube, and a fast transistor to control the discharge current. Two stages of Faraday isolators in each laser beam path reduce the possible optical feedback. After putting FM sidebands on each laser beam, these two orthogonally polarized laser beams are combined spatially via a polarization beam splitter. The acousto-optic modulators (AOM), working at 80 MHz, provide further isolation for the optical feedback, since a reflected beam would have a 160 MHz frequency shift. This 160 MHz frequency shift in the reflected beams prevents the weakened back-scattered beams from re-entering the laser cavities and interacting with the atoms which provide gain for these particular lasing frequencies. Detecting the

transmitted beams of the cavity and using two independent servo systems, we lock these lasers on two adjacent longitudinal modes of the reference cavity. In this setup the heterodyne frequency between these two lasers is 250 MHz while the optical frequency is about 500 THz. The effect of changes of the cavity resonance frequency due to the environmental disturbances is reduced by a factor of ~ 2×10^6 in the heterodyne signal between these two lasers, thus allowing us to test the laser frequency stabilization system itself.

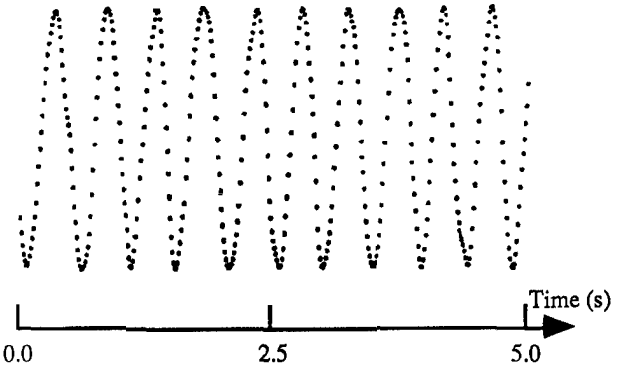


Fig. 8. Sampled waveform of optical heterodyne signal between two He-Ne lasers independently locking on the two adjacent longitudinal modes of a cavity. Optical beat frequency was one free spectral range of the cavity (~ 250 MHz). Resulting beat heterodyned to ~ 2 Hz with a phase-stable frequency synthesizer locked to an excellent quartz reference.

After an additional rf frequency synthesizer lowers the heterodyne signal frequency to the 0-2 Hz range, we are able to visualize the temporal behavior of this diagnostic heterodyne signal. Figure 8 shows the heterodyne signal at ~ 2 Hz. It is seen that this down-converted optical heterodyne waveform simply looks like a sinewave. Fitting this waveform gives a beat linewidth ~ 50 mHz. Since this scheme uses two orthogonal polarizations to lock two lasers, the change in the birefringence, both temporal and stress dependent, of the cavity mirrors limits the optical linewidth in this case.

To test further the capabilities of our laser frequency stabilization system, we chose to stabilize the frequency of a commercial dye laser system.[11] In this system, the rapid thickness fluctuations of the free-flowing dye jet stream lead to much of the observed frequency (and intensity) noise. Figure 9 is a block diagram of the experimental setup. For the optical phase/frequency reference we use a Fabry-Pérot cavity with a free spectral range of about 330 MHz and a linewidth of 250 kHz (FWHM). To obtain information over a large bandwidth we use the Pound-Drever scheme with a 12.5 MHz modulation frequency. Rather than feedback directly to the laser as in the previous system, we correct the optical phase/frequency error in the laser beam with an *external* optical phase/frequency transducer.[16] This transducer consists of an electro-optic modulator (EOM) and an AOM. The unity gain frequency of the servo loop is about 4 MHz which is much higher than the cavity linewidth $\gamma/2\pi$ = 125 kHz (HWHM); thus we are able to use the optical *phase* sensitive mode for times less than $1/\gamma \sim 1.3$ μs. For times longer than $1/\gamma$, the information is *frequency* sensitive.

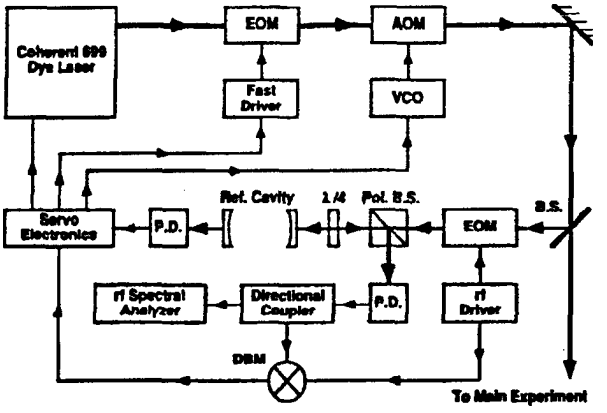


Fig. 9. Block diagram for locking the dye laser's frequency to a high finesse Fabry-Pérot cavity. The transmitted beam from the cavity is used to control the integrators in the servo loop to keep the dye laser frequency locked to the cavity even though there are some power drop outs due to the bubbles in the dye stream. See description in text.

To monitor the result of the experiment, a directional-coupler sends part of the error signal at 12.5 MHz to an rf spectral analyzer. Figure 10 illustrates the noise spectrum, due to the dye laser frequency noise, centered around the 12.5 MHz modulation frequency. The top (reference) curve shows the frequency excursions of the commercial laser locked only to its own reference cavity. For this full-scale calibration the laser is detuned from the resonance frequency of the external high finesse cavity by about a half cavity linewidth. Correcting for (deliberately introduced) amplifier saturation, a 12.5 MHz coherent signal level of -3 dBm results. When the laser is locked via the external transducer system (lower curves), the frequency noise is drastically reduced. The servo gain increases between each successive curve, until the noise level of the system is reduced (lowest curve) to near the measurement noise level (~ - 83 dBm at 1 kHz bandwidth) which is shown as the dashed line in Fig. 10. Below ~ 100 kHz, the shot noise of the measurement has been written as a low-level FM onto the laser's frequency, so that the "error point" shows less noise than this base measurement noise level. With the present cavity, this level is -110 dBc/$\sqrt{\text{Hz}}$, which corresponds to a spectral density of 0.395 Hz/$\sqrt{\text{Hz}}$. According to Eq. (1) this represents a linewidth of 0.49 Hz . The remaining 12.5

error in our preliminary version of the fast servo amplifier.

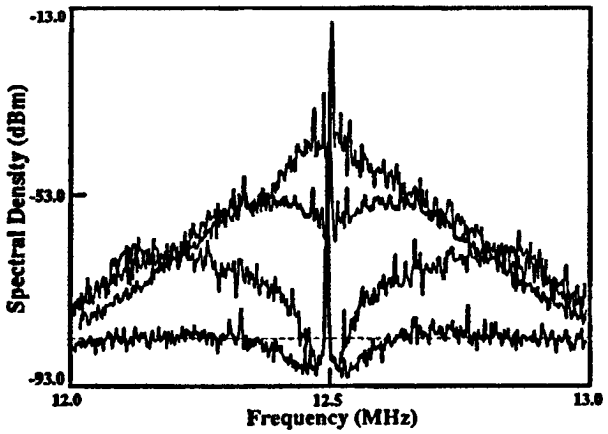


Fig. 10. Result of the cavity locking. Resolution bandwidth is 1 kHz. At the highest gain no intrinsic laser noise remains, and even the shot noise below 100 kHz has been written as an equivalent FM onto the laser. The dashed line shows the shot noise level at 1 kHz resolution bandwidth.

The final laser linewidth relative to the reference cavity in this experiment is limited by the measurement noise (shot noise) level due to the rather large linewidth (250 kHz FWHM) of the cavity used. It would be straightforward to produce a vastly tighter laser locking by employing a (stable) reference cavity of higher finesse, such as the ones employed in other experiments in our labs. For example with our finesse 30000 stable reference cavity, the discriminator slope could be increased by the finesse ratio, a factor $\frac{30000}{1320} = 22.7\times$ (since the optical efficiency of 25% is about the same for both cavities). The projected noise density would then be 22.7-fold smaller than the above-reported experimental value of 0.395 Hz/$\sqrt{\text{Hz}}$, i.e 17.4 milliHz/$\sqrt{\text{Hz}}$, leading to a Lorentzian laser linewidth of just below 1 milliHertz.

II B. Optical phase locking

Now let us consider the case of using a stable laser as a frequency shifter. Thus we could use a frequency-to-voltage (f-V) converter or phase detector in the rf frequency region to detect the *optical frequency difference* or *optical phase difference* between the reference laser and the laser under study. When the optical phase error is used as the correction signal, we can make an optical PLL in which the voltage-controlled-oscillator (VCO) is the laser to be stabilized.

The experimental setup [11] is illustrated in Fig. 11. An amplitude stabilizer, which is not shown in the figure, keeps the amplitude noise in the dye laser beam small enough so that the phase noise dominates the noise sidebands in the heterodyne signal. The dye laser beam is combined with the reference laser, a He-Ne laser working at 612 nm, by a beam splitter for heterodyne detection. An avalanche photodiode detects the heterodyne signal between these two lasers. Then the heterodyne signal, after being pre-amplified, is sent to an ECL phase/frequency discriminator which has a 45 MHz quartz crystal oscillator as its local phase/frequency reference. Optionally, as noted above, an auxiliary synthesizer-driven doubly-balanced-mixer (DBM) allows us to have a variable optical beat frequency while keeping a fixed 45 MHz output signal. This lets us to tune the laser by an rf synthesizer. The phase error developed by the phase detector is sent back to the external optical phase/frequency transducer for suppression of the phase/frequency noise of the dye laser. The heterodyne signal is displayed on an rf spectral analyzer to monitor the optical PLL performance. Additionally the phase error from the phase discriminator is shown on an oscilloscope for monitoring. From the phase-to-voltage (φ-V) conversion curve of the phase detector, we know that the residual phase excursion is limited within the range ± 0.65 rad. For the purpose of further analysis the heterodyne signal is mixed in another DBM with the output signal from an rf frequency synthesizer, thus providing a 5 MHz waveform for digitizing. The sampling rate of the digital oscilloscope is 10^8 samples/s, with a maximum record length of 32,000 samples. In order to avoid aliasing, a low-pass filter with cut-off frequency at 25 MHz is used before digitization to block the mixer's sum frequency output at 85 MHz.

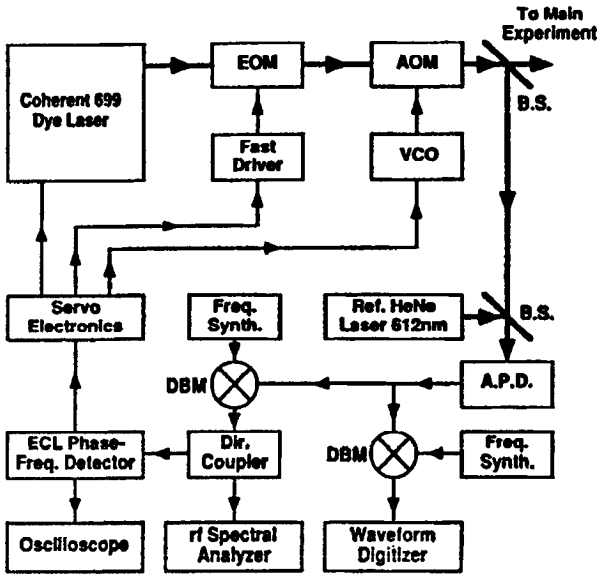


Fig. 11. Block diagram for optical phase locking the dye laser to a reference He-Ne laser at 612 nm. See text for details.

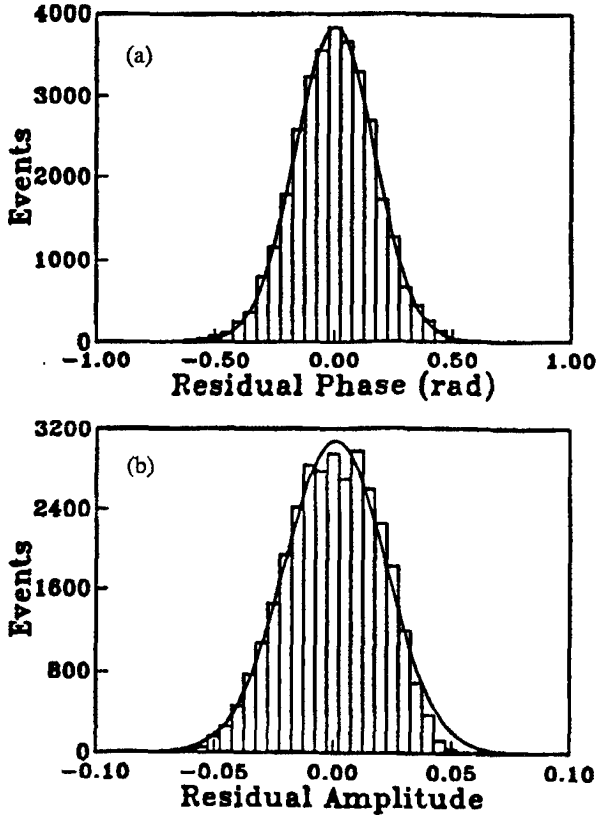


Fig. 12. (a) Histogram of the residual phase fluctuations $\varphi_N(t)$. (b) Histogram of the residual amplitude fluctuations $V_N(t)$. Solid lines are the Gaussian distributions fit to the data.

To interpret further the experimental result,[11] we can write the laser's output electric field as

$$E(t) = E_0[1 + V_N(t)] \exp\{-i\omega_0 t - i\varphi_N(t)\} \quad (2)$$

where E_0 is the nominal amplitude, ω_0 is the nominal angular frequency, $V_N(t)$ is the (noise) amplitude modulation, and $\varphi_N(t)$ is the (noise) phase modulation. We assume that $V_N(t)$ is an ergodic stochastic process and its variance is much smaller than unity. We also assume that $\varphi_N(t)$ is a ergodic Gaussian stochastic process. Using the definition of the autocorrelation function $R_U(\tau)$ for the time-varying quantity U(t) as

$$R_U(\tau) = \langle U(t)U^*(t+\tau) \rangle \quad (3)$$

and assuming that $V_N(t)$ and $\varphi_N(t)$ are uncorrelated, we obtain the autocorrelation function for the electric field

$$R_E(\tau) = E_0^2 e^{i\omega_0\tau} e^{-R_\varphi(0)} + E_0^2 e^{i\omega_0\tau} e^{-R_\varphi(0)} \{[1+R_V(\tau)]e^{R_\varphi(\tau)}-1\} . \quad (4)$$

Using the Wiener-Khintchine theorem,

$$P_E(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} R_E(\tau)e^{-i\omega\tau}d\tau , \quad (5)$$

we know that the first term in Eq. (4) gives the carrier while the second term in Eq. (4) gives noise sidebands. Hence we can obtain the optical power spectral density $P_E(\omega)$ if we know both $R_V(\tau)$ and $R_\varphi(\tau)$.

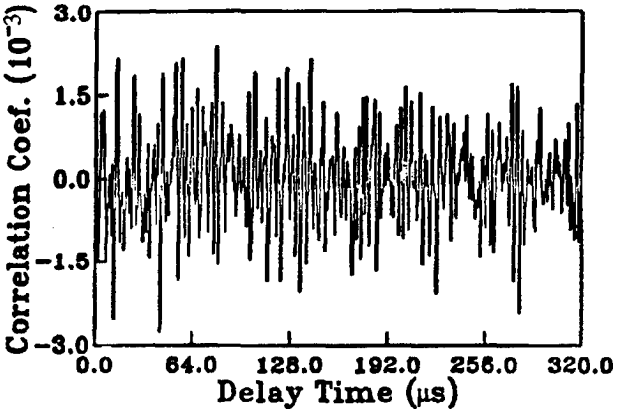


Fig. 13. The correlation coefficient of the phase fluctuations and amplitude fluctuations as defined by Eq. (6). Basically the amplitude and frequency fluctuations are uncorrelated. See text.

To obtain $V_N(\tau)$ and $\varphi_N(\tau)$ (and consequently $R_V(\tau)$ and $R_\varphi(\tau)$), we process the digitized waveform in the following way. First the recorded heterodyne signal at 5 MHz is fit to a sine wave using a nonlinear least-squares fitting routine. Four parameters (dc offset, amplitude, frequency, and initial phase) are included in the global fitting which uses all 32,000 data points for each data set. Keeping the globally-determined dc offset and frequency fixed, we fit the amplitude and phase locally using a subset containing five successive data points. The next subset of data partially overlaps the previous one. This fitting process gives the amplitude fluctuation $V_N(t)$ and phase fluctuation $\varphi_N(t)$. Figure 12 shows the histograms for phase and amplitude fluctuations of a typical data set. It also shows the Gaussian distribution curves which are fit to the data histograms. It can be seen that the Gaussian distribution describes the phase fluctuations rather well, while the amplitude fluctuations conform somewhat less accurately to a Gaussian distribution. Note

that to obtain Eq. (4) we did not require $V_N(t)$ to be a Gaussian process.

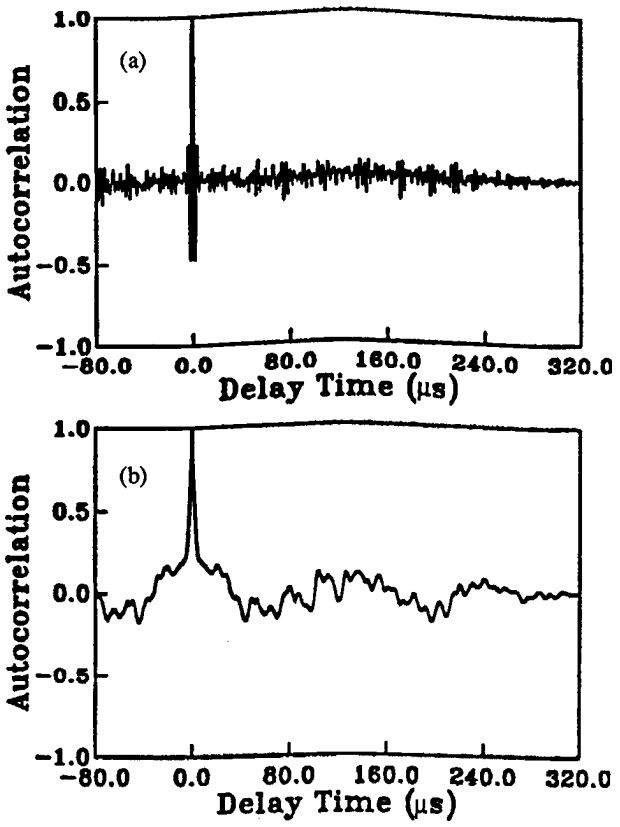


Fig. 14. Normalized autocorrelation functions of residual phase fluctuations, (a), and of residual amplitude fluctuations, (b), under phase-locked conditions. Finite high frequency gain and some remaining settling problems in the AOM/EOM crossover region lead to some residual phase errors. Still, 97% of the emission is captured by the optical phase-locked loop. See text.

The correlation coefficient of the phase fluctuations $\varphi_N(t)$ and the amplitude fluctuations $V_N(t)$ of the heterodyne signal, which is defined as

$$C(\tau) = \frac{\langle V_N(t)\varphi_N(t+\tau)\rangle}{\sqrt{\langle [V_N(t)]^2\rangle \langle [\varphi_N(t)]^2\rangle}}, \qquad (6)$$

is shown in Fig. 13. It is seen that basically $V_N(t)$ and $\varphi_N(t)$ are not correlated. This is not unexpected since the dye laser's amplitude noise and phase noise are suppressed by two separate servo-loops, with the residual fluctuations ideally due to the measurement noise in each loop. The reference He-Ne laser's amplitude noise modulation, which was not suppressed by any servo system, actually dominates the amplitude modulation of the heterodyne signal. Our further data reductions are based on the observation that $V_N(t)$ and $\varphi_N(t)$ are not correlated.

We can compute the autocorrelation functions for both phase and amplitude fluctuations as shown in Fig. 14. For this typical data set the variance for phase fluctuations is ~0.028 rad^2, which implies that more than 97% of power is in the carrier of this heterodyne signal according to Eq. (4). Equivalently this result shows that more than 97% of the dye laser power is put into the linewidth of the reference He-Ne laser. The variance of the amplitude fluctuations is ~4×10^{-4} which corresponds to a 2% amplitude fluctuation in the 25 MHz measurement bandwidth.

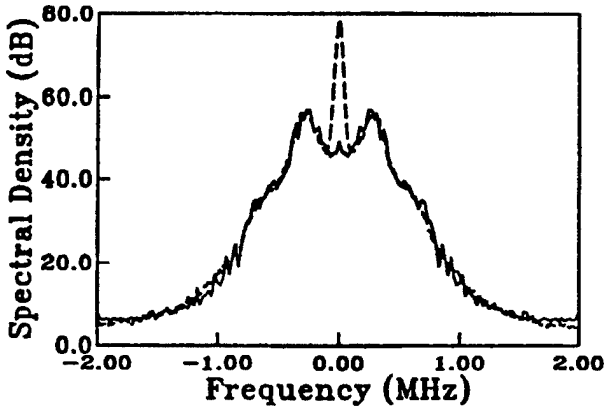


Fig. 15. Solid line is the noise power spectral density reconstructed by using the phase and amplitude fluctuations. Dotted line is the power spectral density of the heterodyne signal taken by the rf spectrum analyzer when the dye laser is phase locked to the reference He-Ne laser. The linewidth of the central peak is limited by the linewidth of the built-in filter of the rf spectrum analyzer. The resolution bandwidth of this display is 30 kHz.

According to the second term in Eq. (4), the power spectral density of the sidebands due to both phase fluctuations and amplitude fluctuations can be calculated using the autocorrelation functions $R_V(\tau)$ and $R_\varphi(\tau)$. The result is shown in Fig. 15. This figure also shows the heterodyne signal recorded by the rf spectrum analyzer. Note that the noise sidebands calculated theoretically from the digitized optical electric field waveform agree very well with the noise sidebands displayed by the rf spectrum analyzer. In our experimental setup, as noted previously, the residual phase fluctuations make the main contribution to the noise sidebands. It is likely that further thought concerning the phase-compensation in the servo amplifier design could reduce the peaks at $\frac{\omega-\omega_0}{2\pi} = \pm 200$ kHz which are mainly due to the phase difference at the cross-over frequency of EOM and AOM. Unfortunately the argon

pumping laser contributes substantial noise in this range which strains the servo loop considerably, and thus reveals some small residual defects in the otherwise smooth-settling of the loop following transient excitation.

III. Intermediate term stability

When a Fabry-Pérot cavity is used as the reference to improve a laser's short term stability, the fluctuations of its resonance frequency are mapped to the laser as frequency changes. The possible sources of these fluctuations include thermal expansion of the spacer, fluctuations of the index of refraction of the air in the cavity, acoustic noise in the laboratory, vibration of the building floor, etc. Therefore it is of great importance to choose a proper design for the reference cavity and to isolate it from the environmental perturbations, especially for stability on a time scale from seconds to hours.

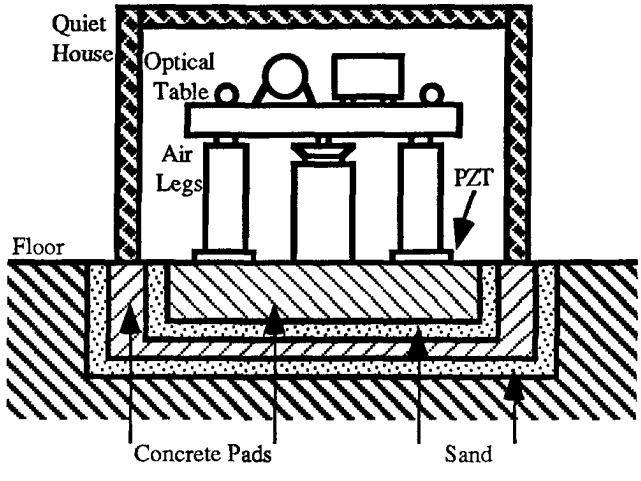


Fig. 16. Experimental setup for reducing environmental perturbations on the optical table. The 12 inch woofer is shown between the airlegs. See text for details.

To reduce the effect of the thermal expansion, materials with a low thermal expansion coefficient, e. g., Zerodur or ULE,[17] are chosen to make the spacer for the reference cavity. In addition, temperature stabilization of the shell of the cavity is also required. In our laboratory we use either Zerodur or ULE for our ultra-stable cavities. The cavity mirrors with Zerodur substrates are optically contacted to the polished ends of the spacers. To eliminate temperature-sensitive elements, such as a PZT or a galvo-motor which drives an intra-cavity Brewster-angled plate, these stable cavities are usually not tunable. With the dual layer cooling/heating temperature control of the aluminium shells within which the cavity is mounted, we reach a couple of mK long term temperature stability, corresponding to a few kHz in optical frequency. Correlation measurements show a prompt term (with ~ 30 minutes delay) plus an integrator term with a time constant of ~ 1 day.

The index of refraction of the air differs from unity by about 3×10^{-4} at 1 atmosphere.[18] Fluctuations in the number of air molecules in the laser beam path inside the cavity cause fluctuations of the cavity resonance frequency. To reduce the sensitivity to this effect, we attach a small ion pump to the reference Fabry-Pérot cavity, thus maintaining the pressure in the cavity to the level $p < 10^{-6}$ Pa. The corresponding index of refraction now differs from unity by $<4\times10^{-15}$, yielding a net frequency shift of < 2 Hz in the optical frequency. Fluctuations in this frequency shift can be safely ignored.

Although the He-Ne laser, as well as other gas lasers, is very sensitive to acoustic noise acting on the laser tube directly, a properly designed servo loop can suppress the frequency fluctuations when the laser frequency is locked to a reference cavity. However the reference cavity length change due to subkilohertz acoustic noise cannot be suppressed by this frequency stabilization servo loop (since the laser follows the cavity.) After a careful examination of all the requirements we constructed a sound-proof housing,[19] the Quiet House, as shown in Fig. 16. This Quiet House can be hoisted to the ceiling to allow personnel to work on the experiment during setup and adjustment stages. Then during data runs the Quiet House can be lowered over the optical table with the concrete isolation pad providing the final side of the acoustic isolation enclosure.

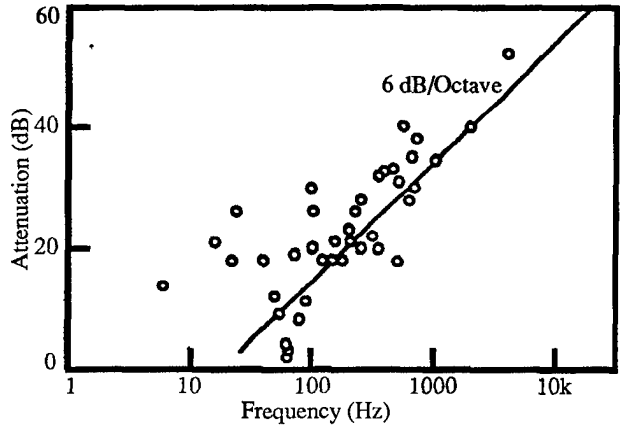


Fig. 17. Attenuation of the "Quiet House" in audible frequency range.

Figure 17 shows the acoustic noise attenuation of the Quiet House. The straight line represents the theoretical "mass law" corresponding to an attenuation of 6 dB/octave. Based on this measurement of the audible frequency attenuation, in the "mass controlled" region,[20] the mass density of the Quiet House is approximately 35.5 kg/m^2. A standing wave resonance at 59 Hz is

indicated. This standing wave is greatly reduced by using the absorbing pads under the optical table.

Building floor vibrations can deform the shape of the spacer of the reference cavity, thus changing the resonance frequency of the cavity. In this way the floor vibrations can cause modulation of the laser frequency. If the vibration consists of one or several discrete frequencies, it generates FM sidebands. However if the vibration is a stochastic process with low frequency components, it broadens the optical linewidth. Consider a Zerodur spacer with a length of l that is mounted vertically on one of its ends. In the Earth's gravity field the relative length change is

$$\frac{\Delta l}{l} = \frac{\rho l}{2Y} g \qquad (7)$$

where $g \approx 9.8\ m/s^2$ is the Earth's gravitational acceleration, and $\rho = 2.53 \times 10^3\ kg/m^3$ and $Y = 90.3$ GPa are the density and the Young's modulus of compression of Zerodur, respectively. For a $l = 30$ cm long cavity the optical frequency change is approximately 20 MHz/g. If we require that the frequency excursion due to the vibration is, say, less than 20 Hz, we have to reduce the vibration acceleration to 1 μg level.

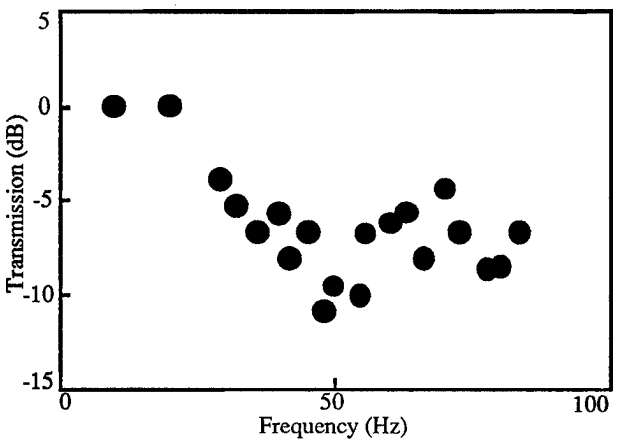


Fig. 18. Transmission of the seismic vibration of the dual concrete isolation pads.

The ideal solution to this problem might be to perform the experiment in a special vibration-free environment, e. g., on the space shuttle. However, this ideal environment is not so easily available because of the vibrations generated by gyros required by the attitude control system. In the laboratory our approach is to provide three stages of vibration isolation control, as shown in Fig. 16: the dual concrete isolation pads, airlegs, and an active vibration servo system.

In Fig. 18 we show the transmission of the concrete vibration isolation pads. Note that the dual isolation pads now provide about 7 dB vibration isolation above 35 Hz. Unfortunately, when the sand is not dry (as after a couple of floods in the laboratory), the attenuation decreases.

To isolate the experimental apparatus further from building vibrations, commercial airlegs are used to support the optical table. The top trace in Fig. 19 shows the vibration of the dual isolated concrete pads while the middle trace shows the effect of the airlegs. It is seen that the isolation at low frequency, < 40 Hz, is limited by the resonance frequency of the airlegs. Therefore active servo control of the vibration in this low frequency range is required.

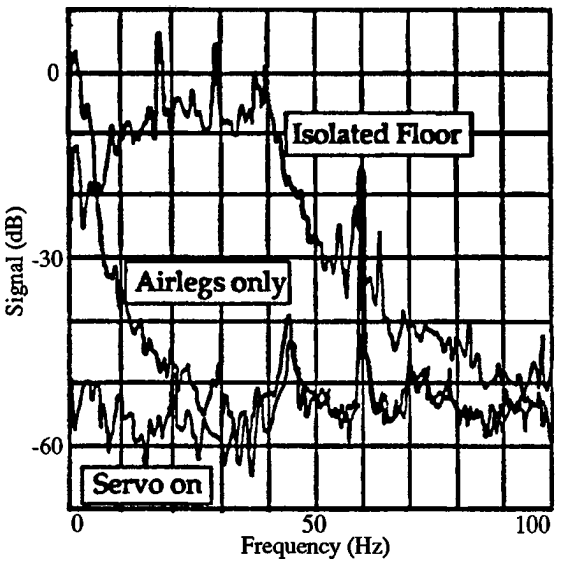


Fig. 19. Seismic vibration of the optical table with isolated floor only (top trace), with isolated floor and airlegs (middle trace), and with active servo system (bottom trace).

The sensor for this servo system is a combination of a tiltmeter and a seismometer. Using electromagnetic pneumatic valves we can control the amount of air pumped into the airlegs; thus we are able to control the level of the optical table at ~ 2 Hz which is limited by the slow response of the tiltmeter. The fast vertical actuator is a 12 inch diameter woofer coupled to the optical table via a compliant member. With the proper design of the driver and the servo-loop we are able to push the unity gain frequency up to ~ 50 Hz. The bottom trace in Fig. 19 shows that ~ 35 dB reduction of the seismic noise at low frequencies by this active servo system. This residual level is limited by the measurement noise (- 55 dB) of the seismometer. In Fig. 19, the calibration of the reference level (0 dB) gives $\frac{1\ \mu m}{s\sqrt{Hz}}$; thus the - 55 dB measurement noise is equal to $1.78 \times 10^{-9}\ \frac{m}{s\sqrt{Hz}}$. Integrating from dc

to 133 Hz we have a total seismic acceleration of 1 μg within this bandwidth. Above the 133 Hz cutoff frequency the modulation index is small enough that it will not seriously degrade the optical linewidth.

IV. Long term stability

By long term stability we mean stability on a time scale from hours to years for a laser frequency locked to a stable Fabry-Pérot cavity whose temperature is stabilized. The cavity is not, however, stabilized to an atomic reference transition. To test the long term stability, we use an iodine-stabilized He-Ne laser as our reference and record the heterodyne signal for diagnostics. This type of long term stability is useful not only for an optical atomic frequency standard system, but also for other precision measurements, e. g., Kennedy-Thorndike experiment.

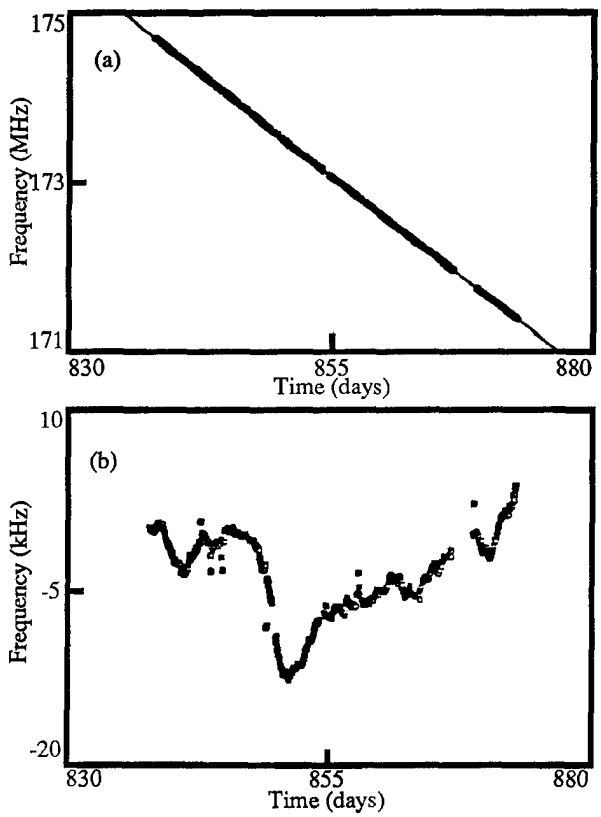


Fig. 20. Zerodur cavity frequency drift (a), and the residual frequency after the linear term 1.1 Hz/s is removed (b).

After we optically contacted the cavity mirrors to the spacer, we observed a series of "micro-glitch" events. A number of events with a frequency jump of ~ 50 kHz occurred in the first day. Then events of ~ 300 Hz occurred in the first couple of months. This type of "micro-glitch" event is probably due to a local interface change after the optical contact is made.

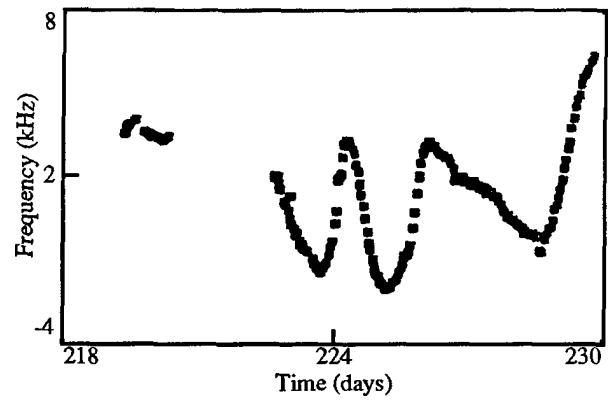


Fig. 21. Residual ULE cavity frequency fluctuations after a 0.25 Hz/s linear trend is removed.

Even after a high vacuum is established and the temperature is stabilized, there remains a residual slow frequency drift. This is due to the aging of the cavity spacer material. Figure 20(a) shows the frequency drift of the Zerodur cavity. After the present linear drift of 1.1 Hz/s is removed, the residual frequency is as shown in Fig. 20(b), which indicates that the frequency is predictable within ~ 1 kHz/day. Figure 21 shows the residual frequency of the ULE cavity after a 0.25 Hz/s linear drift is removed. The large drift rate of Zerodur is believed to be associated with the time dependent equilibrium of glassy and ceramic phases of this mixed-phase material. We found the aging rate is much reduced at ~ 27°C. This will be discussed in detail in another publication elsewhere. The smaller drift rate of ULE is probably due to the more uniform material and/or less stress in the material.

V. Summary

Here we report progress in making stable laser sources which can be used as local oscillators for an optical atomic frequency standard. For short term stability, we stabilized the laser frequency to a resonance frequency of a high finesse Fabry-Pérot cavity. We obtained a linewidth of 50 milliHz (FWHM) for He-Ne lasers and 0.49 Hz (FWHM) for a commercial dye laser system. By locking the optical phase of a commercial dye laser system to a reference laser, we were able to put more than 97% of the power into the carrier which means >97% of the dye laser power is within the linewidth of the reference laser. With both passive and active vibration isolations of the optical table, to which the reference cavities are mounted, we obtained <1 μg seismic vibration and <1 μrad tilt environment, corresponding to <20 Hz laser frequency excursion and <10 Hz frequency shift, respectively. To study the long term drift we obtained a

1.1 Hz/s drift rate for the Zerodur spacer, which agrees well with the 1×10^{-7} /year aging factor provided by the manufacturer.[14] For the ULE spacer, the drift rate is 0.25 Hz/s. Both cavities are shrinking.

For future improvement, an active vibration and tilt servo system with better sensors and better actuators is under development. For example, the woofers will be replaced by solenoids which can provide more force. The PZT sensor/actuator unit between airlegs and the concrete isolation pad could provide a better starting level. Additionally, a reference cavity with a different geometric configuration is under discussion. The Doppler shift, which is caused by the swing of the reference cavity, can be cancelled using a system which monitors the frequency shift of the reflected beam from the cavity and, after properly processing the data, feeds back to the frequency shift unit, e. g., an AOM. All these improvements are in progress. And of course, improvement in locking accuracy is of great importance in the reduction of all systematic errors. We are looking forward to using the stable laser system as the optical local oscillator for atomic fountain measurements.

Acknowledgments

We are indebted to Ch. Salomon and D. Hils for the work on He-Ne laser frequency stabilization, and D. Hils and J. Faller for the design and construction of the Quiet House. The research on laser frequency stabilization is supported at JILA in part by the National Institute of Standards and Technology under its program of research in fundamental metrology and in part by the National Science Foundation and the Office of Naval Research.

* Staff member, Quantum Physics Division, National Institute of Standards and Technology

References

[1] H. Dehmelt has pioneered these ideas about ion trap spectroscopy. See, e. g., Bull. Am. Phys. Soc. 20, 637 (1975).
[2] JOSA B 2, No. 11 (1985), special issue on the mechanical effects of light, P. Meystre and S. Stenholm, eds.
[3] JOSA B **6**, No. 11 (1989), special issue on laser cooling and trapping of atoms, S. Chu and C. Wieman, eds.
[4] J. L. Hall, M. Zhu, and P. Buch, JOSA B **6**, 2194 (1989).
[5] J. R. Zacharias, Phys. Rev. 94, 751 (1954).
[6] N. F. Ramsey, Molecular Beams, Oxford, Oxford U. Press, repr. 1985, p 138 ff.
[7] J. L. Hall, in International Conference on Lasers (Beijing, 1980), D. Wang, ed. (Wiley, New York, 1983).
[8] D. Hils and J. L. Hall, Rev. Sci Instrum. **58**, 1406 (1987).
[9] R. W. P. Drever, J. L. Hall, F. V. Kowalski, J. Hough, G. M. Ford, A. J. Munley, and H. Ward, Appl. Phys. B **31**, 97 (1983).
[10] T. W. Hänsch and B. Couillaud, Opt. Comm. **35**, 441 (1980).
[11] M. Zhu and J. L. Hall, submitted to JOSA B, and the references therein.
[12] D. S. Elliott, R. Roy, and S. J. Smith, Phys. Rev. A **26**, 12 (1982).
[13] Ch. Salomon, D. Hils, and J. L. Hall, JOSA B **5**, 1576 (1988).
[14] Schott Glass Technologies Inc., Duryea, PA. Also see [15].
[15] Mention of a commercial product is for technical communication only. It does not imply endorsement nor does it suggest that other products are necessarily less suitable for the application.
[16] J. L. Hall and T. W. Hänsch, Opt. Lett. **9**, 502 (1984).
[17] Corning Inc., Corning, NY. Also see [15].
[18] J. C. Owen, Appl. Opt. **6**, 51 (1967).
[19] D. Hils, J. E. Faller, and J. L. Hall, Rev. Sci. Instrum. **57**, 2532 (1986).
[20] C. M. Marris, Handbook of Noise Control, New York, McGraw-Hill, 1957.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LASER-COOLED NEUTRAL-ATOM FREQUENCY STANDARDS

William D. Phillips and Steven L. Rolston

National Institute of Standards and Technology, PHYS A167, Gaithersburg, MD 20899

Abstract

Recent developments in laser cooling show that neutral atoms hold significant promise for time and frequency standards. This paper summarizes the material presented at the 1992 IEEE Frequency Control Symposium and provides references to the published literature. Much of the material may be found in more detail in ref. [1].

Many of the most significant sources of uncertainty and inaccuracy in laboratory atomic time and frequency standards arise from the thermal motion of the atoms. Laser cooling is a powerful technique for reducing this thermal motion and improving the performance of such standards. Laser cooled trapped ions have a long history of development as atomic frequency standards and are certain to play a major role in future generations of standards [2].

More recently there has been increased interest in the use of neutral atoms for frequency standards. Compared to ions, neutral atoms are at a disadvantage because they cannot be trapped without seriously perturbing the internal structure of the atoms. Consequently, one considers using free atoms, which implies that the observation time will be short compared to the nearly unlimited time available with trapped ions. On the other hand, trapped ion frequency standards generally operate with a small number of ions because of the strong electrostatic ion-ion interactions. Neutral atom standards are not so severely limited and so should offer better signal-to-noise performance.

Recent advances in laser cooling now allow free atom observation times on the order of one second, so that linewidths on the order of a hertz are possible. Before 1988 it was generally accepted that laser cooling could reduce the temperature of atoms to the "Doppler cooling limit" given by $k_B T_D = \hbar\Gamma/2$, where Γ is the linewidth of the atomic transition used for cooling [3]. For Cs atoms this corresponds to an rms velocity of 9 cm/s.

Recently, temperatures significantly below this limit have been observed [4]. The theoretical explanation requires degeneracy of the ground state of the atoms [5,6] and implies that the new limit of laser cooling is several times the recoil velocity imparted to the atom on absorbing a photon [7]. Indeed, experiments with both Na [8] and Cs [9] confirm this, with the lowest Cs rms velocity being about 1 cm/s at a temperature of 2.5 μK.

Such cold temperatures mean that longer observation times are possible. A one second observation time with 9 cm/s tranverse velocities implies that the atoms spread out to a few hundred cm^2, while with a 1 cm/s velocity, the spread is a few cm^2.

To make a frequency standard having such a long observation time with free atoms one uses an atomic fountain, the idea of Zacharias [10] where atoms are launched vertically through a microwave cavity and fall back again, interacting twice with the microwaves in the manner of Ramsey's separated oscillatory fields. Since the discovery of laser cooling below the Doppler limit, such fountains have been demonstrated with Na [11] and with Cs [12]. The resonance linewidth of 2 Hz and the S/N ratio of 30 for 1 s demonstrated in the Cs fountain imply a short term stability if operated as a frequency standard of $3\times10^{-12}\,\tau^{-1/2}$, about equal to the best of present laboratory standards.

Expected improvements in the noise performance and in the number of atoms launched should lead to short term stability of $2\times10^{-14}\,\tau^{-1/2}$ or 10^{-16} in a day. One limitation in achieving such improvements is that at high Cs densities collisions will shift the resonance. Calculations [13] indicate that at densities of 10^7 cm^{-3} the relative frequency shift would be 10^{-15}. Preliminary experiments confirm this estimate [14].

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

Greater stability can be achieved by using a higher clock frequency, all other things being equal. We have analyzed metastable Xenon as an optical frequency standard [1] and find it to be very promising. The laser cooling of Xe* should be very similar to that of Cs. The clock transition is a two-photon transition at 2.19 μm, or about 10^{14} Hz and should be observable with a linewidth on the order of 1 Hz. Using conservative estimates for obtainable signal, we estimate a short term stability of $2X10^{-17}\ \tau^{-1/2}$. Clearly, considerable improvement in flywheel oscillators would be required to realize such performance, but the possibilities are certainly attractive. The accuracy of such a device is likely to be limited by ac Stark shifts which would require careful evaluation and stabilization [1].

Laser cooled neutral atom frequency standards offer significant opportunities for improvement over present standards, and they offer features that are in many ways complementary to those offered by trapped ion standards. It seems likely that in the future we can look forward to comparing different kinds of advanced, laser cooled, atomic frequency standards with each other.

References

[1] S. L. Rolston and W. D. Phillips, "Laser-Cooled Neutral Atom Frequency Standards," Proc. IEEE, vol. 70, pp. 943-951, 1991, and references therein.

[2] W. N. Itano, "Atomic Ion Frequency Standards," Proc. IEEE, vol. 79, pp. 936-942, 1991, and references therein.

[3] D. Wineland and W. Itano, "Laser Cooling of Atoms," Phys. Rev. A, vol. 20, p. 1521, 1979).

[4] P. Lett, R. Watts, C. Westbrook, W. Phillips, P. Gould, and H. Metcalf, "Observation of atoms laser cooled below the Doppler limit," Phys. Rev. Lett., vol. 61, p. 169, 1988.

[5] J. Dalibard and C. Cohen-Tannoudji, "Laser cooling below the Doppler limit by polarization gradients: Simple theoretical models," J. Opt. Soc. Am. B, vol. 6, p. 2023, 1989.

[6] P. Ungar, D. Weiss, E. Riis, S. Chu, "Optical molasses and multilevel atoms: Theory," J. Opt. Soc. Am. B, vol. 6, p. 2058, 1989.

[7] Y. Castin and J. Dalibard, "Quantization of Atomic Motion in Optical Molasses," Europhys. Lett., vol. 14, pp. 761-766, 1991.

[8] P. Lett, W. Phillips, S. Rolston, C. Tanner, R. Watts, and C. Westbrook, "Optical Molasses," J. Opt. Soc. Am. B, vol. 6, p. 2084, 1989.

[9] C. Salomon, J. Dalibard, W. Phillips, A. Clairon, and S. Guellati, "Laser cooling of Cs atoms below 3 μK," Europhys. Lett., vol. 12, p. 683, 1990.

[10] N. Ramsey, Molecular Beams. Oxford:Oxford University Press, 1985, p. 138.

[11] M. Kasevich, E. Riis, S. Chu and R. DeVoe, "RF spectroscopy in an atomic fountain," Phys. Rev. Lett, vol. 63, p. 612, 1989.

[12] A. Clairon, C. Salomon, S. Guellati, and W. Phillips, "Ramsey resonance in a Zacharias fountain," Europhys. Lett., vol 16, p. 165, 1991.

[13] E. Tiesinga, B. Verhaar, H. Stoof, and D. van Bragt, "Spin-exchange frequency shift in a cesium atomic fountain," Phys. Rev. A, vol 45, p. R2671, 1992.

[14] K. Gibble and S. Chu, private communication 1992.

Hg^+ TRAPPED ION STANDARD PERFORMANCE WITH THE SUPERCONDUCTING CAVITY MASER OSCILLATOR AS L.O.*

J. D. Prestage, R. L. Tjoelker, R. T. Wang, G. J. Dick, L. Maleki

California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

The frequency stability obtained with an ion trap based frequency standard where $^{199}Hg^+$ ions are confined in a hybrid rf/dc linear trap is described. The 40.5 GHz clock transition is measured to be as narrow as 17 mHz representing a quality factor of 2.4×10^{12}. A hydrogen maser interrogation of a 37 mHz Hg^+ resonance is used to measure 1×10^{-15} stability of the ion resonance at 10,000 seconds averaging time. In another measurement the frequency of the superconducting cavity maser oscillator is steered to follow the Hg^+ clock transition again providing frequency stability of 1×10^{-15} at 10,000 seconds.

Introduction

Atomic frequency standards with high stability for averaging times τ longer than 1000 seconds are necessary for a variety of astrophysical measurements and long baseline spacecraft ranging experiments. The CASSINI spacecraft, planned for a mission to Saturn, will carry a Ka-band transponder enabling the most sensitive search yet made for low frequency (0.1 – 1mHz) gravitational radiation. The high frequency Ka-band link will reduce interplanetary plasma scintillation noise to the level where current frequency standard instabilities may be the limiting noise source. The CASSINI gravitational wave search has set a goal for stability of the standards of 4×10^{-16} for averaging times $\tau \geq 1000$ seconds.

The performance of microwave frequency standards in use today are summarized in Fig.1 [1,2]. For short term stability ($\tau < 150$ seconds) the Superconducting Cavity Maser in use at JPL shows stability as good as 2×10^{-15} for averaging times up to 2000 seconds[3]. Hydrogen masers are presently the most stable frequency standard for $150 < \tau < 30,000$ seconds and are the primary standard in use in JPL's Deep Space Tracking system. For averaging times greater than 6 months the millisecond Pulsar PSR 1937 + 21 exceeds the stability of international timekeeping abilities at a level of 1-2 parts in 10^{-14} [2]. For $\tau \approx 10^6$ seconds the most stable clock yet measured is the Hg^+ ion standard based on 2×10^6 ions confined in an rf Paul trap[1,4]. In this paper stability measurements of a Hg^+ frequency standard based on $\sim 5 \times 10^7$ ions in a linear trap are reported.

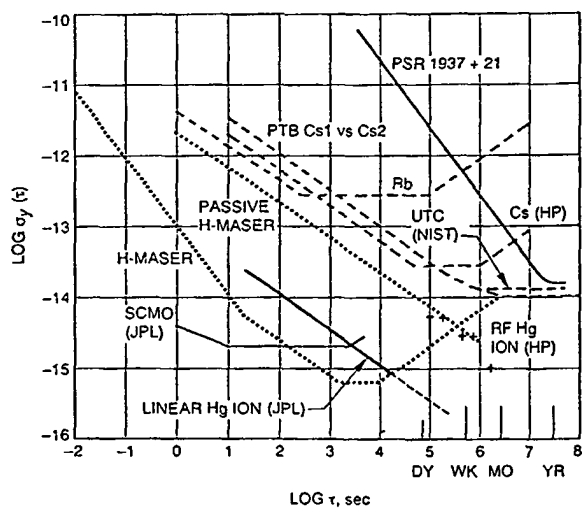


Figure 1. Fractional frequency stability of several precision frequency standards.

*This work represents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract sponsored by the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

Linear Ion Trap Based Frequency Standard

The advantages of the linear ion trap for frequency standard application have been previously described[5-8]. The linear trap used in this work is shown in Fig. 2. Ions are created inside the trap by an electron pulse along the trap axis which ionizes a neutral vapor of ^{199}Hg. A helium buffer gas (10^{-5} torr) collisionally cools the ions to near room temperature. Resonance radiation (194 nm) from a ^{202}Hg discharge lamp optically pumps the ions into the F=0 hyperfine level of the ground state[9]. This UV light is focused onto the central 1/3 of the 75 mm long ion cloud. The thermal motion of the ions along the length of the trap will carry all the ions through the light field so that pumping is complete in about 1.5 seconds for typical lamp intensities.

To minimize stray light entering the fluorescence collection system this state selection light is collected in a pyrex horn. Placement of the LaB_6 electron filament is also chosen to prevent light from entering the collection system.

Microwave radiation (40.5 GHz) propagates through the trap perpendicular to the trap axis satisfying the Lamb-Dicke requirement that the spatial extent of the ion's motion along the direction of propagation of the microwave radiation be less than a wavelength. Radiation enters the trap region through the pyrex horn and propagates in the opposite direction to the UV state selection/interrogation light. This allows collection of atomic fluorescence in both directions perpendicular to the incident pumping light. For the resonance and stability data shown in this paper, fluorescence was collected in only one of these two directions.

At the present time no thermal regulation is incorporated into the ion standard itself. However, for all stability measurements described the ion trap standard together with its support electronics was housed in an environmentally controlled test chamber where temperature variations were regulated to approximately ±0.05°C.

Operation of the ion trap as a frequency standard with a stable quartz crystal oscillator acting as a local oscillator (L.O.) was previously reported[10]. In such operation system performance is limited by fast frequency fluctuations of the L.O.[11-13]. In spite of this degradation, stabilities of $3.3 \times 10^{-13}/\sqrt{\tau}$ were reached which surpassed reference H- maser stability for $\tau \geq 10,000$ seconds[10].

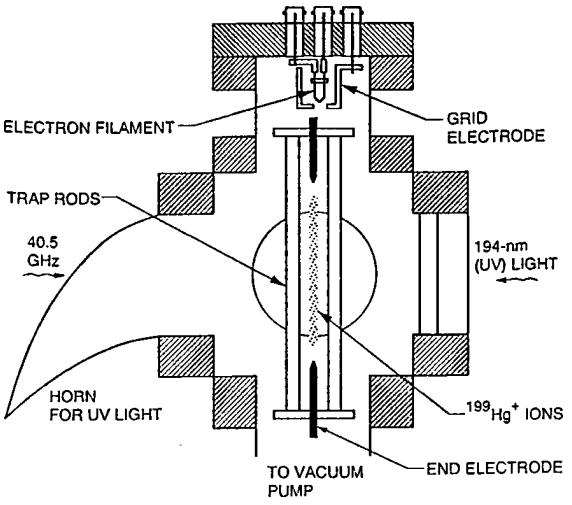


Figure 2. Linear ion trap assembly residing in its high vacuum enclosure. State selection light from the ^{202}Hg discharge lamp enters from the right, is focused onto the central 1/3 of the trap and is collected in the horn. Fluorescence from the trapped ions is collected in a direction normal to the page.

H-Maser Interrogation of the Hg^+ Clock Resonance

The technique of successive oscillatory fields[14] is used to probe the approximately 40.5 GHz hyperfine clock transition in $^{199}Hg^+$ ions confined to the linear trap described above. In the initial measurements the 40.5 GHz signal is derived from an active Hydrogen maser frequency source as shown in Fig. 3. A representative resonance line of the $^{199}Hg^+$ clock transition is shown in Fig. 4. State selection and interrogation is accomplished during the 1.5 seconds following the lamp turn on. It is necessary to switch the UV state selection/interrogation light level to near zero during the microwave interrogation period to prevent light shifts and broadening of the clock transition. A background light level of about 300,000 per 1.5 second collection period has been subtracted to generate the resonance shown. The data shown in Fig. 4 is an average of ten scans with a 10 mHz frequency step size. The central fringe is about 160 mHz wide.

The central portion of the narrowest resonance lines yet obtained with this apparatus are shown in Fig. 5. This line is derived from two 0.275 second pulses separated by a 30 second free precession period. The linewidth of 17 mHz represents a line Q of 2.4×10^{12} on

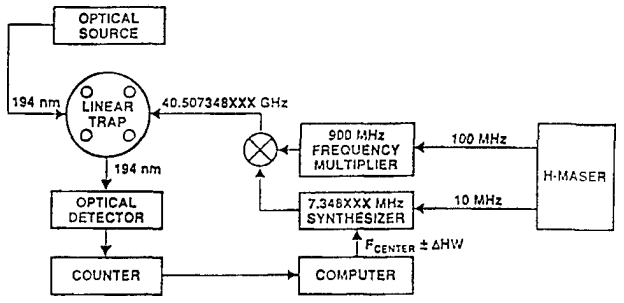


Figure 3. Schematic operation of the mercury ion trap using a hydrogen maser to interrogate the Hg^+ clock resonance.

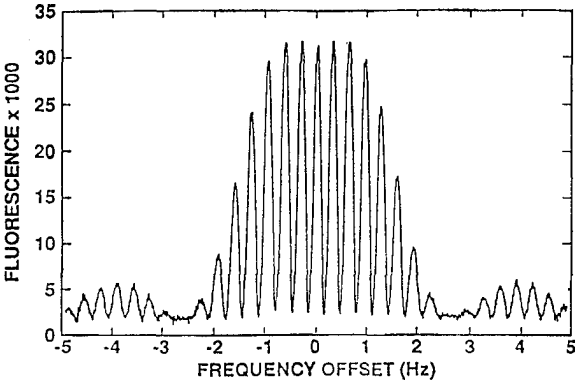


Figure 4. $^{199}Hg^+$ clock transition as measured with method of successive oscillatory fields. The central line is about 160 mHz wide and the data shown is an average of ten scans.

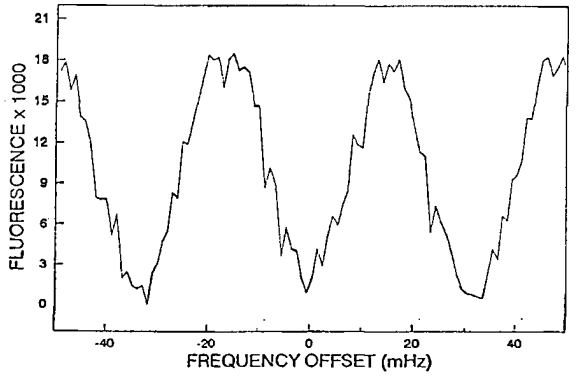


Figure 5. High Q Ramsey fringes resulting from two microwave pulses of 0.275 seconds separated by a 30 second free precession period. The fringes are about 17 mHz wide and the data shown is a single scan.

the 40.5 GHz transition. The data displayed is a single scan and, to our knowledge, is the highest Q transition ever measured in a microwave atomic transition.

To determine the frequency stability of the overall system of ions, trap, microwave source, etc., the output frequency of the 40.5 GHz source is locked to the central peak of a 37 mHz resonance in a sequence of 16384 frequency measurements. The time required for each measurement is about 18 seconds and the loop response time was 5 measurement cycles. By averaging the frequencies of 2^N adjacent measurements (N=1,2,..,13) we calculate the Allan deviation shown in Fig. 6. The stability measured in this mode of operation is 1×10^{-15} for a 10,000 second averaging time as can be seen in Figures 1 and 6. Beyond 10,000 seconds instability in the reference H-maser probably limits the measurement.

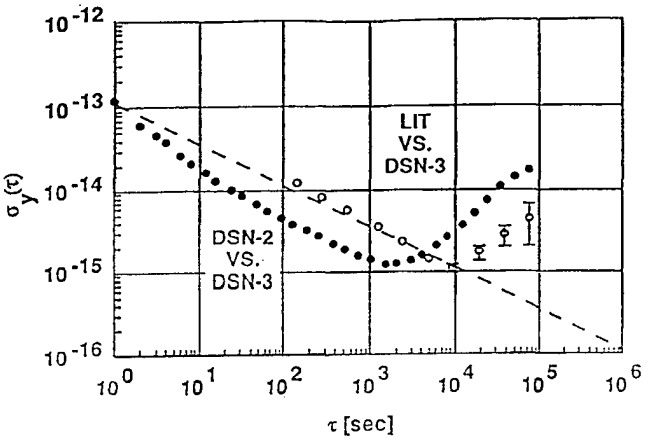


Figure 6. Performance of the Hg^+ system as determined by maser interrogation. The fractional frequency stability of the ion trap system is measured with a 37 mHz Hg^+ resonance line. Also shown is the stability comparison of the reference maser to a another H-maser.

Operation with the Superconducting Cavity Maser Oscillator

The Hg^+ standard has been operated using the superconducting cavity maser oscillator (SCMO) as the L.O. The SCMO consists of a high Q lead-on-sapphire superconducting cavity operated at liquid helium temperature[3]. The cavity excitation is provided by an ultra-low noise cryogenic ruby maser. Oscillation at a frequency of 2.69 GHz results from ruby maser op-

eration with a 13.1 GHz pump frequency to create a population inversion as shown in Fig. 7. Also shown is a magnetic tuning coil installed on the ruby housing to allow frequency adjustments to the SCMO output frequency. The fractional frequency sensitivity of this adjustment is 7×10^{-12} per mA of tuning current with a range of about 10^{-10}.

The uncompensated SCMO performance is $4 - 5 \times 10^{-15}$ for $1 \leq \tau \leq 1000$ seconds with a long term drift rate of 4×10^{-13} per day. The long term stability of the ion trap frequency standard was transferred to the SCMO in the arrangement shown in Fig. 8. The Hg^+ ion resonance was 37 mHz for the operation described here. Stability of the steered SCMO output frequency was measured against a H-maser and is shown in Fig. 9. The full stability as measured with the H-maser passive interrogation of the ion resonance has been transferred to the SCMO with no apparent degradation.

For averaging times in the range $1 \leq \tau \leq 10$ seconds, the measured stability is that of the DSN-2 maser since the SCMO is more stable during that period[3]. For the next decade, $10 \leq \tau \leq 100$ seconds, the H-maser is more stable thus showing the performance the SCMO to be about 10^{-14}. The servo time of the ion standard frequency lock of the SCMO is about 100 seconds so that the stability shown for $100 \leq \tau \leq 10,000$ seconds is that of the ion standard. During this period the performance is about $1.2 \times 10^{-13}/\sqrt{\tau}$ in good agreement with the data shown in Fig. 6. Beyond 10^4 seconds H-maser instabilities probably limit the measurement.

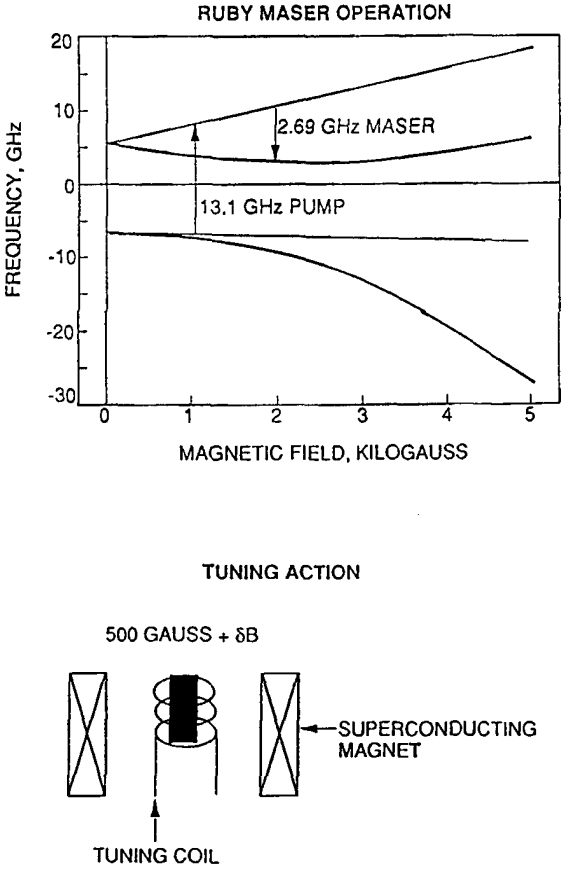


Figure 7. Ruby maser operation and the frequency adjusting current loop in the SCMO.

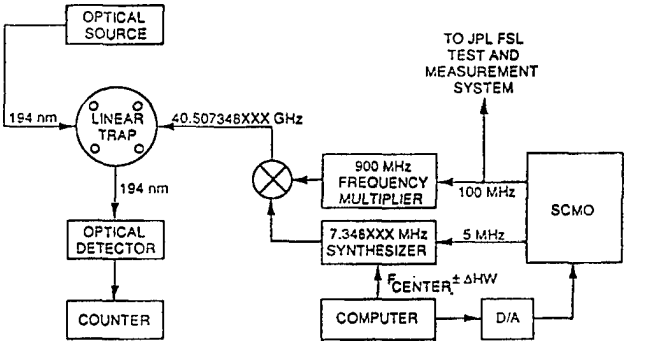


Figure 8. Schematic operation of the mercury ion frequency standard with the SCMO as a local oscillator.

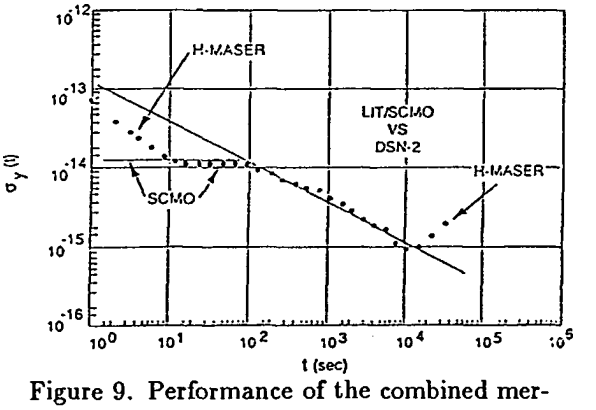


Figure 9. Performance of the combined mercury ion atomic resonator and SCMO local oscillator. For $\tau \geq 100$ seconds the stability is that of the ion standard, $1.2 \times 10^{-13}/\sqrt{\tau}$. For $\tau \geq 10{,}000$ seconds maser frequency instabilities probably limit the measurement of the mercury ion standard frequency stability.

Sources of Frequency Instability

While short term performance of the ion trap standard is determined by signal-to-noise ratio and line Q of the clock transition resonance, the long term stability is limited by the sensitivity of the atomic system to changes in environmental and operating parameters and on our ability to control and stabilize such parameters[15].

The second-order Doppler shift from ion motion driven by the trapping field is determined by measuring clock frequency as ion number, N, decays during the approximately $2000-15000$ second ion storage time for our trapping conditions. The frequency offset between our current operation with about 5×10^7 ions and very few ions where trapping field shifts are minimized is $1.5 - 2.0 \times 10^{-12}$. No active ion number stabilization was used in any of the measurements described here though the present procedure is sufficent to hold ion number variations to below 0.1% over the 10,000 second averaging time required to reach 1×10^{-15} stability.

The fractional sensitivity of the $^{199}Hg^+$ clock transition to magnetic field variations is nearly 1000 times less than that of hydrogen at the same operating field. For the present measurements the field was set at 3.5μT (35 mG). At this operating field the unshielded atomic sensitivity is 1.7×10^{-13} per mG. To prevent ambient field disturbances from influencing the ion frequency the trap region is surrounded by a triple layer magnetic shield with a shielding factor 10,000. With this shielding factor a 5 mG ambient field change would lead to a 1×10^{-16} shift in the atomic resonance frequency. Magnetic gradients must also be minimized in order to reach the highest frequency stability since gradients over the ion cloud can degrade atomic coherence and limit line Q. A partial solution has been to operate the standard at relatively high field settings (35 mG) but this also increases the atomic sensitivity to fluctuations in ambient field.

The temperature sensitivity of the complete system is measured to be less than $10^{-14}/^{\circ}C$ which probably comes from an increase in neutral mercury vapor and a consequent heating and/or ion cloud radius increase. The lamp and its housing must also be temperature controlled as the brightness is highly dependent on temperature.

Long term stability requires controlling these variables to high precision. Since the mercury atom is in general less sensitive to environmental changes than other atoms used in frequency standards an order of magnitude improvement may be obtained in long term stability with the same level of control as existing standards.

Summary

In its first laboratory implementation the JPL linear ion trap based Hg^+ atomic frequency standard has demonstrated the highest stability of any passive frequency standard. At averaging times $\tau = 10,000$ seconds, the Hg^+ standard shows stability competitive with the best H-maser frequency standards. Beyond 10,000 seconds this standard probably exceeds the stability of all frequency standards. A second linear ion trap Hg^+ standard is now being constructed to measure stabilities beyond 10,000 seconds.

Acknowledgment

We thank Bill Diener, Albert Kirk and Roland Taylor for assistance with the Frequency Standards Laboratory Test and Measurement System.

References

[1] D. W. Allan, M. A. Weiss, and T. K. Peppler, "In Search of the Best Clock", *IEEE Trans. Instrum. Meas.* 38 624-630, 1989.

[2] L. A. Rawley, J. H. Taylor, M. M. Davis, D. W. Allan, "Millisecond Pulsar PSR 1937 + 21: A Highly Stable Clock",*Science* 238 761-765, 1987.

[3] G. J. Dick and R. T. Wang, "Ultra-Stable Performance of the Superconducting Cavity Maser", *IEEE Trans. Instr. Meas.*, 174-177, 1991.

[4] L. S. Cutler, R. P. Giffard, P. J. Wheeler, and G. M. R. Winkler, "Initial Operational Experience with a Mercury Ion Storage Frequency Standard", in *Proc. 41st Ann. Symp. Freq. Control, IEEE Cat. No. 87CH2427-3*, 12-19, 1987.

[5] J. D. Prestage, G. J. Dick, and L. Maleki, "Linear Ion Trap Based Atomic Frequency Standard", *IEEE Trans. Instr. Meas.*, 132-136, 1991.

[6] J. D. Prestage, OSTDS Advanced Systems Review, JPL Internal Report, DSN RTOP 310-10-62, June 14-15, 1988, pp. 64-65; 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.

[7] D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano, D. J. Heinzen, S. L. Gilbert, C. H. Manney, and M. G. Raizen, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions", *IEEE Trans. Ultrason. Ferroelec. Freq. Contr.* **37**, 515–523, 1990.

[8] H. G. Dehmelt, "Introduction to the Session on Trapped Ions," *Proc. 4th Symp. Frequency Standards and Metrology*, 286, 1989.

[9] F. G. Major and G. Werth, "Magnetic Hyperfine Spectrum of Isolated $^{199}Hg^{+}$ Ions", Appl. Phys. **15**, 201-208, 1981.

[10] J. D. Prestage, R. L. Tjoelker, G. J. Dick, L. Maleki,"Ultra-Stable Hg^{+} Trapped Ion Frequency Standard," *Proc. 45th Ann. Symp. Freq. Control*, 572-581, May, 1991.

[11] G. J. Dick, "Calculation of Trapped Ion Local Oscillator Requirements," *Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 133–146, 1988.

[12] C. Audoin, V. Candelier, and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards," *IEEE Trans. Instr. Meas.*, April, 1991.

[13] G. J. Dick, J. D. Prestage, C. A. Greenhall, and L. Maleki, "Local Oscillator Induced Degradation of Medium-Term Stability in Passive Atomic Frequency Standards," *Proceedings, 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 487-508, 1990.

[14] N. F. Ramsey, *Molecular Beams* (Oxford University Press, Oxford, 1956).

[15] R. L. Tjoelker, J. D. Prestage, and L. Maleki, "Criteria for Ultrastable Operation of the Trapped Ion Frequency Standard," *TDA Progress Report 42-108, Vol. October-December 1991, Jet Propulsion Laboratory*, pp. 19-30, February 15, 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PROGRESS ON PROTOTYPE SYNTHESIZER ELECTRONICS FOR $^{199}HG^{+}$ AT 40.5 GHz *

C. W. Nelson, F. L. Walls, F. G. Ascarrunz, and P. A. Pond

Time and Frequency Division
National Institute of Standards and Technology
325 Broadway, Boulder, CO 80303

ABSTRACT

The design and specifications of a synthesizer for the $^{199}Hg^{+}$ ion standard will be presented. A fractional frequency stability of $5x10^{-15}$ at 100 s will be required if we are to reach the fundamental linewidth imposed by the shot noise of the atomic signal. This might be achieved by first locking a low-noise 5 MHz oscillator to an active hydrogen maser. A 100 MHz oscillator can then be locked to the 5 MHz oscillator to reduce close-in phase noise. A commercial digital synthesizer can be used to offset lock a 100.018 MHz oscillator to the 100 MHz signal. Frequency resolution and modulation capabilities will be provided by the digital synthesizer. The 100.018 MHz will be multiplied by 5, filtered, amplified, and routed through a step recovery diode (SRD). The 81st harmonic will be extracted to probe the hyperfine resonance of $^{199}Hg^{+}$ ions. A long-term stability of $6x10^{-17}$ is expected from the first prototype.

INTRODUCTION

Several groups have been working on developing a frequency standard based on ions contained in a linear rf ion trap [1-3]. These standards utilize a time-domain Ramsey interrogation scheme. The frequency stability for this configuration varies as

$$\sigma_y(\tau) = \frac{1}{2\pi\upsilon_0\sqrt{\tau n \tau_R}}, \quad \tau > \tau_R \qquad (1)$$

where τ is the measurement time, τ_R is the cycle time, n is the number of ions, and υ_0 is the transition frequency. From Eq. 1, projected frequency stability for approximately 50 ions and τ_R=100 s is $\sigma_y(\tau) = 5x10^{-14}\ \tau^{-1/2}$. To achieve this frequency stability, the local oscillator (LO) is required to have a frequency stability of order 5×10^{-15} at 100 s, unless one has at least two groups of ions that can be alternatively probed, as outlined by Dick et. al [2]. Currently, only active hydrogen masers have a frequency stability that satisfies the stability requirements at 100 s [4-5]. However, there is considerable effort to develop other local oscillators with frequency stabilities that are better than room temperature active masers [4,6-9]. The purpose of this paper is to describe the design and design criteria of a prototype synthesizer that is suitable for such a standard. In doing this, many problems must be solved. The most demanding are to provide the interrogation signal at the atomic line with both frequency and amplitude modulation capabilities, frequency resolution of 1×10^{-16}, and a phase stability within the electronics of approximately 0.5 ps at 100 s and 5 ps at 10^4 s. The design outlined here is a prototype aimed at solving these practical problems. It is hoped that this will lead to a vigorous discussion about the relative merits of various approaches to the many problems.

GENERAL DESIGN

Figure 1 shows the block diagram of the synthesizer. The internal reference is assumed to be phase locked to an external standard such as an active hydrogen maser or possibly a diode laser pumped rubidium standard [9]. The stability requirement on the external reference is roughly 5×10^{-15} at 100 s. The optimum bandwidth for locking the internal 5 MHz reference appears to be approximately 5 Hz. This is large enough to force the internal reference to track the

* Work of U.S. government. Not subject to U.S. copyright.

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

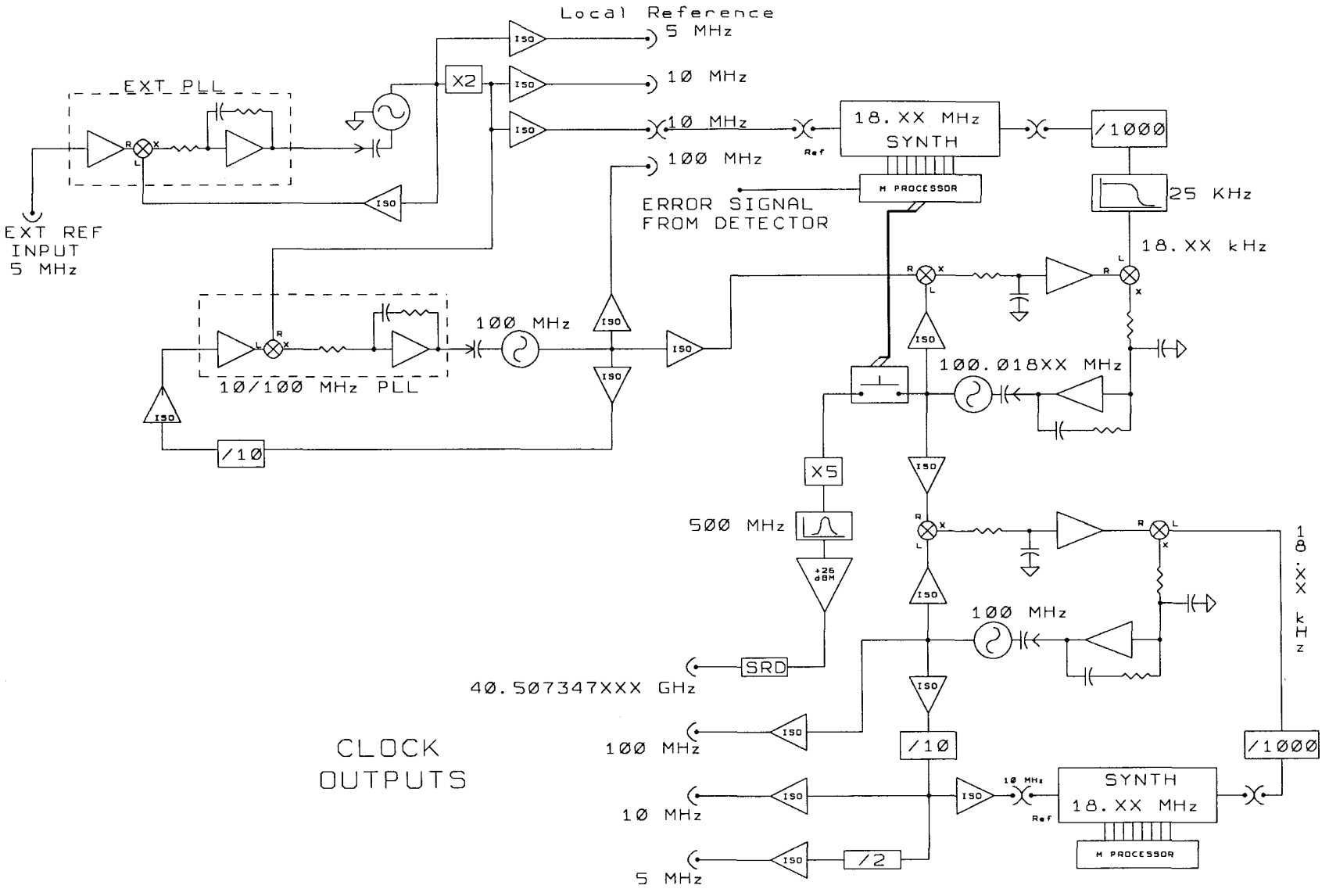


Figure 1 Overall block diagram of $^{199}Hg^{+}$ synthesizer

external reference and slow enough to avoid imposing 60 Hz and its harmonics on the internal reference. The low-noise internal 5 MHz oscillator is multiplied to 100 MHz and used to phase lock a 100 MHz oscillator with a bandwidth of 100 Hz. Noise-filtered reference frequencies at 5, 10, and 100 MHz that are phase locked to the external standard are made available for diagnostic measurements. The 5, 10, and 100 MHz isolation amplifiers follow Felton's general design [10]. The post regulators are used to eliminate the effects of noise from the voltage regulators on the phase noise of the frequency multipliers and rf amplifiers.

A 100.018 MHz oscillator is offset locked to the internal 100 MHz reference using a 18.14 kHz signal derived by dividing the signal from a digital synthesizer by 1000. If the synthesized signal at 18 kHz is added directly to the 100 MHz oscillator, it becomes extremely difficult to attenuate the unwanted side bands at 100 and 99.81 MHz sufficiently. During the multiplication to 40.5 GHz, the amplitude of these side bands increases by 52 dB and could therefore take a large percentage of the power from the desired signal. Figure 2 shows the phase noise of the digital synthesizer at 18 MHz. Dividing the frequency of the digital synthesizer by 1000 reduces the close-in phase noise and spurious lines by 60 dB and increases the frequency resolution to 1 nHz. The divided signal should have a phase noise spectrum equal to that of the input, minus 60 dB, plus the output noise of the divider [11].

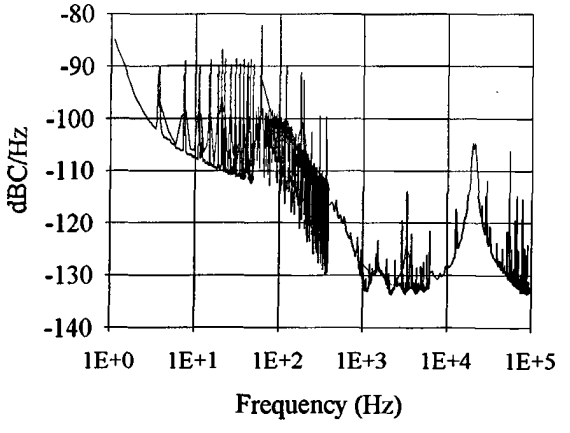


Figure 2 Phase noise of digital synthesizer at 18 MHz

A phase-lock loop with a bandwidth of approximately 100 Hz is used to control the close-in phase noise of the 100.018 MHz oscillator.

The output of the 100.018 MHz oscillator is multiplied by 5 and filtered in a 4 stage interdigital filter with a bandwidth of approximately 4 MHz. The output of the filter is amplified by 26 dB in a balanced amplifier, and routed to a SRD. The 81st harmonic at approximately 40.507 GHz corresponds to the center of the $^{199}Hg^{+}$ ground state hyperfine resonance.

The 40.5 GHz signal is frequency modulated with the digital synthesizer and amplitude modulated with a switch to implement the Ramsey interrogation process and to measure any perturbations that may depend on the various operating parameters of the trapped ions. The error signal from the detector could, in principle, be used to steer the output frequency of the hydrogen maser, and the output frequency of the maser reference at 100 MHz would then be the clock output. This, however, requires a maser dedicated only to the $^{199}Hg^{+}$ standard. An alternative is to use the error signal to steer the average frequency of the high-resolution synthesizer. This requires a separate channel to construct the clock output because all of the frequency modulation would have to be removed. This might be done using another high resolution synthesizer to offset lock the 100 MHz output oscillator to one of the isolated outputs from the 100.018 MHz oscillator. The second synthesizer could be frequency modulated inversely with respect to the first to reduce modulation on the output signal. Using the second synthesizer is advantageous because one maser could be used as the reference for several frequency standards. A potentially attractive alternative to provide improved performance would be to use a diode laser pumped rubidium standard as the reference. See the paper by J. Camparo et. al in these proceedings [9].

The frequency error signal derived from the detector [1-3] controls the average frequency of the 18.14 MHz synthesizer via a microprocessor. The 100.018 MHz oscillator is a direct sub-multiple (405) of the $^{199}Hg^{+}$

hyperfine transition. The microprocessor also controls the modulation of the synthesizer and the switch used to amplitude modulate the interrogation signal.

The primary output frequency of 100 MHz was chosen because of the difficulties of maintaining or even measuring frequency stability at the expected performance level of this clock at 5 or 10 MHz [13].

PRESENT STATUS

The prototype of the 5 MHz to 100 MHz synthesis is complete. Figure 3 shows the phase noise of the free running 100 MHz reference and the phase-locked 100 MHz reference. The excess phase noise from approximately 50 to 300 Hz from the carrier is due to noise in the 100 MHz to 10 MHz frequency divider [11]. This noise could be reduced by using frequency multipliers; however, they would have little effect on the frequency stability of the output [9]. Small phase variations with temperature in this portion of the synthesizer are relatively unimportant since the clock output is derived from the 100.018 MHz oscillator.

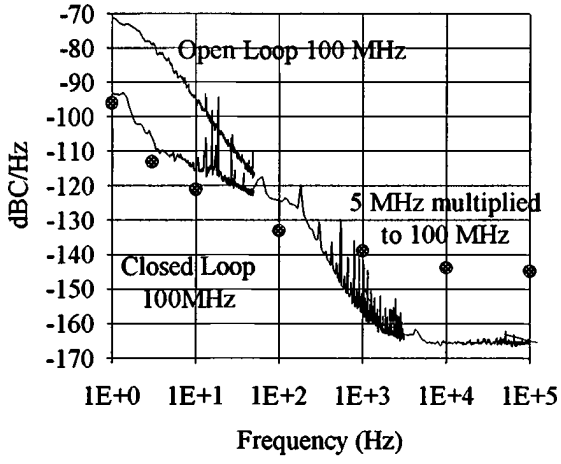


Figure 3 Phase noise of 100 MHz quartz crystal

Phase shifts in the 100.018 MHz circuit contribute directly to instability in the output signal. Figure 4 shows one of the four isolation amplifiers that distributes the 100 and 100.018 MHz signals. The isolation amplifiers are constructed as balanced FET amplifiers with low harmonic distortion. Squaring amplifiers are currently used to drive the divider chain. The phase stability versus temperature of two isolation amplifiers is shown in Figure 5. The power amplifier is a balanced-junction, bipolar amplifier and provides about 26 dB gain, an output power of 0.5 W, low harmonic distortion, low noise, and good phase stability versus temperature.

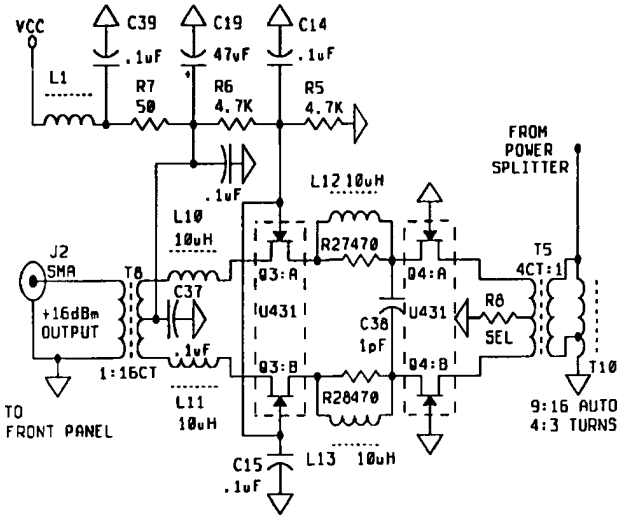


Figure 4 Schematic of 100 MHz distribution amplifiers.

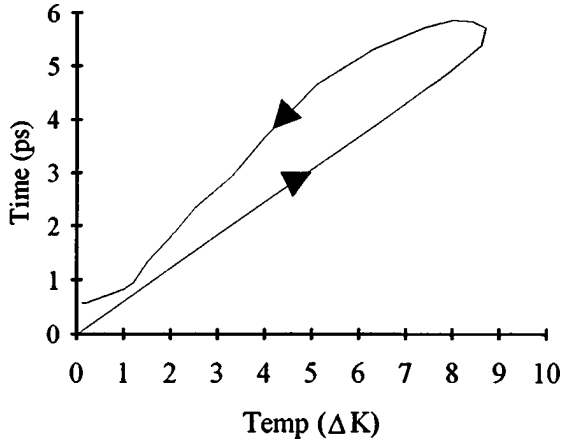


Figure 5 Temperature coefficient of 100 MHz distribution amplifier.

The multiplication chain of 100 MHz to 500 MHz through the SRD has been constructed and tested out to 10.6 GHz. Figure 6, top trace, shows the phase noise of such a signal at 10.6 GHz. The decrease in noise of the

bottom trace above 3 kHz is due to locking the 10.6 GHz signal to a cavity-locked dielectric resonator oscillator (DRO) as described in [12]. The additional multiplication to 40.5 GHz would increase the phase noise by 12 dB. Figure 7 shows the phase variation between two 100 MHz to 10.6 GHz synthesizers over 12 hours. Frequency stability of the synthesis is shown in Figure 8 and is approximately $6x10^{-16}$ at 100 s and $5x10^{-17}$ at 10^4 s. The temperature coefficient for the 100 MHz to 10.6 GHz synthesis is shown in Figure 9. The coefficient of 3.8 ps/K is believed to be dominated by the SRD. By properly biasing and ovenizing the SRD, this coefficient could be greatly reduced.

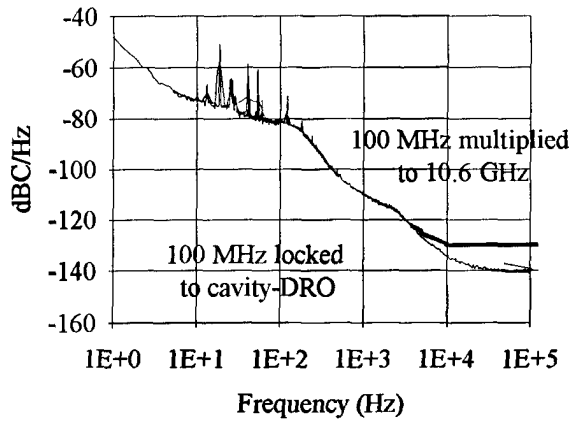


Figure 6 Phase noise of 10.6 GHz synthesis.

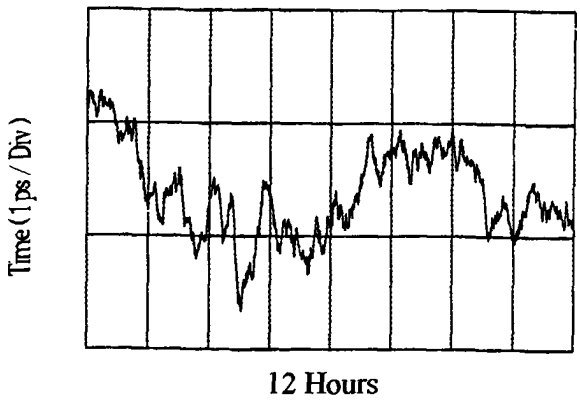


Figure 7. Phase stability of 10.6 GHz synthesis.

The offset lock using the digital synthesizer has not yet been implemented. The frequency stability of the synthesizer is shown in Figure 10. Its contribution to instability is reduced by $5.5x10^3$ because it is an offset on a 100 MHz signal. This brings its contribution to $7x10^{-16}$ at 100 s and $1x10^{-18}$ at 10^4 s.

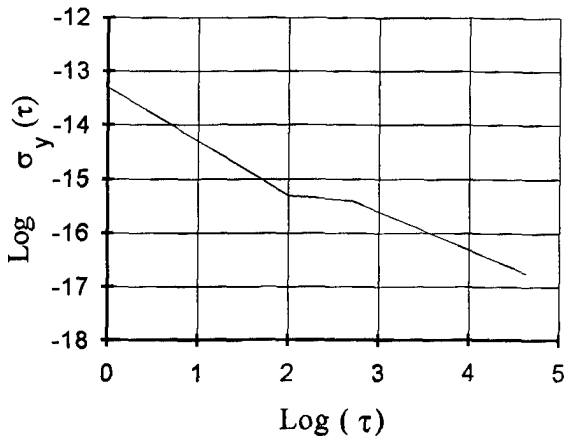


Figure 8 Frequency stability of 10.6 GHz synthesis

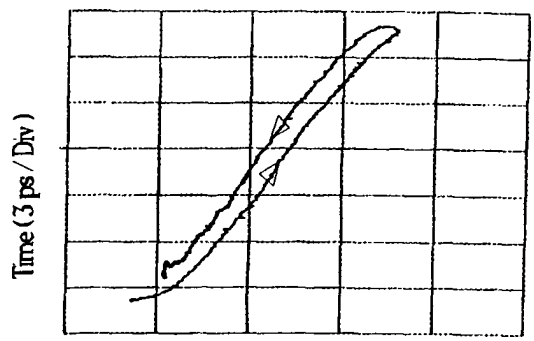


Figure 9 Temperature coefficient of 10.6 GHz synthesis

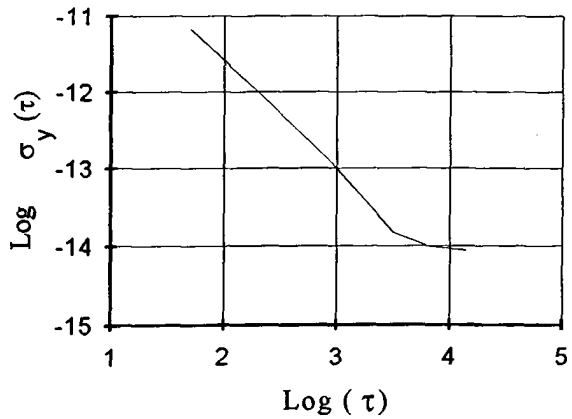


Figure 10 Frequency stability of commercial digital synthesizer

CONCLUSION

Development of a 40.5 GHz synthesizer for a highly stable and accurate $^{199}Hg^{+}$ ion standard is well under way. With existing components, we should be able to produce a synthesizer that is suitable for preliminary testing of the ion standard. By offset locking instead of adding the signal in with a mixer, we eliminate residual sidebands that could cause problems in the multiplication to 40.5 GHz. By analysis of our existing 10.6 GHz synthesis and stability measurements from the digital synthesizer, we can project stabilities for such a system to be 2×10^{-15} at 100 s and 6×10^{-17} at 10^4 s. Redesign of the synthesizer using techniques and materials to minimize environmental effects [13] should bring the stability down to levels needed for the final version of the $^{199}Hg^{+}$ ion standard.

REFERENCES

1. D.J. Wineland, J.C. Bergquist, J.J. Bollinger, and W.H. Itano, "Progress at NIST toward absolute frequency standards using stored ions," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, 37, 1990, pp. 515-523.

2. J.D. Prestage, G.J. Dick, and L. Maleki, "JPL trapped ion frequency standards development," Proc. 41st Ann. Frequency Control Symp., *1987*, pp. 20-24.

3. J.D. Prestage, G.J. Dick, and L. Maleki, "Linear ion trap based atomic frequency standard," Proc. 44th Ann. Frequency Control Symp., 1990, pp. 82-88.

4. F.L. Walls, "Frequency standards based on atomic hydrogen," Proc. 40th Ann. Frequency Control Symp., 1986, pp. 142-146.

5. J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standards, Bristol, Philadelphia, 1989.

6. A.J. Berlinsky and W.N. Hardy, "Cryogenic masers," Proc. 13th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meet., 1981, pp. 547-559.

7. G. J. Dick, "Microwave oscillator (room temperature and cryogenic) for superior short term stability and ultra-low phase noise," These proceedings.

8. A. Luiten, A. Mann, and D. Blair, "High medium-term stability sapphire resonator-oscillator," These proceedings.

9. J. Camparo, R. Drullinger, and C. Szelely, "Strategies for designing a diode laser pumped gas cell atomic standard," These proceedings.

10. C.M. Felton, "Superimposing low-phase-noise, low-drift instrumentation techniques on RF design," RF Design, October 1990, pp. 65-74.

11. F.L. Walls and C.M. Felton, "Low noise frequency synthesis," Proc. 41st Ann. Frequency Control Symp., 1988, pp. 512-518.

12. F.L. Walls and C.M. Felton, "High spectral purity X-band source," Proc. 44th Ann. Frequency Control Symp., 1990.

13. F.L. Walls, L.M. Nelson, and G.R. Valdez, "Designing for frequency and time metrology at the 10^{-18} level," Proc. 8th European Frequency and Time Forum, Netherlands, 1992.

See also L.M. Nelson, F.L. Walls, "Environmental effects in mixers and frequency distribution systems," These proceedings.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

THE DEVELOPMENT OF A YTTERBIUM ION FREQUENCY STANDARD*

D.J. Seidel, A. Williams, R.W. Berends, 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 the accessibility of its first excited electronic energy levels with light from frequency-doubled semiconductor and solid state lasers. In this paper we discuss the desired characteristics of a space-borne instrument and why we have chosen ytterbium to meet these characteristics. We also present the results of spectroscopy carried out on trapped ytterbium ions to determine means for overcoming the population trapping difficulties associated with this ion.

Introduction

Recent progress in the development of trapped ion standards has led to the realization of instruments rivaling the stability of any other microwave standard. The trapped mercury ion standard has achieved fractional frequency stability of $1 \times 10^{-13}/\sqrt{\tau}$ for 150s $\leq \tau \leq$ 10,000s and has made possible stabilities of 2×10^{15} at 24,000 seconds of averaging time, which is unsurpassed by any other frequency standard [1].

Since the inherent stability of the trapped ion standard is inversely proportional to the product of the line Q and the signal- to-noise ratio (SNR), it is desirable to have both high line Q's and high SNR for the highest stability. The mercury ion standard has achieved a line Q of 2.4×10^{12}, the highest ever observed in a microwave transition [1]. Despite this remarkable performance further progress in improving the stability of a field-ready mercury standard is hampered by the requirement of a mercury lamp and associated optics to produce the 194 nm optical pumping light. Optical pumping with lasers greatly improves the SNR, allows higher detection efficiency and replaces complex optical systems with fiber optics. Such an approach is not currently feasible for use with a field deployable mercury ion standard, since generation of light at 194 nm with a laser is complex.

In order to realize the advantages associated with the use of lasers in optical pumping of trapped ions, we have embarked on the development of a standard based on the 171 isotope of singly ionized ytterbium. The hyperfine transition of the ground state of this ion has a relatively large frequency, allowing large line Q's. Furthermore, the lowest lying excited electronic energy levels of the ytterbium ion are accessible with light from frequency doubled semiconductor and solid state lasers, thus making this ion species a desirable candidate as a laser-excited standard.

Another major advantage with laser-excited ytterbium is the potential for the development of a space-borne trapped ion standard for spacecraft science investigations. With semiconductor lasers and optical fibers, it is possible to reduce the size and mass of a trapped ion standard and realize an instrument with stability comparable to hydrogen masers.

Laser excitation also offers the potential for cooling the ions. Reducing the ions' motion in this way causes a decrease in the second-order Doppler shifts, which generate the greatest frequency offset in a trapped ion standard [2]. Reducing this offset equivalently reduces the stringent stability requirements of several trapping parameters, such as trap field strength, endcap voltage and temperature.

Despite its apparent advantages, ytterbium has difficulties associated with population trapping in lower lying metastable states which must be continually depopulated with additional light sources [3-9]. We have

*This work represents the results of one phase of research carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract sponsored by the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

performed a great deal of spectroscopy on trapped ytterbium ions and have determined several effective means to depopulate these lower lying metastable states with available semiconductor lasers.

In this paper we will describe the need for a laser-excited trapped ion standard, with emphasis on design requirements for space applications. The status of the standard and the challenges associated with ytterbium will then be reported.

Requirements of Ultra-Stable Frequency Standards in Space

Placing atomic frequency standards with long-term stability in space will significantly improve a variety of radio science experiments. The long-term stability of the radio link between an earth-based receiver and a spacecraft can be severely degraded due to the delay changes induced by random tropospheric variations. At Ka-band frequencies, the greatest noise source for low frequency gravitational wave searches using the earth and distant spacecraft will be due to the troposphere [10]. The tropospheric noise can be removed by simultaneously combining Doppler data generated by a spacecraft clock with data at the earth station [11].

Other experiments which could benefit from space-borne frequency standards include gravitational red-shift measurements [11], and general relativity tests such as testing the isotropy of the speed of light. Space interferometers consisting of two or more spacecrafts with highly stable and synchronized standards would also benefit from space-borne standards with ultra-high stability.

A science instrument suitable for use in space must be easily integrated into a spacecraft and be highly reliable while still accomplishing its primary function. For a frequency standard this primary function is ultra-high stability, on the order of 10^{-15} at 10,000 seconds of averaging time. The hydrogen maser is the only high-performance frequency standard to fly aboard a spacecraft, achieving stability of 10^{-14} at 400 seconds on a flight whose duration was approximately two hours. This test was conducted in 1976 and confirmed the gravitational red shift to 70 parts/million [11]. The Hg^+ trapped ion standard has already achieved the required stability and high reliability; however, as mentioned earlier, the 194 nm radiation necessary for optical pumping cannot be conveniently generated with laser systems. The use of a mercury lamp provides a simple solution, but requires a significant amount of power and creates a large amount of background noise due to light scattering in the trap.

The desirablilty of a laser-excited trapped ion standard stems from the potential increase in the signal to noise ratio and line Q associated with the use of coherent light. Since the fundamental stability of a trapped ion frequency standard is given by

$$\sigma(\tau) \propto \frac{1}{Q_\ell}\frac{1}{SNR}\frac{1}{\sqrt{\tau}} \qquad (1)$$

it is evident that increasing the line Q and SNR are the only approaches for improving the intrinsic stability of the standard.

In addition, the use of lasers conveniently lends itself to the application of fiber optics which in turn relieves design problems in introducing light into the trap. All this leads to a standard of greatly reduced size without compromising stability. Such a standard could weigh less than 15 Kg, fit in a volume less than 30 liters, only consume $\sim$ 20W of power, and could remain in operation for several years. These physical characteristics are comparable to current spacecraft fields and particles instruments. In comparison, a future space maser being developed by the In Space Technology Experiment Program weighs $\sim$ 40 Kg, occupies $\sim$ 130 liters, consumes $\sim$ 30W, and is being designed to last six months [12].

Characteristics of the Ytterbium Ion Standard

The ytterbium ion has a hyperfine structure similar to the mercury ion, with an energy splitting corresponding to 12.6 GHz (see Fig. 1), as compared to mercury's 40.5 GHz. With laser excitation of ytterbium, very high Q's are obtainable, as has been demonstrated by D. Schnier et. al. [4]. These authors obtained a 16 mHz linewidth for the hyperfine transition in $^{171}Yb^+$, which corresponds to a line Q of 8×10^{11}.

Another major advantage achievable with laser-excited ytterbium is the ability to cycle the ions on the transition. This will allow scattering of many optical photons for each microwave photon absorbed, thus further increasing the SNR. The cycling of the $(^2S_{1/2}$,F=1) $\leftrightarrow$ $(^2P_{1/2}$,F=0) transition is possible due to the narrow linewidth ($\sim$ 10 Mhz) of the laser and the large 2.1 Ghz hyperfine splitting of the $^2P_{1/2}$ state.

In contrast to the case of the mercury ion, the first excited electronic energy levels of the ytterbium ion are accessible with light from frequency doubled semiconductor and solid state lasers. For example, the transition from the ground $^2S_{1/2}$ state to the excited $^2P_{1/2}$ requires light at 369 nm. This wavelength is readily accessible by doubling light from a Ti:Sapphire laser or diode laser [4].

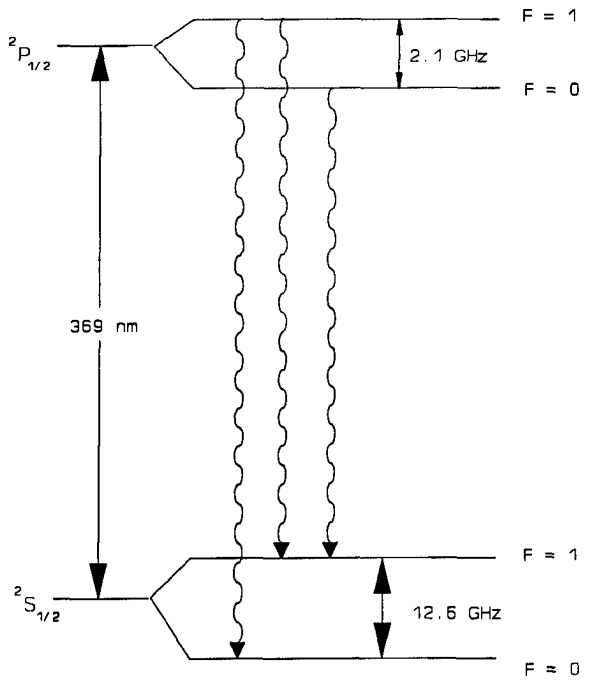


Figure 1: Hyperfine structure of the ground and first excited states in $^{171}Yb^+$.

We have designed a build-up ring cavity for the generation of 369 nm radiation using a 738 nm laser diode and a doubling crystal of lithium triborate (LBO) (Fig. 2). In order to overcome the low doubling efficiency associated with continuous-wave operation and milliwatt input powers, we have designed our cavity to take advantage of the increase in efficiency with tight focusing of the input beam. The two curved mirrors M3 and M4 will focus the fundamental beam to a waist of 25 μm in the doubling crystal. LBO is well-suited to this application due to its large acceptance angle, which allows tight focusing. The custom antireflection

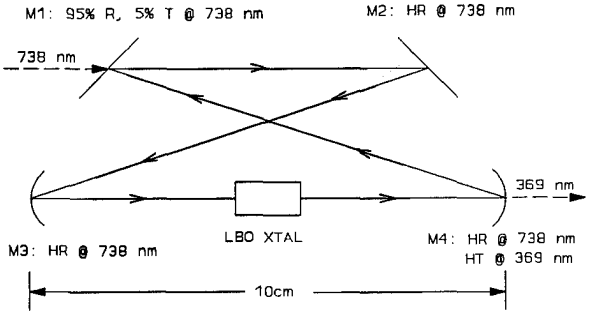


Figure 2: Design of optical buildup cavity for generation of 369 nm light necessary to reach the first excited electronic energy level in Yb^+.

coatings on the mirrors and crystal are expected to allow a buildup factor of approximately 15; therefore, a diode laser with output power of only 5 mW will provide 75 mW to the doubling crystal. With these considerations, we expect greater than 0.5% conversion efficiency, which will produce a minimum of 25 μW of 369 nm radiation with 5 mW of 738 nm input radiation.

Results of Spectroscopy Carried Out on Trapped Ytterbium Ions

Figure 3 shows a partial energy level diagram of singly ionized ytterbium. Using a linear rf/dc trap,

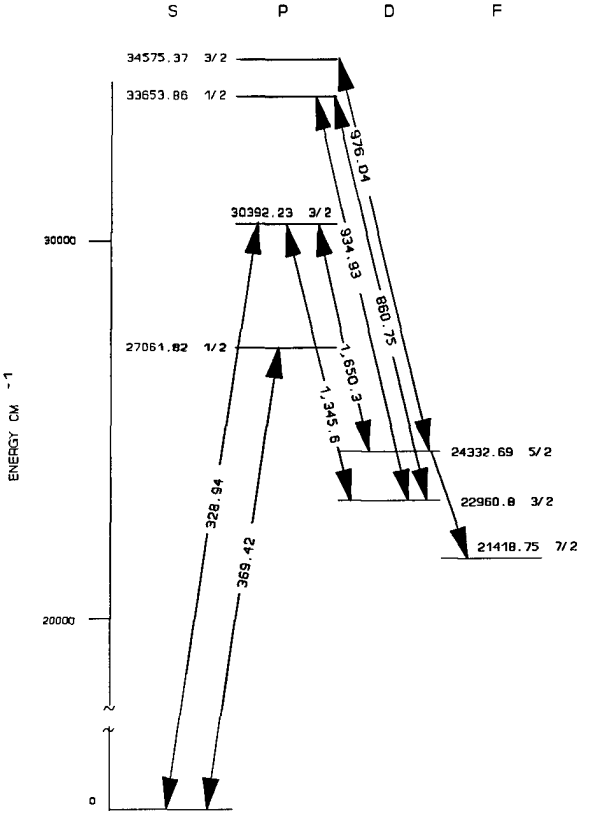


Figure 3: Simplified Yb^+ energy level diagram showing the eight levels involved in our experiment and numerical integrations.

which has been described previously [2], we have carried out extensive spectroscopy on naturally abundant ytterbium ions. We drove the $^2S_{1/2} \leftrightarrow {}^2P_{3/2}$ transition 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 (LIO) doubling crystal. Several milliwatts of uv power was generated this way, and a hollow cathode lamp was used to

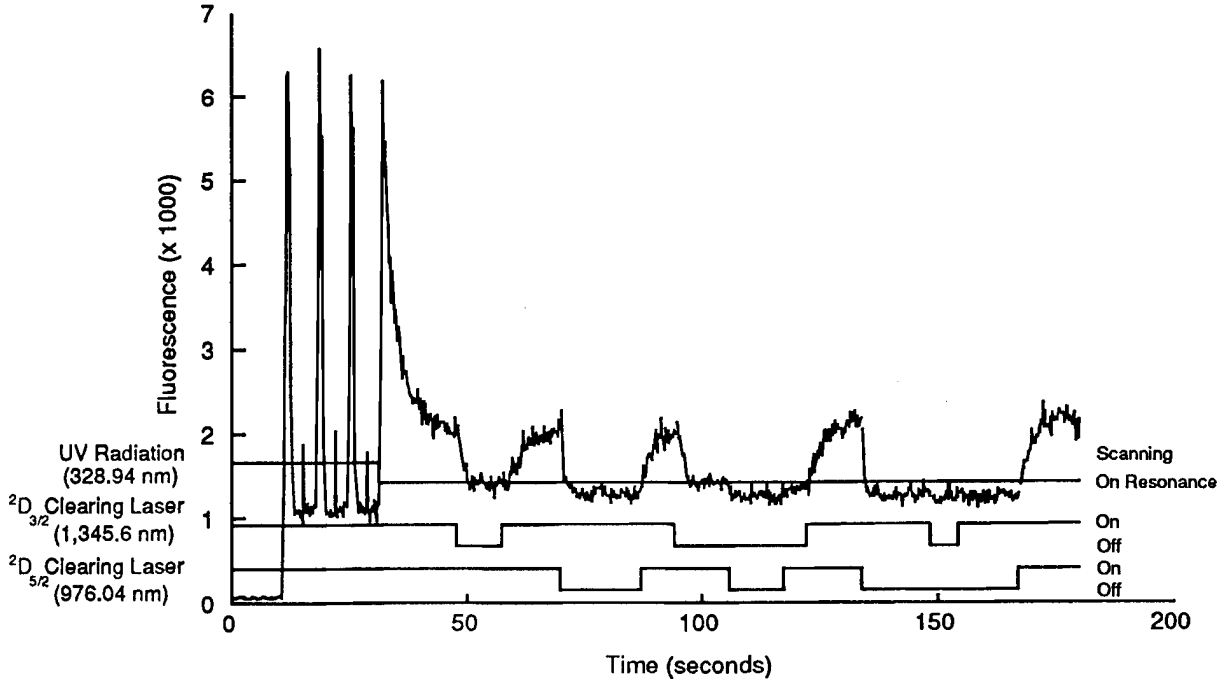


Figure 4: Ion fluorescence for various combinations of clearing lasers.

confirm resonance. We chose to excite the $^2P_{3/2}$ state, as opposed to the $^2P_{1/2}$ state, since it allowed direct decay to both the $^2D_{3/2}$ and $^2D_{5/2}$ states. This is important for the determination of the exact mechanism responsible for the population trapping first described by H. Lehmitz, et. al. [9].

Population trapping causes a loss of ion fluorescence by removing the ions from the cycling process between the ground state and the excited P state. Several mechanisms have been suggested by several different groups, all of which involve the eventual filling of a highly metastable $^2F_{7/2}$ state [3-9]. The $^2F_{7/2}$ state has a theoretical lifetime of 1533 days [13], while Lehmitz et. al have measured a dark period greater than 8 days [9].

In addition to pumping the $^2S_{1/2} \leftrightarrow {}^2P_{3/2}$ transition with 328.9 nm radiation we also drove the $4f^{14}(^1s)5d^2D_{3/2} \leftrightarrow 4f^{13}(^2F_{7/2})5d6s(^3D)^3[3/2]_{1/2}$, $4f^{14}(^1s)5d^2D_{5/2} \leftrightarrow 4f^{13}(^2F_{7/2})5d6s(^1D)^1[3/2]_{3/2}$, $4f^{14}(^1s)5d^2D_{3/2} \leftrightarrow 4f^{13}(^2F_{7/2})5d6s(^1D)^1[3/2]_{3/2}$, and $4f^{14}(^1s)5d^2D_{3/2} \leftrightarrow 4f^{14}(^1S)6p^2P_{3/2}$ transitions using 935 nm, 976 nm, 860 nm, and 1,345 nm radiation respectively. The 935 nm and 976 nm lasers are the best candidates for depopulating the $^2D_{3/2}$ and the $^2D_{5/2}$ levels since these are available semiconductor lasers. Semiconductor lasers at 860 nm are also readily available but could cause coherent population trapping problems when used with the 976 nm laser, as this would effectively connect the $^2D_{3/2}$ and the $^2D_{5/2}$ states. A similar problem exists between the 1,345 nm laser and the 329 nm laser since both these lasers drive transitions to the $^2P_{3/2}$ state.

Figure 4 shows the effect of simultaneous laser pumping with 328.9 nm, 976 nm, and 1,345 nm lasers. The 976 nm radiation was generated with a Spectra Physics 3900S Ti:Sapphire laser with output power of 300 mW in a 30 GHz linewidth. The 1,345 nm laser was a hand-selected Ortel LW300-E01 InGaAsP diode laser. A single fiber carried light from both lasers into the trap. An important point to notice in Fig. 4 is the rapid return of fluorescence after cycling the D state clearing lasers off and then back on again, implying no long-term loss of ions to the $^2F_{7/2}$ state. We are currently investigating the source of discrepancy between these results and the results obtained by others [3-9]. Our experimental setup is more sensitive to any population trapping that might take place in the $^2F_{7/2}$ state, since we populate the $^2D_{5/2}$ state directly. This state decays to the $^2F_{7/2}$ with a decay time of 6.3 msec or to the ground state with a decay time of 28.4 msec, according to the theoretical calculations of Fawcett and Wilson [13].

Figure 5 shows recent work using broadband semiconductor lasers at 935 nm and 976 nm fabricated by the Micro Devices Laboratory at JPL. Research in progress compares these four experimental

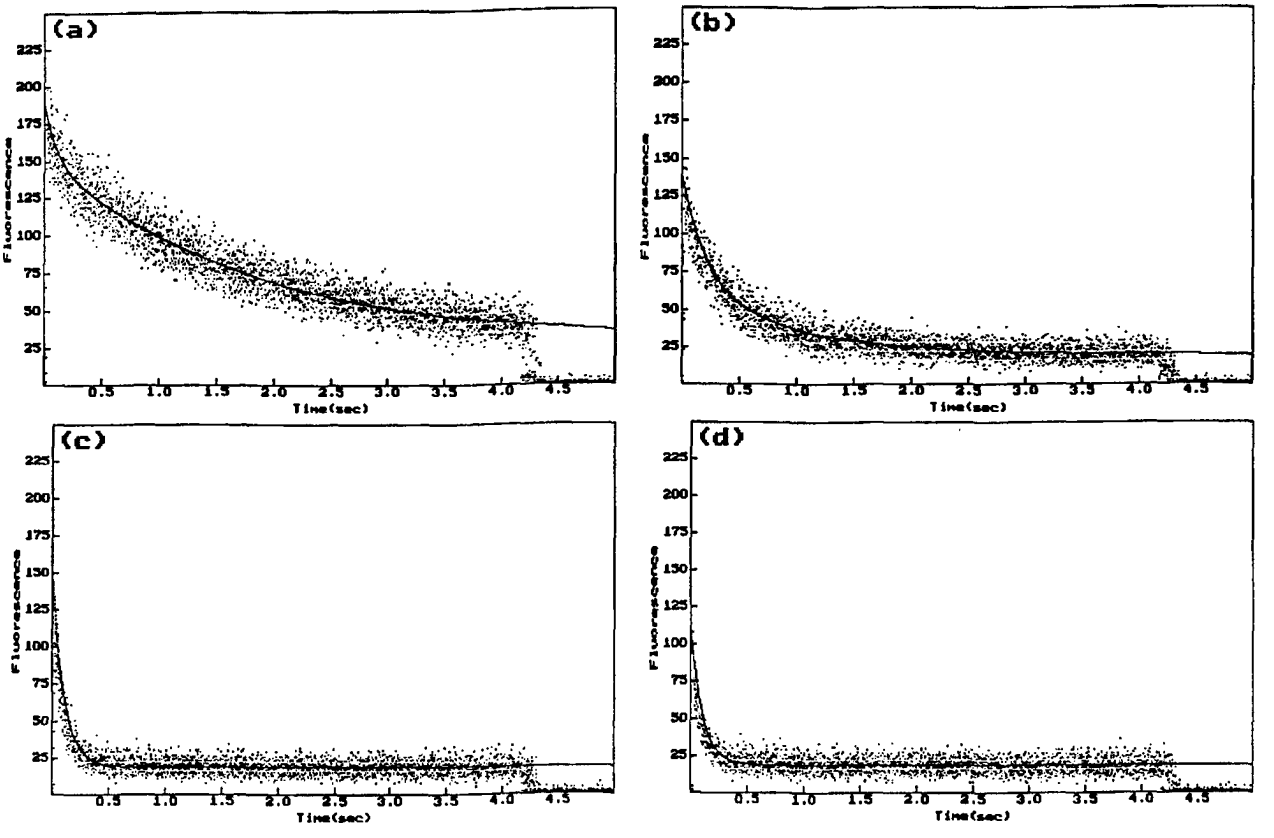


Figure 5: Experimental fluorescence decay curves under various laser excitations and best fits to the data. (a) 329 nm laser power $\sim$ 100μW; 935 nm laser power $\sim$ 1mW; 976 nm laser power $\sim$ 2 mW; fluorescence=$121e^{-t/1769ms} + 28e^{-t/102ms} + 30$. (b) 329 nm laser power $\sim$ 100μW; 935 nm laser power off; 976 nm laser power $\sim$ 2 mW; fluorescence=$49e^{-t/899ms} + 61e^{-t/204ms} + 20$. (c) 329 nm laser power $\sim$ 100μW; 935 nm laser power $\sim$ 1mW; 976 nm laser power off; fluorescence=$103e^{-t/106ms} + 19$. (d) 329 nm laser power $\sim$ 100μW; 935 nm laser power off; 976 nm laser power off; fluorescence=$74e^{-t/105ms} + 19$.

fluorescence decay curves with the P-state level populations obtained by numerically integrating the rate equations for the eight levels involved on a Cray supercomputer. The Einstein A and B coefficients for the various transitions are derived using the oscillator strengths quoted in Fawcett and Wilson [13]. This integration yields a longer decay time in the configuration of Fig. 5(b) than with both clearing lasers on as in Fig. 5(a), in conspicuous disagreement with the experimental results. The source of this discrepancy is presently under investigation.

Summary

We have pursued the development of a trapped ion standard based on singly-ionized ytterbium isotope 171. The choice of this ion is compatible with the use of semiconductor and solid- state lasers for optical pumping. We have demonstrated laser excitation of various levels of ytterbium ions trapped in a hybrid rf/dc linear trap using such lasers. Because there is considerable uncertainty associated with the role of low-lying metastable states of the ytterbium ion in population trapping, we have carried out spectroscopic investigations addressing this issue.

In the course of our experiments we have demonstrated the use of semiconductor lasers and fiber optics with trapped ytterbium ions. This is a crucial step in the development of small, reliable frequency standards for spacecraft deployment in support of various radio science investigations.

Acknowledgements

We wish to thank Kavoos Deilamian, John Prestage, and Bob Tjoelker for helpful conversations and assistance with the ytterbium experiments, and to acknowledge the use of the JPL Cray Supercomputer to perform our numerical simulations.

References

[1] J.D. Prestage, R.L. Tjoelker, G.J. Dick, and L. Maleki, "Ultra-Stable Hg^+ Trapped Ion Frequency Standard," *J. Mod. Opt.*, vol. 39, no. 2, pp. 221-232, Feb. 1992.

[2] 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.

[3] 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.

[4] D. Schnier, A. Bauch, R. Schröder, Chr. Tamm, "Ion Trap Frequency Standard Research at PTB," to be published in *Proc. of the 6th European Frequency and Time Forum*, March 1992.

[5] Chr. Tamm and D. Schnier, "A Tunable Three-Level Neodymium-Doped Fiber Laser and its application for Depletion of the $4f^{14}5d\ ^2D_{3/2}$ Level in Optically-Excited, Trapped Ytterbium Ions," *Opt. Comm.*, vol. 87, no. 5/6, pp. 240-244, Feb. 15, 1992.

[6] 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, July 1991.

[7] H.A. Klein, A.S. Bell, G.P. Barwood, P. Gill, and W.R.C. Rowley, "Studies of Laser-Cooled Trapped Yb^+," *IEEE Transaction on Instrumentation and Measurement*, vol. 40, no. 2, pp. 129-131, April 1991.

[8] R. Blatt, R. Casdorff, V. Enders, W. Neuhauser, and P.E. Toschek, "New Frequency Standards Based on Yb^+," in *Frequency Standards and Metrology*, A. DeMarchi, Ed. Berlin: Springer-Verlag, pp. 306-311, 1989.

[9] 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

[10] S.J. Nelson and J.W. Armstrong, "Gravitational Wave Searches Using the DSN," *TDA Progress Report 42-94*, vol. April-June 1988, Jet Propulsion Laboratory, Pasadena, California, pp. 75-85, August 15, 1988.

[11] R.F.C. Vessot, "Applications of Highly Stable Oscillators to Scientific Measurements," *Proceedings of the IEEE Special Issue on Time and Frequency*, vol. 79, no. 7, pp. 1040-1053, July 1991.

[12] R.L. Sydnor, private communication.

[13] B.C. Fawcett and M. Wilson, "Computed Oscillator Strengths, Landé g Values, and Lifetimes in Yb II," *Atomic Data and Nuclear Data Tables*, vol. 47, no. 2, pp. 241-317, March 1991.

FREQUENCY MEASUREMENT OF A SINGLE BARIUM ION TRANSITION WITH AN AMMONIA LASER

K.J.SIEMSEN, A.A.MADEJ, J.D.SANKEY, B.G.WHITFORD and R.F.CLARK
Institute for National Measurement Standards
National Research Council of Canada, Ottawa, Canada, K1A 0R6
G. MAGERL
Technische Universität Wien, A1040 Wien, Austria

The $5d\ ^2D_{3/2}$ - $5d\ ^2D_{5/2}$ clock transition at 24 THz (12μm wavelength) of a single, trapped and laser cooled $^{138}Ba^+$ ion was probed by an optically pumped ammonia laser. The clock transition frequency is low enough to be measured against the Cs standard frequency using NRC's existing phase locked frequency chain.

The mid-infrared fine-structure transitions of trapped ions are interesting candidates for ultra-high-resolution spectroscopy as has been proposed by Dehmelt and co-workers [1], Werth [2] and our group [3]. One advantage of these transitions is their high transition frequency to natural line width ratio (Q up to 10^{15}) which gives the possibility for extremely high precision in a spectral range where frequency chains phase-locked to a Cs frequency standard already exist. In this laboratory, barium ions have been trapped for some time and recently the D-D fine-structure transition of a single barium ion has been observed using a tunable lead salt diode laser at a wavelength of 12.5 μm [4]. The spectral brightness of the diode laser (0.1 mW at 20 MHz line-width over 1 s) was sufficient to drive the D-D transition well into saturation. The transition frequency was found to coincide with the frequency of the sP(8,7) absorption line of $^{15}NH_3$. These results encouraged us to pursue the investigation of the D-D transition of Ba^+ with better resolution by using a frequency stabilized $^{15}NH_3$ laser.

The mid-infrared ammonia laser, optically pumped by a carbon dioxide laser, can be a sub-kilohertz line-width source when operated under proper pumping conditions [5]. Although ideally suited the sP(8,7) $^{15}NH_3$ transition has not been used because lasing on this line would have required a radioactive $^{14}C^{18}O_2$ laser for pumping. However, laser action on the spectrally closest sP(8,6) transition was recently obtained. An upgraded version of the ammonia laser operating on this transition is described in this work. The frequency of the laser operating on the sP(8,6) transition is not widely tunable. Therefore an electro-optic sideband modulator was built to down-shift the laser frequency by 17.2 GHz and tune it into resonance with the D-D clock transition of Ba^+. The performance of the modulator is also described here.

The energy-level diagram of Ba^+ in Fig.1 shows the transitions used for cooling and detection (650 and 493 nm) and for excitation of the D-D transition at 12.48 μm. Two visible radiation fields are applied to sustain continuous fluorescence of the ion. Excitation by infrared light on the

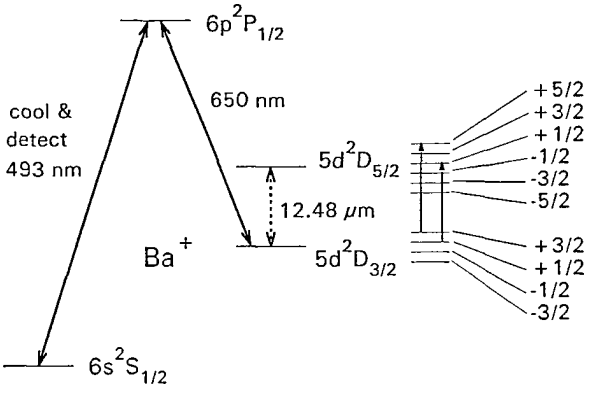


Fig.1. Energy-levels of the lowest electronic states of Ba^+ showing the transitions for detection and cooling. The right hand side indicates the Zeeman splitting of the levels for the infrared clock transition.

0-7803-0476-4/92 $3.00 © 1992 IEEE

$5d^2D_{3/2}$-$5d^2D_{5/2}$ transition results in the cycling electron being shelved in the metastable $5d^2D_{5/2}$ level and interruption of the fluorescence. The fluorescence appears again after spontaneous emission out of the metastable state (lifetime 35 s) and by stimulated emission for sufficiently strong infrared radiation. In order to avoid AC Stark splitting of the lower $5d^2D_{3/2}$ level and thus broadening of the infrared transition, the 650 nm radiation is blocked off by a chopper while the 12.48 μm radiation is on. Periodic interruption for a time of 80 ms or shorter did not effect the ion temperature as was experimentally verified through the absence of motional sidebands. If a static magnetic field is present inside the trap the infrared transition will split into 12 components. Two of these are very close together so that a total of 10 Zeeman components is expected at the present spectral resolution of a few kilohertz. The strength of the individual components will depend on the orientation of the magnetic field relative to the polarization of the laser beam.

Figure 2 shows an overview of the experimental arrangement. The 493 nm Coumarin dye laser is pumped by a Kr^+ laser, the 650 nm DCM dye laser uses an Ar^+ laser as the optical pump. Both lasers are manually set to the low frequency side of the 493 and 650 nm transitions for maximum cooling. The interruptions (quantum jumps) in the 493 nm fluorescence were detected by a photon counting photomultiplier, recorded and displayed with help of a computer. The probe frequency was moved in steps of 10 kHz. Typical observation times were 30 s for each probe frequency setting. The synthesizer supplying the probe frequency was locked to the Cs time standard. The NRC frequency chain, as shown in fig.2, will be used to link the ammonia laser frequency to a primary NRC frequency standard, but for the experiments described herein the chain was not employed. Instead the ammonia laser frequency was set to the saturated absorption dip of the sP(8,6) absorption line of $^{15}NH_3$ as an intermediate frequency reference.

A more detailed view of the arrangement to generate the 24 THz probe frequency is shown in Fig.3. In order to get lasing on the sP(8,6) transition the $^{15}NH_3$ laser requires an optical pump source with 1 W or more power at a frequency coinciding with the center of the sR(6,6) absorption line of $^{15}NH_3$. This pump radiation is supplied by a $^{16}O^{12}C^{18}O$ laser. The 2.6 m long laser cavity

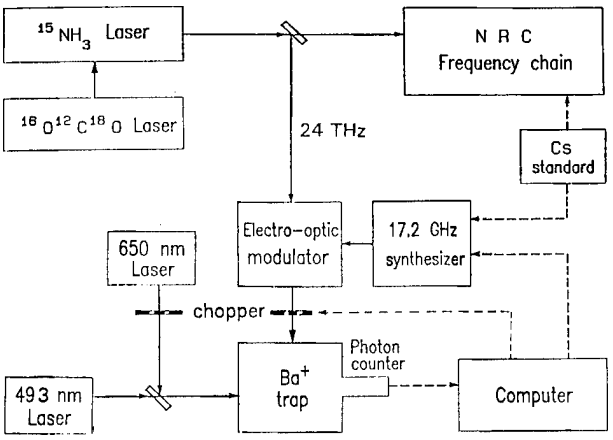


Fig.2. Overview of the experiment to probe the Ba^+ ion.

contains in addition to the 2 m long gain tube a cell filled with $^{12}C^{18}O_2$ at a pressure of 1 kPa. The purpose of the intra-cavity absorber cell is to prevent laser action on the high gain transitions R(20) and R(22) of $^{12}C^{18}O_2$ which is formed through dissociation in the gain tube. Suppression of these transitions enables the much weaker R(40) line of $^{16}O^{12}C^{18}O$ at 9.1 μm wavelength to lase. Competition from the adjacent rotational transitions was eliminated by a 200 grooves per millimetre grating specially blazed for this purpose [6]. The gain of the R(40) transition measured only 5.6 %/m, but with suitable output coupling the laser delivered up to 3 W power. The frequency of the R(40) transition was down-shifted by an acousto-optic modulator (72 MHz, 60% efficiency) into exact coincidence with the sR(6,6) absorption line of $^{15}NH_3$. A small fraction of the pump beam (5 mW) is used to lock the frequency of the carbon dioxide laser to the saturated absorption dip sR(6,6) of $^{15}NH_3$ as shown on the top part of fig.3. The saturation dip was about 400 kHz wide and required a 50 kHz wide modulation of the pump laser frequency for an adequate signal to noise ratio in the first derivative lock. A fraction of this pump laser frequency modulation could have been carried over into the ammonia laser frequency through imperfect operating conditions at the present state of the experiment.

The ammonia laser cavity was made of a CERVIT tube, .35 m long and partially surrounded by dry ice for cooling. The laser does lase at room temperature, however, a five fold power increase

was obtained with cooling. The dichroic mirror in front of the cavity has 97% transmission for the pump radiation and 85% reflection for the generated 12.48 μm wavelength. The gold coated mirror at the end of the cavity was mounted on an piezo-electric transducer for length tuning. The transducer, specially made for vacuum use, was mounted inside onto the end of the CERVIT tube. The ammonia gas pressure was adjusted to give a maximum output power of 40 mW at a pump power of 1 W. The 40 mW of 12.48 μm radiation emerged collinearly with the residual pump beam of about .5 W power. A second identical dichroic mirror provided a convenient way of reducing the unwanted pump power to a level where a 100 grooves/mm grating sufficed to separate the residual pump beam completely from the 12.48 μm beam. A small fraction (2 mW) of the 12.5 μm beam was directed into a 1 m long absorber cell to set or lock the ammonia laser frequency onto the saturated absorption dip of the same sP(8,6) transition as the laser is operating on. The ammonia laser frequency is only tunable over $\pm$2 MHz from line center and the laser was initially locked to peak power. However subsequent experiments with the barium ion trap showed a much better frequency reproducibility when the laser was locked to the approximately 300 kHz wide saturation dip despite a relatively poor signal to noise ratio.

The frequency (f_1) of the sP(8,6) lasing transition was measured by heterodyning against two known stabilized CO_2 laser frequencies (f_2 and f_3) in a tungsten-nickel point-contact diode. The fourth order beat signal $f(beat)=f_1 - (2f_2 - f_3)$ was measured to be 313.8 MHz, where the CO_2 laser frequencies are $f_2=P_I(20)=28\ 306\ 224.91$ MHz and $f_3=R_{II}(34)=32\ 583\ 536.61$ MHz, resulting in a value of 24 029 227.0 $\pm$.2 MHz for the sP(8,6) frequency containing the uncertainties of setting the ammonia laser to peak power and the reference lasers to the 4.3 μm fluorescence dip of CO_2. The passive frequency stability [7] of the ammonia laser was only partially exploited at the present stage of the experiment. The laser frequency is isolated from changes in the length of the cavity by a factor of 8, but the isolation of the laser frequency from pump frequency fluctuations was not yet optimised.

The sP(8,6) $^{15}NH_3$ laser frequency requires a down-shift of 17.2 GHz to probe the D-D transition of Ba^+. This was accomplished by a

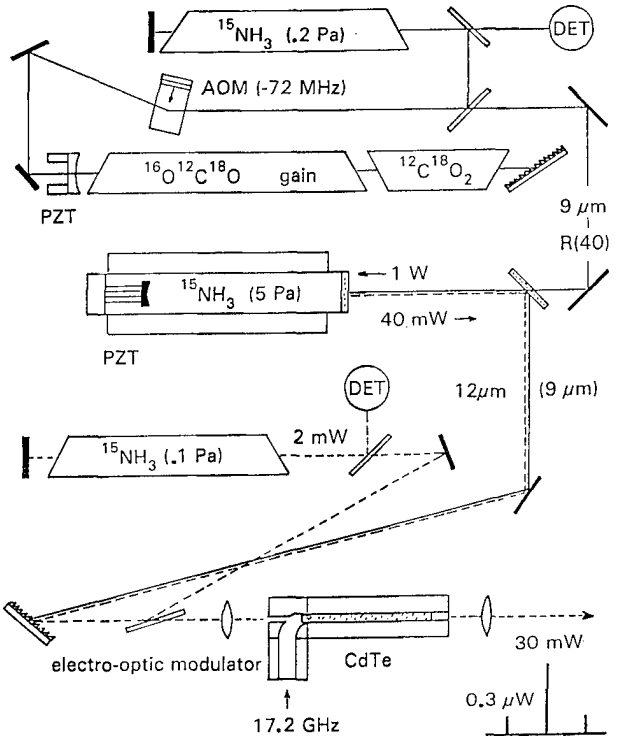


Fig.3. The ammonia laser arrangement, followed by an electro-optic modulator for probing the Ba^+ clock-transition.

specially built electro-optic modulator shown on the bottom part of Fig.3. A CdTe crystal with the dimensions 2.5 mm x 5.55 mm x 50 mm for velocity phase matching and cut for AM modulation was mounted inside a well fitting copper waveguide. The microwave radiation is fed into the crystal through a 90° H-bend which contains a tapered waveguide section for connecting to the standard K_u-band waveguide. The ammonia laser beam is focused through a 2.3 mm hole in the H-bend into the crystal with the beam polarization parallel to the [110] crystal axis which is also parallel to the 2.5 mm crystal dimension and to the E-field vector of the microwave. The crystal is 10 mm shorter than the enclosing waveguide so that the microwave is retro-reflected at the end of the crystal. No metallic reflector is required and the laser beam leaves the crystal unobstructed. The front of the crystal is protruding slightly into the tapered section of the waveguide to minimize the microwave reflection at this interface. A slide tuner in front of the H-bend (not shown) is used to tune

the microwave cavity into resonance. The modulator produces two orthogonally polarized side bands on the laser carrier frequency, the lower frequency side band was used for probing the ion. The sideband power was measured by using a calibrated HgCdTe detector in conjunction with a $^{15}NH_3$ absorber cell, which completely absorbs the carrier frequency, but allows transmission of one or both side bands depending on the microwave frequency chosen [8].

With a microwave power of 1 W and an ammonia laser power of 30 mW (at the modulator) the sideband power measured 0.34 μW, being almost a factor of 10 smaller than reported in the literature for a similar modulator [9]. There was no measurable absorption loss of the 12.48 μm beam in the 50 mm long, AR-coated crystal, so that air gaps between the crystal and waveguide wall may be responsible for the difference in performance. Nevertheless the sideband power was adequate for the present experiments. The power of the laser carrier frequency was reduced as a precautionary measure before it entered the trap by placing two thin GaAs Brewster plates into the beam emerging from the electro-optic modulator.

Fig.4 gives an overview of the clock-transition spectrum of a single barium ion obtained with the ammonia laser probe. Ten distinct narrow lines distributed over the a spectral range of about 80 MHz are observed, whereas a smeared out spectrum was obtained with the tunable diode laser before [4]. The lines are identified as Zeeman components by observing their frequency dependence on externally applied magnetic fields. From the line spacing in Fig.4 a residual magnetic field of 16 gauss is calculated, which is 16 times higher than originally measured at the trap location and is probably caused by the nickel-barium oven or nearby nickel trimming electrodes. Subsequently a number of permanent magnets were placed outside the vacuum chamber to reduce the total distribution range of the Zeeman lines to 42 MHz. The present study was mainly concerned with the $^{138}Ba^+$ isotope. We observe a significant isotope shift when the less common $^{136}Ba^+$ isotope, shown in Fig.4, is occasionally loaded into the trap. The coincidence of two Zeeman components (m_j',m_j'') = (-1/2,-3/2) and (3/2,3/2) confirms the value of the Landé factor. A closer inspection of the spectrum shows that the line width of the magnetically most sensitive Zeeman components ($\pm$5/2,$\pm$3/2) is no greater than that of the most

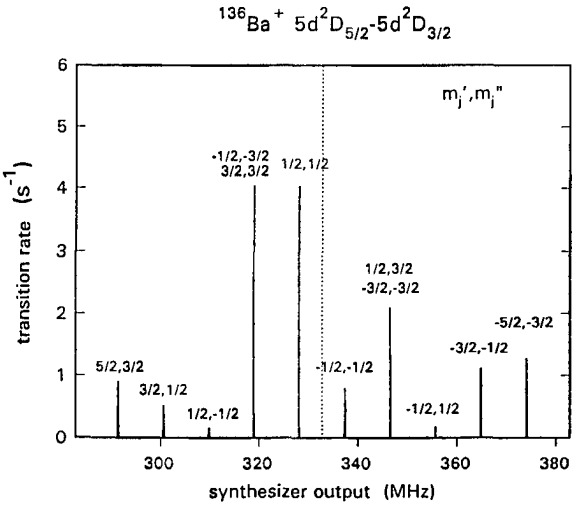


Fig.4. Measured Zeeman spectrum of the $^{136}Ba^+$ clock transition.

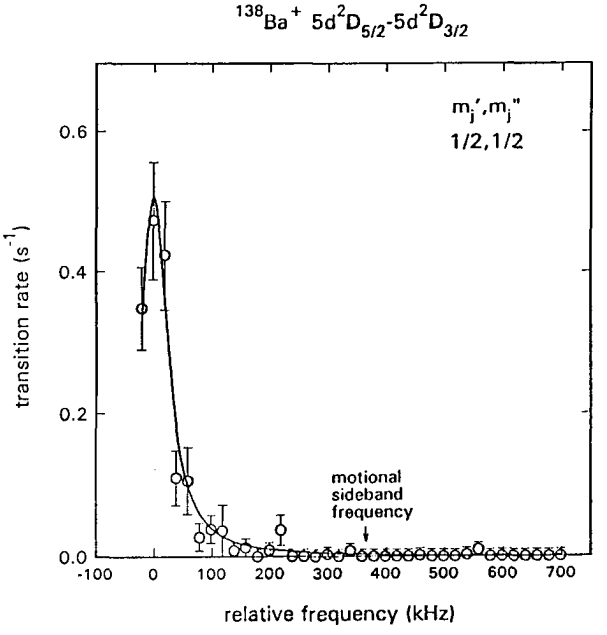


Fig.5. Frequency scan around the (m_j', m_j'') = (1/2, 1/2) Zeeman component in search for motional side bands.

insensitive components ($\pm 1/2, \pm 1/2$), indicating the absence of any significant AC magnetic fields.

Fig.5 shows a scan around the $(m_j', m_j'') = (1/2,1/2)$ Zeeman component of the $^{138}Ba^+$ isotope in search for motional sidebands. Trim electrodes had already been installed inside the vacuum chamber prior to these experiments to minimize any static electric fields. Their performance had been tested through optimising the 493 nm fluorescence signal. The motional side band frequencies were calculated from the trap parameters. From the absence of such side bands at the present resolution the ion temperature can be estimated to be below 45 mK. The width of the (1/2,1/2) Zeeman line in Fig.5 is caused by the relatively large frequency modulation applied to the ammonia laser in order to lock it onto the sP(8,6) saturated absorption dip. The line width reduces considerably when the ammonia laser is free running as shown below in Fig.6.

Fig.6 shows the line shape of the (1/2,1/2) Zeeman component obtained by averaging 5 scans over the transition. The ammonia laser was free running but drifted slowly enough in frequency to complete each scan without resetting to the saturated absorption dip. The reproducibility of the line center frequency as measured on several days was about ± 20 kHz, given by the different techniques of locking or setting the laser to the saturated absorption dip. By measuring the (-1/2,-1/2) Zeeman component in the same fashion and taking the arithmetic mean of the two frequencies the $5d^2D_{3/2}$ - $5d^2D_{5/2}$ center frequency was obtained to be 24 012 048 507 kHz with an uncertainty of 20 kHz against the saturated absorption dip of sP(8,6) $^{15}NH_3$, but a 1σ uncertainty of 0.2 MHz for the absolute value of the frequency.

The preliminary results reported here are encouraging. They show the potential of the D-D transition in Ba^+ as a frequency standard. We are not approaching a limit to this potential at the present stage of the experiment. In fact all that we are seeing right now seems to be the probe laser's own limitations. We are expecting a significant improvement in the frequency determination of the clock-transition center as the ammonia laser will be adjusted to the frequency stability previously attained with ammonia lasers [5,7], and will be linked to a primary standard.

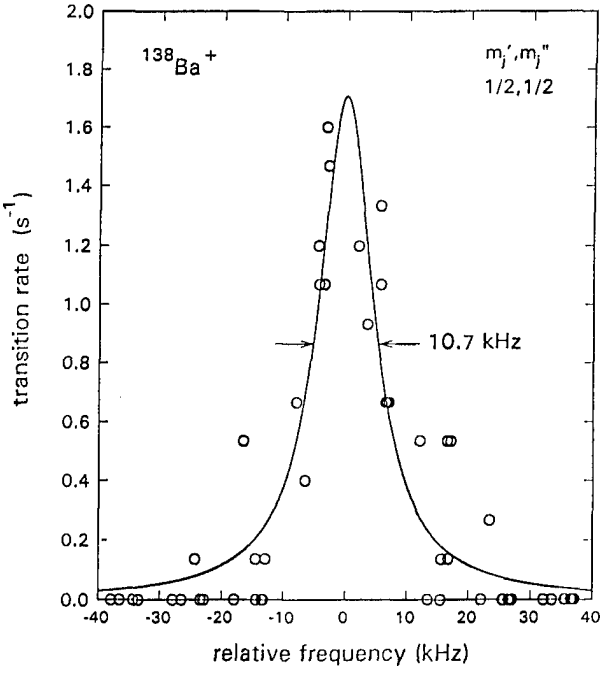


Fig.6. Frequency scan of the (1/2, 1/2) Zeeman component using a free running ammonia laser. The solid line is a Lorentzian fitted to the data.

REFERENCES

[1] H. Dehmelt, W. Nagourney and G. Janik, Bull. Am. Phys. Soc. 27, 402 (1982).

[2] G. Werth, in Frequency Standards and Metrology, p. 293, edited by A. DeMarchi, Springer Verlag, Berlin, 1989.

[3] K. J. Siemsen , A. A. Madej, J. D. Sankey, J. Reid and G. Magerl in Laser Spectroscopy IX, edited by M. S. Feld, J. E. Thomas and A. Mooradian (Academic, San Diego, 1989), p. 292.

[4] A. A. Madej, J. D. Sankey, G. R. Hanes, K. J. Siemsen and A. R. McKellar, Phys. Rev. A 45, 1742 (1992).

[5] K. J. Siemsen, A. A. Madej, G. R. Hanes and J. Reid, Appl. Phys. B 54, 126 (1992).

[6] The grating was made by Hyperfine, Inc., Boulder, Col., USA.

[7] K. J. Siemsen, E. Williams and J. Reid, Opt.Lett. 12, 879 (1978).

[8] The sP(8,7) line of $^{15}NH_3$ absorbs one of the sidebands for a microwave frequency of the difference sP(8,6)-sP(8,7)=17.18 GHz.

[9] G. Magerl, W. Schupita and E. Bonek, IEEE QE-18, 1214 (1982).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DOPPLER EFFECTS IN A MULTI-WAVELENGTH RABI CAVITY FOR THE EXCITATION OF ATOMIC BEAMS IN MICROWAVE FREQUENCY STANDARDS

E. Bava † P. Tavella ‡ A. De Marchi *

†Politecnico di Milano
‡IEN "Galileo Ferraris", Torino,
* Politecnico di Torino

Abstract

An analysis is here presented on the feasibility of using a single multiwavelength Rabi cavity for high resolution and accurate interrogation of an atomic beam in frequency standard applications. Reasons for interest in this solution are mainly the compactness and the ease of construction that such a cavity would offer, but major questions must be first answered on the expected magnitude of Doppler related effects before a resonator of this kind can be seriously considered.

An analytical approach and numerical calculations are given here, which show that the problem may not be as dramatic as one might think. In particular it seems that a cylindrical cavity in the high intrinsic Q TE_{01} mode would be capable of compensating Doppler effects to a very high degree.

1 Introduction

The excitation of microwave transitions in an atomic beam by interaction with the resonant field within a single multi-wavelength Rabi cavity has been recently proven to be capable of producing high resolution interference patterns similar to those obtained with a multiple interaction Ramsey scheme. A cylindrical TE_{013} cavity was used in [1] to confirm experimentally the theoretical analysis, and a line Q of about 10^7 was observed with a thermal optically pumped beam.

Advantages of using such a Rabi cavity, rather than a traditional Ramsey structure, may be seen in the extremely simple design of the cylindrical cavity, with the resulting ease of symmetrization, and in the low losses of the proposed mode. On both accounts it should turn out to be easy to minimize systematic effects related to the existence of a net residual travelling wave in the cavity, such as the first order Doppler effect.

However, while in a Ramsey cavity the atomic beam drifts at right angle to the direction of power propagation, and is exposed to the exciting field only for a fraction of the drift time (in this way minimizing the first order Doppler effect), in the considered Rabi cavity it drifts longitudinally, along the direction of power propagation, and is exposed to the exciting field throughout the drift time. As a result it can be expected that the effects of asymmetries may be maximized.

In this paper the outlined problem is studied in detail for a cylindrical TE_{017} cavity, in view of the possible use of such a cavity in a primary frequency standard. Atoms are followed as a two-level system in their travel through the cavity, and numerical results for the Doppler effect are derived, as a function of cavity asymmetry and microwave power, for a monokinetic beam of 100 m/s. The bias turns out to be linear with asymmetries and negligible for small coupling holes. Total net biases below 10^{-13} appear obtainable with mechanical tolerances which are feasible in cavity fabrication. It is obvious that the bias scales as the inverse of the atomic velocity, so that it would fall below 10^{-15} in a measurement using a 1 m/s beam.

0-7803-0476-4/92 $3.00 © 1992 IEEE

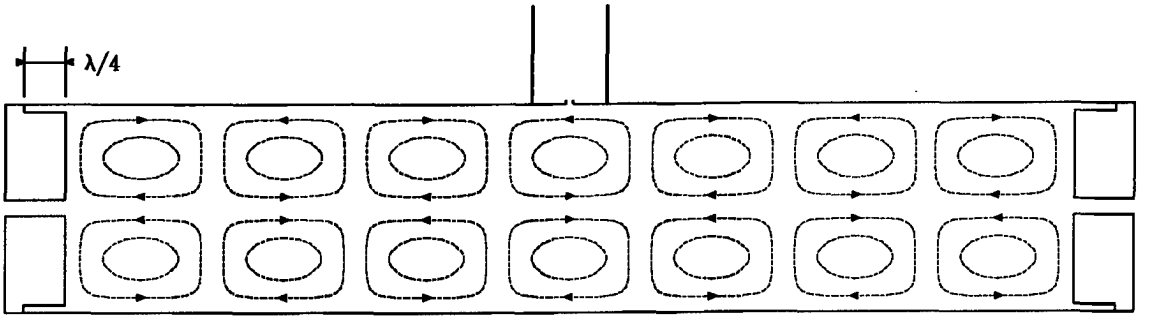


Figure 1: Cross section of the considered cavity with magnetic field lines.

2 Microwave cavity and field description

A cross section of the TE_{017} cavity which is being considered as a possible candidate for a Cesium beam frequency standard is shown in Fig. 1.

It was shown in [1] that at low excitation power an $N\lambda_g/2$ Rabi cavity with small N yields a narrow resonance linewidth, comparable to that of a Ramsey cavity of equal length. It is then clear that N should not be too great. The diameter of the circular waveguide in Fig. 1 is chosen not far from cutoff, so that a small number of half wavelengths (seven) can provide a long enough interaction region for the atomic beam, and yield a narrow resonance linewidth. It is reported below that a linewidth of about 150 Hz is expected from a 100 m/s beam and a 0.37 m long $7\lambda_g/2$ cavity. The price of a greater attenuation constant is paid in this configuration, with respect to the very small values obtained in a normally operated guide. However, because of the intrinsically low attenuation of the TE_{01} mode, the latter is still three times smaller than in a standard rectangular waveguide operated in the normal mode.

In order to excite the desired π transitions in the atomic beam, such a cavity is obviously placed in a longitudinal quantization field, so that the relevant rf field is the longitudinal rf magnetic field component close to cavity axis. The analysis of the interaction, which is illustrated in the following paragraph, requires an analytical description of that field component, which can be derived in the scattering matrix formulism with the model shown in Fig. 2.

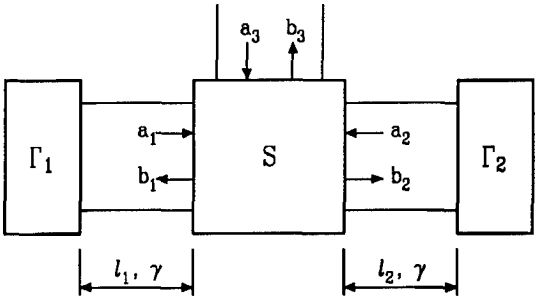


Figure 2: Model used for field calculations.

Only the length of the two sides and the power dissipation in the two end shorts are allowed to be slightly asymmetric in the proposed model. It will then be, for the length: $\ell_1 - \ell_2 = 2d$ and $\ell_1 + \ell_2 = 2\ell = L$; and for the end caps: $\Gamma_1 = \Gamma e^{\chi}$ and $\Gamma_2 = \Gamma e^{-\chi}$, with $\Gamma = -e^{-\xi}$ and $\chi = \chi' + j\chi''$, $\xi = \xi' + j\xi''$. The propagation constant $\gamma = \alpha + j\beta$ instead is considered equal and uniform along the two sides of the cavity; and the scattering matrix of the center feed is also considered symmetric:

$$S = \begin{pmatrix} R & M & T \\ M & R & T \\ T & T & K \end{pmatrix}$$

with R and K representing the reflections from the two guide sections and from the center feed respectively, and M and T the transmissions between the two sides of the cavity and between one side and the feed. In the small losses approximation for the feed, the following ties are derived for the S parameters from the unitarity conditions: $2|T|^2 + |K|^2 = 1$, $M + R = -K^* \frac{T}{T^*}$, and $M - R = e^{j\delta}$ with δ real. In the limit of small

coupling the reference planes where T is real coincide at the center of the coupling hole and δ vanishes (this is obvious, because then $M \to 1$ and $R \to 0$). It will be shown below that this δ plays an important role in the determination of the Doppler effect due to the combined action of feed mispositioning and losses in the guide. The latter in fact turns out to be proportional to $\beta d\delta$. For a circular coupling hole of 4 mm in diameter it can be calculated that $\delta \simeq 2 \cdot 10^{-3}$ rad [3].

A travelling wave analysis of the model of Fig.2 shows that waves b_1 and b_2 entering the two halves of the cavity from the feeding junction can be expressed as a sum of symmetric and antisymmetric components b_s and b_a. These are to be understood as the waves which result from perfect symmetry, and the (small) deviations from them which derive from the existence of some asymmetry in the structure. It is then: $b_1 = b_s + b_a$ and $b_2 = b_s - b_a$, where

$$b_s = \frac{a_3 T}{\Delta_s}$$

$$b_a = (e^{2\gamma d - \chi} - 1)\left(1 - \frac{1}{\Delta_a}\right) b_s = \epsilon b_s$$

$$\Delta_s = 1 - (M + R)\Gamma e^{-2\gamma\ell}$$

$$\Delta_a = 1 + (M - R)\Gamma e^{-2\gamma\ell}$$

As a consequence also the longitudinal magnetic field component H_x, which is relevant to the problem at hand, is the sum symmetric and antisymmetric terms with the same meaning as for the waves. As a matter of fact $j\omega H_x$ is proportional (with a purely real constant) to the transverse electric field, that is to $a + b$, and, with proper signs for the propagation in the two cavity halves ($x < 0$ in half #1 and $x > 0$ in half #2), is:

$$H_{x1}(x) \propto b_1 \left(e^{\gamma x} + \Gamma_1 e^{-\gamma(2\ell_1 + x)}\right)$$

$$H_{x2}(x) \propto b_2 \left(e^{-\gamma x} + \Gamma_2 e^{-\gamma(2\ell_2 - x)}\right)$$

These expressions must be used with actual cavity dimensions and operation frequency, which implies that the considered mode is on resonance. Under these conditions $\Delta_a \simeq 1 + e^{j\delta}$. Reference planes at the junction are assumed to make T real.

Explicit time dependance must be introduced in order to obtain the actual instantaneous value of the field, which is needed to study the interaction with the atomic beam. This is readily done by taking the real part of the complex field, multiplied by $e^{j\omega t}$. The result is a time dependent real magnetic field.

It can then be seen that the asymmetric part of the field, and so the Doppler effect, is linearly dependent on $\epsilon = \epsilon' + j\epsilon''$. Linearization of the latter highlights the quantities which are responsible for the Doppler effect. It is:

$$\epsilon' \simeq \alpha d - \chi'/2 - \frac{\delta}{2}(\beta d - \chi''/2) \simeq -(\chi'/2 + \frac{\delta}{2}\beta d)$$

$$\epsilon'' \simeq \beta d - \chi''/2 + \frac{\delta}{2}(\alpha d - \chi'/2) \simeq \beta d - \chi''/2$$

and $\epsilon = \beta d(j - \frac{\delta}{2})$ in the case the two end caps can be considered symmetric.

It turns out that the real part of ϵ is the important component, and it is then apparent that for the length asymmetry it is the quantity $\beta d\delta$ that counts, as already anticipated above. This means that the asymmetry of the feed position is weighted by δ, and therefore becomes unimportant in the limit of small coupling holes. For these, in fact, $\delta \to 0$ as the third power of the hole's diameter [3].

On the other hand the asymmetry of power dissipation in the two caps produces directly a Doppler effect, independently of cavity asymmetry and coupling hole position. This contribution is accounted for by χ', which is the difference in normalized surface resistance of the end caps (the latter is $\simeq 3 \cdot 10^{-5}$ in the present case, so that $\chi' \simeq 3 \cdot 10^{-6}$ if the two caps differ by 10%). It turns out that the latter is the most severe effect.

3 Interaction with the atoms

A frequency shift due to first order Doppler effect results from the combined effect of losses and structure asymmetry.

The magnitude of the effect can be evaluated by calculating the atoms' transition probability as a function of detuning from the natural resonance ω_0 of the atom at rest.

By using the interaction representation, in which we indicate the Hamiltonian with $\mathrm{H}(y)$, and the rotating wave approximation, after introducing adimensional quantities, we can write for the two-level system (the upper sign refers to half #2 of the cavity, where $x > 0$):

$$\mathrm{H}(y) = -\frac{\hbar \mathrm{b_c} \tau}{4} \begin{pmatrix} 0 & (1 \pm \epsilon^*) e^{-j\theta y} \frac{q^*}{2} \\ (1 \pm \epsilon) e^{j\theta y} \frac{q}{2} & 0 \end{pmatrix}$$

where $y = \beta x = \beta v t$, $\tau = \frac{\lambda_g}{\pi v} = \int_{-\pi/2\beta v}^{\pi/2\beta v} cos\beta v t dt$, $\theta = (\omega - \omega_0)\frac{\tau}{2}$ and q is the complex expression for the H_x field given above, except for the amplitude, which is included in $\mathrm{b_c}$. By $\mathrm{b_c} = 2 b_s \mu_0 \mu_{12}/\hbar$ we indicate the Rabi angular frequency corresponding to the standing wave amplitude[1] (μ_{12} is the magnetic dipole moment of the transition, and μ_0 is the vacuum permeability).

Numerical solutions are obtainable by integration of the differential equations:

$$\begin{cases} \frac{\partial C_1}{\partial y} = j \frac{\mathrm{b_c} \tau}{4} (1 \pm \epsilon) e^{j\theta y} \frac{q}{2} C_2 \\ \frac{\partial C_2}{\partial y} = j \frac{\mathrm{b_c} \tau}{4} (1 \pm \epsilon^*) e^{-j\theta y} \frac{q^*}{2} C_1 \end{cases}$$

where $C_i C_i^*$ is at any chosen time the probability of level i.

4 Numerical results

The evolution in time of the probability amplitudes C_1 and C_2 along the atoms' flight through the cavity was studied by numerical integration of the reported differential equations with a 4th order Runge Kutta method.

[1]This is the same notation used in [2]

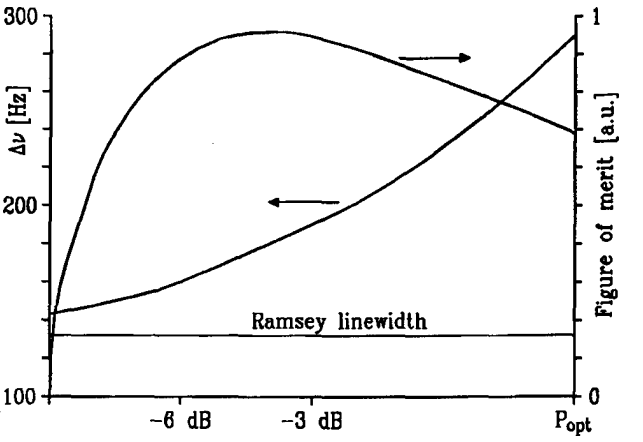


Figure 3: Calculated linewidth and figure of merit as a function of power. The linewidth of a Ramsey resonance in a cavity of equal length is also reported for comparison. Velocity is 100 m/s.

The error tolerance for the algorithm was set to 10^{-10} in all cases, except in the case of $\delta = 0$, for which it was set to 10^{-11}. The same level of confidence was verified at each integration step by comparing to 1 the calculated value of $|C_1|^2 + |C_2|^2$. Initial values are $C_1 = 1$, and $C_2 = 0$ in all cases. In this way the transition probability $|C_2|^2$ at the exit of the cavity is the final value that it reaches at the end of its evolution.

Calculation of this quantity as a function of mistuning from the unperturbed natural resonance frequency yields directly the lineshape of the beam response. Characterizing parameters of these lineshapes are reported in Fig. 3. The linewidth of a Ramsey resonance in a cavity of equal length is also reported.

From the lineshapes obtained in this way, for various cavity asymmetry and microwave power conditions, the total net Doppler effect was calculated as the displacement of line center. The latter was inferred from the difference in probability at the same mistuning on the two sides from unperturbed line center, and from the computed values of the slope of the probability curve.

Results of these calculations are reported in Fig. 4 and 5, where the Doppler effect has been normalized for generality with respect to atomic velocity.

It can be seen that asymmetry in the end caps

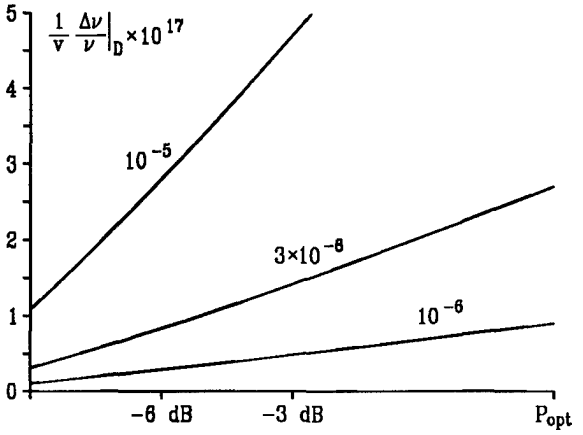


Figure 4: Normalized Doppler effect for equal end caps as a function of power for different values of $\beta d\delta/2\pi$.

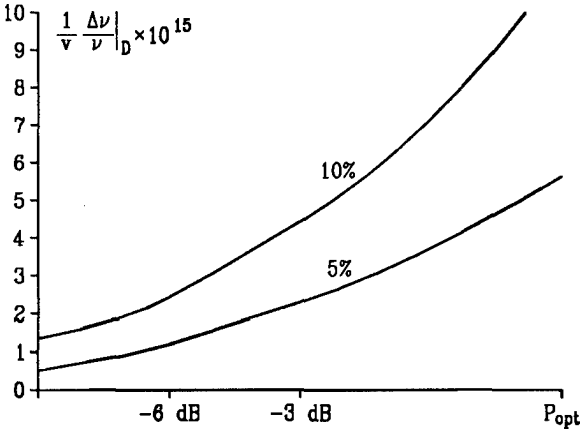


Figure 5: Normalized Doppler effect for a symmetric cavity as a function of power for different relative end caps asymmetry ($\chi'/(1-|\Gamma|)$).

gives the most important contribution, at the level of 10^{-13} for 5% cap asymmetry, a beam of 100 m/s, and a power level 6 dB below optimum power. If end caps can be made symmetric to 0.5%, a net Doppler effect of 10^{-14} is expected under the same conditions.

5 Conclusions

The conclusion of the present study seems to be that the considered multiwavelength Rabi cavity, has great potential in the excitation of the clock transition in primary atomic beam frequency standards, both from the point of view of resonance linewidth and from the point of view of compensation of Doppler related effects. This conclusion goes against the first impression, that such an excitation arrangement would be very sensitive to cavity asymmetries because the atoms travel along the direction of power propagation. The unexpectedly good compensation can be understood in relation with the very low attenuation of the waveguide mode and the vanishing effect of a small feeding hole's position on field asymmetry. The limiting effect to the reduction of Doppler shift seems to be the expectedly limited capability of symmetrizing losses in the two end caps, particularly in view of the fact that one of them may be more subject than the other to Cesium deposition in Cesium standard applications.

A number of cautionary comments should be added at this point about the fact that the cavity model used for calculations may not be adequate to represent actual realizations. In particular, no effect of disuniformity in waveguide diameter or attenuation has been allowed. However, the effect of asymmetries in propagation constant variations should be smaller than the net calculated effects, which are already small enough.

The proposed cavity could conceivably be realized in glass with silver deposition, and allow in this way optical surface quality of the inner wall as well as high temperature stability.

Acknowledgements The authors wish to acknowledge the valuable help of G. P. Bava and E. Rubiola, of the Politecnico di Torino, and fruitful discussions with J. Shirley of NIST and L. Cutler of HP.

References

[1] A. De Marchi, R. E. Drullinger, J. H. Shirley, "Interference fringes from single cavity excitation of an atomic beam", *Proc. 44th Symp. on Freq. Control*, p.34 (1990).

[2] J. Vanier, C. Audoin, *The quantum physics of atomic frequency standards*, Adam Hilger, Bristol (1989).

[3] R. E. Collin, *Field theory of guided waves*, IEEE Press, New York (1991).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

FREQUENCY, PHASE AND AMPLITUDE CHANGES OF THE HYDROGEN MASER OSCILLATION

Claude AUDOIN
Laboratoire de l'Horloge Atomique, Unité Propre de Recherche du CNRS associée à l'Université Paris-Sud - Bât. 221 - Université Paris-Sud
91405 ORSAY Cedex - France
and
William A. DIENER
California Institute of Technology - Jet Propulsion Laboratory
PASADENA, California 91109 - USA

Abstract

The frequency, the phase and the amplitude changes of the hydrogen maser oscillation, which are induced by the modulation of the cavity resonant frequency, are considered. The results obtained apply specifically to one of the H-maser cavity auto-tuning methods which is actually implemented, namely the cavity frequency switching method. The frequency, the phase and the amplitude changes are analyzed theoretically. The phase and the amplitude variations are measured experimentally. It is shown, in particular, that the phase of oscillation is subjected to abrupt jumps at the times of the cavity frequency switching, whose magnitude is specified. The results given can be used for the design of the electronic system associated with the hydrogen maser, with the aim of minimizing the transfer of the phase modulation to the slaved VCXO.

Introduction

In the hydrogen maser [1, 2], the cavity pulling effect translates any microwave cavity frequency offset, $(\omega_c - \omega_o)/2\pi$, into a corresponding frequency bias, $(\omega - \omega_o)/2\pi$, of the maser oscillation. This effect is described by the equation :

$$\omega - \omega_o = Q_c(\omega_c - \omega_o)/Q_\ell \qquad (1)$$

where ω_o, ω_c and ω are the angular frequencies of the atomic resonance, of the cavity resonance and of the maser oscillation, respectively. Q_c and Q_ℓ are the quality factors of the microwave cavity and of the atomic line. In the following, we shall neglect any frequency offset due to the spin-exchange interaction between hydrogen atoms and we shall assume that the value of Q_c does not depend on the cavity resonant frequency.

Different cavity auto-tuning systems have been implemented. Their aim is to control the long term frequency stability of the hydrogen maser [3]. Among them, that based on the cavity frequency switching method has proven its efficiency [4,5]. It ensures an excellent long term frequency stability, without perturbing significantly the short and medium terms frequency stability. This method is founded on the inertia property of the atom assembly, which is related to the large value of the transverse relaxation time T_2, of the order of 0.3 s. When the cavity resonant frequency is modulated, with a period of about 0.1 s or shorter, the amplitude of the atomic magnetic moment does not change significantly. If the mean value of the cavity resonant frequency is equal to the atomic transition frequency, the amplitude of the magnetic field sustained by the alternating magnetic moment remains practically a constant. If, in contrast, the mean value of the cavity resonant frequency is not equal to the atomic transition frequency, then the amplitude of the microwave field is modulated, which provides a signal depending on the cavity mistuning.

However, the cavity resonant frequency changes produce a periodic modulation of the frequency and of the phase of oscillation, which has been noticed [4,6]. The amplitude is also very slightly modulated. The purpose of this paper is to give a quantitative information on these effects. They are analyzed theoretically in Section 1 and experimental results are given in Section 2.

1) Theoretical derivation of the maser behaviour

1.1. General equations : The properties of the amplitude and of the phase of oscillation can be described using a radioelectric model [3], which takes into account the resonance property of the microwave cavity , as well as the characteristic properties of the atomic medium, i.e. its frequency selectivity and its saturation by the microwave field.

We introduce the same notations as in [3]. The hydrogen maser output voltage is proportional to $b\cos(\omega t + \varphi)$, where b is the normalized oscillation amplitude and φ is the phase of oscillation. The quantity b is actually the Rabi angular frequency of the stimulated atomic medium. The oscillating magnetic moment of the atom assembly is represented by $m\sin(\omega t + \psi)$, where m is the amplitude of the magnetic moment referred to the Bohr magneton and ψ is its phase. The population difference of the two atomic levels involved is M_3. The cavity response time is T_c and the longitudinal and the transverse relaxation times of the atoms are T_1 and T_2, respectively. We have

$$T_c = 2Q_c/\omega_c \text{ and } T_2 = 2Q_\ell/\omega_o . \qquad (2)$$

In practice, we have $T_c \approx 10\ \mu s$, T_1, $T_2 \approx 0.3$ s and, therefore $T_c \ll T_1, T_2$. Furthermore $(\omega_c - \omega_o)/\omega_o$ is always a very small quantity. In general, the parameters b, m, M_3, φ and ψ can be time dependent. The oscillation angular frequency ω will be assumed constant during each half-period of modulation.The instantaneous oscillation angular frequency will thus be given by $\omega + \dot{\varphi}$, where the dot means a time derivative.

0-7803-0476-4 $3.00 © 1992 IEEE

In the present case, where we neglect the cavity thermal noise and where no external signal feeds the maser cavity, the equations describing the maser behavior are:

$$T_2 \dot{m} + m = M_3 \, b \, T_2 \cos\theta \qquad (3)$$

$$T_1 \dot{M}_3 + M_3 = - T_1 \, b \, m \cos\theta + T_1 \tilde{I} \qquad (4)$$

$$T_c \dot{b} + b = K Q_c \, m \cos\theta \qquad (5)$$

$$\dot{\psi} = \omega_o - \omega + M_3 \, (b/m) \sin\theta \qquad (6)$$

$$\dot{\varphi} = \omega_c - \omega - (K Q_c \, m / T_c b) \sin\theta \qquad (7)$$

$$\theta = \varphi - \psi \ , \qquad (8)$$

where $\tilde{I}$ is the flux of state selected atoms entering the maser storage bulb and K is a constant defined in [3].

Neglecting, for simplicity, any time constant in the setting of the cavity resonant frequency, we shall consider the following cavity frequency modulation :

$$\omega_c(t) - \omega_{co} = \omega_m \, g(t) \qquad (9)$$

where ω_{co} is the mean value of the cavity resonant angular frequency, ω_m is the amplitude of modulation and g(t) represents a square wave of period T_M, whose value is + 1 for $0 < t < T_M/2$ and - 1 for $T_M/2 < t < T_M$. We shall assume T_M small compared to T_1 and T_2. The variation of the cavity resonant frequency is shown in figure 1a.

1.2. Stationary regime : We shall show, in Section 1.3, that the duration of the transient response following an abrupt change of the cavity resonant frequency is of the order of T_c. Therefore, during most of each half-period of modulation, a stationary regime is attained and the parameters of equations (3-8) can then be assumed constant. The subscript s will characterize the two steady state values of the oscillation frequency and of the oscillation amplitude occuring during the longest part of each half-period.

Assuming, at first, that the microwave cavity is tuned in the mean sense, i.e. that we have $\omega_{co} = \omega_o$, it can be shown from the preceding equations that the oscillation angular frequency is given by:

$$\omega_s(t) - \omega_o = \frac{T_c}{T_2} \omega_m \, g(t) \ , \qquad (10)$$

in agreement with the cavity pulling effect. The amplitude of oscillation, b_s, is the same during the two half-periods of modulation. We have :

$$T_1 T_2 b_s^2 = \alpha - [1 + T_c^2 \, \omega_m^2] \qquad (11)$$

where $\alpha = K Q_c T_1 T_2 \tilde{I}$ is the oscillation parameter. We see that the amplitude of oscillation is decreased in the presence of modulation, to a degree which depends on the value of α and of $\omega_m T_c$. One may also note that equation (11) provides a simple means of obtaining the value of α by measuring the ratio of the maser output powers for two different values of ω_m. Equation (11) shows that the amplitude of oscillation is the same during the two half-periods of modulation. This follows from a general property of this amplitude, which is a quadratic function of the cavity resonant frequency.

If, now, the cavity frequency is not tuned in the mean sense, the oscillation angular frequency becomes :

$$\omega_s(t) - \omega_o = \frac{T_c}{T_2} [\omega_{co} - \omega_o + \omega_m \, g(t)] \qquad (12)$$

The amplitude of oscillation is no longer a constant. For a small mistuning, such that $T_c(\omega_{co} - \omega_o) << 1$, the change $\Delta b_s(t)$ is given by :

$$\frac{\Delta b_s(t)}{b_s} = - \frac{T_c^2 \omega_m (\omega_{co} - \omega_o)}{\alpha - (1 + T_c^2 \omega_m^2)} \, g(t) \qquad (13)$$

It is periodic, with period T_M, and it is proportional to the mean cavity mistuning $\omega_{co} - \omega_o$. Consequently, a synchronous detection of the modulation of the oscillation amplitude provides an information which makes it possible a feedback control of the cavity resonant frequency [4].

1.3. Phase transient : It can be shown easily, from equations (6-8) that the angle θ varies during a transient regime, even in the case where the cavity is tuned to the atomic transition. This greatly complicates the exact solution of the whole set of differential equations. However, a sufficiently accurate expression of the time variation of the phase angles θ, φ and ψ can be obtained by assuming that the quantity $T_c \dot{b}/b$ remains smaller than unity during the transient. Then, the values of θ and φ obtained will enable us to derive simply the time variation of the measure b of the oscillation amplitude and to verify that $T_c \dot{b}/b$ remains actually small during the transient.

It is appropriate to consider, at first, the differential equation for the angle θ. Neglecting $T_c \dot{b}/b$, we have from equation (3) and (5-8) :

$$\dot{\theta}(1 + \tan^2\theta) = \omega_c(t) - \omega_o - \frac{1}{T_c}(1 + R) \tan\theta \ , \qquad (14)$$

where $\omega_c(t)$ is given by equation (9) and where we have set :

$$R = T_c/T_2 = Q_c/Q_\ell \ . \qquad (15)$$

The parameter R is typically equal to 3×10^{-5}. The periodic solution of equation (14) is :

$$\tan\theta_1(t) = \frac{T_c}{1 + R} (\omega_{co} - \omega_o + \omega_m) - 2 T_c \omega_m \exp(-t/T_c) \qquad (16a)$$

$$\tan\theta_2(t) = \frac{T_c}{1 + R} (\omega_{co} - \omega_o - \omega_m) + 2 T_c \omega_m \exp[-(t - T_M/2)/T_c] \ . \qquad (16b)$$

In these equations and in the subsequent ones, the subscripts 1 and 2 refer to $0 < t < T_M/2$ and $T_M/2 < t < T_M$, respectively. It is worth noting that the factor $(1 + R)$ has been omitted in the second term of the right-hand side of equations (16a) and (16b). This does not change the main conclusions of this paper. In contrast, it cannot be omitted in the first term, to account for the cavity pulling effect. The variation of $\tan\theta$ versus time is shown in figure 1b.

The differential equation describing the behavior of the phase of oscillation is obtained from equations (5) and (7). We have :

$$\dot{\varphi} = \omega_c(t) - \omega_s(t) - \frac{1}{T_c} \tan\theta(t) \qquad (17)$$

where $\omega_c(t)$, $\omega_s(t)$ and $\tan\theta(t)$ are given by equations (9), (12) and (16), respectively. It follows that the instantaneous angular frequency, $\omega_i(t) = \omega_s(t) + \dot{\varphi}(t)$ is given by :

$$\omega_{i1}(t) - \omega_o = \frac{T_c}{T_2} (\omega_{co} - \omega_o + \omega_m) + 2\omega_m \exp(-t/T_c) \quad (18a)$$

$$\omega_{i2}(t) - \omega_o = \frac{T_c}{T_2} (\omega_{co} - \omega_o - \omega_m) - 2\omega_m \exp[-(t - T_M/2)/T_c] \ . \quad (18b)$$

Therefore, the instantaneous oscillation angular frequency shows brief, but very large, deviations, practically equal to $2\omega_m$, at the times of the cavity frequency switching. They are depicted in figure 1c. Typically, $2\omega_m$ is of the order of $2\pi \times 10^4$ rad s^{-1}.

The mean value of the instantaneous oscillation angular frequency is $\overline{\omega_i(t)}$ given by :

$$\overline{\omega_i(t)} - \omega_o = \frac{T_c}{T_2} (\omega_{co} - \omega_o) = \frac{Q_c}{Q_\ell} (\omega_{co} - \omega_o) \ . \quad (19)$$

Equation (19) shows that the mean oscillation frequency obeys the cavity pulling equation (1) and that it is not biased by the cavity frequency switching.

The time variation of the phase of oscillation is obtained by integrating equation (17). We assume that the phase is measured with respect to that of an oscillator having a constant angular frequency equal to $\overline{\omega_i(t)}$ given by equation (19), and that the mean value of the phase difference observed is set equal to zero. We thus have, φ_o being a constant, and for $T_c << T_M/2$:

$$\varphi_1(t) - \varphi_o = \omega_m T_c [1 - 2\exp(-t/T_c)] + \omega_m \frac{T_c}{T_2} (t - \frac{T_M}{4}) \quad (20a)$$

$$\varphi_2(t) - \varphi_o = \omega_m T_c \{2\exp[-(t - T_M/2)/T_c] - 1\} - \omega_m \frac{T_c}{T_2} (t - \frac{3}{4} T_M) \ . \quad (20b)$$

The variation of the phase of oscillation versus time is shown in figure 1d. We see that phase jumps of magnitude $2\omega_m T_c$ arise at the times of the cavity frequency switching. These phase variations occur with a time constant equal to the cavity response time, assuming abrupt cavity resonant frequency changes. Afterwards, the phase variation follows a ramp of slope $\pm \omega_m T_c/T_2$ during each half-period of modulation. It is due to the cavity pulling effect which causes an offset of the oscillation angular frequency, equal to $\pm \omega_m T_c/T_2$, with respect to $\overline{\omega_i(t)}$ Equations (20a) and (20b) are valid without an approximation on the magnitude of $\omega_m T_c$, provided of course that the maser oscillation subsists at the given offset $\pm \omega_m$ of the cavity resonant angular frequency.

In a simple, but efficient, model of the H-maser, it has been assumed that the oscillating magnetic moment, which excites the cavity electromagnetic field, behaves like an independent source driving a tuned circuit [4]. If η is the total change of the phase shift across this circuit, occuring when its angular resonant frequency is symmetrically offset from one side of the fixed value ω_o to the other, with a total range of $2\omega_m$, then we have:

$$\eta = 2 \arctan(\omega_m T_c) \ . \quad (21)$$

For practical values of $\omega_m T_c$, of the order of 0.3, we have $\eta \approx 2\omega_m T_c$. η is thus an excellent estimate of the oscillation phase jump.

The phase ψ of the magnetic moment can be obtained from equations (8), (16) and (20). For practical values of $\omega_m T_c$, and assuming $\omega_{co} = \omega_o$, we have :

$$\psi_1(t) - \varphi_0 \approx \omega_m \frac{T_c}{T_2} (t - \frac{T_M}{4}) + \frac{1}{3} \tan^3 \theta_1(t) \quad (22a)$$

$$\psi_2(t) - \varphi_o \approx -\omega_m \frac{T_c}{T_2} (t - \frac{3}{4} T_M) + \frac{1}{3} \tan^3 \theta_2(t) \ . \quad (22b)$$

Actually, the variation of ψ is noticeably smaller than that of θ and φ. This is shown schematically in figure 1e.

1.4. Amplitude transient : We define ε, the relative change of the amplitude of oscillation by :

$$\varepsilon = \frac{b}{b_s} - 1 \ . \quad (23)$$

Let us consider, at first, the beginning of each half-period of modulation, during a time interval equal to several times T_c. Since the time constant associated with the amplitude m of the magnetic moment is $T_2 >> T_c$, the value of this amplitude does not change significantly and it remains equal to that attained at the end of the preceding half-period, i.e. m_s. We thus have, from equation (5) :

$$T_c \dot{\varepsilon} + \varepsilon = \frac{\cos\theta}{\cos\theta_s} - 1 \ , \quad (24)$$

where $\cos\theta_s$ is the steady state value of $\cos\theta$. Assuming $\omega_{co} = \omega_o$, figure 2a shows the small variation of ε versus time, for three different values of the parameter $\omega_m T_c$, around the experimental value $\omega_m T_c \approx 0.3$. The amplitude transient repeats itself identically after each cavity frequency abrupt change, since $\cos\theta$ is an even function. Therefore, its period is $T_M/2$ Figure 2b shows that the value of $T_c \dot{b}/b$ remains much smaller than unity. Consequently, the assumption made to derive the equations for the phase angles θ, φ and ψ was actually justified.

After the transient is finished, the angle θ is a constant and the variation of ε is given by :

$$\ddot{\varepsilon} + \frac{1}{T_1} \dot{\varepsilon} + 2 b_s^2 \varepsilon = 0 \ . \quad (25)$$

However, the initial values of ε and $\dot{\varepsilon}$ are equal to zero, so that the oscillation amplitude is a constant during the remainder of each half-period.

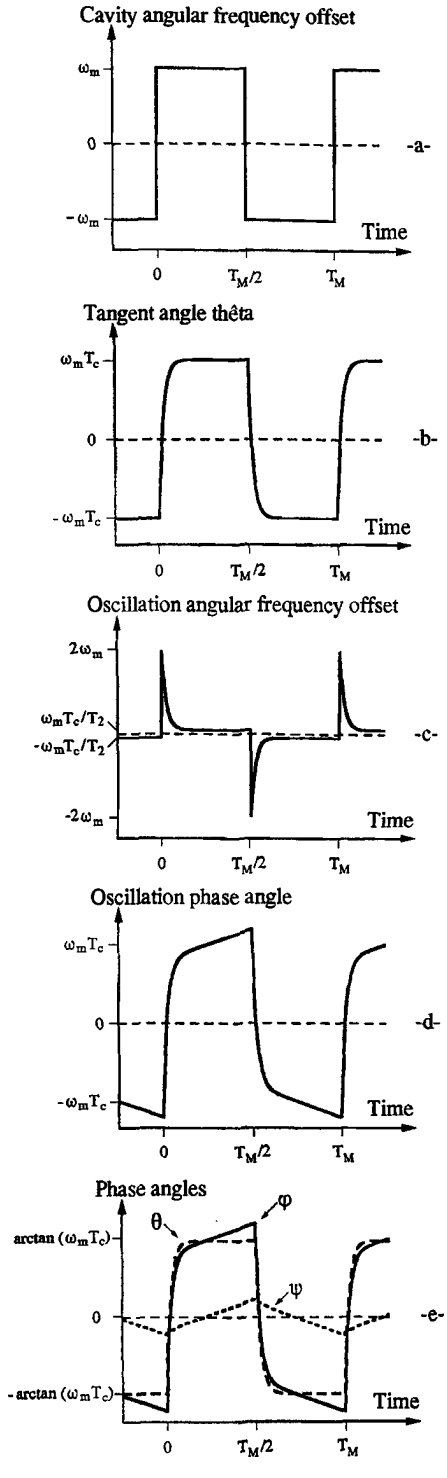


Fig. 1. Periodic variation of parameters of interest. The value of the ratios T_c/T_M and T_c/T_2 has been greatly increased, for the purpose of illustration. It has been assumed $\omega_{co} = \omega_o$.

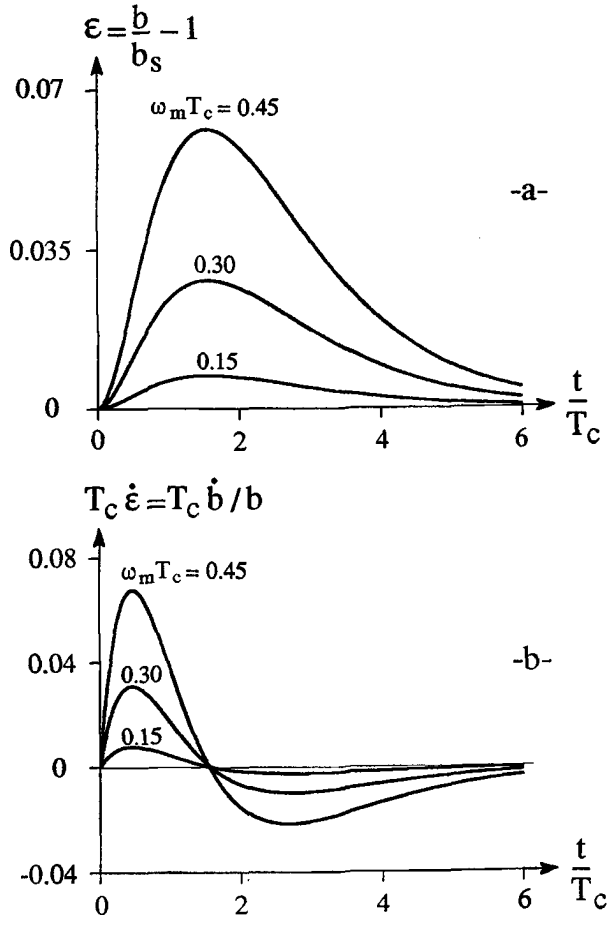


Fig. 2a. Amplitude transient for three different values of the parameter $\omega_m T_c$. The jump of the cavity resonant frequency occurs at t = 0. It is assumed $\omega_{co} = \omega_o$.

Fig.2b. Variation of the quantity $T_c \dot{\varepsilon} = T_c \dot{b}/b$ versus time after a cavity frequency jump, for three different values of the parameter $\omega_m T_c$. It is assumed $\omega_{co} = \omega_o$.

2) Measurement of the phase and of the amplitude transients

2.1 Experimental set-up : The maser used in this experiment, owned by JPL, was manufactured by the Sigma Tau Standards Corporation. Its cavity quality factor is Q_c = 47 000 and thus the cavity response time is T_c = 10.5 μs. The atomic flux is set "high" and the atomic line quality factor is $Q_\ell = 1.48 \times 10^9$. The frequency of the modulation of the cavity resonant frequency is 8.5 Hz. The peak to peak cavity frequency change is 9.5 kHz, giving :

$2\omega_m = 2\pi \times 9.5 \times 10^3$ rad s^{-1} and $\omega_m T_c = 0.31$.

The experimental set-up is represented in fig. 3. In its principle, it makes it possible to compare the modulated phase of the maser oscillation at 1.4 GHz to the mean value of that phase. For that purpose, an auxiliary high quality 5 MHz VCXO is phase locked to the 5 MHz output of the maser, but with a bandwidth of about 0.5 Hz, which is sufficiently narrow to attenuate greatly any residual phase modulation of the maser 5 MHz output. The frequency of the VCXO is multiplied up to 1 400 MHz and mixed with the amplified maser output, at 1420. 405 MHz. The gain is of the order of 120 dB. The signal obtained at 20.405...MHz feeds a balanced mixer used as a phase detector, which also receives a signal from a frequency synthesizer driven by the auxiliary VCXO. The frequency setting of this synthesizer is the same as that of the build-in maser synthesizer. Moreover, the phase of the maser signal can be shifted, so that a voltage with a zero mean value can be obtained at the output of the phase detector. This voltage varies in the linear part of the phase detector characteristics. It represents the oscillation phase change related to the cavity resonant frequency switching. The phase measurement is calibrated as usual by offsetting slightly the frequency of the signal delivered by one of the synthesizers.

The changes of the maser oscillation amplitude are observed simply by connecting a rectifier to the amplified maser output, in place of the mixer at 20.4 MHz. Calibrated attenuators enable to measure the relative amplitude changes.

2.2 Experimental results : Figure 4a shows a recorded phase transient. Assuming an exponential decay, the time constant is of the order of 12 μs, in satisfactory agreement with the theoretical value of 10.5 μs. The slight lengthening of the transient is likely related to the time constant -of the order of 3 μs- of the control voltage applied to the varactor coupled to the microwave cavity.

The amplitude change is shown in figure 4b (the voltage detected was negative). The amplitude increases slightly after the cavity frequency switching and returns to its initial value after a few times T_c, in agreement with theory. However, the relative size of the bump is of the order of 1 % instead of about 3 % as predicted from figure 2a. This disagreement might be due to the additional time constant of the varactor control voltage and to a possible mismatch of the calibrated attenuator used to measure the level change. It has been verified that, besides this bump, the modulation of the cavity resonant frequency does not induce oscillation amplitude changes.

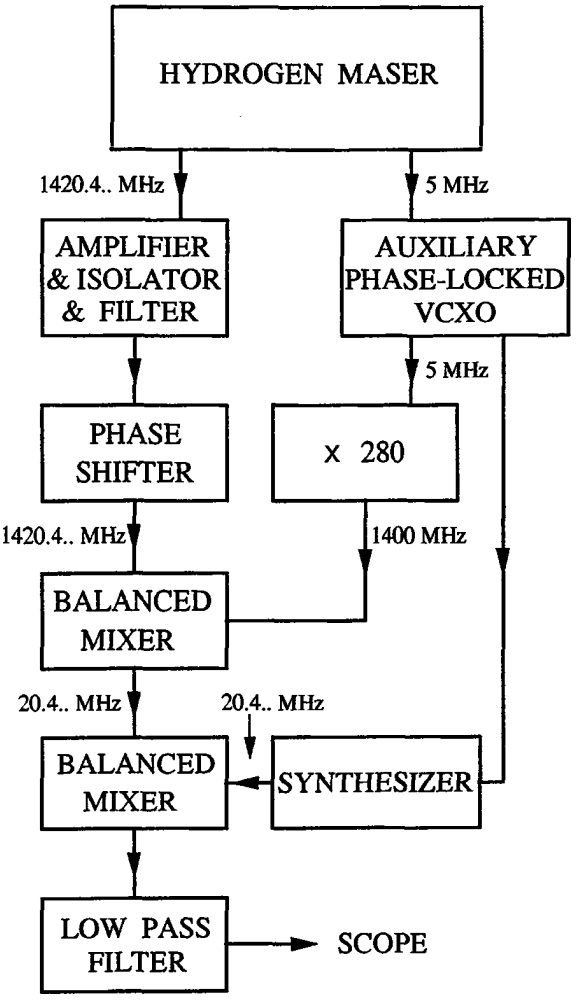


Fig.3. Schematic of the experimental set-up.

Of most interest is figure 4c, which shows an example of recorded periodic phase changes. A phase jump occurs at the times of the cavity frequency switching and the phase varies linearly with time in the intervals between the cavity frequency changes. The measured magnitude of the phase jumps is 0.66 radian and the measured slope of the linear phase variations is 0.92 radian per second. The estimated relative accuracy of these values is 5 % for the jumps and 10 % for the slopes. The corresponding theoretical values, known with an accuracy of a few %, are 0.63 rad and 0.94 rad s^{-1}, respectively, in good agreement with the measured values.

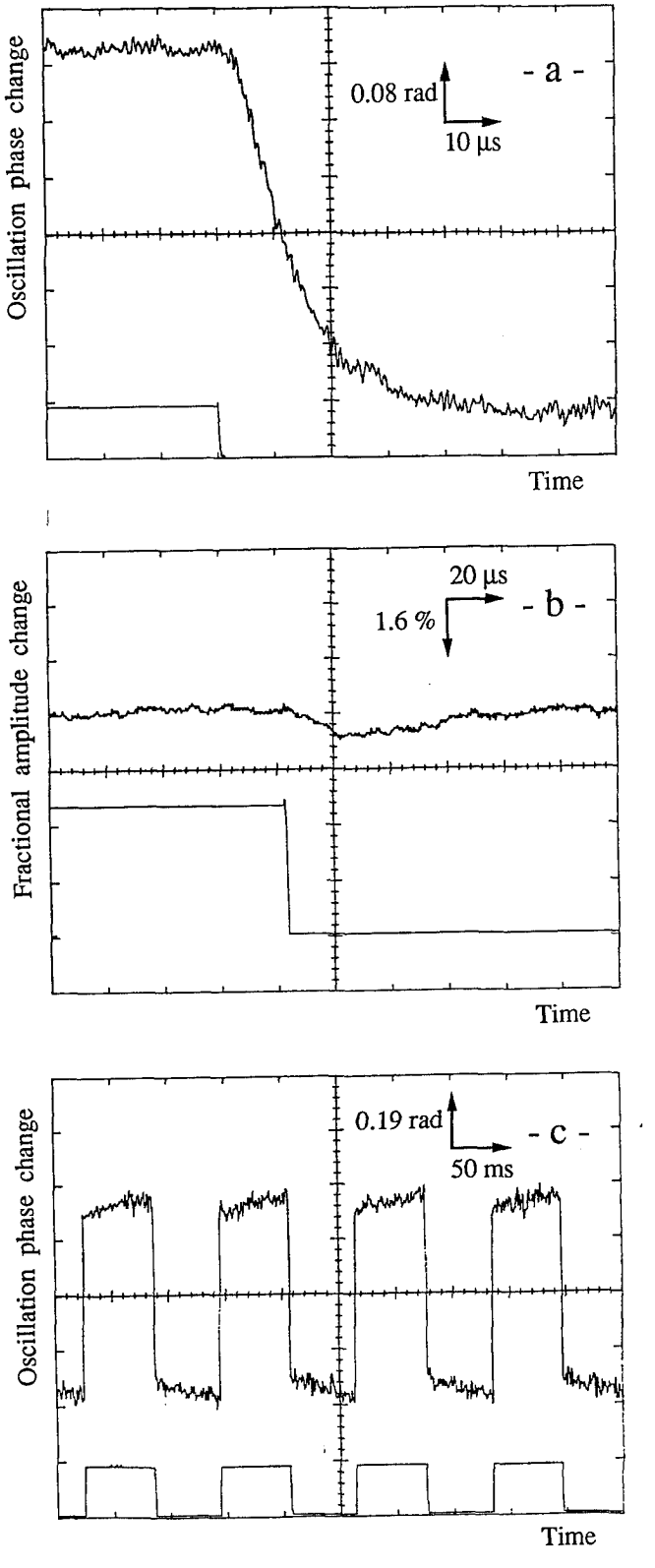


Fig. 4. Examples of experimental records. The observation bandwidth is 1 MHz. The lower trace shows the control voltage applied to the varactor.
a. Oscillation phase transient.
b. Oscillation amplitude transient.
c. Periodic change of the oscillation phase

Conclusion

The phase and amplitude changes occuring in the cavity frequency switching method of the hydrogen maser cavity control have been analyzed and measured. Experimental results support the theoretical ones. The latter can thus be used for the design of the electronic system associated with the hydrogen maser, with the aim of minimizing the transfer of the phase modulation of the maser to the slaved VCXO [7].

Acknowledgements

One of the authors (C.A.) is pleased to acknowledge Dr. Lute Maleki and JPL for their hospitality and for a partial support of his stay at JPL.

References

[1] D. Kleppner, H.M. Goldenberg and N.F. Ramsey. "Theory of the Hydrogen Maser". Physical Review, vol. 126, pp. 603-615, April 1962.

[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, May 1965.

[3] J. Vanier and C. Audoin. *"The Quantum Physics of Atomic Frequency Standards"*. Bristol and Philadelphia : Adam Hilger, 1989, ch. 6, pp. 949-1256.

[4] H.E. Peters. "Design and Performance of New Hydrogen Masers Using Cavity Frequency Switching Servos" in Proc. of the 38th Annual Symposium on Frequency Control, pp. 420-427, 1984.

[5] H.E. Peters, H.B. Owings and P.A. Koppang. "Hydrogen Masers with Cavity Frequency Switching Servos" in Proc. of the 22nd Precise Time and Time Interval (PTTI). Applications and Planning Meeting, pp. 283-292, 1990.

[6] T.K. Tucker. "Operating and Environmental Characteristics of Sigma Tau Hydrogen Masers Used in the Very Long Baseline Array (VLBA)". TDA Progress Report 42-97, pp. 72-80, January-March 1989.

[7] H.B. Owings, P.A. Koppang, C.C. Mac Millan and H.E. Peters. "Experimental Frequency and Phase Stability of the Hydrogen Maser Standard Output as Effected by Cavity Auto Tuning" in Proc. of the 46th Annual Symposium on Frequency Control, this issue, 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

EXPERIMENTAL FREQUENCY AND PHASE STABILITY OF THE HYDROGEN MASER STANDARD OUTPUT AS AFFECTED BY CAVITY AUTO-TUNING

H. Bryan Owings, Paul A. Koppang, Collin C. MacMillan and Harry E. Peters

Sigma Tau Standards Corporation
1711 Holt Road, PO. Box 1877
Tuscaloosa, Alabama 35403-USA

Abstract

This paper describes the Sigma Tau Standards Corporation (STSC) atomic hydrogen maser cavity auto-tuner system and the effect of auto-tuning on the time and frequency domain stability. Measurements of very long term stability and drift of STSC masers were made at several international standards laboratories. Time and frequency domain measurements were made at STSC.

Introduction

It has been shown that phase, frequency and amplitude variations of the maser signal, coupled directly from the resonant cavity at 1,420,405,751.xxxx Hz, result from the cavity resonant frequency-switching method of auto-tuning [1,2,3]. However, compensation for the resulting phase shifts in the maser signal is made by inserting precise, synchronous phase shifts into the maser receiver-synthesizer signal. After mixing the IF with the synthesizer signal in a phase detector, there remains only a small residual phase perturbation. The receiver system's phase-locked loop VCO filter further reduces the residual phase, frequency and amplitude perturbations.

In the time domain, the added short term instability due to frequency-switching is less than two parts in 10^{14} for measurement intervals between one and 20 seconds. For longer measuring times there is negligible perturbation of the system output and the auto-tuner effectively removes cavity related systemic instabilities with typical uncertainties of one to two parts in 10^{15}. In the frequency domain, the principal observable effect on the system output signals is a small coherent peak (order -100 dBc) in the spectral density of phase at the 28.7xx Hz cavity modulation frequency.

Each maser produces a stable frequency output (in both long and short term) that is unavailable from any other standard or standard ensemble in real time. Auto-tuning Sigma Tau masers stabilize the cavity resonant frequency with respect to the hydrogen emission line, and thus are not significantly subject to the major limitation of other hydrogen masers for use in long term standards applications.

There are 31 STSC masers operating world-wide today. These masers are currently installed at various international time and frequency standards laboratories and at several Very Long Baseline Interferometry radio astronomy installations. In addition to their extremely good long term stability, these masers demonstrate the exceptional short term stability inherent in the classical active oscillator hydrogen maser design.

Reference [1], a paper presented at the 38th Annual Symposium on Frequency Control (1984), describes the general cavity frequency-switching servo design principles and gives the results of measurements on the first two prototypes using the servo. Reference [2] gives performance data on several more recent masers against the BIPM and illustrates the servo transient response as well as the method used to compensate for spin-exchange pulling.

The Cavity Auto-Tuner & Phase Compensation System

Figure 1 is a block diagram illustrating the overall hydrogen maser physical and electronic design. Areas of particular interest to this paper are the cavity frequency control servo and the phase compensation circuitry of the receiver system. The essential features of each are illustrated in Figures 2a and 2b, respectively.

0-7803-0476-4/92 $3.00 © 1992 IEEE

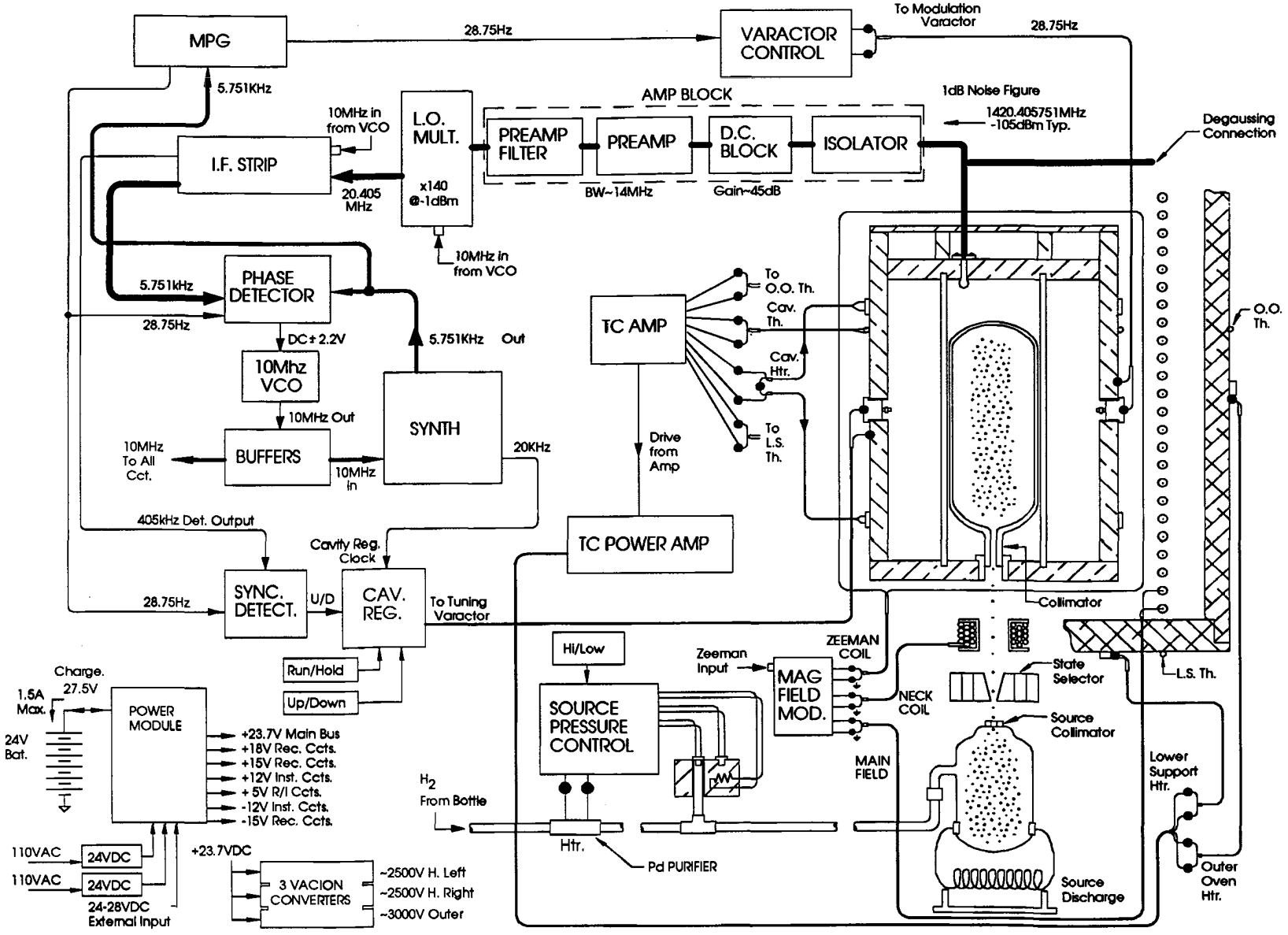


Figure 1. STSC Hydrogen Maser Block Diagram.

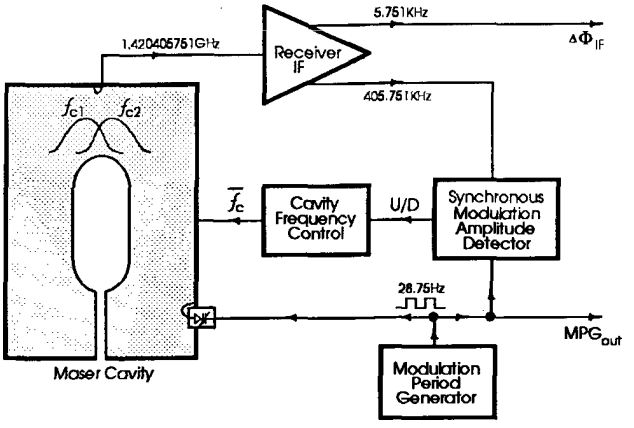


Figure 2a. Cavity Frequency Control Servo.

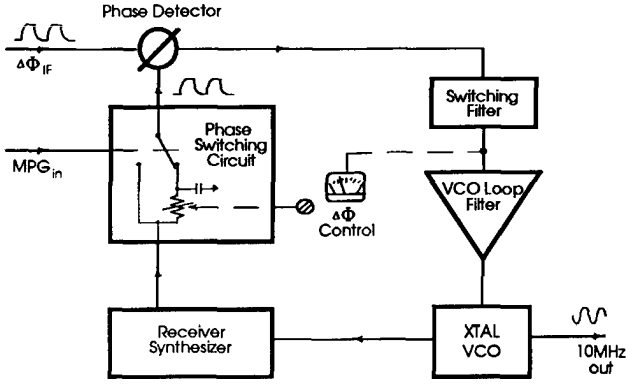


Figure 2b. Phase Modulation Elimination Circuitry.

In Figure 2b, phase modulation is present on the maser IF signal at 5.751xxx KHz. This is not a random or noise-based modulation, but is exactly periodic and constant in amplitude, with the frequency of the cavity modulation produced by the Modulation Period Generator (MPG, typically 28.7xx Hz). A phase shifting circuit in series with the receiver-synthesizer signal introduces adjustable phase shifts with the periodicity of the maser IF signal modulation.

The IF and phase shifted synthesizer signals enter the VCO phase detector (where phase modulations on the two signals cancel) and then pass through a switching filter circuit to the VCO loop filter. The output of the loop filter goes to the Electronic Frequency Control (EFC) of a low noise crystal oscillator. This phase-locks the crystal to the maser signal and produces the stable standard output frequencies of the maser.

Auto-Tuner Transient Response

If a maser has been turned off and back on - or if some other large disturbance to the cavity frequency occurs - there will be a transient period with a time constant of about one hour during which the cavity, and hence the maser output, settles to the original frequency. The maser output frequency and phase transient response to an induced offset is shown in Figure 3. Figure 4 shows the typical performance after such transients have settled down.

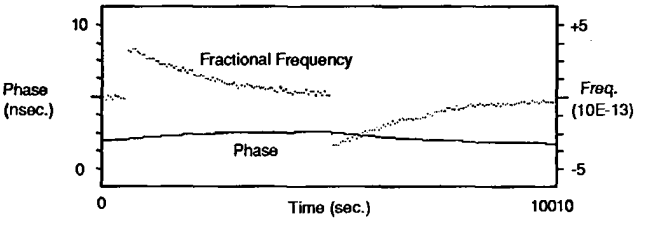


Figure 3. STSC Cavity Switching Servo Transient Response.

Spectral Purity Data

Spectral purity data between masers was obtained using a Femtosecond Systems FSS-600 phase noise detector and a HP-3561A Dynamic Signal Analyzer. The signal analyzer computed the spectral density of phase fluctuations and the output was plotted as $\mathcal{L}(f)$, the single side-band mean square phase fluctuations in dBc, versus Fourier frequency.

Most of the spectral density measurements were made at 10 MHz, the frequency of the crystal VCO supplying the other frequency outputs of the maser. Spectral measurements on the other standard outputs of 5 MHz and 100 MHz, when adjusted for the differences in frequency, gave very similar results. Data was obtained from several pairs of masers using different VCO bandwidths and with the cavity servo modulations on and off.

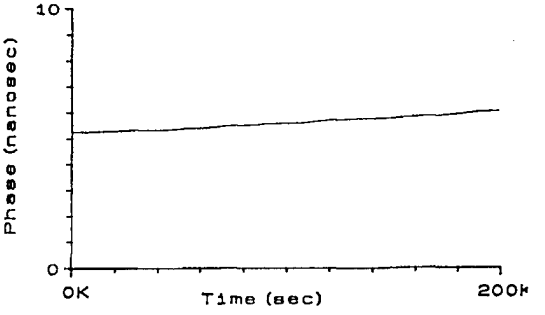
Figure 5 is a plot of the data obtained between masers SN 105-18 and 105-23 with cavity modulation on -23 "ON" and "OFF" on -18. The -111 dBc spur at 60 Hz is typical of all spectral data gathered using the present instrumentation and is purely system noise and not maser related. The -108 dBc peak at 28.75 Hz is due to the cavity modulation.

Figure 6 is a composite plot showing the results using different phase-locked loop bandwidths, again with modulation on maser -23 "ON" and "OFF" on -18. This

TAU = 200000 105-18 vs 105-23 04/21/92

Averaging Time (sec)	Stability
1	1.114e-013
2	5.968e-014
5	2.828e-014
10	1.704e-014
20	1.07e-014
50	6.179e-015
100	4.211e-015
200	2.849e-015
500	1.944e-015
1,000	1.424e-015
2,000	1.442e-015
5,000	1.479e-015
10,000	1.568e-015
20,000	9.499e-016
50,000	7.558e-016
100,000	5.54e-016

Initial Phase = 5.262
Phase = 6.036
Ave Freq Offset = 3.977e-015
Change Synth. by 85.



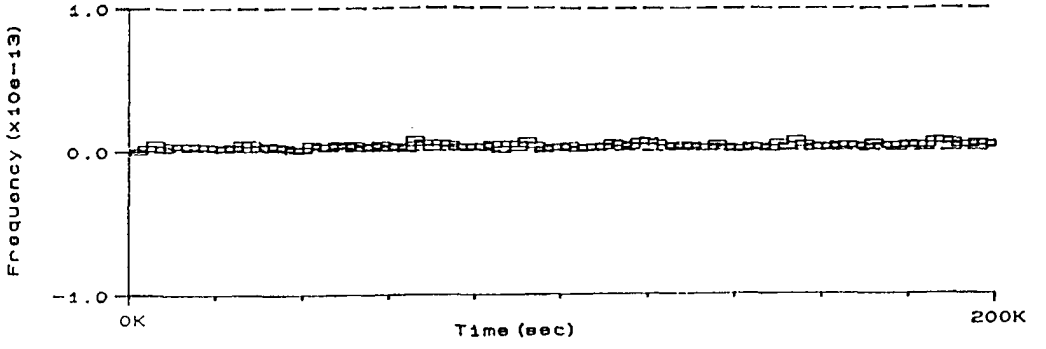


Figure 4. Stability data for two STSC Hydrogen Masers: Allan Variance, Frequency Offset and Time Differences, 200,000 seconds measurement

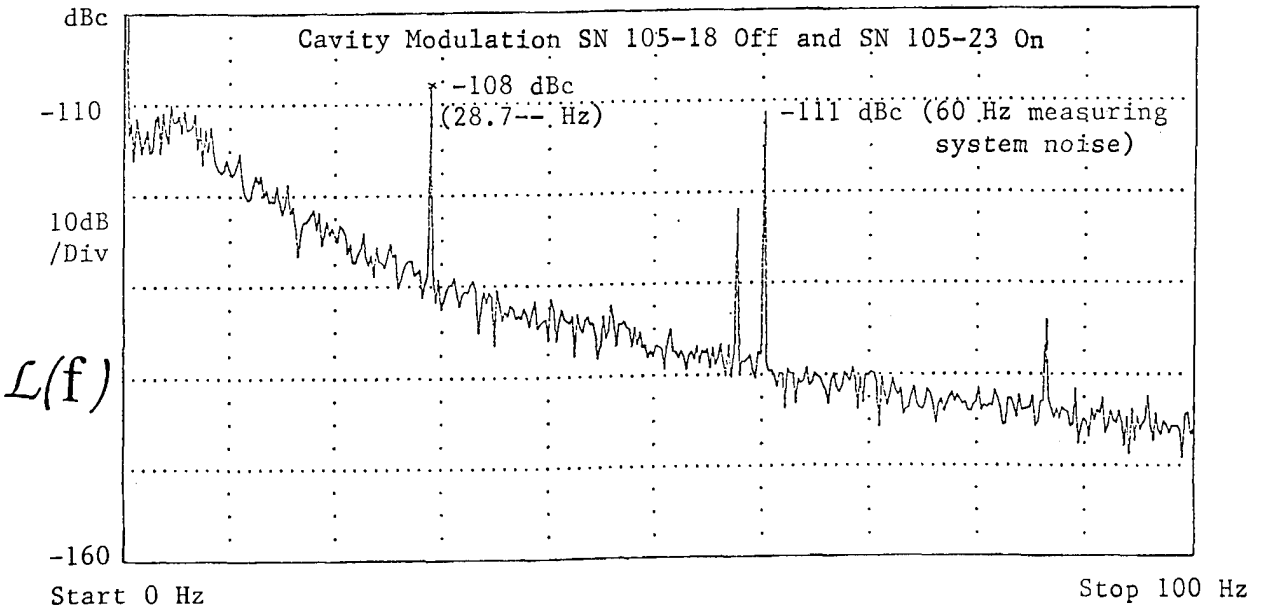


Figure 5. STSC Hydrogen Masers SN 105-18 vs 105-23 Spectral Density Plot

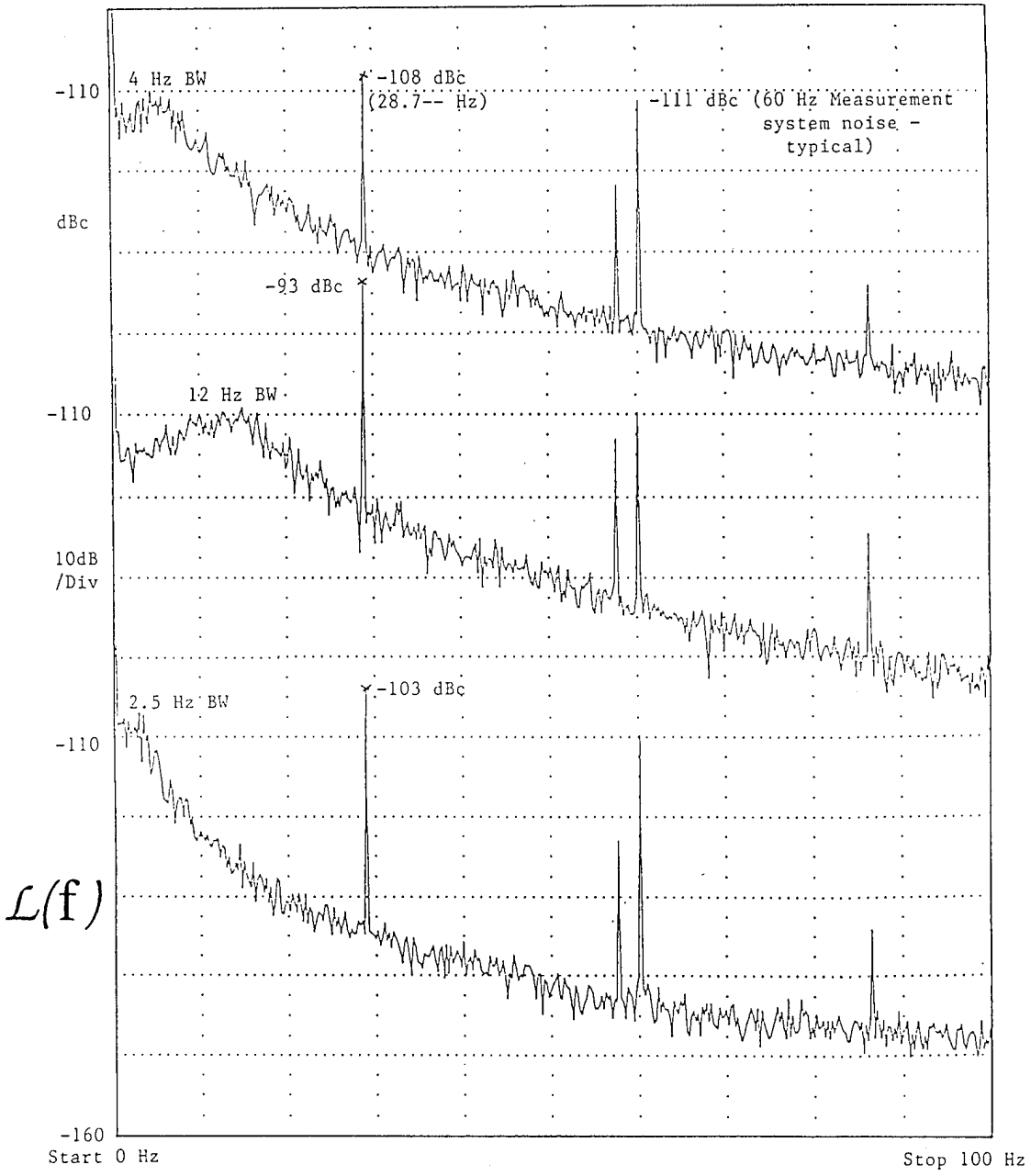


Figure 6. STSC Hydrogen Maser SN 105-18 vs 105-23 Spectral Densities as a Fucntion of Crystal VCO Loop Bandwidth. Modulation ON -23, Off -18.

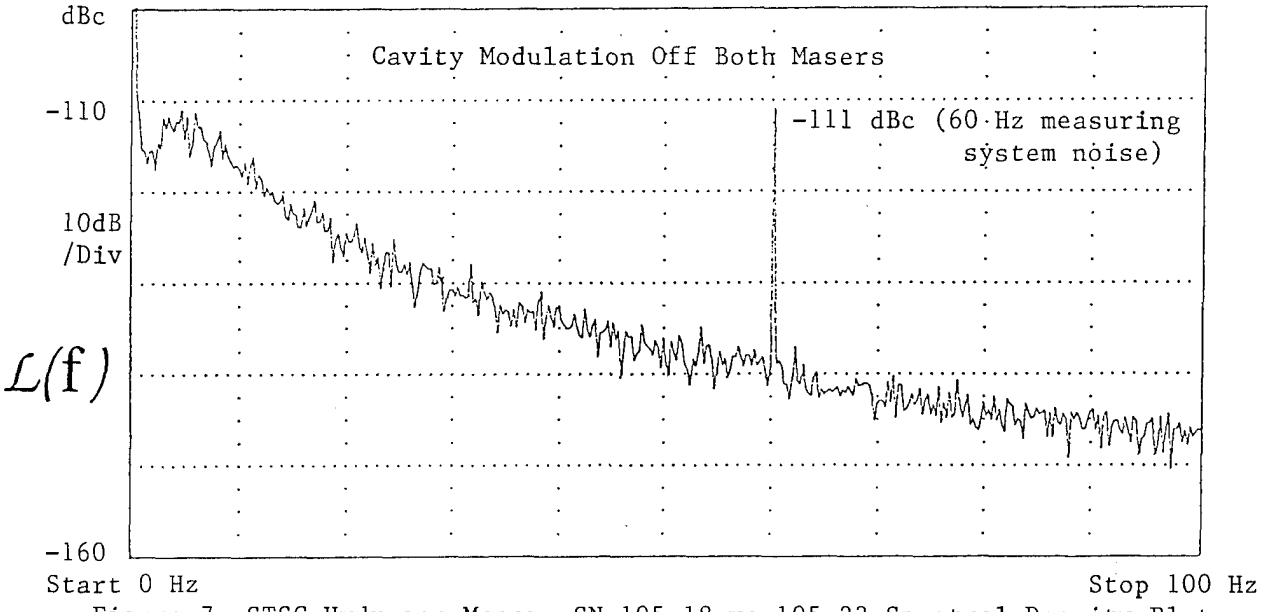


Figure 7. STSC Hydrogen Masers SN 105-18 vs 105-23 Spectral Density Plot

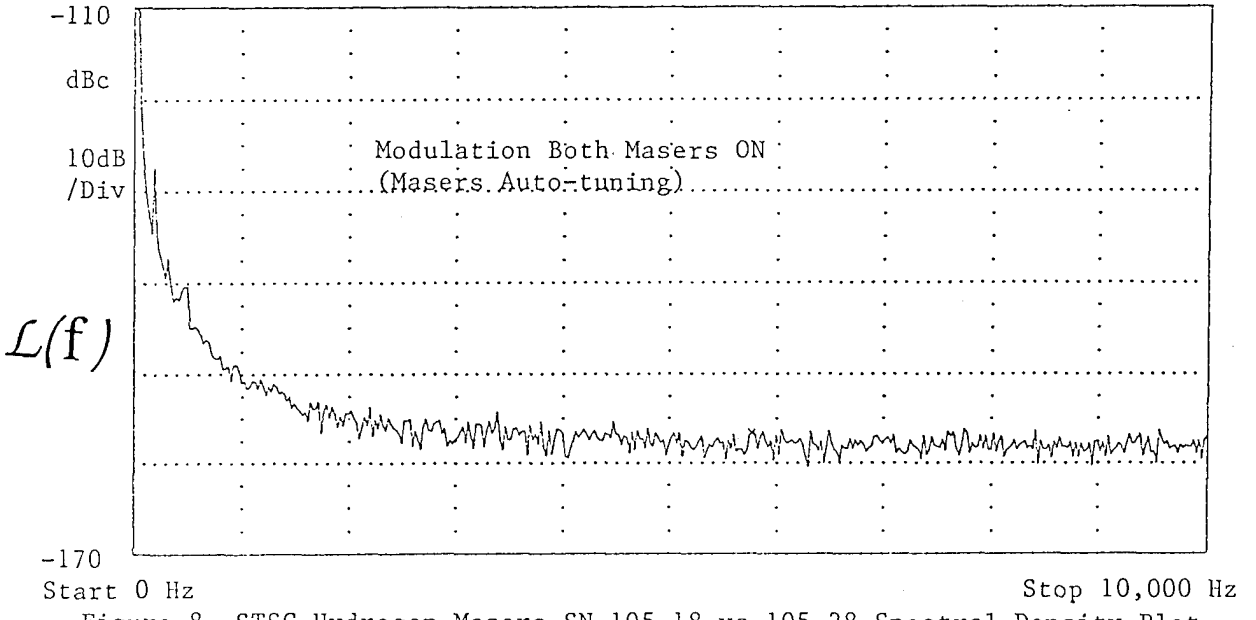


Figure 8. STSC Hydrogen Masers SN 105-18 vs 105-28 Spectral Density Plot

figure shows that the peak in the phase noise plot is a minimum at a loop band-width between 3 and 5 Hz and the peak at the cavity modulation frequency is also near its lowest point at this bandwidth. In this range it will later be seen that the time domain stability is also near optimum.

Figure 7 is similar to Figure 5, but with modulation "OFF" on both masers. Comparison of Figures 5 and 7 illustrate that the phase noise of the maser signals over the range of Fourier frequencies is not significantly affected by the cavity modulation, except for the spur at the 28.7-- Hz modulation frequency. Figures 8 shows the spectral data at Fourier frequencies up to 10 KHz.

It is clear from the above data, as well as from similar measurements on the other STSC hydrogen masers, that the spectral purity of the output signals for Fourier frequencies above the VCO loop bandwidth frequency is essentially limited by the noise of the crystal VCOs. All VCOs use 10 MHz SC cut crystals and have noise properties near the practical limit for presently available commercial crystal oscillators.

Time Domain Stability vs. VCO Bandwidth

The time domain stability data taken at STSC is all non-drift-removed data. The system uses factor of 10 multiplication of the 10 MHz VCO signals with a synthesized offset of 1 Hz at the resultant 100 MHz comparison frequency. The 1 Hz signal goes to a system that computes and plots the Sigma Tau curves (Allan Variances), the fractional frequency offset and the relative time.

The maser short term time domain stability is not highly sensitive to the operation of the cavity-switching auto-tuner. Table 1 shows the time domain stability for measuring intervals from one second to 20 seconds, with modulations both off and on, and different VCO phase-locked loop bandwidths for the comparison masers. The third column represents the data at the nominal VCO bandwidth: 3.5 - 5 Hz; the range used in the present STSC hydrogen masers. This was chosen in consideration of the poorer time domain stability when 2.5 Hz or lower VCO band-width is used and the poorer spectral purity when using band-widths over about 5 Hz.

TABLE 1. STABILITY PERFORMANCE OF MASER 105-18 & 105-23 FOR VARIOUS VCO BANDWIDTHS WITH AND WITHOUT MODULATION.

Measurement Interval (sec.)	Stability (parts in 10^{14})			
1	9.2	10.0	11.1	16.6
2	5.0	5.5	6.4	7.8
5	2.5	2.6	2.6	3.5
10	1.6	1.5	1.9	2.3
20	1.0	1.0	1.1	1.7
Modulation: SN 105-23	Off	On	On	On
Modulation: SN 105-18	Off	On	On	On
VCO Bandwidth (Hz)	12	12	3.5-5	2.5

Long Term Stability and Drift

The data in Figure 4 shows stability performance for two STSC hydrogen masers operating in a room with normal building heating and cooling that maintains the room temperature within ±1°C. Figures 9 and 10 show stability data for other pairs of STSC masers under similar conditions.

The performance without auto-tuning can be inferred from the corrections that the auto-tuner makes to the cavity frequency. Cavity register data for six USNO masers and one maser located at Joddrel Bank, England provide the basis for the data in Table 2.

TABLE 2. IMPLIED DRIFT WITHOUT AUTO-TUNING COMPUTED FROM CAVITY REGISTER CORRECTION DATA

Maser	Fractional Correction by Tuner Per Year (parts in 10^{12})	Fractional Correction by Tuner Per Day (parts in 10^{14})
NAV-1	1.63	0.46
NAV-2	2.63	0.72
NAV-3	0.47	0.13
NAV-4	4.83	1.32
NAV-5	0.24	0.06
NAV-6	2.24	0.64
AL-1	2.12	0.58

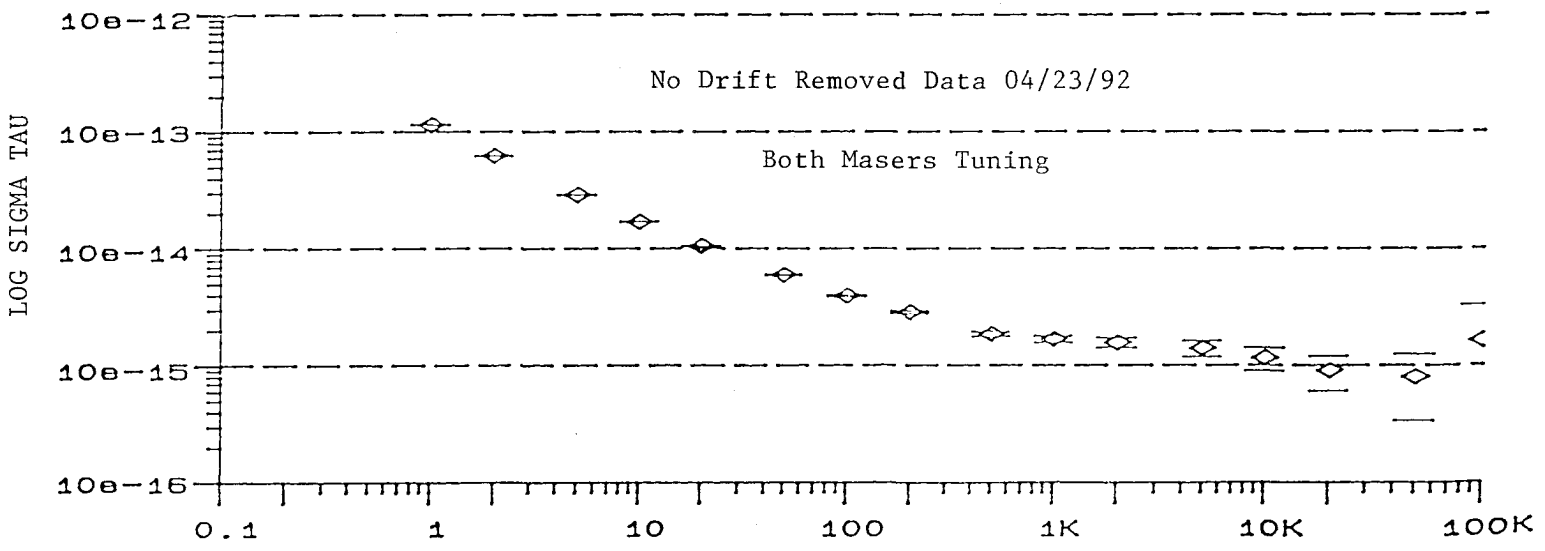


Figure 9. Stability Plot, STSC Hydrogen Masers SN 105-18 vs NAV-6

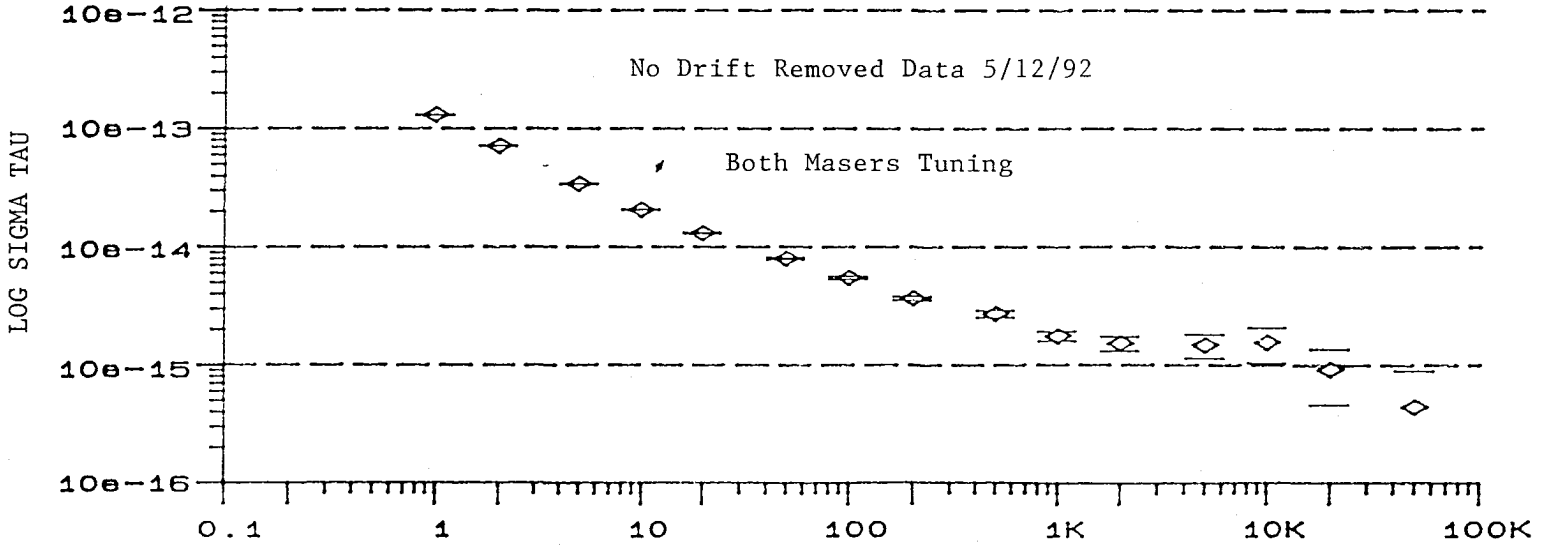


Figure 10. Stability Plot, STSC Hydrogen Masers SN 105-22 vs NAV-6

Because these masers must be delivered to the users soon after construction, it is not feasible to obtain very long term frequency stability or drift data at STSC. Fortunately, the following organizations using STSC masers have provided us with long term operational data:

NRL Masers N-1 and N-2

The STSC hydrogen masers designated N-1 and N-2 by NRL have been operating almost continuously since delivery in September, 1985. The performance of these masers was reported at the 46th Symposium on Frequency Control [4]. N-1 is reported to be approximately 2.5 x 10^{-14} higher than the BIPM at present, and within the precision and accuracy of the defined standard of frequency it has the same frequency now as it had when delivered.

N-2 has not performed quite as well, but has had a nearly continuous linear drift since installation; it is reported to be drifting at a rate of -3.3 x 10^{-15} per day at present. The cause of the drift in N-2 has not been determined since its operation has never been interrupted for analysis of the system.

NASA (APL, JPL) Maser (ST-1)

The STSC maser ST-1 was delivered to APL in January, 1985 and has subsequently operated without major interruption. This was the first operational maser constructed that incorporated the cavity frequency-switching servo. It has been used as a frequency standard at APL and as a traveling clock for time coordination [5]. Including operation at STSC before delivery, the maser has operated essentially continuously for over eight years at this writing.

JPL acquired the maser in April, 1989, and it has since been used for metrology and research such as that reported in reference [3]. No continuous record of frequency has been maintained, however, JPL measurements during two weeks in May, 1992 indicate that the frequency is within 1.2 x 10^{-13} of the frequency measured two years ago against NIST(AT1) [6]. On this basis, the frequency drift during this period is within ±1.6 x 10^{-16}/day relative to NIST(AT1).

NIST STSC -1

One STSC hydrogen maser has been operating at the US National Institute of Standards and Technology for over a year at this writing. Present data indicates that the drift relative to NIST(AT1) over a 100 day period beginning December 21, 1991 is 3.0 x 10^{-16}/Day. The stability plot for this period is given in Figure 11.

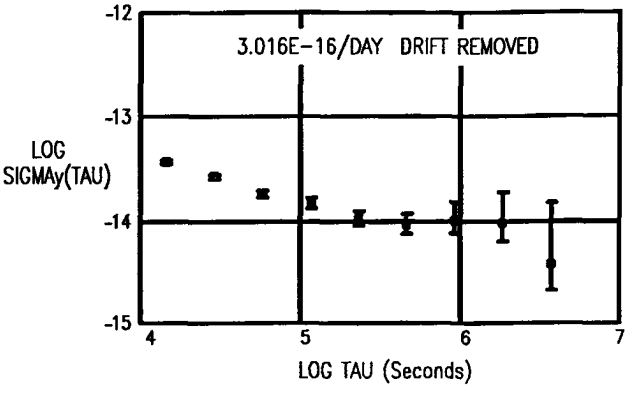


Figure 11. Stability: STSC/NIST-1 vs. NIST(AT1).

The data presented in this paper suggests that the auto-tuned atomic hydrogen maser is well below the noise level of the NIST(AT1) cesium ensemble reference and that the removed drift is also inconsequential in view of the statistics of the measurement.

Observatory of Paris OP-1

One STSC hydrogen maser was installed at the Observatory of Paris in November, 1991, and the frequency has since been monitored continuously against

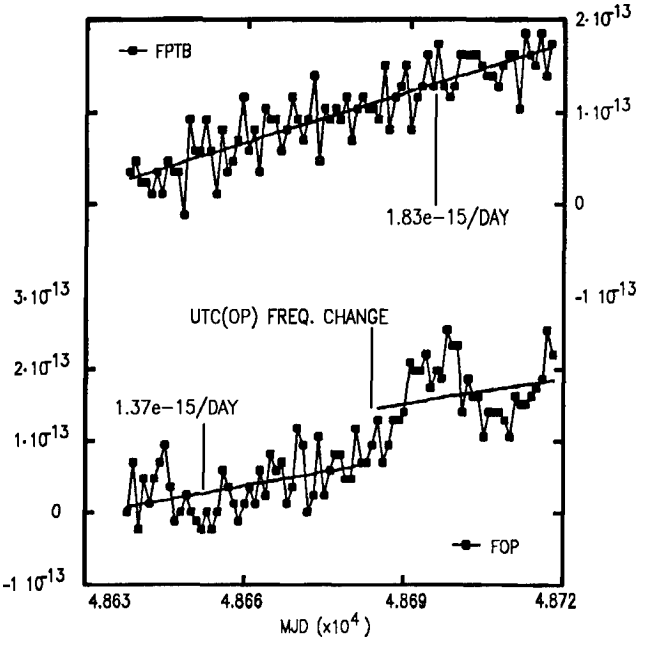


Figure 12. Frequency STSC/PARIS-1 vs. UTP(OP) and UTC(PTB) via GPS.

both UTC(OP) and CS1(PTB). Figure 12 shows the frequency data from January 17, 1992 to April 8, 1992, a period of 81 days.

In the CS1(PTB) plot, there is an apparent average upward drift of 1.8 x 10^{-15} per day (indicated in the figure). Analysis of the data for the last 10 days of this period indicates that the slope has changed to approximately 1.2 x 10^{-15} per day, however the statistical significance of this change in rate is small.

National Physical Laboratory NPL-1

One STSC hydrogen maser was installed at the National Physical Laboratory (England) on November 30, 1990 and has been monitored against UTC(NPL) since then. The frequency of NPL(STSC) vs. UTC(NPL) is plotted in Figure 13 for the 441 day period from December 13, 1991 through February 27, 1992.

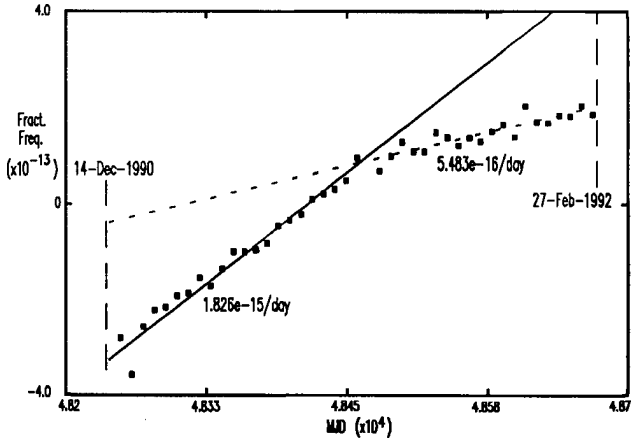


Figure 13. Frequency NPL(STSC) vs. UTC(NPL).

The data for the first seven months of this period indicates that the maser frequency was drifting upward at a rate of 1.8 x 10^{-15} per day, and that the upward drift has slowed to approximately 5.4 x 10^{-16} during the last 220 days. This apparent slight upward drift of the STSC hydrogen masers, and the change in drift rate, is similar to that observed at the Observatory of Paris and at the USNO.

NOAA STSC-1

One STSC hydrogen maser was delivered in March, 1991 to the National Oceanic and Atmospheric Administration at the Mojave radio telescope site at NASA's facility at Goldstone Dry Lake in California. After tests that included placing the maser in a Mobile Radio Telescope trailer and traveling between many observing sites in California, the maser was shipped by sea via the Panama Canal and the north Atlantic ocean to Norway with the maser fully operational during the trip.

The maser was then moved to an observing site in Sweden and at last report the mobile unit with maser was taken by ship to the Azore Islands. A section of timing data relative to GPS for 150 days obtained while in Norway has been provided by NOAA and is shown as a time plot in Figure 14. The offset frequency during this period is calculated as 1.18 x 10^{-13}. This frequency difference is just a synthesizer offset, but the significance of the data is that there is no frequency change within the error of measurement.

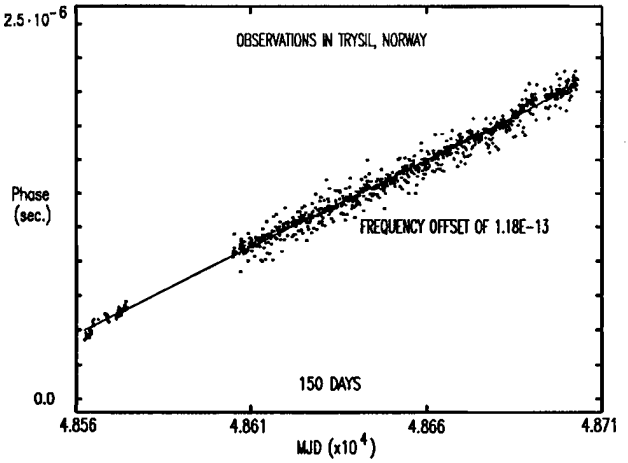


Figure 14. Phase NOAA/STSC-1 vs. GPS.

USNO Masers NAV-1 -- NAV-6

Seven STSC hydrogen masers have been delivered to the USNO. Four of these are presently used in conjunction with the other standards involved with the USNO time and frequency standards operation, the other three are located more remotely at USNO radio astronomy installations involved with geodesy, polar motion and similar applications.

Figure 15 shows the relative frequency of four of these masers during a 90 day period starting in November, 1989 versus the BIPM [2,4]. The frequency starting points on the vertical axis are synthesized and arbitrary. All four of the masers are increasing smoothly in frequency relative to the BIPM by approximately two parts in 10^{15} per day, while the masers are varying amongst themselves by one part in 10^{15} or less. Figure 16

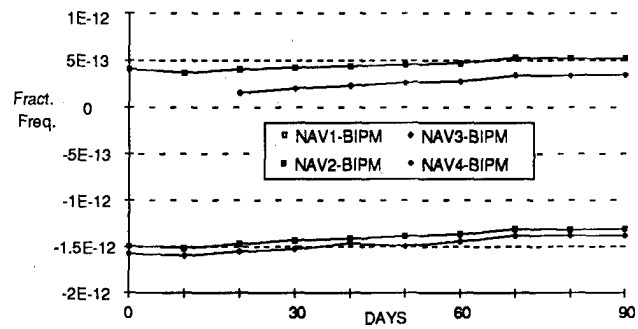


Figure 15. Frequency Stability of Naval Observatory STSC Masers vs. BIPM. (November 29, 1989 to February 27, 1990.)

shows the relative frequency of four USNO-STSC masers during a period of 500 days beginning November, 1990. STSC maser NAV-5 is the reference in Figure 16. The drift among the masers is almost unnoticeable in this figure and calculations indicate that the variation is on the order of one part in 10^{16} per day or less.

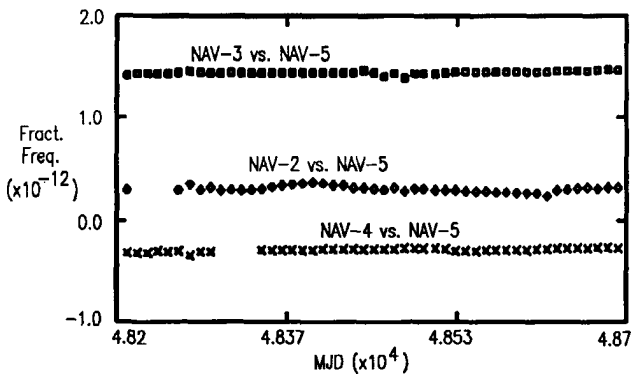


Figure 16. STSC Masers at USNO, Frequency Plots (Synthesizer frequency and phase changes removed).

The drift of the four USNO STSC masers relative to the BIPM over a period of 170 days ending in April, 1992 is given in Table 3. As stated in the USNO message containing this data, which was received by modem on June 1, 1992, measures less than one part in 10^{16} are not significant due to measurement system uncertainties, transmission errors and uncertaintiers in the fundamental standards used for defining absolute frequency.

TABLE 3. STSC HYDROGEN MASER DRIFT VS. BIPM. (Data from USNO message received June 1, 1992)

Maser	Drift-Fractional Frequency
N2	+9.4 x 10^{-17}/day
N3	+0.4 x 10^{-17}/day
N4	-6.4 x 10^{-17}/day
N5	+9.4 x 10^{-17}/day
AVG.	3.2 x 10^{-17}/day (± see text)

Conclusion

The main purpose of this paper was to illustrate the effect of the automatic cavity tuner on the maser stability. It is clear from the data presented that the spectral purity of the maser output signal and the time domain stability are not significantly degraded by cavity frequency-switching; the auto-tuned maser retains the exceptional short and medium term stability of the classical active oscillator maser.

By far the most important result is that cavity pulling, the most serious problem with attainment of very long term stability with a hydrogen maser, is now under precise control, and the exceptional long term stability capability of the maser can be attained with a single, field operable, standard.

ACKNOWLEDGMENTS

The authors gratefully acknowledge past R&D support or provision of operational data on installed STSC atomic hydrogen masers from the following:

- The National Aeronautics and Space Administration
- The Naval Research Laboratory (two masers)*
- The National Institute of Standards and Technology (one maser)*
- The Johns Hopkins University Applied Physics Laboratory
- The United States Air Force
- The United States Naval Observatory (seven masers, four at USNO in Washington, DC)*
- The National Oceanic and Atmospheric Administration (one maser)*
- The (UK) National Physical Laboratory (one maser)*

- The Observatory of Paris (one maser)*
- The National Radio Astronomy Observatory (11 masers)
- The Jet Propulsion Laboratory (one maser)*

* Contributed data for this paper

REFERENCES

[1] H. E. Peters, "Design and Performance of New Hydrogen Masers Using Cavity Frequency Switching Servos," in Proceedings of the 38th Annual Symposium on Frequency control, 1984.

[2] H. E. Peters, H. B. Owings and P. A. Koppang, "Hydrogen Masers with Cavity Frequency Switching Servos," in Proceedings of the 22nd Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1990.

[3] Claude Audoin and William A. Diener, "Frequency, Phase and Amplitude Changes of the Hydrogen Maser Oscillation," (Paper submitted for the present meeting) in Proceedings of the 1992 Frequency Control Symposium (48th Annual Symposium on Frequency Control).

[4] E. Powers, A. Gifford and P. Wheeler, "Hydrogen Maser Performance at the United States Naval Observatory and the Naval Research Laboratory," in Proceedings of the 47th Annual Symposium on Frequency Control, 1991.

[5] L. Rueger, M. C. Chiu, S. D. Deines, J. T. Broomfield and C. O. Alley, "Portable Hydrogen Maser Clock Time Transfer," in Proceedings of the 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1987.

[6] W. A. Diener, JPL, Private communication, May 18, 1992

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DIODE-LASER-PUMPED, GAS-CELL ATOMIC CLOCKS

R. E. Drullinger and Csaba Szekely
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder CO

and

J. C. Camparo
M2-253
Electronics Technology Center
The Aerospace Corporation
P O Box 92957
Los Angeles CA 90009

Abstract

Recent theoretical calculations suggest that the short-term stability of rubidium, atomic, gas-cell frequency standards could be improved by several orders of magnitude if diode lasers were used for optical pumping. Moreover, the theoretical results predict that rubidium, as opposed to cesium, is the best alkali atom to choose for gas cell atomic clock operation. This paper describes strategies for developing such clocks for space applications and as local oscillators for other high-accuracy atomic clocks.

Introduction

The rubidium gas cell atomic clock has been around for 30 years and is usually thought of as a "mature" technology. Recent theoretical analysis, [1] however, indicates that replacing the rubidium lamp with a laser could yield several orders of magnitude improvement in short-term stability, $0.1\ s < \tau < 100\ s$. How has this tremendous potential gone untapped in lamp-pumped standards? The spectroscopic complexity of the lamp-pumped system makes a detailed analysis of the physics almost impossible. However, the following somewhat simplified description gives clues as to the problem.

The conventional lamp-pumped, buffer-gas cell standard operates in a spectroscopic cacophony composed of two isotopes each with two principal resonance lines (D_1 and D_2) which are further split into two hyperfine components each composed of several unresolved additional hyperfine components. These many lines are each characterized by a width, a position, and a degree of self reversal or optical depth that are all functions of the operational conditions. It is the job of the designer to concoct a witch's brew of buffer gases, different in the lamp, (filter) and cell, to broaden and shift these lines in such a way as to improve upon a coincidental, near overlap of a line in one isotope with a line in the other. It is this overlap that allows the combination of lamp,(filter), and cell to produce the optical pumping necessary for the operation of the standard. Not surprisingly, however, this complex spectrum leads to a situation where the degree of optical pumping achieved is small (some of the undesirable light leads to depumping) and the level of background light is high (contributing noise).

The ultimate stability of a rubidium cell standard can be glimpsed through a very simple argument. Consider only how many rubidium atoms are in a cell, how often each of them can contribute to the signal, and what the resonance linewidth might be. At 50°C, the vapor density of rubidium is about $10^{11}\ cm^{-3}$ and the spin exchange rate is the order of 100/s. This could result in $> 10^{12}$ "clock" transitions per second and a linewidth < 100 Hz. We can then model the atomic shot-noise limited, short-term stability as roughly $\sigma_y(\tau) \approx (Q \cdot S/N)^{-1} \approx 10^{-14}/\sqrt{\tau}$. More elaborate calculations [1] optimizing the microwave power, the optical power and taking into account the effects of laser noise predict similar

Contribution of the U. S. Government, not subject to copyright.

0-7803-0476-4/92 $3.00 © 1992 IEEE

performance. Previous experiments, [3] confirmed in the current NIST work, have shown that lasers can indeed produce much higher optical pumping and correspondingly higher microwave/optical double resonance signals while at the same time reduce the amount of light falling on the detector by an order of magnitude or more. These simple experiments provide proof, in principle, of the available potential. The reason they have not produced eye-popping stability is, in part, related to the fact that they have been done with available commercial standards in which the microwave synthesis is not capable of supporting such performance. [4]

Comparing the optimized short-term stability of laser-pumped standards with that actually achieved in lamp-pumped standards may not seem totally fair because conventional lamp-pumped standards have compromised their short-term stability somewhat in order to achieve better long-term stability. However, the laser-pumped standard affords other possibilities (to be discussed) for the control of light shift and, hence, long-term stability. Therefore, it seems justified to discuss their optimized short-term potential.

In Section II we will briefly review the detailed theory, its predictions, and limitations. In section III we will outline a planned experimental project which will build off from existing technology with an eye toward space applications. In Section IV we will outline a project which is directed toward ultimate performance in the 1-200 s region for use as a local oscillator in future "super" clocks.

Theory

The theory as published in Ref. 1 assumes a conventional device in which the lamp is simply replaced by a laser. The laser is assumed to operate continuously while the cell is of the buffer-gas type and is constrained to a minimum cavity volume. A one-dimensional model is developed to account for the laser-induced optical pumping, atomic diffusion to the deactivating walls, collisional relaxation, and the expected noise behavior of the laser.

For a given laser intensity and noise, the model is used to predict clock stability as cell temperature and microwave power are optimized. The locus of such points calculated for ^{87}Rb is plotted in Figure 1 as a function of laser power density. The line labeled "with light shift" assumes the FM noise of a solitary laser diode (linewidth $\simeq$ 50 MHz). The curve labeled "shot noise alone" is presented to suggest the results that may be achievable with a system in which the laser FM noise is reduced; e.g. enhanced cavity Q laser, extended cavity laser or gated light operation.

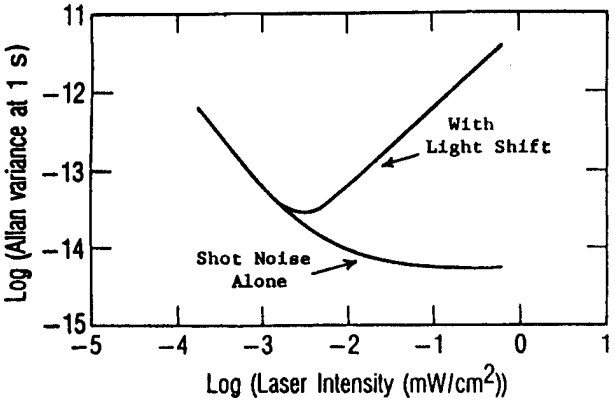


Figure 1. The Allan variance at 1 s for diode-laser pumped rubidium frequency standard with resonance cell temperature and microwave Rabi frequency chosen in such a way that the clock's stability is optimized. The curve labeled "with light shift" assumes the FM noise of a solitary laser diode couples to the clock stability through the light shift term. The curve labeled "shot noise alone" is displayed to represent the clock stability when an advanced, FM controlled laser is used.

With lamp pumping, only rubidium has been developed into a commercial cell standard. This is because the needed filtering, is conveniently available only through the coincidental, near overlap of spectroscopic lines in rubidium. With diode lasers, no such filter is needed, and other atoms can be considered. [2] Figure 2 looks at predicted clock stability achievable with several isotopes of both rubidium and cesium. Rubidium remains the atom of choice for optically pumped cell standards. This results from the fact that its lower nuclear spin more than compensates for the higher transition frequency and hence line Q in cesium.

The model calculations presented here and in Ref. 1 have a number of limitations, but their general predictions should not be altered. In fact, preliminary experiments with lasers verify the potential. Although a full 3-D gas-cell clock signal model now exists, [5] the diode laser calculations employed a 1-D signal model. Consequently, the optical and microwave radial field distributions were not included. This should have only minimal effect on the predicted values. On the other hand, the

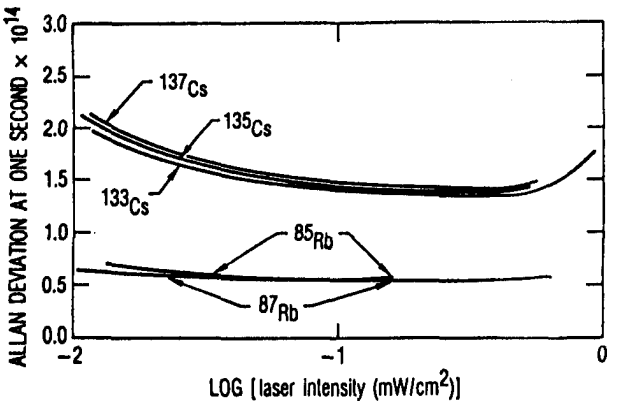


Figure 2. Comparison of predicted, optimized stability (Allan variance at 1 s) for several isotopes of cesium and rubidium as a function of laser power.

calculations were based on a minimum microwave cavity volume, and theoretical predictions [2] suggest that increasing the cavity volume improves short-term performance. Finally, the model assumes cw excitation in a buffer gas cell. Large regions of parameter space remain unexplored. Evacuated, wall-coated cells allow the atoms to average over some of the field inhomogeneities that contribute to long-term instability problems in existing standards. Furthermore, the laser can be more easily gated than a lamp and in an evacuated cell this leads to a fundamental elimination of the light shift effect. Also, the rapidly advancing field of diode laser technology has already produced solitary lasers which have substantially lower noise than those used in these calculations. [6]

Rubidium Clocks for Space Applications

The Aerospace Corporation is studying the space applications of diode-laser-pumped clocks. The project will, therefore, be principally aimed at small, robust devices that can withstand relatively high temperatures. These restrictions limit the regions of parameter space that can be used: *small* puts limits on the cavity and cell size, *robust* requires a certain simplicity of design, and *high temperature* may limit the use of wall coated cells. The program intends to make changes in the rubidium clock physics package incrementally. Phase 1 will simply replace the lamp in a commercial clock and attempt to improve on past results. As pointed out above, previous experiments have shown that it is easy to achieve dramatic improvement in the degree of optical pumping and reduction in the background light, whereas utilizing the much higher signal-to-noise ratio will require new electronics. Following phase 1, a second phase will examine the use of more radically changed devices perhaps employing pulsed pumping and/or evacuated cells with chemisorbed coatings to improve upon the long-term stability of the clock.

Rubidium Clocks as Local Oscillators

Future "super clocks" will probably employ trapped atoms or ions and exhibit extremely narrow atomic resonances with very high signal-to-noise ratios. Such clocks are expected to have atomic limited, short-term performance characterized by $\sigma_y(\tau) < 10^{-14}\tau^{-1/2}$ and use modulation rates from 0.1 to 0.001 Hz. The NIST laser-pumped rubidium cell program is primarily investigating development of a "local oscillator" for such a standard. It is directed toward a laboratory device in which ultimate performance in the $1\ s < \tau < 1000\ s$ range is the goal; size, weight, complexity and power consumption are of secondary importance.

In the first phase, extended cavity diode lasers will be used to control the FM noise of the laser and reduce the light shift noise in the rubidium cell standard. A new microwave synthesis scheme has been developed to provide the ultra-low phase noise microwave signal necessary to support this standard. [7] The enhanced levels of optical pumping already observed at low light levels together with the new synthesizer should produce dramatic stability results. Advancing several orders of magnitude in stability will probably reveal problems not encountered before. Therefore, future directions (pulsed optical pumping, wall coated cells, etc.) will await the outcome of the initial tests.

Summary

Theoretical models indicate that laser optical pumping in rubidium gas cell atomic clocks afford the possibility for several orders of magnitude improvement in short-term performance over conventional lamp-pumped devices. We have begun projects to investigate this potential. Preliminary experiments indicate that the short-term stability is achievable. Additionally, the spatial and temporal coherence of the laser affords possibilities for control of the long-term instability caused by light shift; that is, new geometries and cell types are possible because of the lasers spatial coherence while the temporal coherence and tunability allow one to operate at carefully chosen conditions to minimize light shift. All of this may allow total performance

from a rubidium cell standard that rivals active hydrogen masers.

References

[1 J. C. Camparo and R. P. Frueholz, "Fundamental stability limits for the diode-laser-pumped rubidium atomic frequency standard," J. Appl. Phys., vol. 59, pp. 3313-3317, 1986.

[2] James C. Camparo and Robert P. Frueholz, "A Comparison of various alkali gas-cell atomic-frequency standards," IEEE Trans. Ultrason. Ferroelectr. Freq. Control, vol. UFFC-34, pp. 607-611, 1987.

[3] L. L. Lewis and M. Feldman, "Optical pumping by lasers in atomic frequency standards," in Proc. 35th Ann. Freq. Control Symposium, Ft. Monmouth, NJ, May 1981, pp. 612-624.

[4] C. Audoin, V. Candelier, and N. Dimarcq, "A limit to the frequency stability of passive frequency standards due to an intermodulation effect," IEEE Trans. Instrum. Meas., vol. 40, pp. 121-125, 1991.

[5] J. C. Camparo and R. P. Frueholz, "A three-dimensional model of the gas cell atomic frequency standard," IEEE Trans Ultrason. Ferroelectr. Freq. Control, vol. 36, pp. 185-189, 1989.

6) Motoichi Ohtsu, Highly Coherent Semiconductor Lasers, Artech House, Boston, 1992, and references cited therein.

7) F. L. Walls, John Lowe, and R. E. Drullinger, "Ultra-high stability synthesizer for diode laser pumped rubidium," to appear in Proc. 1992 IEEE Freq. Control Symposium.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

RUBIDIUM FREQUENCY STANDARD WITH A HIGH RESOLUTION DIGITAL SYNTHESIZER

A. Stern, B. Levy & M. Bootnik
Time & Frequency Ltd.
P.O BOX 1792, Holon 58117, ISRAEL

E. Detoma, G. Pedrotto
Sepa S.p.A.
Corso G. Cesare 294-300 - 10154 Torino, ITALY

Abstract: The development and updated results of a newly developed Rubidium Frequency Standard with a High Resolution Digital Synthesizer are described.
A fractional frequency synthesizer developed by SEPA is being implemented into the Frequency-Lock-Loop (FLL) of TFL's Rubidium Frequency Standard. This allows a frequency settability in steps of $1.5x10^{-12}$ over a wide range of $6x10^{-9}$, without changing the C-field.
A square-wave frequency modulation/demodulation is applied to the FLL. This is combined with a C-field modulation which is being alternating coherently with the frequency modulation. The "ac" C-field method is being used to reduce the magnetic field sensitivity.
Both, square-wave frequency modulation as well as C-field modulation give rise to transient signals which are related to adiabetics and non adiabetics phenomena that occur in the Rb spin system. These transients and the means to eliminate them are also being described.

Inroduction

This paper presents updated results of a joint venture between TFL and SEPA, where a high resolution synthesizer developed by SEPA is incorporated into TFL's rubidium frequency standard.
The unit developed allows for digital frequency control with a resolution of $1.5x10^{-12}$ over a range of $6x10^{-9}$.
Actually the range can be enlarged and is limited only by the VCXO control range - parts per $1x10^{7}$.
We also plan to have a finer resolution of $<1x10^{-13}$, to match the expected frequency stability of the Rubidium.

A few years ago the first author suggested the method of the ac "C" field, for reducing the sensitivity to external magnetic field. [1,2]. This method has also been incorporated into the device which was developed.

A preliminary report of the development was given in a previous paper [3]. The present paper also describes the device, gives updated results, and put different emphasises on several aspects. These include a discussion of the transient signals, that arises from the square FM and the ac C-field, as well as a short analysis of the relation between frequency control, time control and the noise level.

Frequency Control in Rubidium (and Cesium) Frequency Standards

Fig. 1 shows a conventional loop of a Rubidium frequency standard. The 10MHz from the VCXO is multiplied to 90MHz and is then combined with a 5.3125MHz in a Step-Recovery-Diode to produce the "Clock-Transiation" frequency nominal at 6.834687500MHz. The Rubidium Physics Package responds with an error signal which is being used to control the VCXO.

The conventional method to control the frequency in a Rubidium (and Cesium) Frequency Standard is via the magnetic "C-field".
The clock transition has a quadratic dependence on an external magnetic field. Thus, one controls the frequency by controlling the current in a coil that induces a magnetic field. The advantage of this method is its simplicity and ease of implementation. However due to the quadratic C-field-to-frequency-dependence this method is non-linear over a wide range and the voltage-to-frequency coeficiant is non reproducible from unit to unit.
In addition interfering with the magnetic field absets the magnetic equilibrium and can produce hysteresis effects that will deteriorate the frequency stability.

However, there exists another scheme to control the frequency in a Rubidium and Cesium Frequency Standards. This is achieved by controlling the output of the SYNTHESIZER located within the Frequency-Lock-Loop.
The synthesizer (in our Rubidium) produces a 5.3125MHz output which is combined with the 90MHz-76th harmonics to produce the final clock transition frequency:

$$76 \times 90\text{MHz} - 5.3125\text{MHz} \approx 6834.6875\text{MHz}$$

A change of 1Hz in the 5.3125MHz results in a change of 1Hz x 10MHz / 6834MHz in the 10 MHz. Thus the resolution is inreased by about 683.

This scheme, which has been introduced lately in Cesium and Rubidium standards has several advantages over the "C" field method. Frequency control is performed digitally with a very high resolution and is linear over a wide frequency range. Furthermore, it is reproducible from unit to unit and the frequency step is fixed and does not vary with time. There is no interference with the physics through the magnetic field it does not affect the stability.

0-7803-0476-4/92 $3.00 © 1992 IEEE

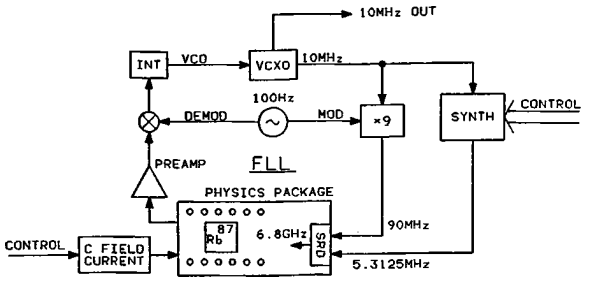


Fig.1: Frequency Control in Rubidium Frequency Standard.

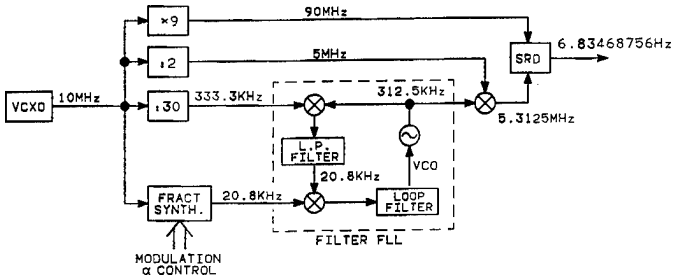


Fig.2: The Synthesizer Architecture.

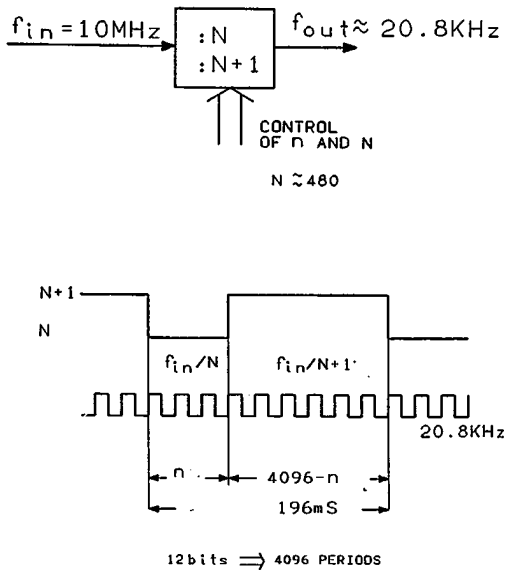


Fig.3: Principles of operation of the fractional synthesizer.

The Synthesizer Architecture

The Architecture of the synthesizer that was dveloped is described in fig. 2.
The 5.3125MHz is produced as follows:

10MHz/2 + 10MHz/30 - 20.83KHz = 5.3125MHz

Frequency control is achieved by controlling the output of the fractional synthesizer which provides approximately 20.8KHz.

The principles of operation of the fractional synthesizer are described in fig.3. It divides the 10MHz input by N during n periods of the output and by N+1 during 4096-n periods of the output. Then the output frequency is a weighted average:

$$\bar{f}_{out} = \frac{n}{4096} \cdot \frac{f_{in}}{N} + \frac{4096-n}{4096} \cdot \frac{f_{in}}{N+1} \qquad (1)$$

Where nominal N=480. Fine frequency tuning is achieved by changing "small n".

The resolution is obtained by calculating the difference:

$$\frac{1}{f_0}\left[\bar{f}_{out}(n+1) - \bar{f}_{out}(n)\right] = \frac{1}{f_0}\frac{f_{in}}{4096 \cdot N^2} \qquad (2)$$

Here f_{in} is the input 10MHz, f_0 is the clock transition 6.834GHz and small-n counter.
This resolution is obtained over an integration time of 196ms which is about the duration of 4096 periods of the 20.8KHz.
The range of frequency variation that results from changing the small-n is:

$$\frac{f_{in}}{f_o} \cdot \frac{1}{N^2} \approx 6.3\text{x}10^{-9} \qquad (3)$$

However this range can be further increased by allowing for the change of large N. It is, in principle, limited only by the voltage control range of the OCXO- parts per $1\text{x}10^{7}$!

Square-Wave Frequency Modulation

The synthesizer allows a straight-forward implementation of a square-wave frequency modulation, instead of a sinewave modulation, which is commonly used in the Frequency-Lock-Loop, to detect and lock the VCXO. An analisys of the two methods is given in [4]. The method of square wave modulation is depicted in fig.4.
A square wave modulation should in principle improve the signal to noise ratio[4]. However a square-wave from the rapid passage through the resonance region [5].

Fig. 5 exhibits transients that were measured at the output of the photo-diode, in a lock, and in an unlock condition. The modulation frequency is 80Hz. The width of the transients is about 0.7ms which is related to the Rubidium spins relaxation time [5]. In order to filter the transient signals we use a Sample and Hold device which samples the signal just before the time the transient occurs and hold it at a steady level during the transient time.
The result of operation of the Sample and Hold device is shown in fig.

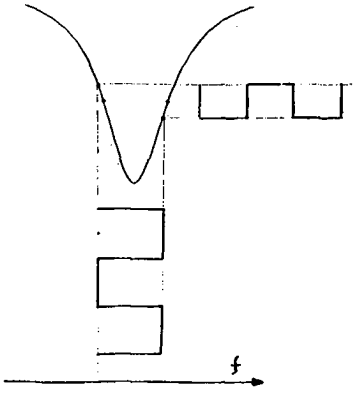


Fig.4: Square-wave FM in the Frequency Lock Loop.

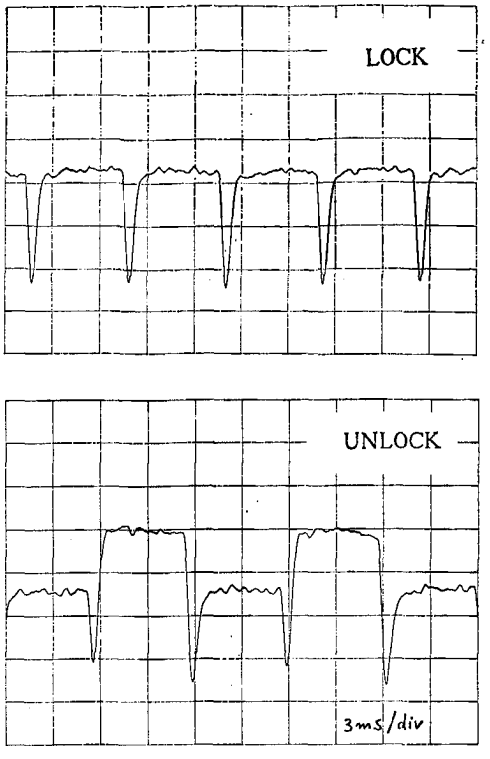


Fig.5: Transients at the photo-diode output that result from a square wave FM.

Results of Frequency Settability

Fig.7 presents examples of frequency settability. The upper graph shows a frequency step of about 1.6×10^{-9} obtained by changing the 10th bit that controls "small-n". The middle graph shows a frequency step of 4xE-10 where the 8th bit is changed. The bottom graph depicts a finer step of 5×10^{-11} that results from the change of the 5th bit.

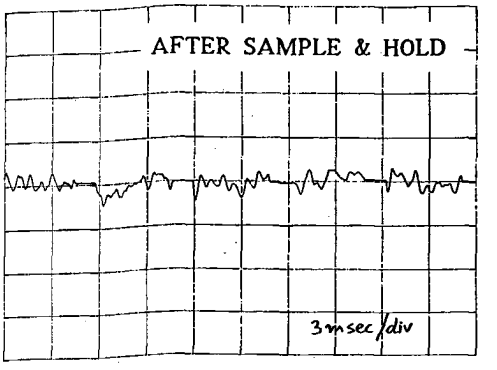


Fig.6: Operation of the Sample and Hold to filter the transients.

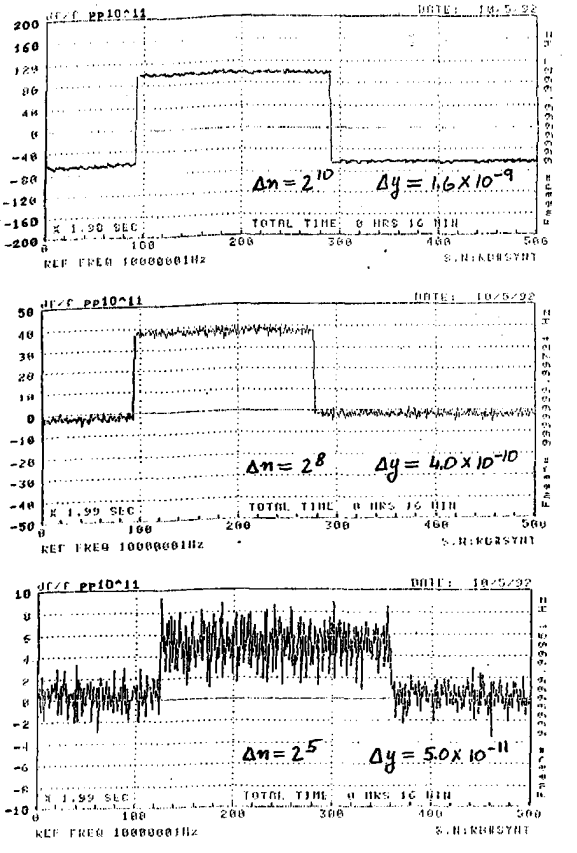


Fig.7: Examples of frequency settability.

Frequency Settability and Noise

An even finer step of 1.2×10^{-11} is exhibited by fig. 8, the upper graph. This is obtained by changing the 3rd bit. Here, however, the frequency step is almost overshadowed by the noise.

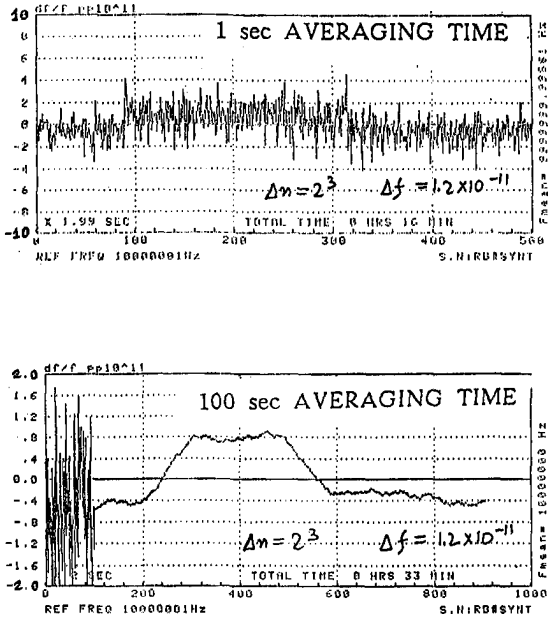


Fig.8: Frequency settability and noise.

The peak to peak noise in the upper graph of fig. 8 is about $3x10^{-11}$ and is related to a 1sec averaging time. The bottom graph depicts a same step of $1.2x10^{-11}$ but the averaging time of the measuring system was increased to 100sec. The step is clearly observable here since the peak to peak noise is reduced to few parts in $1x10^{12}$, in consistent with the Allan deviation measurement.

The Allan deviation plot is exhibited in fig. 9. This is a measurement of the breadboard versus another Rubidium Standard model TF-4000B, which is used as a reference. The measured graph is about $1x10^{-11}/\sqrt{tau}$. Taking into account the reference stability, the actual stability of the breadboard is estimated to be $7x10^{-12}/\sqrt{tau}$. Thus, we expect that for tau=500 sec we shall have a short-term-stability of about $3x10^{-13}$ which is what is normally measured with our Rubidium Standard model TF-4000B.
Hence, a finer resolution in frequency settability can be achieved. Therfore we plan to increase the number of bits that control the "small n" to 16, in order to obtain a finer resolution of $9x10^{-14}$ in the frequency settability.

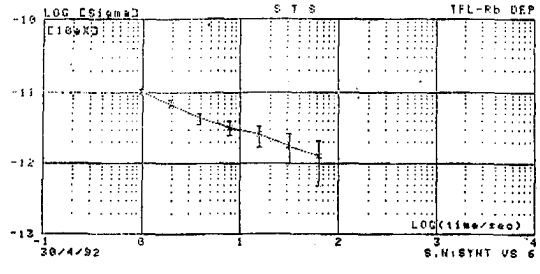


Fig.9: Allan deviation plot: Breadboard vs Rubidium Standard model TF-4000B.

Phase/Time control

The phase or time error, $x \equiv \Delta t$ is an integral of the fractional frequency error $y \equiv \Delta f/f$

$$x(t) = \int_0^t y(t)dt' \qquad (4)$$

Therefore a precise frequency control also allows for a precise phase/time control: apply a frequency step over a period T; this results in time step given by:

$$\Delta x = T \cdot \Delta y \qquad (5)$$

This is illustrated in fig. 10.
In our case the minimum time-step is given by:

$$\Delta x = 1.5x10^{-12} \times 196ms = 0.3ps \qquad (6)$$

Where the period of time is the minimum integration time required for averaging the output of the fractional synthesizer, as discussed above.
However the rms time-error noise is given roughly by [6]:

$$\sigma_x(\tau) \sim \tau \cdot \sigma_y(\tau) = 7x10^{-12}\sqrt{\tau} \qquad (7)$$

where τ is measured in seconds. The minimum time error noise is obtained at the smallest integration time which is 196ms, where:

$$\sigma_x(\tau = 196ms) \sim 3psec \qquad (8)$$

i.e., in order to resolve the time step above noise one has to perform a step larger than 3psec.

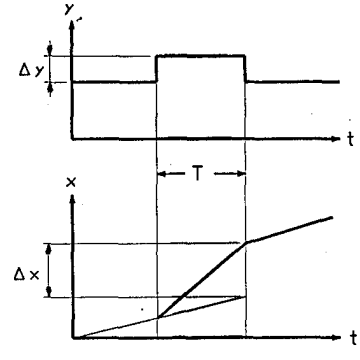


Fig.10: Relation of a frequency step to a phase/time step.

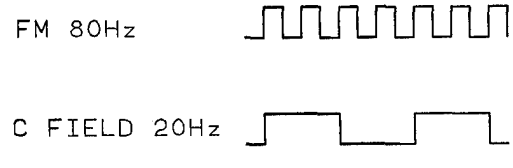


Fig.11: Timing relations between frequency modulation and C-field modulation.

Reducing the Sensitivity to Magnetic Field by the A.C. C-field Method

The a.c. C-field scheme was described in [1] and [2]. The idea is to alternate the C-field between $+H_c$ and $-H_c$, and to average. The alternating field can take the form of a square-wave or any periodic function such as a sinewave. The alternating frequency should be smaller than the Rubidium linewidth (1200Hz) but higher than the servo-loop bandwidth (2Hz).
We present here a simplified calculation that demonstrates the method. The dependence of the frequency, f, on the magnetic field H, is given by [4].

$$f = f_0 + aH^2 \qquad (9)$$

with: $a = 575 Hz/Gauss^2$ and $f_0 = 6.8 GHz$
Writing:

$$H = H_c + H_D/\alpha \qquad (10)$$

where, H_c is the alternating C-field, H_D is the parallel component of the disturbance field and α is the attenuation shielding factor.

Substituting (10) in (9) and averaging:

$$\bar{f} = f_0 + aH_c^2 + (\overline{\mp 2aH_cH_D/\alpha}) + a(H_D/\alpha)^2 = f_0 + aH_c^2 + a(H_D/\alpha)^2 \qquad (11)$$

The fractioanl frequency change due to the disturbance is given by:

$$\Delta f/f = a(H_D/\alpha)^2/f_0 \qquad (12)$$

Substituting $\alpha = 7000$ we obtain:

$$\Delta f/f = 1.7 \times 10^{-15} \qquad (13)$$

For $H_D = 1$ Gauss, and

$$\Delta f/f = 1.7 \times 10^{-13}$$

For $H_D = 10$ Gauss.

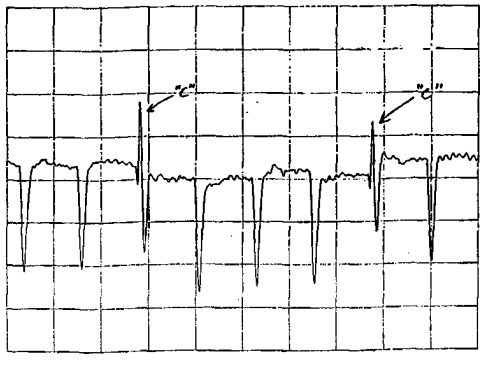


Fig.12: Transient signals that arise (positive) and from square-wave FM.

This is compared with 1×10^{-11}/Gauss for the case of constant C-field. However alternating the magnetic field induces transients signals that arise from the zero-crossing of the C-field [2]. In our case, the C-field modulation is being synchronized with the frequency modulation as shown in fig. 11. Then, the C-field transients occur at the same time as the transients that arise from the frequency switching. The Sample-and-Hold stage, is then being used to block both transients.

Fig. 12 shows the output of the photo-diode with the applied C-field modulation. One observes additional positive transients that occur every fourth negative transients. These are related to the zero-crossing of the magnetic C-field and result from non-adiabetic phenomena as discussed in ref.2. Their width is about 1ms or less (in fig. 12 the zero-crossing transients are superimposed on the square FM transients and are difficult to resolve; however, in ref.2 the zero-crossing transients are measured independently).

The response of the breadboard to an external magnetic disturbance field is shown in fig. 13. The disturbing field is applied parallel and perpendicular to the C-filed (only parallel mesurements are shown in fig. 13).
No change is observed in frequency when applying the ac C-field, as expected, with external magnetic fields up to 20Gauss.

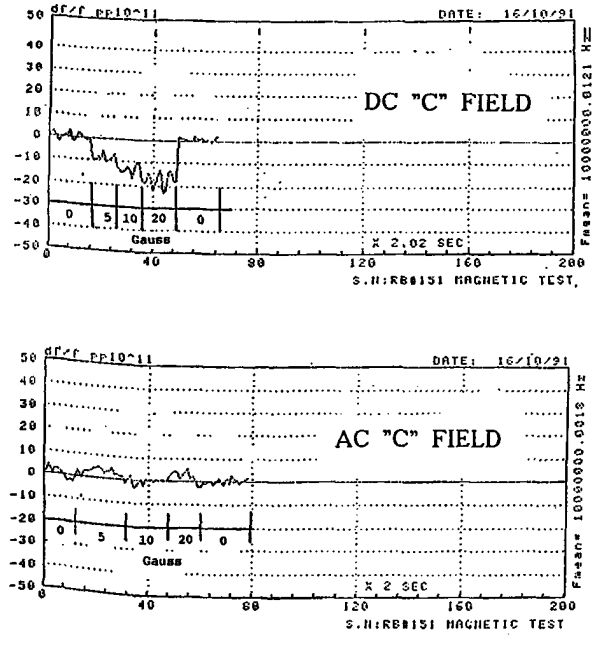


Fig.13: Reduction of the maggnetic sensitivity by the AC C-field compared with the DC C-field.

Applications

The following is a non complete list of applications envisioned for the "synthesized" Rubidium-frequency-Standard:

a. Disciplined oscillators in PLL with long time constant:
- Tracking oscillator in communication networks.
- GPS receivers
- Loran receivers

b. Realization of software clocks

c. Remote control of frequency and time:
- Spaceborne oscillator
- Communication networks

d. GPS navigation:
- Improved navigation solutions

We intend to devote a special paper to the applications which shall be published shortly.

References

1. A. Stern et al 43rd Ann Symp on Freq Control, 1989, p. 124.
2. A. Stern and M. Golosovski, 44th Ann Symp on Freq Control.
3. E. Detoma et al, Proc. of the EFTF 92.
4. C. Audin and J. Vanier, "The Quantum Physics of Atomic Frequency Standards".
5. J.C. Camparo & R.P. Frueholz, Phys Rev. A, 1984 p. 803.
6. D. W. Allan et al, 45th Ann Symp on Freq Control 1991, p. 667.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

TOWARDS A MINIATURE LASER-PUMPED CESIUM CELL FREQUENCY STANDARD*.

P. J. Chantry, B. R. McAvoy, J. M. Zomp, and I. Liberman.

Westinghouse Science & Technology Center, Pittsburgh, PA 15235

Abstract

With the goal of minimizing the overall size of a gas cell frequency standard we have characterized cylindical Cs cells having inner dimensions R=2 mm, L=18 mm. Under conditions of present interest the lifetime in the absence of a tuned microwave field tends to be dominated by Cs-Cs collisions. Wall collisions have only a marginal effect on the lifetime provided we satisfy the approximate condition $pR^2 > 80$ Torr. mm^2 where p is the N_2/Ar buffer gas pressure. Using a TE_{101} mode rectangular cavity the microwave resonance lineshape signal parameters were measured over a range of optical and microwave powers at various cell temperatures, and used to predict the achievable short term stability. For the optimum combination of optical and microwave powers the linewidth is dominated by power broadening, and the short term stability, $\sigma(\tau)$, predicted from these measurements decreases from $\sigma(\tau).\tau^{1/2} = 6.4 \times 10^{-12}$ to 4.2×10^{-12}, where τ is in seconds, as the cell cold spot temperature is raised from 50°C to 65°C.

1. Introduction

The gas cell frequency standard has matured in a form which employs rubidium as the working medium, optically pumped by a microwave discharge lamp, with isotopic filtering of the light to provide selectivity. Various forms of this device have evolved, including "miniature" versions where emphasis has been placed on reducing the total package size. The smallest of these [1,2] have volumes of 0.4 liter, and short term stabilities of $\sigma(\tau)\tau^{1/2}$ = few x 10^{-11} where τ is in seconds. The emphasis in our work is similarly to minimize the total package size, including taking advantage of the availability of compact laser diodes for highly selective optical pumping of the medium in order to maintain the short term stability performance $\sigma(\tau)\tau^{1/2} \leq 1 \times 10^{-11}$. A number of authors have discussed the potential advantages of substituting laser diode pumping for the conventional discharge lamp plus filter, and the improvement in Signal-to-Noise (S/N) achievable in this way has been demonstrated [3,4,5]. The potential for improving the overall performance has been explored in depth by Camparo and Frueholz [6] using a theoretical model of the device, taking into account the noise characteristics of available laser diodes.

If one plans to minimize the total package size, in part by integration of much of the electronics, one is faced by the fundamental contraints on the size of the microwave resonator, two or three dimensions of which necessarily scale as the microwave wavelength. Laser diodes are available for pumping either Rb or Cs, allowing one to consider either option. For Rb and Cs the microwave wavelengths are respectively 4.4 and 3.3 cm. While Rb theoretically has greater potential for ultimate stability performance [7] by virtue of having fewer magnetic sublevels, the relative advantage is only a factor of 1.6 and other differences may well reduce this factor in a practical device. We have chosen to employ Cs in place of Rb, since by so doing we are able to reduce the resonator volume by at least a factor of two. For example, a dielectrically loaded rectangular cavity operating in the TE_{101} mode designed for operation with Rb [8] has an internal volume of 4.6 cm^3. We have adapted this design for use with Cs, and obtained an internal volume of slightly less than 1 cm^3.

Thus we are faced with determining the net effect of three major changes from existing commercial devices having established performance figures, namely (i) the use of laser diode pumping, (ii) the use of Cs in place of Rb, and (iii) reduction of the resonator volume to a sensible minimum. The substitution of laser diode pumping is clearly expected to provide improved S/N. On the other hand, minimization of the microwave resonator size necessarily implies reduction of the gas

* Work supported by Westinghouse IR&D funds.

0-7803-0476-4/92 $3.00 © 1992 IEEE

cell volume, imposing a corresponding penalty in achievable stability.

The penalty from cell size reduction is expected to occur in two ways. The reduced number of accessible Cs atoms will necessarily reduce the optical signal strength, and consequently the achievable S/N. Also, a pumped Cs atom will interact sooner with the cell walls, reducing its effective lifetime, increasing the effective linewidth of the hyperfine transition, and consequently reducing the "Q" of the transition. Both effects can be mitigated by increases in the buffer gas and Cs densities, but even with re-optimization of these parameters a penalty will remain. Since the product of Q and S/N determines the achievable short-term stability, the net penalty from the cell size reduction could be unacceptably severe even with the benefits derived from the laser diode pumping. The work described here addresses the trade-off between these various factors in order to determine net achievable stability.

In Section 2 we describe experiments and calculations to determine the relative importance of the wall interactions in determining the line Q. In Section 3 we describe optical-microwave double resonance measurements which demonstrate the effect of power broadening on the linewidth, and draw conclusions regarding the relative importance of wall interactions in determining the line Q. In Section 4 we use these measurements to predict the achievable short term stability of a device employing a minimum size resonator and gas cell similar to those employed in the present study.

2. Lifetime vs. Cell Size, Pressure, and Cs Density

In order to determine the effect of the increased wall interactions on the hyperfine lifetime we fabricated a number of cylindical cells having inside dimensions R=2 mm, L=18 mm. These dimensions were chosen to be compatible with the expected dimensions of minimum sized resonators tuned to the Cs hyperfine frequency of 9.263..GHz. Three cells were vacuum processed and filled with a mixture of N_2 and Ar at pressures of 40, 70, and 200 Torr, and a similar set was filled with N_2 and Kr. In addition, all cells contained a small amount of Cs. Each cell was installed in turn in the experimental arrangement shown in Figure 1. The cell temperature was controlled such that the cold spot was close to 43 ° C, and the envelope is approximately 5 ° C higher. These temperatures were chosen to be relatively low compared to that expected in an operating cell, in order to keep the effects of Cs-Cs collisions

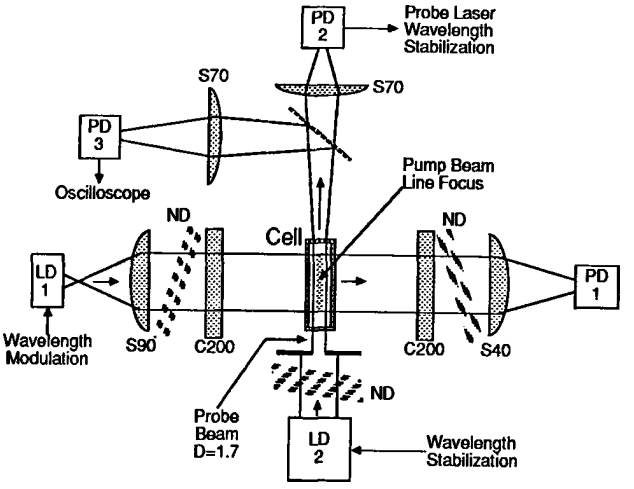


Figure 1. Experimental arrangement used to measure the time constant for relaxation of the imbalance in the hyperfine populations produced by selective optical pumping.

relatively small. Two laser beams pass orthogonally through the cell, allowing their strengths to be monitored without cross talk.

A typical current tuning curve for the type of laser used is shown in Figure 2, together with the relevant Cs energy levels. This data is taken by holding the laser diode junction temperature at an appropriate value, and ramping the laser drive current through the appropriate range. Under the conditions of all the present measurements the individual absorption lines have a Voigt profile due to the combined effects of Doppler and collision broadening by the buffer gas. For this reason the hyperfine components of the upper (6P) state are not resolved, but separation of the lower state components is essentially complete, indicating that either may be pumped with very high selectivity.

Laser diode 1 (LD 1), the pump laser, is operated from a current stabilized supply which permits current modulation. A slow square wave is used to switch the laser in and out of resonance at the selected absorption dip. The laser is kept on resonance for a fraction of a second, sufficient for the hyperfine populations to reach a steady state in the presence of the pump light. The pump beam is brought to a line focus at the cell's axis using a combination of a 90 mm focal length spherical lens (S90) and a cylindical lens (C200). This beam had sufficient intensity to significantly depopulate Cs atoms in the absorbing hyperfine level.

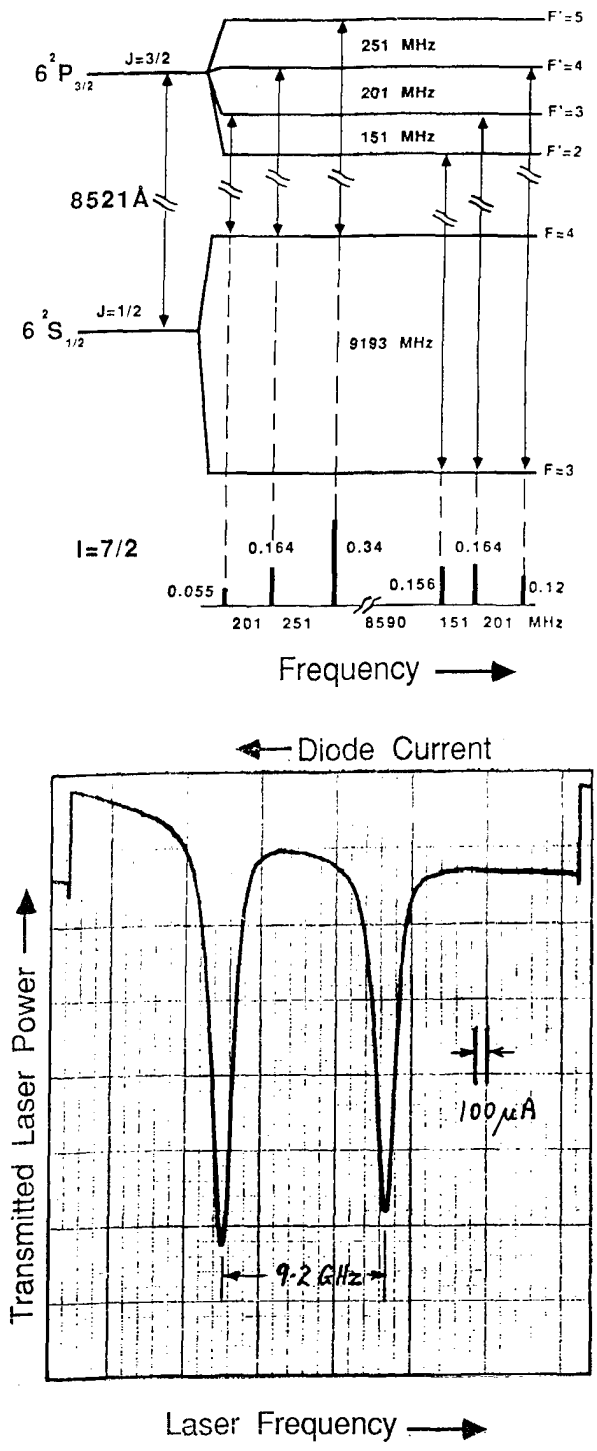


Figure 2. The Cs energy levels relevant to the present work, and an example of the transmitted laser signal as it is current tuned through the absorption lines. Transitions to the individual hyperfine levels of the upper 6P state are not resolved.

The second laser (LD2) was similarly operated at an appropriate junction temperature, with the wavelength stabilized to the selected absorption dip by feedback control of the current. A small current "dither" applied to the laser provided the required error signal via standard techniques of phase sensitive detection [9]. The signal detected by photodiode PD2 was used for this purpose. This laser provided the probe signal for monitoring the time dependence of the selected absorbing hyperfine level, and was kept at sufficiently low power to avoid any pumping. The measured relaxation time constants were shown to be independent of the probe power at the levels used.

An example of the data recorded in this way is shown in Figure 3. The time scale is 2 ms per major division. The upper trace is the transmitted probe signal following detuning of the pump. The lowest trace is the corresponding baseline, recorded by blocking the probe beam beyond the cell. The Tektronix 7854 oscilloscope used allows us to operate on the stored data in order to extract the time constant of the relaxation observed in the recorded data. In this example the probe laser is tuned to the pump wavelength, and thus is monitoring an increasing density N(t) of absorbers during the relaxation period. We assume that this occurs with a single time constant τ

$$N(t) = N_{eq} - \Delta N.\exp(-t/\tau) \quad (1)$$

The transmitted probe signal is an exponential function of N(t)

$$S(t) = S_0.\exp\{-\sigma L N(t)\} \quad (2)$$

It follows that

$$\log\{\log S(t) - \log S(t>>\tau)\} = \log\{\sigma L \Delta N\} - t/\tau \quad (3)$$

Thus we reduce the data by first subtracting the baseline, then taking the natural log of the net signal, then subtracting its value at late times, and then taking the log of the result. This is shown by the middle trace in Figure 3. Finally, two cursors are set to be vertically separated by a convenient integer, 3 in the example shown, in which case they are separated horizontally by that multiple of the time constant, allowing it to be recorded directly from the oscilloscope screen.

The relaxation time constants measured in this way for three separate cells filled respectively with 40, 70, and 200 torr of a N_2/Ar (30/70) mixture, are shown

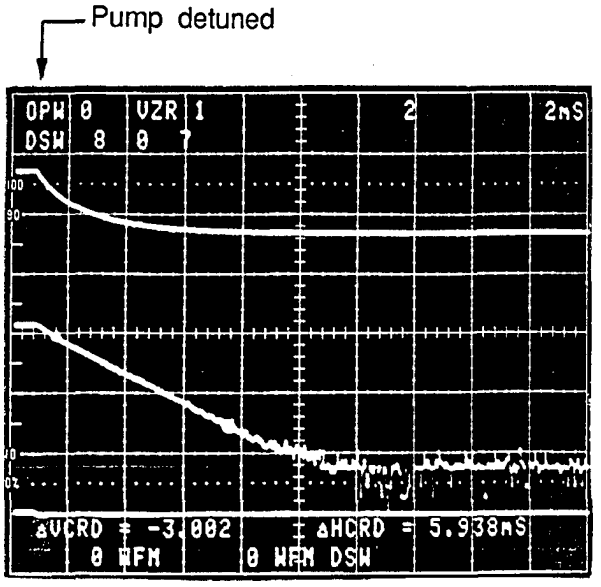


Figure 3. Time dependence of the transmitted optical probe signal following detuning of the optical pump, recorded and analysed on a Tektronix 7854 oscilloscope equiped with a waveform calculator.

in Figure 4, together with the dependence predicted from the equation

$$(1/\tau) = (DN)_m/ N_m\Lambda^2 + N_m\Sigma x_i v_i \sigma_i + [Cs]v\sigma_x \quad (4)$$

On the right of this equation the first term represents the rate of diffusion of pumped Cs atoms to the cell wall, where we assume they are equilibrated. Assuming fundamental mode diffusion we calculate the cell diffusion length Λ from

$$(1/\Lambda^2) = (\pi/L)^2 + (2.405/R)^2 \quad (5)$$

The second term of Eq. 4 represents relaxation of the pumped population imbalance due to collisions with the buffer gas constituents. The final term represents the effect of Cs-Cs collisions, where σ_x is usually referred to as the spin exchange cross section. The total density of buffer gas molecules is N_m, and $(DN)_m$ is the net density normalized diffusion coefficient related to the individual coefficients by

$$1/(DN)_m = \Sigma\{x_i/(DN)_i\} \quad (6)$$

where x_i is the mole fraction of constituent "i".

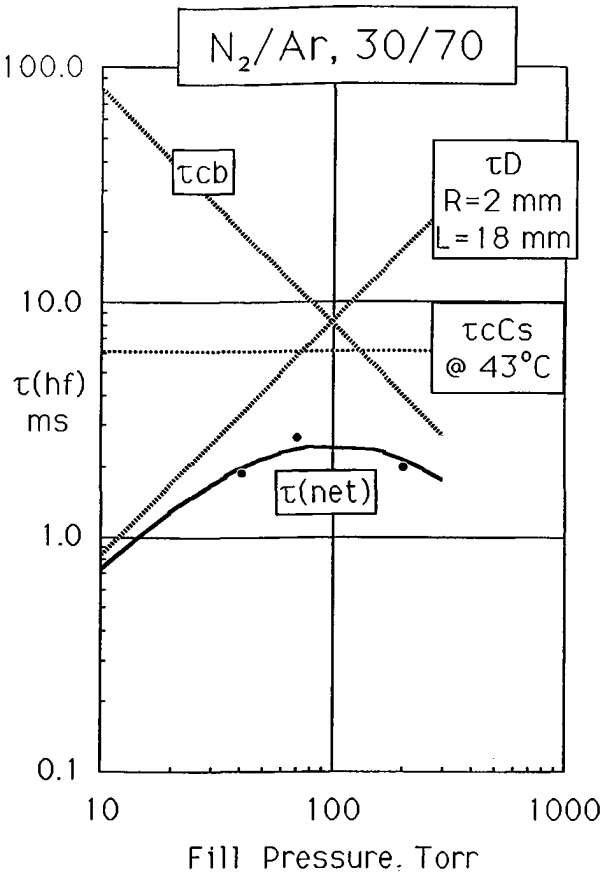


Figure 4. Comparison of the measured hyperfine relaxation time constants, shown by the points, and theory represented by the full curve labelled 43°C. The broken lines show the lifetimes corresponding to the individual processes represented in Eq. 1.

Using available gas data [10] Eq. 4 has been evaluated to give the individual and total contributions to τ, and plotted in Figure 4 as a function of buffer gas fill pressure. The straight lines indicate the relaxation time constants that would result from the processes acting independently, and the curve gives the net effect of their concerted action. At the cold spot temperature of 43°C indicated the Cs density is 2.4×10^{11} cm^{-3}. There is good agreement between theory and our experimentally measured time constants. Similarly good agreement was obtained for a set taken with a N_2/Kr mixture, confirming the much larger cross section for quenching by Kr atoms compared to Ar [10]. For this reason this mixture is of less practical interest.

The effect of raising the Cs density to values in the range of interest in an operating cell is shown in Figure 5, where the predictions for coldspot temperatures of 50, 60, and 70 • C are shown, corresponding to Cs densities of 4.2, 9.4 and 20 x 10^{11} cm^{-3} repectively. Figure 5 also shows the linewidths corresponding to the predicted relaxation times, given by $(\pi\tau)^{-1}$. For reference we also show the narrowed "Dicke" linewidth [11], given by $4\pi D/\lambda^2$, where λ is the RF wavelength. It is clear that in these small cells there is no possibility of approaching this limit.

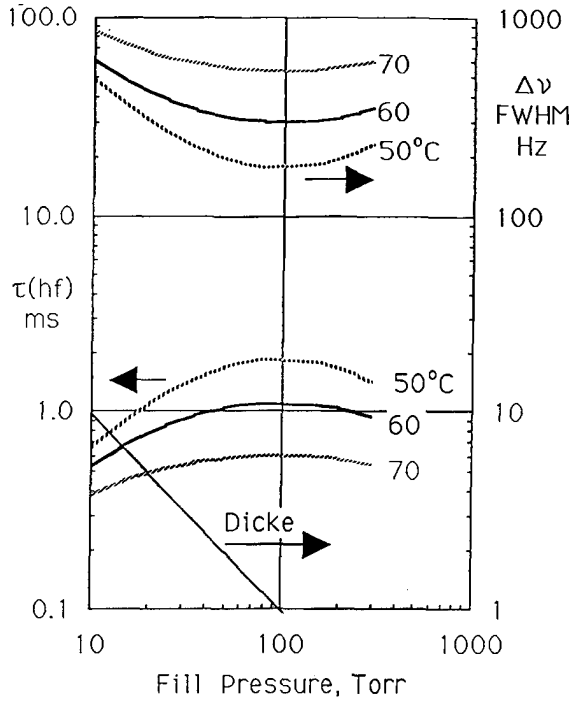


Figure 5. Predicted hyperfine relaxation time constants for Cs in a cell having inside dimensions R=2 mm, L=18 mm, as a function of pressure of a N_2/Ar (63/37) buffer gas mixture. Also shown are the corresponding linewidths, and for reference the linewidth predicted for a Dicke narrowed line.

The flattening of the curves in Figure 5 with increasing Cs density reflects the growing dominance of Cs-Cs collisions in determining the relaxation rate. This process will dominate over wall interaction effects provided

$$(DN)_m / N_m \Lambda^2 < [Cs] v \sigma_x \qquad (7)$$

For cells having L>>R we see from Eq. 5 that Λ = R/2.405, and for a cold spot temperature of 60 • C Eq.7 becomes

$$p_m R^2 > 65 \text{ Torr mm}^2 \qquad (8)$$

for a N_2/Ar (63/37) buffer gas mixture. To provide a margin for more general applicability we strengthen this constraint to >80 Torr mm^2. The dependence of τ on the scaling variable pR^2 is shown in Figure 6 for three temperatures, and treating the radius R as a parameter. The two sets of curves, calculated for R=2 and 5 mm respectively, are seen to closely overlap when plotted in this way.

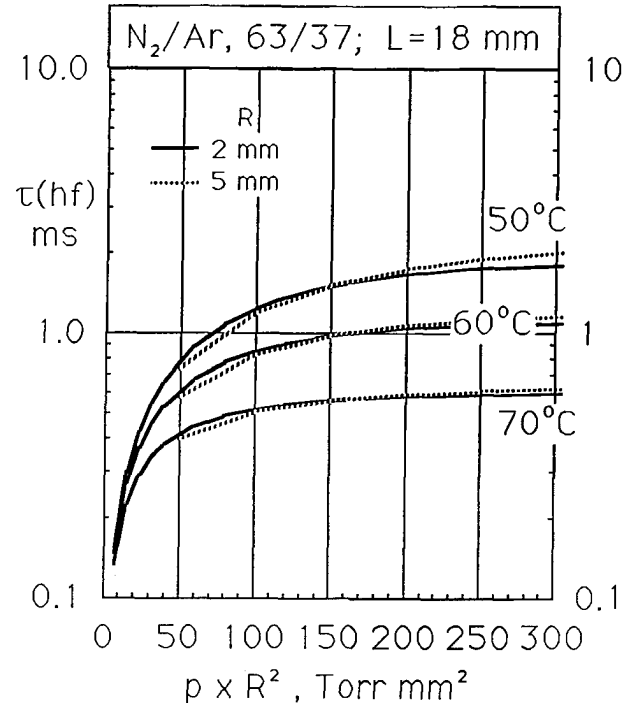


Figure 6. Showing the dependence of the predicted relaxation time constant on the paramenter pR^2 which controls the diffusion length of cells having L>>R. All three processes represented in Eq. 4 are included.

We conclude from these calculations, that at the Cs densities of interest, the effect of wall interactions is relatively small despite using such small cells, provided the buffer gas pressure is kept above 20 torr. The relative importance of wall collisions is further diminished when power broadening effects are taken

into account, as seen from the results presented in the next section.

3. Double Resonance Measurements

Two of the cells used for the lifetime measurements were subsequently used for microwave/optical double resonance measurements of the 9.192 GHz hyperfine line profile, using the arrangement shown in Figure 7. The cell was housed in a rectangular TE_{101} rectangular cavity made from X band copper waveguide. The loaded Q of the cavity including the cell was approximately 100. The magnetic field at the cell was controlled by adjusting the currents in three pairs of Helmholz coils. The ambient field was first nulled by adjusting the coils to collapse the Zeeman pattern, and the longitudinal "C" field was then applied to obtain ample separation of the Zeeman lines.

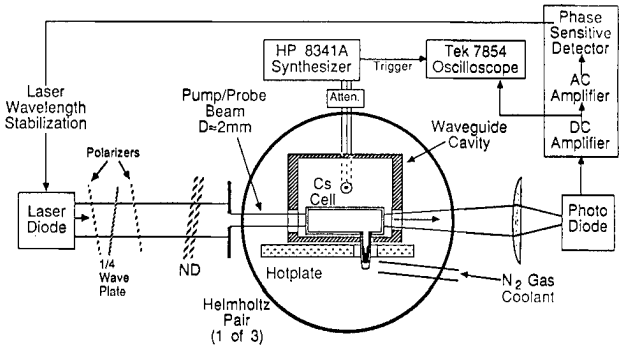


Figure 7. Experimental arrangement used for microwave/optical double resonance measurements of the hyperfine lineshape.

Measurements were performed at various cold spot temperatures, monitored by a thermocouple. The laser diode wavelength was locked to the absorption line being used, using a current dither frequency of approximately 2 kHz. The laser diode power incident on the cell was controlled by interposing neutral density filters, and measured with a power meter. The microwave power was supplied by a synthesizer which could be set to scan over the required frequency range with good reproducibility, allowing data averaging of multiple sweeps to be used when needed. Frequency scan rates were kept kept low to ensure undistorted line profiles [12].

Examples of microwave resonance dips recorded at various microwave powers are shown in Figure 8. This data was taken at an optical power near that giving the optimum predicted short term stability, at a cold spot

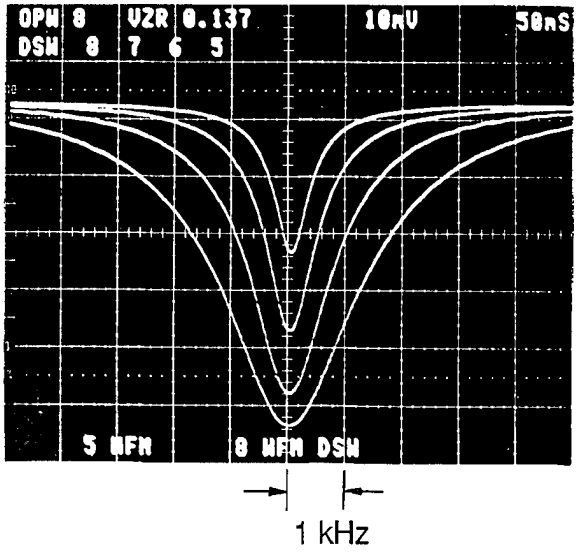


Figure 8. Examples of the lineshapes measured at applied microwave power levels of 0.3, 1, 3, and 10 μW. The baseline optical signal is 550 mV, and the dip amplitudes are 26, 40, 51, and 57 mV, increasing with the microwave power.

temperature of 55 • C. Note that signal dips exceeding 10% of the off resonance signal are straightforwardly obtainable because of the highly selective pumping. The dependence of the measured linewidths on the applied microwave power is shown in Figure 9 for two values of the optical signal. We plot the square of the linewidth vs. microwave power to display the expected linearity [13].

Data of the type displayed in Figure 8 was systematically recorded over a range of microwave and optical powers, with the coldspot temperature fixed at a number of different values. The width and fractional depth of the dips was noted, and the absolute strength of the signal. The relevant instrumental constant is the transimpedance gain of 10^6 of the amplifier used to convert the detected photocurrent to a voltage. Thus, for example the 550 mV signal indicated for Figure 8 represents a photocurrent of 0.55μ A, produced by 1μW of optical power absorbed by the Si photodiode. The corresponding laser power incident on the cell would typically be a factor of ~3 greater than this.

4. Prediction of Short Term Stability

With this information we can predict the achievable short-term stability, based on the formulation

of Vanier and Bernier [14]. These authors derive the relationship

$$\sigma(\tau)\tau^{1/2} = (3/2)^{1/2}\,(3/2^5) / Q(S/N) \qquad (9)$$

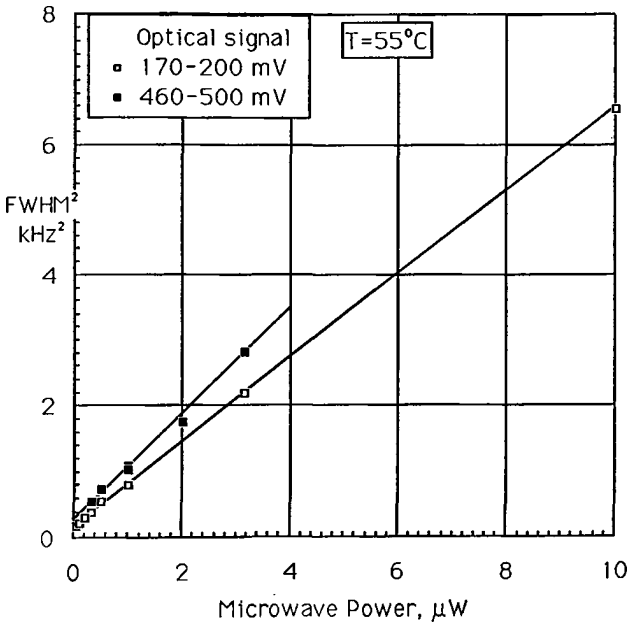


Figure 9. Dependence of the measured FWHM of the microwave resonance line on the microwave power applied to the resonator.

for the short term Allan variance, where $Q = \nu_0/\Delta\nu$, $\Delta\nu$ is the FWHM of the dip, and (S/N) is the signal-to-noise ratio of the peak discriminator signal, i.e. at $\nu_0 \pm \delta\nu/2$, where $\delta\nu = \sqrt{3}\,\Delta\nu$ is the separation of the peaks in the discriminator pattern if small amplitude modulation is applied with a Lorentzian lineshape. In order to use Eq. 9 or an equivalent we assume that the signal can be detected at the shot noise limit, and that we employ a peak to peak frequency modulation $= m\delta\nu$ where $m \ll 1$. With these assumptions it is convenient to recast Eq. 9 in the following forms

$$\sigma(\tau)\tau^{1/2}$$

$$= (3e/2)^{1/2}\,(1/2)(\nu_0)^{-1}(m)^{-1}\{\Delta\nu\,\sqrt{A_0}/\Delta A\} \qquad (10a)$$

$$= (e/2)^{1/2}\,(3/2)(\nu_0)^{-1}(m)^{-1}\{\delta\nu\,\sqrt{A_0}/\Delta A\} \qquad (10b)$$

$$= (e/2)^{1/2}\,(9/4)(\nu_0)^{-1}(m)^{-1}\{\sqrt{A_0}/\delta A'\} \qquad (10c)$$

$$= 2(2e)^{1/2}(\nu_0)^{-1}(m)^{-1}\{\sqrt{A_0}/\delta\nu\,/A_0''\} \qquad (10d)$$

where A_0 is the photosignal at the dip in amperes, ΔA is dip amplitude, $\delta A'$ is the peak to peak change in the first derivative of A, occurring between $\nu_0 \pm \delta\nu/2$, and A_0'' is the peak value of the second derivative of A, occurring at ν_0. The values of $\delta\nu$ and $\delta A'$ are read from the $A'(\nu)$ curve derived numerically from the digitally stored $A(\nu)$ curve. The value of A_0'' is similarly read from the $A''(\nu)$ curve obtained by further numerical differentiation. All forms of Eq.10 are equivalent if one has a noise free Lorentzian lineshape. Conversely, agreement between the various values of $\sigma(\tau)\tau^{1/2}$ is an indication that the lineshape is close to Lorentzian. If it is not, Eq. 10d provides the most realistic estimate of the achievable stability in that it is based on the slope of the discriminator curve - the property actually used to provide the error signal. When the S/N of the raw data was poor, as for example when the applied optical power and microwave powers were puposely made small, it was necessary to signal average, and perform some smoothing of the data between taking derivatives. In extreme cases Eq. 10d was not usable.

Examples of the data reduced to predicted values of $\sigma(\tau)\tau^{1/2}$ are shown in Figure 10 as a function of applied microwave power, and in Figure 11 as a function of the transmitted optical signal. Both sets of data refer to a cold spot temperature of 60° C. The stability predicted at this temperature with both power levels optimized is shown together with other similar data taken at other temperatures in Table 1. The parameter $\Delta\nu/\sqrt{3}\,\delta\nu$ should be exactly unity for a Lorentzian lineshape.

5. Discussion and Conclusions

The values of $\sigma(\tau)\tau^{1/2}$ were derived from Eq.10 using a value of m=0.1, chosen sufficiently small to ensure the validity of the "small signal" analysis used to derive Eq. 9. A "large signal" analysis has been made for low modulation frequencies, representing the response as a Fourier series, from which we conclude that the sensitivity peaks for a value of m=0.85, giving approximately a fivefold improvement over that predicted for m=0.1. Thus, in this respect, the predictions of $\sigma(\tau)\tau^{1/2}$ are conservative.

The assumption of shot noise limited signal detection remains to be fully tested. Preliminary measurements of the noise spectrum, under conditions essentially the same as used for the double resonance measurements, indicate that in the region of ~200 Hz the recorded noise was a factor of 8 above the shot noise

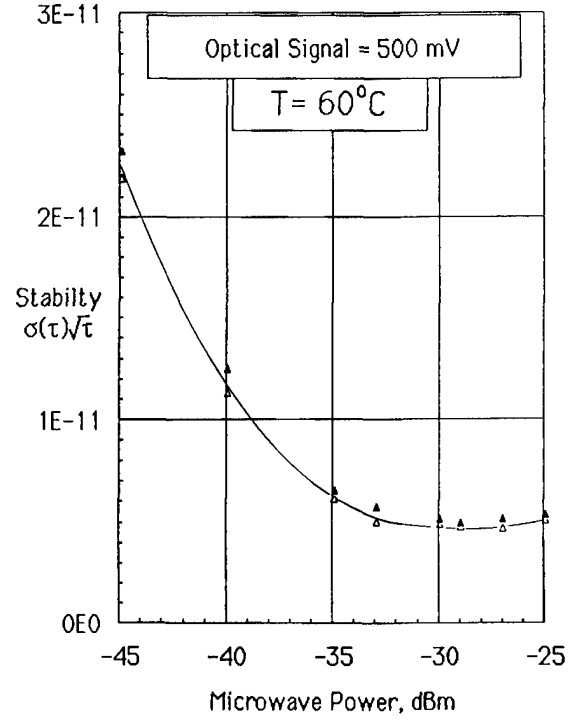


Figure 10. The short term stability predicted from measurements such as are shown in Figure 7 using Equation 10. This data was taken at an optical pump power close to the optimum for this cell temperature.

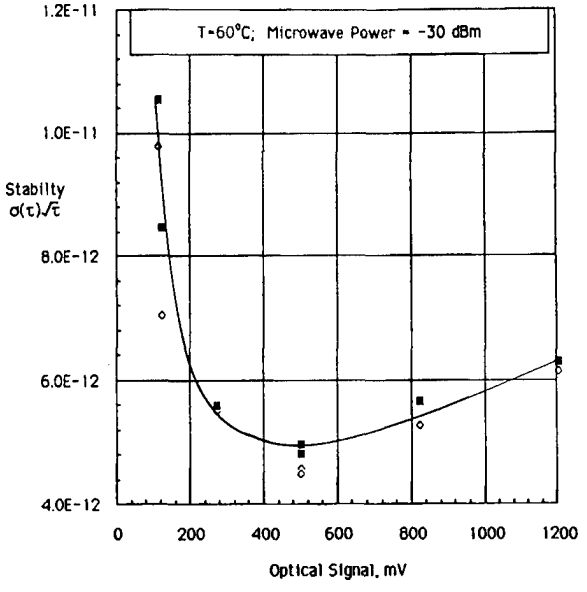


Figure 11. The short term stability predicted from measurements taken as a function of transmitted optical power. This data was taken at a microwave power close to the optimum for this cell temperature.

Table 1. Summary of the microwave resonance profile measurements for a cell having inside dimensions R=2 mm, L=18 mm, filled with 70 torr of N_2/Ar buffer gas. For each temperature the optimized conditions are listed for minimum value of the predicted short term stability, $\sigma(\tau)\tau^{1/2}$.

Coldspot Temp. °C	50	55	60	65
Inc. laser power, μW	2.5	3.4	5	6.3
Transmitted Optical Signal, S_0,mV	400	500	500	680
"Dip" Signal ΔS, mV	19	38	45	78
FWHM, Δν , Hz	760	1050	1080	1480
Discriminator FW, δν , Hz	430	600	645	762
$\Delta\nu/\sqrt{3}\delta\nu$	1.02	1.01	0.97	1.12
$\sigma(\tau)\tau^{1/2} \times 10^{12}$	6.4	5.2	4.7	4.2

limit, and appeared to be resulting from vibration effects in the widely spread optical setup. This source of noise should yield to improved mechanical mounting of the components in a much more compact form currently being implemented.

The data reported in Section 3 were obtained with a cell containing 70 Torr of N_2/Ar buffer gas mixture. The equivalent data from a 40 Torr cell gave predicted stabilities of 7×10^{-12}. We conclude from this, and the predictions from the theoretical analysis in Section 2, that adequate short term stability should be achievable with cells of this diameter filled with 20 torr of buffer gas or higher.

6. Acknowledgements

We are indebted to L. Jasper for his expert glassblowing, and to W. E. Dooley for his technical assistance in processing and filling the cells.

7. References

[1] K. Chiba and T. Hashi,"An Ultra-Miniature Rubidium Frequency Standard", Proc. 39th Ann. Freq. Control Symp., 1985, pp. 54-58.

[2] W. Weidemann, "Subminiature Rubidium Oscillator Model FRS", Proc. 40th Ann. Freq. Control Symp., 1986, pp 470-473.

[3] L.L. Lewis and M. Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards",1981, Proc. 35th. Ann. Freq. Control Symp., 612-624.

[4] M. Hashimoto and M. Ohtsu, "Experiments on a Semiconductor Laser Pumped Rubidium Atomic Clock", 1987, IEEE J. Q. Electr., QE-23 446-451.

[5] M. Hashimoto and M. Ohtsu, "Modulation Transfer and Optical Stark Effect in a Rubidium Atomic Clock Pumped by a Semiconductor Laser",1989, J. Opt. Soc. Am. B6, 1777-1789.

[6] J. C. Camparo and R. P. Frueholz, "Fundamental Stability Limits for the Diode-Laser Pumped Rubidium Atomic Frequency Standard", 1986, J. Appl. Phys.59 , pp. 3313-3317.

[7] J. C. Camparo and R. P. Frueholz, "A Comparison of Various Alkali Gas Cell Atomic frequency Standards", Proc. 40th Ann. Freq. Control Symp., 1986, pp. 447-451.

[8] H. E. Williams, T. M. Kwon and T. McClelland, "Compact Rectangular Cavity for Rubidium Vapor Cell Frequency Standards". Proc. 37 th. Ann. Freq. Control Symp., 1983, pp 12-17.

[9] T. M. Shay and Y. C. Chung, "400 Hz Frequency Stability of a GaAlAs Laser Frequency Locked to the Rb (D2) Line",1990, Opt. Eng. 29, pp.681-683.

[10] F. Strumia, N. Beverini, A. Moretti, and G. Rovera, "Optimization of the Buffer Gas Mixture for Optically Pumped Cs Frequency Standards",1976, Proc. 30th. Freq. Control Symp. pp. 468-472.

[11] R. H. Dicke, "The Effect of Collisions upon the Doppler Width of Spectral Lines", 1953, Phys. Rev. 89, pp.472-473.

[12] J. C. Camparo and R. P. Frueholz, "Parameters of Adiabatic Rapid Passage in the 0-0 Hyperfine Transition of 87Rb", 1984, Phys. Rev. A30 pp.803-811.

[13] J. C. Camparo and R. P. Frueholz, "Linewidth of the 0-0 Hyperfine Transition in Optically Pumped Alkali-Metal Vapors", 1985, Phys. Rev. A31 pp.1440-1448.

[14] J. Vanier and L-G. Bernier, "On the Signal-to-Noise Ratio and Short-Term Stability of Passive Rubidium Frequency Standards", 1981, IEEE Trans. Instr. Meas. IM-30, 277-282.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

WAYS OF IMPROVEMENT OF LONG-TERM FREQUENCY STABILITY OF ONBOARD ATOMIC BEAM FREQUENCY STANDARD FOR GLONASS SYSTEM

Prof. Arvid G.Gevorkyan, Dr. Vadim S.Zholnerov,
Dr. Arkady B.Basevich

Russian Institute of Radionavigation and Time (RIRT)
2, Rastrelli square, St.-Petersburg, 193124, Russia

Abstract

The results are presented concerning with RIRT's experimental and theoretical investigations to be aimed for improvement of long-term frequency stability of passive atomic frequency standards. A brief review is given with regard to parametric frequency shifts of the atomic clock transition in a cesium beam tube discriminator. The transformation of fluctuations in an atomic beam tube and in passive atomic frequency standard is considered. The investigations performed have permitted to improve the accuracy parameters of frequency standard by an order of magnitude.

Introduction

Frequency stability of a passive atomic frequency standard (AFS) is determined by fluctuation behaviour of specific assemblies of the AFS as an automatic control system. With regard to the influence of specific elements' fluctuations on the stability of the AFS output frequency, it is necessary to determine two types of fluctuations. The first type produces fluctuations of signal controlling the frequency of the quartz oscillator (QO) and thereby initiates the random fluctuations of the QO's conversed frequency with respect to the frequency of atomic transition in an atomic beam tube. The other type of fluctuations results, while the parametric shifts of atomic transition frequency take place, in random changes of this frequency and, through the automatic frequency locking system, gives rise to the fluctuations of the AFS output frequency. The fluctuations of the first types include intrinsic fluctuations of the QO frequency, fluctuations of frequency multiplier's phase and power as well as fluctuations of atomic beam intensity.

In order to estimate a contribution of those fluctuations to the stability of the AFS output frequency, the nature of its conversion in quantum discriminator, i.e. atomic beam tube, was investigated. As to the estimate of contribution of second-type fluctuations to the AFS output frequency stability, parametric shifts of atomic transition frequency in an atomic beam tube were investigated. It was ascertained that those frequency shifts depend on corresponding interfacing parameters between the assemblies of AFS and atomic beam tube. The behaviour of those parameters' fluctuations was investigated, as well as the response of AFS automatic frequency locking system to those fluctuations. Partial contributions of every fluctuation process to the AFS total frequency stability were determined.

All the practical results were derived for GEM AFS installed the GLONASS satellites onboard. The investigations performed have given for AFS breadbard a value of daily frequency stability not worse than 5.10^{-14}.

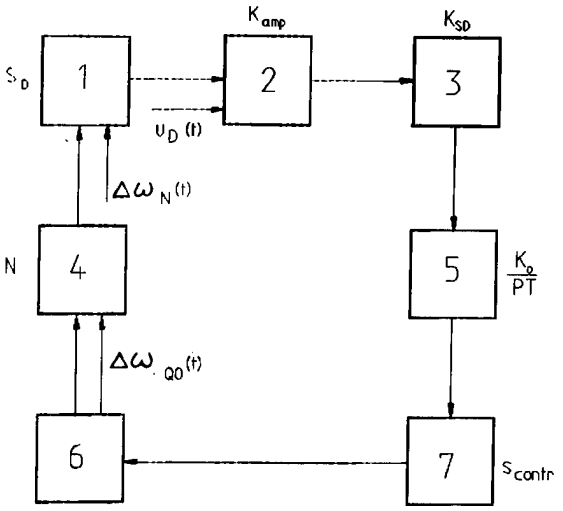
Review of parametric frequency shifts of the atomic clock transition in passive atomic frequency standards

Frequency shifts of the atomic transition in cesium atomic beam tube (ABT) can be divided into three groups taking into account the behaviour of shifts under external perturbations, that is the shifts connected to: (1) constant magnetic field in Ramsey resonator, (2) temperature of atomic beam source (or ambient temperature), (3) microwave excitation power. Temperature variations of the atomic beam source may be responsible for frequency shifts due to the thermal detuning of the microwave cavity or to the effect of stray current flowing through the cavity under the action of thermo-emf which

0-7803-0476-4/92 $3.00 © 1992 IEEE

arises at intermetal junction in ABT of adopted construction. The slope of the "transition frequency vs. temperature" curve of the beam source is 2.10^{-13} 1/K for the ABT in the first sets of GEM AFS. This value had been reduced to 2.10^{-14} 1/K as a result of improving ABT while inserting a dielectric isolation to interrupt the thermocurrents.

Some kinds of frequency shifts belong to the category of the microwave power-dependent shifts. These are the following: shifts arising from frequency pulling of an atomic clock transition by adjacent Zeeman transitions and by microwave cavity; shifts due to phase detuning of interaction regions of Ramsey resonator, and shift resulting from the second order Doppler effect. The shift produced as a result of frequency pulling by adjacent Zeeman transitions is of the most significant value [1 - 2].



Equivalent blok - diagram of passive atomik frequency standard being a part of automatik control system, where:
1 - quantiem discriminator; 2 - error signal amplifier;
3 - synchronous detektor; 4 - non - multiple frequency converter - frequency multiplier; 5 - low - pass filter; 6 - quartz oscillator;
7 - control element.

Fig 1

Fig.1 presents an equivalent block-diagram of a passive AFS with one ideal integrator. This block-diagram shows the transfer functions of AFS assemblies in linear approximation and random perturbations which are supposed being statistically independent.

Experimental investigations show that spectral power densities (SPD) of QO intrinsic fluctuations, of non-multiple conversion (NMC) assemblies and of quantum discriminator (QD) output signal fluctuations comprise flicker components and may be presented with the sufficient accuracy with following expressions:

$$W_{QO}(\Omega) = \frac{A1}{\Omega^2} + \frac{B1}{\Omega} + C_1$$

$$W_M(\Omega) = A_2 + B_2\Omega + C_2\Omega^2 \qquad (1)$$

$$W_D(\Omega) = W_B(\Omega) + S_{T_S}(\Omega)\left(\frac{\partial\nu}{\partial T_S}\right)^2 + S_{T_{PE}}(\Omega)\left(\frac{\partial\nu}{\partial T_{PE}}\right)^2 + S_{P_M}(\Omega)\left(\frac{\partial\nu}{\partial P_M}\right)^2 + S_H(\Omega)\left(\frac{\partial\nu}{\partial H}\right)^2 = \sum_{i=1}^{n}\left[A_{3i} + \frac{C_{3i}}{\Omega^{\alpha_i}}\right] \qquad 1 \le \alpha \le 2,$$

where W_B is SPD of atomic beam fluctuation and $S_i(\Omega)$ are fluctuations of beam source temperature, of platform temperature, of frequency multiplier output power and of magnetic field strength.

Below presented are the experimental results of determination of $S_i(\Omega)$ - values.

Fig.2, 3, 4, 5 present correspondingly the fluctuations of SPD for platform temperature, for microwave power of frequency multiplier, for magnetic field strength and for beam source temperature.

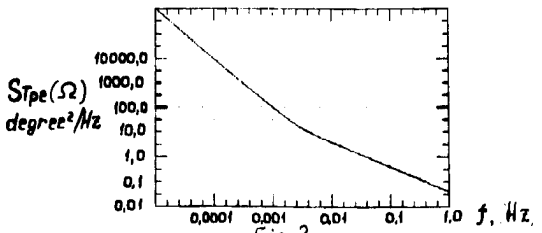


Fig.2

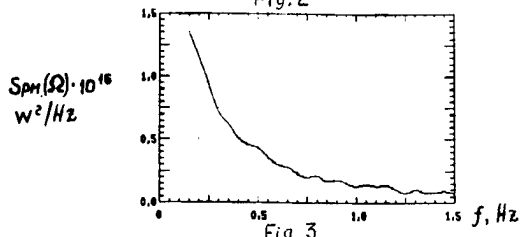


Fig.3

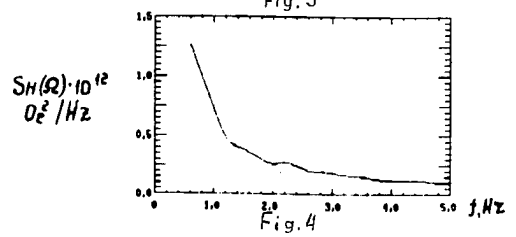


Fig.4

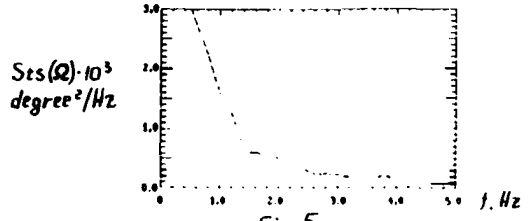


Fig. 5

Let us choose as a measure for AFS frequency stability the relative root-mean-square two-sample frequency variance.

$$\sigma(\tau) = \frac{1}{\omega_o}\sqrt{\frac{8}{\pi}\int_0^\infty W(\Omega)\frac{\sin^2\frac{\Omega\tau}{2}}{(\Omega\tau)^2}d\Omega}, \quad (2)$$

where ω_o is QO output frequency,
τ is time measurement interval,
$W(\Omega)$ is SPD of QO frequency fluctuation for the closed frequency lock loop of AFS.

$W(\Omega)$ consists of three parts:

$$W(\Omega) = \left|G_{QO}(j\Omega)\right|^2 \cdot W_{QO}(\Omega) + \left|G_M(j\Omega)\right|^2 \times$$

$$\times W_M(\Omega) + \left|G_{QD}(j\Omega)\right|^2 W_{QD}(\Omega), \quad (3)$$

where $G_{QO}(j\Omega)$, $G_M(j\Omega)$, $G_{QD}(j\Omega)$ are closed loop transfer ratios from the point of application of random perturbations, such as $\Delta\omega_{QO}(t)$, $\Delta\omega_N(t)$, $U_D(t)$, to the output.

By substituting (3) into (2) we get:

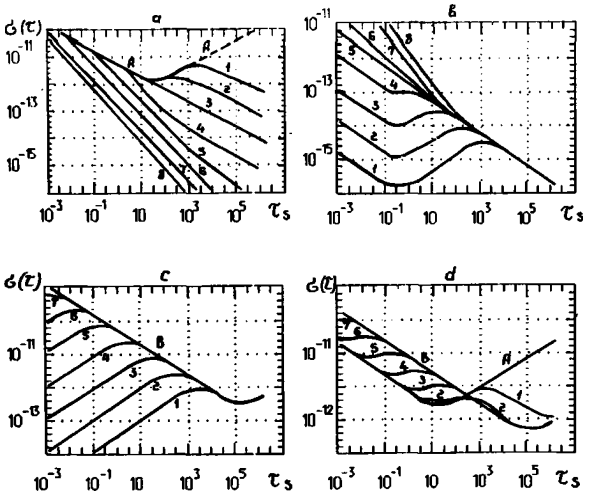
$$\sigma(\tau) = \sqrt{\sigma^2_{QO}(\tau) + \sigma^2_M(\tau) + \sigma^2_{QD}(\tau)} \quad (4)$$

The results of (2) - equation solution for corresponding values of $W_{QO}(\Omega)$, $W_M(\Omega)$, $W_{QD}(\Omega)$ as well as (4)-equation solution are presented on Fig.6, where one can see the contributions of fluctuations of QO(a), of multiplier (b), of QD(c) into the AFS instability and the total frequency instability $\sigma(\tau)$ (d) versus of measurement time duration (1-8 - plots correspond with $\Omega_c = 10^{-3}$, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4 °C^{-1}). Here $\Omega = \beta/T$, that is cut-off frequency of AFS lock loop, $\beta = NS_{QD}\ K_A\ K_{SD}\ K_O \cdot S_{CE}$, that is regulation ratio being the product of transfer ratios for elements of the frequency lock loop. The solution was derived for values of A, B, C - coefficients determined experimentally for assemblies and ABT of GEM AFS.

The plots presented above show that contributions of frequency multiplier and quantum discriminator to the total frequency instability $\sigma(\tau)$ increase, when increasing cut-off frequency, whereas the QO contribution decreases. Under those conditions, when $\Omega_c \to 0$, $\sigma_{QO}(\tau)$ approaches the inherent stability of QO frequency (A - curve); when $\Omega_o \to \infty$, $\sigma_{QD}(\tau)$ approaches the B-curve of the most instability, and as to the curve of the total instability $\sigma(\tau)$, it is always above the least value derived from the values of curves A and B for various Ω_o - values. The linear part of B - curve corresponds with τ^{-1} law, and for the linear segment a value of frequency stability is determined by coefficient

$F = \frac{S_{QD}}{C_{QD}}$ [3] which is called QD figure of merit.

As a result of investigations performed, it was stated that the greatest contribution to the daily frequency instability is the consequence of fluctuations of the magnetic field strength and of temperature fluctuations of beam source (platform temperature).

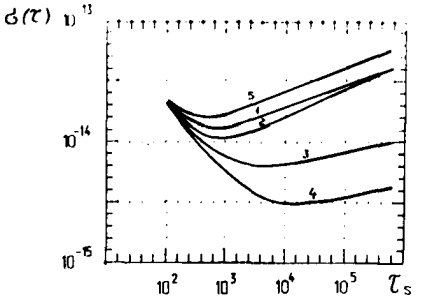


Contributions of fluctuations of quartz oscillator (a), non-multiple frequency converter (b), quantum discriminator (c), total frequency instability $\sigma(\tau)$ (d) vs. duration of time measurement interval, 1-8, for $\Omega_o = 10^{-3}$; 10^{-2}; 10^{-1}; 1; 10; 10^2; 10^3; 10^4 sec^{-1}

Fig. 6

It was induced by the first version of ABT design used in GEM AFS. That version was a tube made of copper throughout with sealed deflecting magnets and Ramsey cavity. Its disadvantage was the warming of the cavity due to the temperature of the beam source. In addition, no care was taken to eleminate the thermo-currents arising at the places of contacts and sealings of various metals, for instance, copper and permalloy, and that was the reason of emerging additional non-stable magnetic fields in a region of the microwave cavity. The up-to-date ABT design has no such disadvantages due to the sealed ceramic insertions mounted along ABT length, these insertions providing with electrical and thermal isolation.

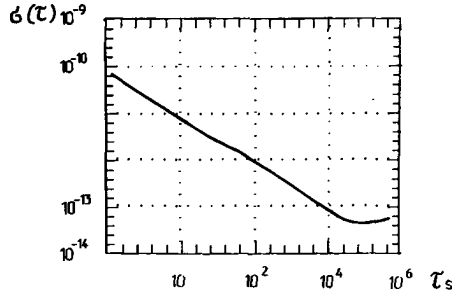
Fig.7 presents the partial contributions of various fluctuations, i.e. those of constant magnetic field strength (curve 1), of beam source temperature (curve 2), of microwave excitation power (curve 3), external temperature (curve 4) into the long-term instability of AFS with the new ABT with an improved figure of merit. The total effect of all the fluctuations is also shown.



Partial contributions of fluktuations of constant magnetic field strength (1), atomic beam source temperature (2), microwave excitation power (3), external temperature (4) with respect to long-term stability of atomik frequency standard and a total effect of fluctuations (5).

Fig. 7

Fig.8 gives an instability behaviour for the new AFS breadboard with the improved ABT. Realized is the daily instability value not worse than 5.10^{-14}.



Instability of sample.

Fig. 8

The investigations performed have permitted to determine partial contributions of AFS specific assemblies and ABT into the long-term stability. Developed is the AFS breadboard with the improved ABT having a daily frequency instability not worse than 5.10^{-14}.

References

[1.] Yanke E., Emde F., Lesh F. Spetsialnye funktsii (Special functions). Moscow, Nauka, 1977, p.60.

[2.] Gradstein I., Ryzhik I. Tablitsy integralov, sum, ryadov u proizvedeny (Tables of integrals, sums, series and products). Moscow, GIFML, 1962, p.107.

[3.] Lacey R., Helgesson A., Holloway J., Proc. IEEE, 1966, v.5, N 2, p.170.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ARCHITECTURE AND ALGORITHMS FOR NEW CESIUM BEAM FREQUENCY STANDARD ELECTRONICS

Leonard S. Cutler and Robin P. Giffard
Hewlett-Packard Laboratories
P. O. Box 10350, Palo Alto, CA 94303-0867

Abstract-The main contributions to frequency error and long term instability in conventional cesium beam frequency standards are Ramsey and Rabi pulling, cavity phase shifts and cavity mistuning when there are microwave magnetic field amplitude variations in the microwave cavity. C-field instability is another large contributor. Changes in effective amplitude level or C-field with time, temperature, humidity, magnetic field, etc. are the main sources of frequency drift and instability with time or environmental changes. Careful design of magnetically deflected beam tubes can greatly reduce Ramsey pulling, and properly designed optically pumped tubes have virtually no Ramsey or Rabi pulling but the remaining frequency shifts with amplitude level and C-field changes can still be large. Finally, microwave spectral asymmetry and even order distortion of the line-center-finding modulation are also very important sources of frequency shift.

A description is given of a new electronics design that addresses microwave amplitude shift, C-field, Rabi pulling, microwave spectral purity, and modulation distortion. The design is microprocessor based with automatic lock-up and long term operation without operator intervention. The basic architecture and algorithms are covered as well as the design considerations necessary to limit error contributions due to electronics effects to the level of a few parts in 10^{14}. Frequency standards using the new electronics have demonstrated the expected stability and accuracy improvements [1].

Introduction

It is well known that one of the main contributions to frequency error and long term instability in cesium beam atomic frequency standards can be amplitude level changes in the microwave cavity magnetic field. When Ramsey pulling, Rabi pulling, cavity phase shifts, and/or cavity detuning are present, the frequency produced by the tube/electronics combination can depend on the microwave amplitude [2-5]. Then if the amplitude changes with time and/or environmental influences such as temperature or humidity, corresponding frequency drifts or instability can occur. In some cases the variation of frequency with amplitude is an oscillatory function of C-field value and can vanish at certain C-field values [6].

Asymmetry in the microwave spectrum can also cause frequency shift [7,8]. Asymmetry can arise from even order distortion in the intentional line-center-finding modulation or from the presence of unwanted sidebands produced in the frequency synthesis process used to obtain the desired microwave frequency. Also, if there is amplitude modulation that is coherent with the intentional line-center-finding phase or frequency modulation, spectrum asymmetry can occur. A simple example of this is sinusoidal frequency modulation with coherent amplitude modulation. If the relative phase of the two modulations is not exactly $\pi/2$ then the spectrum will be asymmetric. Coherent amplitude modulation can occur, for example, with mistuning of the tube's microwave cavity or a filter associated with it. There will then be a first order dependence of amplitude on frequency so that as the frequency changes during the modulation cycle there will be a corresponding amplitude change. Such modulation will be either in-phase or in antiphase with the intentional frequency modulation and thus lead to an asymmetric spectrum. This is the real origin of frequency pulling due to cavity mistuning.

There is, however, a saving feature. Due to the non-linear behavior of the beam tube's response, there is always a particular microwave magnetic field amplitude for which the frequency pulling due to coherent amplitude modulation vanishes. As will be discussed, this optimum amplitude depends on the type of intentional frequency or phase modulation. If the amplitude were to be stabilized at that level then pulling due to cavity mistuning would essentially vanish and, since the amplitude is stabilized, all the other amplitude dependent frequency shifts would either not be present or would be stabilized.

Note that since the cavity Q can change, constancy of microwave power does not imply constancy of magnetic field amplitude. The field amplitude is what is important and what must be measured.

Another main contributor is C-field instability. Clearly, if the C-field changes for any reason there will be a corresponding frequency shift. The main causes of

0-7803-0476-4/92 $3.00 © 1992 IEEE

C-field changes are instabilities in the C-field power supply, expansion or contraction of the actual tube structure induced by temperature change, etc., and changes in the magnitude or direction of the external environmental magnetic field that are imperfectly removed by the tube's magnetic shields.

With all this in mind, the design of a new electronics package was undertaken with an error budget goal of less than 1 x 10^{-14} frequency shift for the effects mentioned above. Digital circuity was to be used as much as possible along with a microprocessor, and all important parameters were to be servo-controlled to remove the need for operator intervention and adjustments. A further goal was to make the standard capable of being completely controlled and monitored remotely through an RS232 interface. The architecture and algorithms of the design are described in the following sections.

Design and Algorithms

Modulation Type

The first choice to be made was the type of modulation to be used to find the line center. A modified type of slow square-wave frequency modulation [9,10] was selected. The microwave probe frequency is switched sequentially between various points on the resonance line. After each frequency change the tube output current shows a transient whose length is roughly the average transit time of atoms through the cavity smeared out by the atomic velocity distribution, the spread of the travel times to the detector, and the spread of cesium sticking times at the detector. After the transient has settled, the tube output current is averaged for a relatively long interval while the probe frequency is held constant. For the tubes used, the duration of the transient is 2 to 3 ms, and the frequency is changed each 12.3 ms, corresponding to a frequency of 40.7 Hz for pure square-wave modulation. There are several advantages gained by using this type of modulation which will now be discussed.

First, slow square wave frequency modulation allows the microwave amplitude to be controlled to produce the maximum signal from the beam tube at symmetrical points on either side of the line, the inflection points being preferred [14]. The microwave amplitude necessary to maximize the signal on the sides of the transition is exactly that amplitude for which the effects due to coherent amplitude modulation vanish so shifts due to cavity detuning are reduced to a negligible value. Detailed computer modeling verified these conclusions. Since the amplitude is now controlled to keep the microwave magnetic field in the transition regions of the tube constant, all the other amplitude dependent shifts are stabilized to a high degree giving excellent long-term stability as well as low sensitivity to temperature and humidity.

While there is also clearly an optimum amplitude [9] for other types of modulation such as sinewave phase or frequency modulation, or square-wave phase modulation, the optimum value can not be determined from tube signal measurements. For these types of modulation, computer simulation and experiments show, for example, that maximizing the current at the Ramsey peak, with or without modulation, or maximizing the recovered second harmonic does not give the optimum amplitude. Slow square wave frequency modulation is superior in this respect. Using the tube signal itself to determine and stabilize the proper amplitude is clearly optimum. Any other detection means would have to be calibrated and could drift with time or fluctuate with ambient conditions.

The second reason for the choice is that it is not too difficult to generate the required frequency modulation pattern with adequately low even-order distortion by using a direct digital synthesizer. The requirement is -120 dB with respect to the fundamental component to cause less than 1 x 10^{-14} frequency shift. The details of the frequency synthesizer are covered in a companion paper [11].

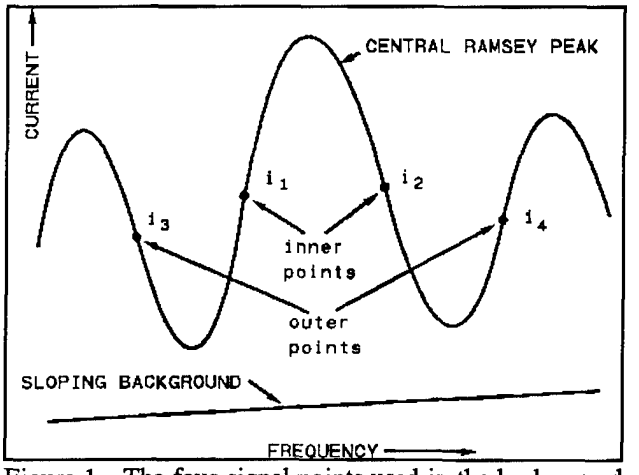


Figure 1. The four signal points used in the background slope removal technique.

The third reason is that it gives the best error-signal to noise ratio and thus the best asymptotic short term frequency stability for a given tube. Also, the error

signal is very easy to handle and can be converted at an early stage to the digital domain for processing.

Digital Synthesizer

One side benefit of using a digital frequency synthesizer is that the modulation can be extended easily to look at other points with opposite slope on the Ramsey pattern such as shown in Fig. 1. This is one of the modifications to the pure square-wave frequency modulation. Combining the error signals from the "inner" and "outer" pairs of points in the proper proportion and with the right sign gives an error signal that is free of any contribution from linear slope with frequency [12,14].

The total frequency error signal used for the case shown in Fig. 1 is given by:

$$i_{f\,err} = i_1 - i_2 - \frac{F_{p\,inner}}{F_{p\,outer}} (i_3 - i_4) \qquad (1)$$

where $F_{p\,inner}$ and $F_{p\,outer}$ are the frequency modulation deviations for the inner and outer points and i_j is the beam tube current at the point j. Such a background slope is produced by unbalanced neighboring $\Delta M=0$ transitions, the source of Rabi pulling. Using this algorithm greatly reduces the Rabi pulling. It must be mentioned that the optimum amplitude for the outer points differs from that for the inner points so a separate amplitude servo loop is required. A correction for this difference is applied in the algorithm that removes the background slope. Measurements made on the tubes used with this electronics show very little Rabi pulling. The calculated pulling without correction is somewhat less than 1×10^{-13}.

A major benefit of the digital synthesizer is that its frequency is programmable in precisely known steps thus allowing the cesium standard to be offset or steered by exactly known amounts either by means of its keyboard or remotely through the RS232 port. This is extremely useful for controlling the members of an ensemble of clocks.

Other benefits of the digital synthesizer arise from the use of its frequency agility. During part the modulation sequence it is set to the peak of the central line allowing a measure of tube signal size for setting the gain of the main frequency servo loop. It is also sequenced to the sides of the upper adjacent zeeman transition line. This is used for servo control of the C-field to keep that frequency constant. This is extremely important since it effectively removes C-field variations (except for any effects of field inhomogeneities which must be carefully considered in the design of the tube).

Frequency Architecture

The next main choice was the frequency architecture. To meet the goal of 1×10^{-14} requires that any unbalance in sideband pairs or any individual sidebands in the tube microwave probe signal be lower than about -70 dBc if they fall anywhere in the range of $\pm$ 130 Khz from the M=0 to M=0 clock transition [7,8]. This takes into account all the pulling effects including a sideband directly exciting any of the field dependent $\Delta M=0$ transitions at a point of maximum slope. Sidebands outside this range have a pulling effect that decreases inversely with the frequency separation and so become less important [7,8]. To ease the requirements on the synthesizer with regard to sideband levels, its output is translated to the microwave frequency in two stages with no frequency multiplication giving the additional benefit of no degradation in its frequency resolution. The first stage of the translation scheme uses a moderately narrow band phase-locked quartz crystal oscillator with the synthesizer output as one of its reference signals. The output of this oscillator at 87 MHz is used as the one of the references for the second stage of the translation, a phase-locked microwave dielectric resonator oscillator that provides the actual probe signal to the beam tube. The synthesizer frequency is thus effectively added to a 9192.5 MHz reference to obtain the tube probe signal. Both the 9192.5 Mhz signal and the synthesizer signal are derived from the same reference so they are coherent and the synthesized probe frequency is exactly known with respect to the reference. The two phase-locked oscillators performing the frequency translation provide symmetrization and filtering of undesired sidebands from the synthesizer and the other reference signals. With these techniques, the -70 dBc requirement for unbalanced sidebands is easily met without the need for a high cavity Q or any other microwave filtering. Considerably more detail is given in a companion paper describing the RF architecture [13].

The microwave adaptor that completes the cavity external to the tube and couples the microwave probe signal into the tube was designed to have minimal asymmetry about the plane normal to the E field in the waveguide. This should help reduce the phase shift associated with the unwanted mode excited by these asymmetries and thus reduce frequency shift.

The basic frequency source for the microwave chain is a clean, stable 10 MHz crystal oscillator which is

frequency controlled by the main cesium standard frequency servo loop. The output of this oscillator is not used for the output signal of the cesium standard for the following reasons. The frequency that is actually being controlled and stabilized is that of the microwave input to the cesium beam tube. To get to the tube input from the 10 MHz oscillator requires a lot of frequency multiplication with possibly poor phase shift versus temperature or humidity characteristics. Thus the phase of the 10 MHz oscillator with respect to the stabilized microwave frequency could change with time as the temperature or humidity changes. This would lead to transient frequency shifts since frequency is proportional to the time derivative of phase. To keep the user's output signals from the cesium standard tied as tightly as possible to the cesium tube input, a signal well up the frequency multiplier chain at 320 MHz is digitally divided by 4 to get to 80 MHz which is then effectively divided by 16 or 8 to get the actual 5 or 10 MHz output signals. The phase stability behavior of the digital division is considerably better than frequency multiplication [13].

Frequency and Microwave Amplitude Servo Loops

After amplification, the signal from the beam tube is converted to a digital signal and all further processing is done digitally up to the output DACs (Digital-to-Analog Converter). The frequency loop is completed by feeding the frequency control DAC output through a small amount of analog filtering to the 10 MHz crystal oscillator electronic frequency control input.

The error signal is processed using the background linear frequency slope removal algorithm mentioned earlier. The sense of the frequency modulation is periodically reversed and a form of second differencing is performed in order to obtain a frequency error signal that is free from the effects of signal and background linear drift with time. The resulting digital error signal is then processed with a software digital filter that performs a double integration with some damping to provide loop stability. Double integration removes the offset that would occur with linear drift of the quartz oscillator with time if only single integration were used. A digital servo loop monitors and keeps the current at the center of the line constant by controlling electron multiplier voltage and amplifier gain. The size of the actual signal is determined by the difference in current between line center and the sides of the line. After conversion to the digital domain, the signal size is kept constant by the software. This keeps the overall frequency servo loop gain constant. The closed loop time constant is several seconds.

The signal from the beam tube also contains the error information necessary to optimize the microwave amplitudes for the inner and outer interrogation points. Separate software signal processing is done to develop the two control words for the amplitude control DAC and these are sequenced into the DAC in synchronism with the frequency modulation so that the correct amplitude is always applied to the tube cavity. The amplitude modulation signal necessary to determine the optimum amplitudes is also generated by the DAC. The amplitude servo loops are closed by feeding the output of the DAC to a carefully designed amplitude modulator at the output of the microwave chain.

Microprocessor System

The microprocessor system controls almost all functions of the standard. It does all the digital signal processing and filtering and controls the cesium tube oven temperature and the C-field current. It also provides the interface between the operator and the instrument by means of a small keypad located on the front panel as well as the remote interface through the RS232 serial port. The SCPI (Standard Commands for Programmable Instruments) command set is used for remote control and communication.

Many subsystems provide monitor signals to the microprocessor and these signals are scanned frequently to see if they are in the proper range. The servo loops implemented in software are also continuously monitored. The basic status of the standard is indicated by a pair of lights controlled by the microprocessor on the front panel. One of these is a green continuous operation light. Most of the conditions and levels are available as data on a front panel LCD dot-matrix display or through the RS232 port. Conditions that require attention but do not cause improper operation are flagged as warnings. Anything that is interpreted as improper operation leads to a fatal error state. Either a fatal error state or microprocessor failure prevents the continuous operation light from being on. Warnings and fatal errors are documented in a log produced by the microprocessor and kept in non-volatile memory so the user can tell what occurred at some later time.

A TTL open collector status output is provided on the rear panel and the state for which it pulls down can be user programmed. Consequently, a number of standards can be wire-ORed together on one line giving the user a simple way to tell that at least one of the standards is in its programmed pull-down state.

Start-up is automatic and completely under microprocessor control. After initial warmup of ovens,

etc., levels and gains are set and a frequency scan is done to locate the line center and to set the C-field. If everything is normal, the servo loops are then closed and the system is in full normal operation.

All the information about the cesium tube that is necessary for proper operation is contained in a ROM attached to the tube. The microprocessor reads and uses this information to set up the proper operating conditions. This feature makes changing tubes extremely simple. There are no adjustments or measurements needed.

1 Pulse-per-Second circuitry

The circuitry for the 1 PPS (Pulse Per Second) is driven from an 80 Mhz signal derived from the same divider that runs standard's RF output circuitry thus making the 1 PPS coherent with the 5 or 10 MHz outputs. The timing or delay of the output pulse can be set manually or synchronized automatically to an external pulse with a resolution of 50 ns.

Power Supplies

The main internal supplies that are distributed to all the modules are +5, +12, and -12 volts DC. These are generated from high efficiency regulated switching supplies which are very carefully shielded and filtered. The input to these supplies can come from an internal supply powered by the AC line, or an external DC supply, or an internal battery. Power selection and internal battery charging are automatic and under control of a separate power steering logic circuit.

Many of the modules must have extremely clean power sources to avoid frequency pulling effects or sideband generation. Since the main distributed voltages are not perfectly clean most of the modules have additional internal post-regulator/filters or power conditioners. The worst offending signal on the power sources is at 81.4 Hz, the "second harmonic" of the basic interrogation frequency (the situation is complicated because the modulation is not just a simple square wave). Any extraneous frequency modulation of the microwave probe at this frequency can cause large frequency errors. One generator of this frequency on the power supplies is the microprocessor itself since it is processing chunks of data at that rate. Very careful filtering is done to keep this effect under control. Tests of complete standards include measures of both the susceptibility to this frequency on the supplies as well as the amount of signal actually there thus allowing the worst-case pulling to be calculated.

High voltage for the beam tube's electron multiplier and ion pump are developed by low power fly-back circuitry which is synchronized at a submultiple of the basic standard frequency. As mentioned earlier, the electron multiplier voltage, and consequently gain, is controlled by the microprocessor.

Software

The software for the whole standard was written in C. It is fairly complex since it includes all algorithms for the servos, signal processing, start-up, run, and general control. It also contains all the code that manages the interfaces, SCPI commands, diagnostics and logging.

Block Diagram

Fig. 2 is the basic block diagram of the cesium standard. Most of the features have already been described. Not shown are any of the monitoring connections or power supply details.

Summary

A new design for cesium beam atomic frequency standard electronics addresses most of the causes of frequency instability and error that are associated with circuity. The error budget for the various contributions is 1×10^{-14}. A microprocessor controls essentially everything so that start-up and operation is automatic without operator intervention. Complete remote control of the standard through an RS232 port, including precise frequency steering or offsets, makes it a valuable and versatile system component. Demonstrated performance with time and varying environmental conditions is excellent [1]

Acknowledgements

John Sadler contributed to a number of the algorithms, particularly the start-up algorithm, and wrote most of the fairly large amount of software. Rick Karlquist contributed to the frequency architecture, particularly to the standard's unique RF output circuits. Jim Johnson is credited with the efficient, small size, and low noise power supplies as well as the power steering logic.

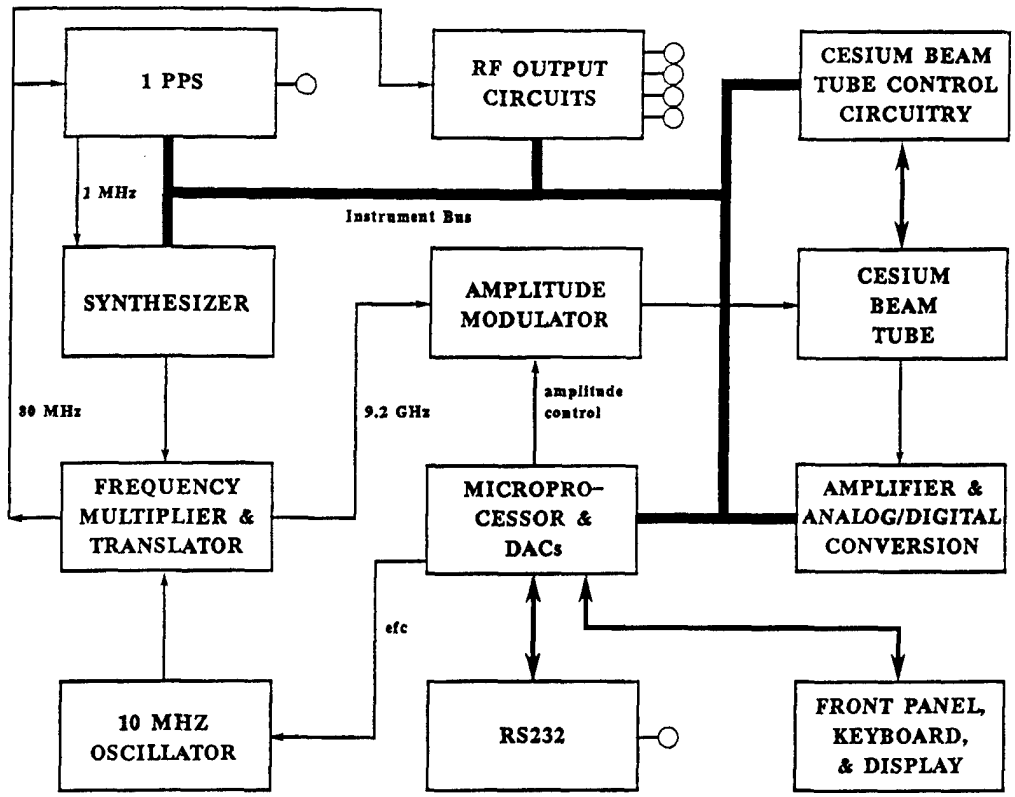


Figure 2. Basic block diagram for the new cesium standard electronics. Not shown are any power supply details or any of the monitoring connections.

References

[1] J. Kusters and J. Johnson, "A new Cesium beam standard: performance data", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[2] A. de Marchi, "Understanding Environmental sensitivity and ageing of Cesium bean frequency standards",in Proceedings of the 1st European Forum of Time and Frequency, Besancon, France, 18-20 March 1982.

[3] A. de Marchi, G.D. Rovera and A. Premoli, "Pulling by neighboring transitions and its effects on the performance of Caesium-beam frequency standards", Metrologia 20,37-47, 1984.

[4] L.S. Cutler, C.A. Flory, R.P. Giffard and A. de Marchi,"Frequency pulling by hyperfine sigma transitions in Cesium beam atomic frequency standards", J. Appl. Phys., 69(5), pp. 2780-2792, March 1991.

[5] L.S. Cutler, C.A. Flory, R.P. Giffard and A. de Marchi,"Frequency pulling in Cesium beam frequency standards due to delta M= ± 1 (sigma) transitions", in Proceedings of the 45th Annual Symposium of frequency Control, pp. 544-553, IEEE 1991.

[6] A. de Marchi, "Rabi pulling and long-term stability in Cesium beam frequency standards", IEEE Transactions on Ferroelectrics and frequency control, UFFC-34, 6, pp. 598-601, November 1987.

[7] J.H. Shirley, "Some causes of resonant frequency shifts in atomic beam machines", J. Appl. Phys., 34, pp 783-791, April 1963.

[8] C. Audoin, M. Jardino, L.S. Cutler, and R.F.

Lacey, "Frequency offset due to spectral impurities in Cesium-beam frequency standards" IEEE transactions on Instrumentation and Measurement, Vol IM-27 No.4, pp.325-329, December 1978.

[9] A. de Marchi, G.D. Rovera and A. Premoli, "Effects of servo loop modulation in atomic beam frequency standards employing a Ramsey cavity", IEEE Transactions on Ferroelectrics and Frequency Control, UFFC-34, 6, pp. 582-591, November 1987.

[10] Y. Nakadan and Y. Koga, "A squarewave F.M. servo system with digital signal processing for cesium frequency standards", in Proceedings of the 36th Annual Frequency Control Symposium, pp. 223-229, IEEE 1982.

[11] R.P. Giffard and L.S. Cutler, "A low frequency, high-resolution digital synthesizer", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[12] A. Bauch, K. Dorenwendt and T. Heindorff, "The PTB's atomic frequency standards CS2 and CSX: frequency shifts by pulling due to neighboring transitions", Metrologia 24, pp. 199-203, 1987.

[13] R. Karlquist, "A new RF architecture for Cesium frequency standards", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[14] U.S. Patent pending.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A NEW RF ARCHITECTURE FOR CESIUM FREQUENCY STANDARDS

Richard K. Karlquist
Hewlett-Packard Company, Santa Clara Division
5301 Stevens Creek Blvd., MS 52U/7, Santa Clara, CA 95052

ABSTRACT

Recent advances in cesium beam tube technology, modulation and signal processing, and control system techniques have decreased the errors due to these sources to a level comparable to the errors due to RF chains typically used in atomic frequency standards. Improved RF chains are necessary in future standards to obtain optimum performance. However, the implementation of conventional architectures has already been pushed to its practical limits, requiring substantial changes to achieve the necessary results.

A novel RF chain is described which overcomes many of the limitations of previous RF chains making possible cesium standards with unprecedented performance. The new architecture greatly reduces the four main problems in existing RF chains: excessive spurious sidebands, temperature-induced phase instability, inadequate isolation of internal and external subsystems, and significant modulator distortion.

A new cesium standard realizes higher performance by utilizing amplitude modulation in addition to frequency modulation and controls the RF level with a servo loop based on atomic interactions. The RF chain described here implements these techniques. It also has sufficient available drive power for the beam tube to allow considerably lower Ramsey cavity Q.

The new RF chain starts with a surface acoustic wave (SAW) oscillator (locked to a conventional quartz flywheel oscillator) which is split into two signals. One signal drives a relatively short frequency multiplier chain ending at the tube. The other signal drives a direct digital synthesizer (DDS) that provides the user's outputs. The frequencies of the outputs are programmable and have 120 dB of output to output isolation.

The microwave output section uses a 9192 MHz dielectric resonator oscillator (DRO), which is phase locked to the low frequency signals. Frequency modulation is applied at this point by adding a programmable frequency offset to the phase locked loop. A linear amplitude modulator has been added to support the new servo systems. This results in a spectrum with extremely low spurious sidebands.

Introduction

The RF architecture used in commercial cesium frequency standards has been relatively unchanged over the last decade or two. However, recent advances in understanding of cesium beam tubes (CBTs) and computer controlled servo systems have created a need for a corresponding improvement in RF architectures. New servo techniques using stepped frequency and amplitude modulation greatly decrease errors due to incorrect RF and C-field amplitude, CBT cavity detuning, and effects such as Rabi pulling. [1] This has created a need for a new RF architecture to implement the new modulation scheme and have reduced error contribution commensurate with the CBT improvements. The principal error mechanisms in the RF chain include insufficient spectral purity, phase instability over temperature and modulation distortion.

Requirements

Error budget

Current commercial CBTs are accurate to at best 1×10^{-13} fractional frequency error. In order that the RF chain contribute negligible additional error, it is reasonable to limit the principal errors in the electronics to an order of magnitude below this, or 1×10^{-14}, which will be assumed in the discussion below.

Modulation

The new servo techiques employ precise steps of frequency and amplitude to interrogate the central and adjacent lobes of the principal Ramsey line, plus the Zeeman resonance approximately 40 kHz. away. The frequency needs to settle to better than 1×10^{-14} in 3 ms to obtain maximum measurement time utilization. 3 ms is available without contributing additional dead time because it takes that long for transients related to CBT atomic transit time to die out. [1]. The amplitude steps require about 2 dB of amplitude modulation capability and compensation of CBT cavity detuning loss requires about 4 dB additional range (based on a Q of 900 over 0 to 50 degrees C).

The frequency and amplitude stepping imposes a settling time requirement that arises from the need to minimize any dead time in the frequency measurement process. Following a step change in the rf frequency, the output of a

0-7803-0476-4/92 $3.00 © 1992 IEEE

typical CBT shows a transient lasting about 3.5 ms caused by transit time effects in the cavity and detector. This transient is allowed to settle out before the tube output is measured to determine the frequency error. The microwave signal phase must have settled accurately in about 1.5 ms, however, because the average time taken for an atom to pass from the entrance of the cavity to the detector is about 2ms. Detailed calculation shows that if the slowest time-constant associated with phase-locked loops in the frequency modulation path is shorter than 150 μseconds, the error due to settling will be negligible. Accurate settling for these purposes is defined as frequency settling to within 1×10^{-14} and amplitude settling to within .01 dB. [2].

Spectral purity

Any sideband power on the interrogating signal that is not balanced will cause the servo to move off the center of the atomic resonance. This applies to both phase noise and discrete spurious sidebands (spurs.) If the sidebands are perfectly balanced, they will cause equal and opposite pulling and cancel out. However, it is poor engineering practice to depend on perfect balance because there are many AM-to-PM or PM-to-AM conversion processes that can unbalance the sidebands. Mistuning of the CBT cavity (or auxiliary filters) induces PM to AM conversion by slope detection. These effects worsen at increasing offset from the carrier and with increasing loaded cavity Q. It is impractical to keep the cavity tuning under servo control because present tuning diodes have insufficient Q and/or variation of capacitance.

For error budget purposes, it is convenient to divide the spectral purity requirements into three categories according to the offset from the carrier: the spurs outside the Zeeman band, the spurs in the Zeeman band, and close-in spurs related to the modulation processes. For spurs outside the Zeeman band (more than 130 kHz offset from the carrier) the allowable unbalanced sideband level for an error contribution of 1×10^{-14} varies linearly with offset frequency from -37 dBc. at 130 kHz to -20 dBc at 10 MHz offset. However, it is good engineering practice to keep all spurious sidebands down to -40 dBc to prevent them from competing with the carrier for power. For spurs inside the Zeeman band, an upper bound has been established that corresponds to about -70 dBc for a pulling error of 1×10^{-14} [3,4].

For spurs that are coherent with the rate of measurements (about 82 Hz in the new design to be described below), the effect of coherent modulation will cause errors that correlate from measurement to measurement instead of averaging to zero, which is the case for 120 Hz spurs. A mechanism by which the RF chain could be modulated at the measurement rate is the presence of 82 Hz noise on the power supply. This can be caused by periodic switching of bus drivers, multiplexers, etc. by the microprocessor. The effect is software dependent and can be minimized by proper coding. In the worst case, a fractional frequency deviation of 1×10^{-14} at the measurement rate will translate to a 1×10^{-14} error for the standard. For a 9.2 GHz carrier, this corresponds to an absolute peak deviation of 92 μHz, so the modulation index (β) for incidental 82 Hz modulation must be held to about 1×10^{-6} . This corresponds to an 82 Hz bright line at a maximum level of -126 dBc. Note that since the 10 MHz quartz flywheel oscillator is ultimately multiplied by a factor of 928, any 82 Hz spurs on the 10 MHz signal will increase by 20 Log 928 = 59 dB so they must not exceed a level of -185 dBc at 10 MHz.

Phase stability

When the standard is required to operate in environments with temperature transients, phase instabilities in the RF chain can result in frequency errors in the standard. The are also self-induced temperature transients during warm-up. These transients last longer than the time it takes for modern commercial cesium standards to warm up and operate to high accuracy. [5] This error mechanism arises because angular frequency, by definition, is the time rate of change (derivative) of phase. If phase is a non-constant function of temperature, and temperature is changing with time, then phase will change with time. This represents a frequency offset. If the standard takes 1 hour to reach thermal equilibrium resulting in an internal temperature rise of 15 degrees, then there is a temperature ramp of about .004 degrees per second (taking a linear approximation.) A 10 MHz output circuit with a temperature coefficient of phase of 150 μradians per degree C would drift at a rate of 0.6 μradians per second. Since there are 2π radians per cycle, this is equivalent to a frequency offset of 0.1 μcycles per second, or 0.1 μHz, which is a fractional frequency error of 1×10^{-14}. The phase stability requirement of 150 μradians per degree C at 10 MHz can also be expressed as a delay stability of 2.5 ps per degree C, which is independent of frequency. This would be 125 ps over 0 to 50 degrees C.

Conventional architecture

In order to provide a proper context for the description of the new architecture, a brief review of prior art will be presented. Figure 1 shows the block diagram of a representative example (Hewlett-Packard model 5061B.) A 10 MHz quartz oscillator is divided by two to obtain 5 MHz to which phase modulation is applied. This is multiplied to 90 MHz with transistor multipliers which then drive a step-recovery-diode (SRD) comb generator to produce harmonics of 90 MHz. The 102nd harmonic at 9180 MHz is used as a carrier that is phase modulated at a rate of about 12.6 MHz to produce a sideband at about 9192.6 MHz. A filter attenuates the carrier and spurious sidebands

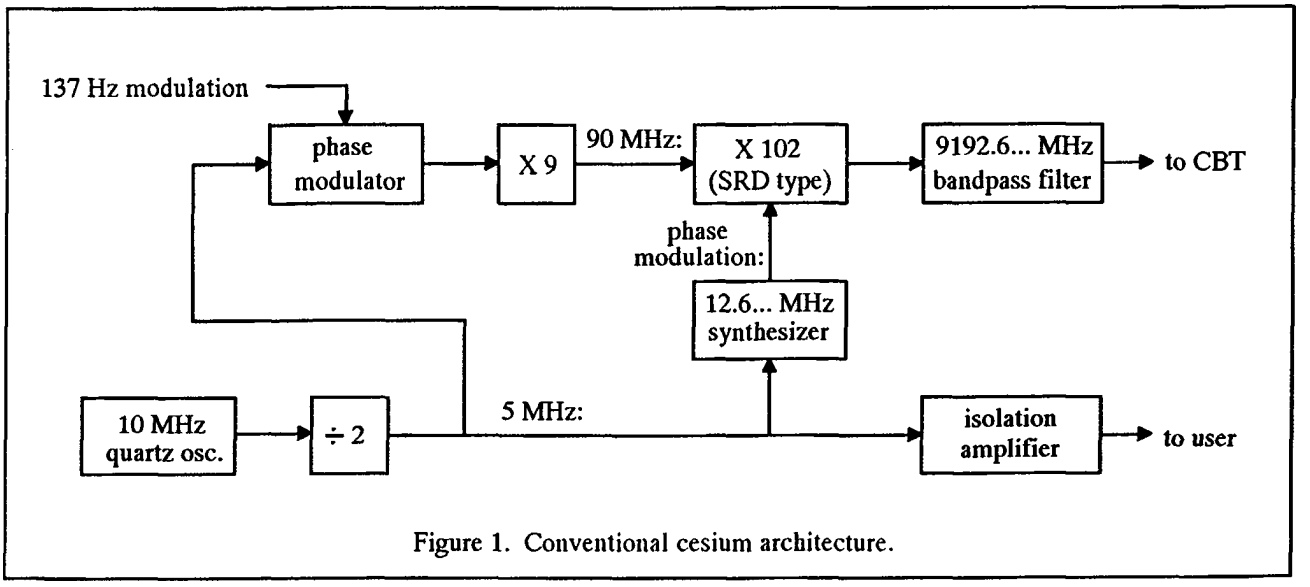


Figure 1. Conventional cesium architecture.

leaving just the 9192.6... MHz sideband to excite the cesium resonance in the CBT. The sideband is generated by modulating the SRD bias voltage with a high resolution synthesizer locked to the quartz oscillator. The center of the resonance is found by detecting the response of the CBT to the audio FM with a sychronous demodulator. The quartz oscillator signal is buffered and supplied as the user's output.

Most cesium standards and some non-cesium standards are similar to this. The deficiencies of this approach are difficulties with spurious sidebands, insufficient power output, no convenient or accurate way to do amplitude modulation, excessive phase modulation distortion, poor phase stability with temperature, and undesireable manufacturability and reliability characteristics.

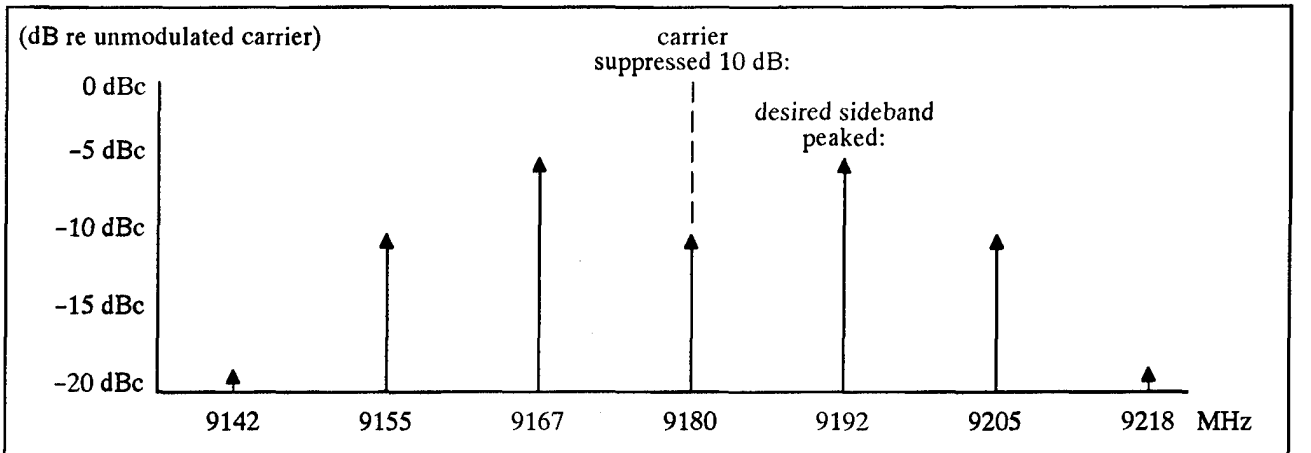


Figure 2. Typical microwave spectrum before filtering (conventional architecture)

Microwave subsystem

Figure 2 shows the microwave spectrum after the SRD multiplier and before the microwave bandpass filter. 0 dB. refers to the unmodulated carrier level. The modulation index (β) is set to the value (1.85) that maximizes the first sidebands (at 9192.6... and 9167.4... MHz.) Under this condition, the desired sideband is down 5 dB from the unmodulated carrier. The carrier is suppressed 10 dB below its unmodulated level. This leaves the carrier down only 5 dB relative to the desired sideband. The second harmonic sidebands (25.3 MHz from the carrier) are also down only 5 dB relative to the desired sideband. The carrier and second harmonic sideband must be suppressed an additional 14 dB in order to achieve an error contribution of 1×10^{-14}. This requires a high-Q auxiliary filter made of

invar in combination with a high-Q CBT cavity. The auxiliary filter adds to the complexity and the high-Q CBT cavity requires temperature compensation.

Manufacturability

The conventional architecture dictates extensive use of waveguide construction. This leads to practical problems with corrosion, contact resistance, and difficult -to-make adjustments. Also, there are many opportunities for RF to leak out and the atmosphere to leak in. the It is difficult to prevent stray 9.2 GHz fields from impinging on the drift region in the CBT and cause errors due to extraneous quantum transitions. Atmospheric leakage can cause sensitivity to humidity and exascerbate corrosion problems in areas with air pollution.

Modulation

The architecture of Figure 1 does not include amplitude modulation capability, because conventional servo systems did not use it. Adding it would require insertion of a waveguide-type PIN diode modulator between the SRD multiplier and the bandpass filter. This is inconvenient and runs the risk of thermal transients in the PIN diodes causing phase transients that result in frequency errors. The analog phase modulator would have to have an impractically low second harmonic distortion of -120 dBc in order to hold its error contribution below 1×10^{-14} with a conventional servo system. [1]

Phase stability

The conventional architecture suffers from poor phase coherence between the signal at the CBT and the user's output. This is due to phase shifts in the initial multiplier stages and the long chain of isolation amplifiers necessary to prevent output loading from affecting the accuracy of the standard. The initial doubler from 5 to 10 MHz must have considerable filtering to attenuate spurious output signals at 5 and 15 MHz before driving the next stage, because these would result in 5 MHz sidebands on the CBT drive signal. In order to keep these sidebands down to -40 dBc after multiplication to 9.2 GHz, it is necessary keep the 5 and 15 MHz spurious signals down to -90 dBc. Phase drift in the filter results in a frequency shift when the ambient temperature is changing. Additional phase drift occurs in the path from the quartz oscillator to the CBT.

New RF chain

The new architecture is shown in Figure 3. [6] There are four main areas of difference. The 9192.6 MHz signal for the CBT is generated by a dielectric-resonator oscillator (DRO) running directly at that frequency. The signal path now splits between the CBT chain and the output chain at the 320 MHz point rather than the 10 MHz point, and a direct-digital-sythesizer (DDS) is used to generate the user's outputs. The modulation type and technique is different to accommodate the needs of the new servo loops.

The new architecture, like the conventional one, retains a 10 MHz quartz oscillator that is controlled by the servo loop. However, it is more convenient to think of the signal flow originating with the 640 MHz surface-acoustic-wave (SAW) oscillator which drives a frequency divider to generate 320 MHz. (A SAW oscillator frequency of 640 MHz rather than 320 MHz was chosen because 640 MHz SAW resonators were readily available, not for any fundamental reason.) The 320 MHz signal splits into two paths, one leading to the CBT and the other leading to the user's output. To excite the CBT, the 320 MHz signal is multiplied by 29 to produce a reference at 9280 MHz. This is used in an offset phase locked loop to tune the DRO to 9192.6... MHz, which is the difference frequency between the 9280 MHz reference and the upconverted agile synthesizer (near 87.4 MHz). The output of the DRO is amplitude modulated by the servo loop before driving the CBT. To produce the user's output, the 320 MHz signal is also divided by 4 to clock a DDS at 80 MHz. This produces sine wave outputs at 5 and 10 MHz after only minimal filtering. The short term stability of the SAW oscillator is inadequate for an atomic frequency standard, hence it is locked to the 64th harmonic of the 10 MHz quartz oscillator to clean up close-in phase noise.

Microwave subsystem

A detailed block diagram of the microwave subsystem is shown in Figure 4. The 320 MHz is multiplied using an SRD and the 29th harmonic is extracted by a 9280 MHz bandpass filter. This is mixed with the DRO frequency to produce an 87 MHz intermediate-frequency (IF). A limiting IF amplifier is used to get a reliable input for the phase detector that compares the IF frequency with the 87.4 MHz reference from the agile synthesizer. A PLL with 1 MHz bandwidth tunes the DRO to the correct frequency. The output of the DRO is amplitude modulated by the servo loop using a mixer before being sent to the CBT.

The DRO has a power output of over 10 mW, which allows it to drive the local-oscillator (LO) ports of the two mixers. Since the SRD drives the RF port of a mixer, only a small amount of SRD output power is needed. Only -20 to -10 dBm is needed to overcome the IF noise floor. Hence a simple, non-critical, low-efficiency SRD circuit can be used. The limiting IF can tolerate wide variations in SRD output. Also, the 9280 MHz filter is not critical; only minimal filtering of the comb spectrum is necessary to keep from overloading the mixer with the combined power of dozens of harmonics. Nearby spurious signals

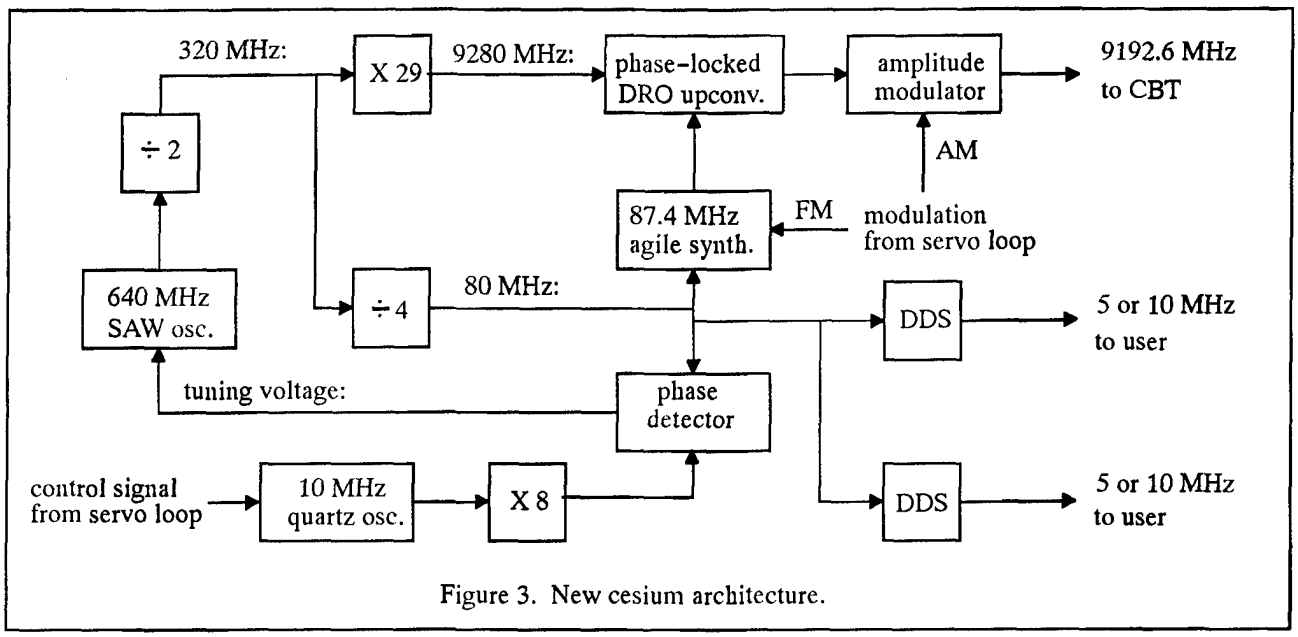


Figure 3. New cesium architecture.

such as the 28th and 30th harmonic can be tolerated because they are outside the bandwidth of the PLL. Insertion loss of the filter is also non-critical because it can be made up with extra gain in the IF amplifier. Hence it is acceptable to implement the filter with coupled-line microstrip construction.

The DRO also drives the LO port of the modulation mixer, which is a conventional single-balanced diode mixer with a DC-coupled IF port. Since the DRO is driving the diodes hard enough to act as switches, there is little AM-to-PM conversion or thermal effects in the diodes. There is an inherently linear up-conversion process from the IF port to the RF port so the same mixer can serve as the stepped amplitude modulator and also the average power controller (to compensate for RF amplitude changes due to CBT cavity drift) without interaction. Also, if the IF port is driven with a 50 ohm source impedance, then the RF port will automatically have very close to a 50 ohm source impedance as well. This contrasts to the extreme difficulty of maintaining low VSWR at the ports of a PIN diode modulator.

Another feature of the DRO is that sidebands above 1 MHz on the 320 MHz or 87.4 MHz inputs are attenuated by the PLL. This relaxes the spec for 10 MHz sidebands on the 320 MHz signal and 21.9 MHz sidebands on the 87.4 MHz signal. This coupled with the fact that the

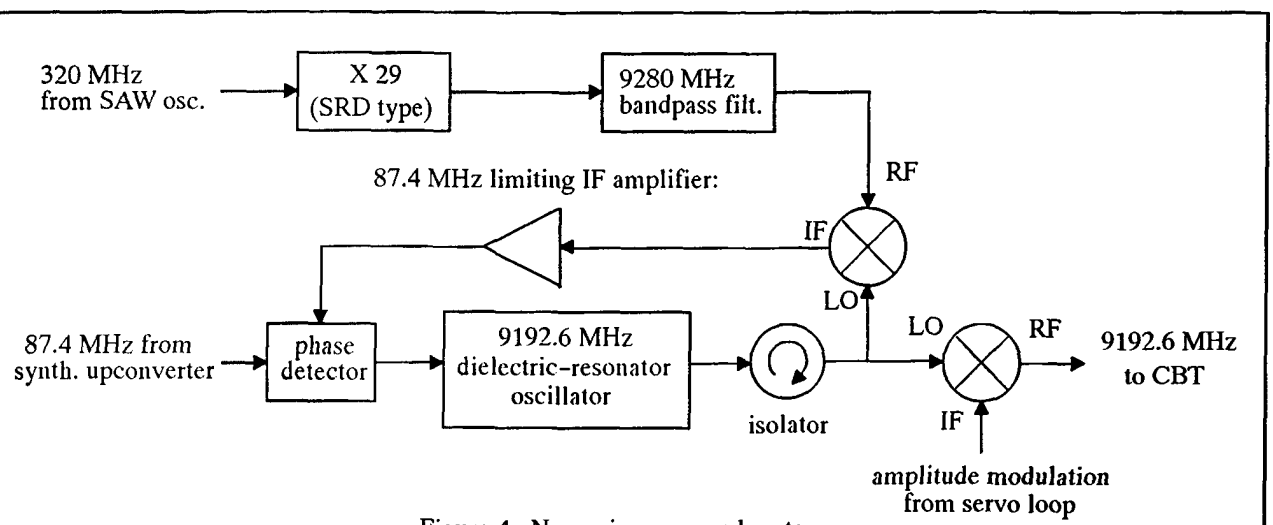


Figure 4. New microwave subsystem.

DRO frequency is used to excite the atoms directly (rather than with a sideband) eliminates spectral clutter.

Saw oscillator

In designing this part of the architecture, a number of alternatives were investigated. It would be possible to multiply directly to 320 MHz using five cascaded frequency doublers. However, good engineering practice would dictate that all spurious sidebands (such as 10 MHz) be held to 40 dBc after the 320 MHz is multiplied by 29, which requires 70 dBc at 320 MHz. The 10 MHz sidebands on the output of the first doubler would be enhanced by 24 dB by the subsequent four doublers hence requiring 95 dBc at 20 MHz. This would require extensive filtering and multiple shielded chambers.

Starting with a signal at 320 MHz reduces parts count and simplifies shielding to two chambers: one for the SAW oscillator and one for everything else.The noise floor of the SAW oscillator is lower than the multiplied-up noise floor of the quartz oscillator. Any spurious sidebands on the quartz oscillator more than a few kHz. away from the carrier, such as 200 kHz from the switching power supply, are stripped off by the PLL. Any drift in the phase of the quartz oscillator with respect to the SAW oscillator will affect the CBT drive signal and the user's output equally and hence drop out of the error budget as a common mode error.

Output subsystem requirements

A fresh look was taken at the output amplifier problem and a new set of requirements generated that were commensurate with the high performance in the rest of the standard. It was decided that two isolated high performance outputs were necessary. One could be permanently connected to the user's distribution system. The other one would then be available for testing, development work, phase comparison, etc. In order to make this scheme useful, the two outputs should be isolated by at least 100 dB so that connecting test equipment to one does not appreciably affect the phase of the other output that is being distributed. Also, the phase of the two outputs should track very closely. The outputs should be extremely well behaved to minimize interactions with the distribution system: the harmonic distortion should be below 1% (-40 dBc), the source impedance should be 50 ohms, not just at the output frequency, but harmonics as well. This will assure that any reflected signals from the user's equipment including distortion generated by phase detector inputs will be absorbed by the output rather than reflected back to the user. This is a common problem in existing hardware. The output circuit should have enough dynamic range to avoid voltage clipping when driving an open circuit and current clipping when driving a short circuit. A short circuit load would arise if an electrical quarter wave of coax (about 5 meters at 10 MHz) were connected from the output to a high impedance load. Finally, as mentioned above, the delay (phase) drift due to temperature should ideally be held to 2.5 ps per degree C or 125 ps for an ambient change of 50 degrees, to keep the phase drift error below 1×10^{-14}.

Output subsystem design

Figure 5 shows a block diagram of an output subsystem that meets these requirements. The 80 MHz clock signal is divided by 8 or 16 to produce a 10 or 5 MHz. digital ramp. The ramp drives a DAC to produce a 4-times oversampled sine wave as shown in Figure 6. If the DAC were perfect, there would only be outputs at 5, 35, 45, 75, 85, etc. MHz (for a 10 MHz output). In practice, suppression to about -50 dBc is reliably obtained. A first order hold after the DAC reduces the amplitude of the 70 and 90 MHz signals to -17 and -19 dBc respectively before filtering. (See Figure 7) This allows much less filtering than would be necessary for a square wave, reducing phase drift in the output filters. Extremely high speed, temperature-compensated logic with 500 ps propagation delay is used to keep delay drift below 125 ps over temperature in the digital sections. When the output is at 5 MHz, the first harmonic requiring filtering is at 35 MHz. This allows a common reconstruction filter to be used for either frequency by designing it to pass frequencies below 10 MHz and cut frequencies above 35 MHz. Hence the output frequency can be programmed to be either 5 or 10 MHz simply be changing the divide ratio. A dissipative filter is used to provide a 50 ohm output impedance at the output frequencies and all harmonics. A continuously-counting auxiliary divide-by-16 is used to assure phase coherency between the two outputs even if the output frequencies are reprogrammed. The digital nature of the frequency dividers and DACs provides for high reverse isolation with low complexity.

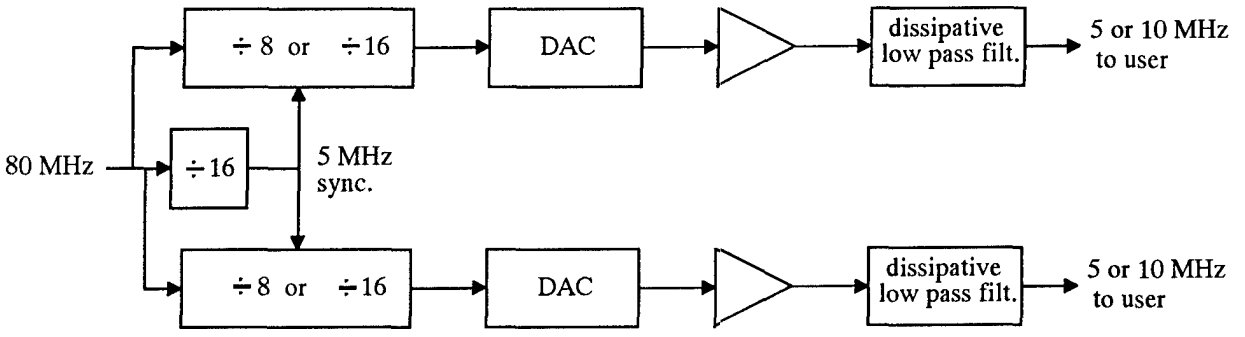


Figure 5. New output subsystem.

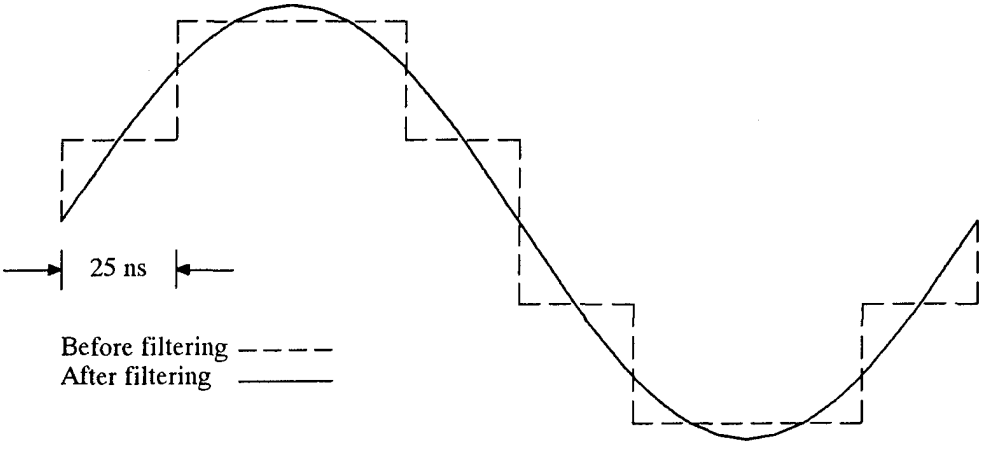


Figure 6. Output waveform before and after reconstruction filter (5 MHz output).

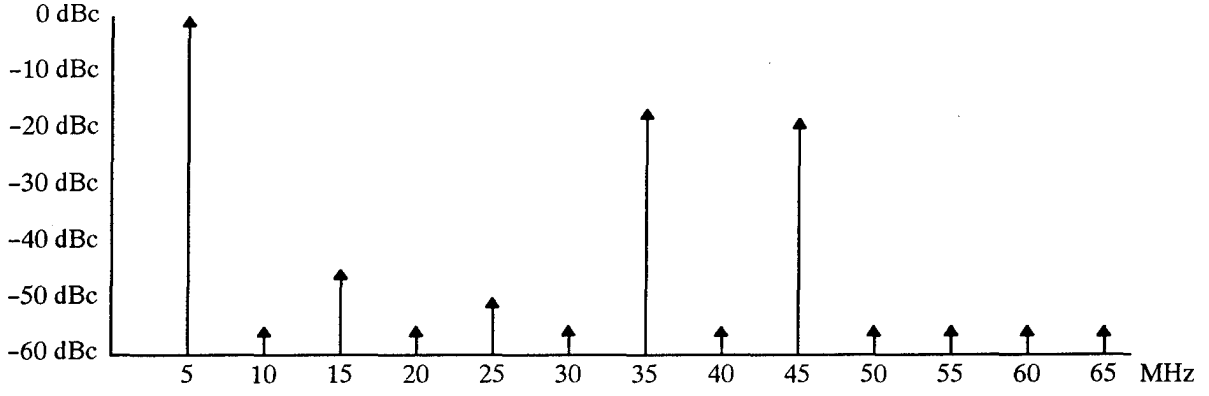


Figure 7. Output spectrum before reconstruction filter (5 MHz output).

Synthesizer upconverter

A high resolution agile synthesizer [2] with an output from 125 to 175 kHz (nominal frequency, about 131 kHz) is used to generate the frequency modulation needed by the servo loop. [1] The synthesizer upconverter raises the synthesizer frequency to around 87.4 MHz, where it is upconverted again by the microwave subsystem described above. In both cases, it is not multiplied, which would require additional resolution by the DDS.

Figure 8 shows the block diagram of the synthesizer upconverter. A 21.837 MHz VCXO with a tuning range of at least .1% is quadrupled to 87.368 MHz and sampled at 2.5 MHz by an SRD sampler. This has the effect of mixing it with the 35th harmonic of 2.5 MHz (87.5 MHz) to produce an IF frequency of 125 to 175 kHz. A phase detector compares the IF with the DDS signal to phase lock the VCXO.

Because the VCXO loop has only 3 msec. to settle to within 92 μHertz, a rather wide bandwidth of 10 kHz is necessary. The most severe requirement occurs when the DDS jumps about 40 kHz after interrogating the Zeeman line for the C-field servo. This requires very careful analysis of loop acquisition and settling time characteristics. The maximum loop bandwidth is limited by spurious crystal responses located a few tenths of a percent above series resonance. These degrade loop stability by adding extraneous gain and phase shift to the VCXO tuning voltage input.

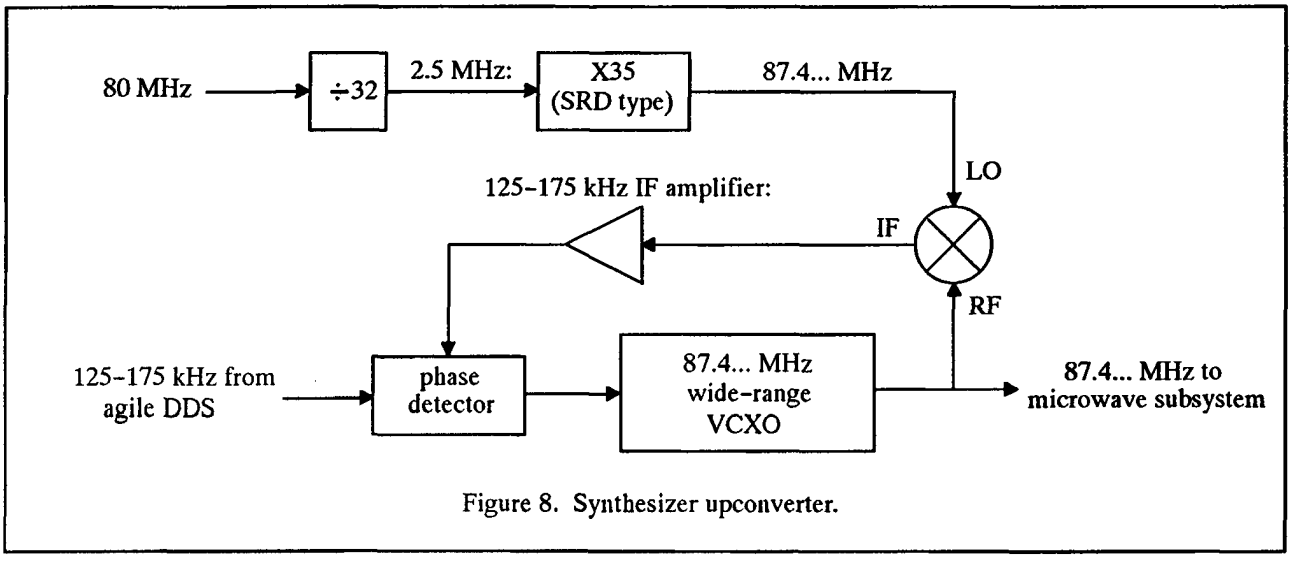


Figure 8. Synthesizer upconverter.

Summary

A design has been shown which meets the objective of keeping errors due to RF problems such as spectral impurities, modulation distortion, and phase instability below a budget of 1×10^{-14}. The resulting RF chain is more manufacturable and repeatable than previous ones. Demonstrated performance with time and varying environmental conditions is excellent [5].

ACKNOWLEDGEMENTS

L. Cutler suggested the microwave mixer technique for applying amplitude modulation. R. Giffard contributed to the design of the power conditioning and the microwave system design. They also acted as technical advisers. J. Kusters assisted with writing and presenting this paper. T. Parisek did the mechanical design of the RF shielding. J. V.Brzeski characterized some of the microwave hardware and contributed to the synthesizer upconverter.

REFERENCES

[1]: L.S. Cutler and R.P.Giffard, "Architecture and algorithms for a new cesium beam frequency standard electronics", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[2]: R.P.Giffard and L.S. Cutler, "A low-frequency, high resolution digital synthesizer", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[3]: J.H. Shirley, "Some causes of resonant frequency shifts in atomic beam machines", J. Appl. Phys., 34, pp. 783-791, April 1963.

[4]: C. Audoin, M. Jardino, L.S. Cutler, and R.F. Lacey, "Frequency offset due to spectral impurities in cesium-beam frequency standards" IEEE Transactions on Instrumentation and Measurement, Vol IM-27 No. 4, pp. 325-329, December 1978.

[5]: J. Kusters and J. Johnson, "A new Cesium beam standard: performance data", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[6]: R. Karlquist, U.S. Patent Pending, filing date Aug. 1, 1991.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A NEW CESIUM BEAM FREQUENCY STANDARD PERFORMANCE DATA

John A. Kusters
James L. Johnson

Hewlett-Packard Co.
Santa Clara, California 95052-8059

Introduction

Recent papers have discussed the principles of a new architecture for cesium beam frequency standards.[1-4] The design goal of the new architecture was to develop an atomic frequency standard that is essentially independent of environmental effects. Reduction in Ramsey pulling due to cesium beam tube design effects was discussed in 1991[1,2]. With the addition of electronics that address power shift, C-field effects, and Rabi pulling, the new cesium standard shows the expected reduction in overall instabilities due to environmental factors and electronics errors. This paper reports actual performance data obtained from a number of standards designed incorporating the new architecture as measured by various measurement laboratories. The data substantiates the design theories, plans, and goals to produce a frequency standard whose sensitivity to environmental factors is greatly reduced when compared to existing standards.

Background

Previous cesium beam frequency standards generally exhibit a substantial response to environmental influences, such as temperature, pressure, humidity, and external magnetic fields[5-9]. An example of this is shown in Figure 1 where data collected over a year shows that frequency changes measured in several frequency standards maintained in a constant temperature environment, show a strong correlation with measured humidity.

Responses to environmental effects have become the major limitation on overall accuracy and on long-term stability in atomic frequency standards. This in turn requires metrology laboratories using these devices to design elaborate physical facilities to house the instruments in an attempt to minimize the environmental effects.

In addition, warm-up of a cesium frequency standard requires 24-36 hours of stabilization time to achieve final stability. This is a consequence of the typical cesium beam tube design, which conventionally consists of a copper Ramsey cavity weighing more than 1 kg. with a cesium oven at one end operating at 90-130°C, and a hot-wire ionizer at the other end operating in excess of 900°C. Typical thermal response times of a cesium tube are on the order of 6-10 hours, resulting in overall thermal equilibrium times of at least 24 hours.

Primarily through an expanded understanding of the underlying causes of the cesium standard's sensitivity to outside environmental factors, and extensive computer modeling of the atomic physics and of the electronic architecture, we became convinced that with the proper design, environmental factors could be greatly reduced. At the same time this should result in a more useful instrument that requires no special operating environment, and that should be fully operational meeting all specifications within 30 minutes of power-up. The latter condition is especially desirable for field transportable, fast set-up metrology and calibration facilities.

Consequently, a multi-facility, multi-year program was undertaken to develop a cesium standard based on the new architecture.

Environmental Testing

As a part of the investigation of the new architecture and as a qualification process for instruments developed during the project, the response of the new cesium beam frequency standards to external effects had to be accurately determined.

0-7803-0476-4/92 $3.00 © 1992 IEEE

Reference Standard

The fundamental requirement for determining the response of any standard to an outside influence is to have an appropriate reference standard against which to measure. Given the expected thermal response time of the cesium beam tube and other components in the standard, full temperature equilibrium takes about 18 hours to achieve. This dictated that temperature changes in the environmental chamber should occur no more often than once every 24 hours. This in turn requires that the reference standard have sufficient stability over extended time periods to be able to measure the environmental effects. Early in the testing program, we determined that our house frequency standard, composed of several HP5061 cesium standards, did not have the needed accuracy or stability to be used as a reference standard. Given its sensitivity to temperature and humidity, the expected stability of the house standard over several days of measurement was on the order of a part in 10^{13}. During the first temperature run, no change in frequency of the new standard was observed within the allowable measurement errors of the house standard.

The first challenge was to create a reference standard that was sufficiently stable and had sufficient intrinsic accuracy to be used as a reference. As a result, three of the new standards were constructed and connected as an active ensemble as shown in Figure 2. In this configuration, the controller drives the three standards to operate at exactly the same frequency and phase, producing a frequency that is the mean frequency of the three standards, and where the absolute sum of the corrections applied to the individual standards are at a minimum. Operating in this manner, for identical standards, the overall accuracy and stability improve as the square root of the number of units in the ensemble, while the phase noise spectral density (in dB) improves as 10 log(number of units). This results in a reference standard whose expected accuracy and stability is a factor of 1.7 greater than that of a single unit.

Testing at facilities outside of Hewlett-Packard was done using various models of hydrogen masers. Up to sampling times in the vicinity of 10^5 seconds, this proved adequate. Beyond this, corrections had to be made to the data to account for the drift of the maser used as the reference frequency standard.

Temperature Sensitivity

During early temperature measurements, analysis of the data using correlation techniques showed that there was a frequency change response that occurred every 24 hours, the same period as the temperature cycling. The correlation also showed peaks with time periods less than 24 hours, indicating that some other process was occurring. Given that at this time we were looking for effects on the order of parts in 10^{14}, after a difficult and time-consuming process the phase detector used to measure the phase difference between the unit under test and the reference standard was found to be temperature sensitive and varying with the room air conditioning. After also environmentally controlling the phase detector, the upper bound of the temperature sensitivity of the standard was determined to be 1×10^{-15}/°C.

Further testing at JPL was performed to refine the data. This facility was used because of the excellent environmental chambers used, the existence of several hydrogen masers as reference standards, and the outstanding work done by this group in evaluating and understanding the environmental effects seen in our older standards[8,9]. Figure 3 is a plot of the Root Allan Variance of the standard under test. This is a conventional method of expressing oscillator stability as a function of averaging time[10]. In this test, temperature was cycled between 15° and 35°C. once per day. If there had been a significant response to temperature, this would have shown as a peak in the stability curve at a period corresponding to 24 hours. Given that there is none, the upper bound of the temperature sensitivity over the 20°C temperature range was determined to be less than 1×10^{-14}, or correspondingly, less than 5×10^{-16}/°C.

Humidity Sensitivity

Early testing for humidity effects was performed in our facility with no detectable results. An extended run was done using NRL facilities. Figure 4 is a plot of the relative humidity and observed frequency. During this run, the temperature was also being cycled from -17° to 57°C. exceeding the design specification at both ends. Figure 5 is a plot of the correlation function between relative humidity and frequency. Two peaks are shown, but they are too weak to draw any conclusion. NRL data indicates that both temperature and humidity effects are bounded by the measurement uncertainty of 1×10^{-14}.

Pressure Sensitivity

No sensitivities to changes in ambient pressure were seen within the limits of the reference standard. Tests were run over pressure ranges equivalent to 0 - 12,000 m. altitude. An interesting sidelight is that although most frequency standards are specified in terms

of altitude change, the new unit is sufficiently stable that relativistic effects dominate all other effects in a change in altitude.

Magnetic Field

No frequency shifts within the limits imposed by the reference standard (in this case, a hydrogen maser) were seen in a ±2 gauss field. Data was taken in both DC and 55 Hz magnetic fields in all three axes.[11]

Further Characterization

Time Domain Stability

Figure 6 is a plot of the time domain stability in the range of 0.01 to 20 seconds. Figure 3 is a similar plot extending the sampling time out to 10^6 seconds. Of interest here is that at roughly 10^5 seconds, the cesium standard uncertainty crosses the locus for the hydrogen maser used as the reference standard. Freedom from environmental effects results in a flicker-floor substantially below that of conventional cesium frequency standards.

Warm-up Time

Figure 7 is a plot of warm-up time and frequency offset from a hydrogen maser as measured by NIST. From this plot, measured on an instrument that was flown from Santa Clara to Denver, driven to NIST Boulder, then plugged in with no stabilization time, the warm-up time to meet specifications was 10 minutes, 16 seconds.

Phase Noise

Characteristics of the new standard in the frequency domain is shown in the phase noise plot in Figure 8. Apparent here is the improvement in phase noise especially in regions closer to the carrier, from 1 Hz to 1000 Hz offset.

Other Effects

Tests for all other effects, including EMI and radiated and conducted susceptibilities were conducted with no observable results within the measurement uncertainty of the reference standard. Tests also included verification to conformance to MIL-STD-461C, EC-92, and MIL-T-28800D.

Conclusions

The experimental data indicates that it is possible to design and construct frequency standards that are environmentally insensitive, at least to the level of 1 part in 10^{14}. A sound understanding of the physics of processes involved also leads to a rapid warm-up of the cesium instrument.

Acknowledgments

The authors acknowledge the technical direction and continued contributions of Drs. Len Cutler and Robin Giffard of HPL, and the joint HPL and Santa Clara R&D team for the development of the new standard. We also gratefully acknowledge the contributions and assistance of Dr. Richard Sydnor and associates at the Jet Propulsion Laboratory, California Institute of Technology., Dr. Joseph White and associates at the Naval Research Laboratory, Dr. Gernot Winkler and associates at the United States Naval Observatory, and David W. Allan and associates at the National Institute of Standards and Technology for making the majority of the highly precise measurements of the new standard contained in this report.

References

[1] Cutler, L.S., *et.al.*, "Frequency Pulling in Cesium Beam Frequency Standards due to ΔM=±1 (Sigma) Transitions," *Proceedings of the 45th Annual Symposium on Frequency Control*, pp. 544-553, IEEE Publication CH2965-2, 1991.

[2] Mueller, L.F., *et.al.*, "A New High-Performance Cesium Beam Tube Compensated for Ramsey Pulling," *Proceedings of the 45th Annual Symposium on Frequency Control*, pp. 554-559, IEEE Publication CH2965-2, 1991.

[3] Cutler, L.S. and Giffard, R.P., "Architecture and Algorithms for New Cesium Beam Frequency Standard Electronics," *Proceedings of the 1992 IEEE Frequency Control Symposium.*

[4] Karlquist, R., "A New RF Architecture for Cesium Frequency Standards," *Proceedings of the 1992 IEEE Frequency Control Symposium.*

[5] DeMarchi, A., "New Insights into Causes and Cures of Frequency Instability (Drift and Long-Term

Noise) in Cesium Beam Frequency Standards," *Proceedings of the 41st Symposium on Frequency Control*, pp. 54-58, NTIS Document No. ADA-216858, 1987

[6] DeMarchi, A., "Rabi Pulling and Long-Term Stability in Cesium Beam Frequency Standards," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, UFFC-34(6), pp. 598-601, November, 1987.

[7] Karuza, S.K., *et.al.*, "Determining Optimum C-Field Settings that Minimize Output Frequency Variations in Cesium Atomic Frequency Standards," *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 385-400, 1989.

[8] Sydnor, R.L., "Environmental Testing at the Jet Propulsion Laboratory's Frequency Standards Laboratory," *Proceedings of the 43rd Annual Symposium on Frequency Control*, pp. 289-295, IEEE Publication 89CH2690-6, 1989.

[9] Sydnor, R.L., *et.al.*, "Environmental Tests of Cesium Beam Frequency Standards at the Frequency Standards Laboratory of the Jet Propulsion Laboratory,", *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, pp. 409-420, 1989.

[10] Allan, D.W., "Statistics of Atomic Frequency Standards," *Proc. IEEE*, vol. 54, pp. 221-230, February, 1966.

[11] Private communication, Jet Propulsion Laboratory

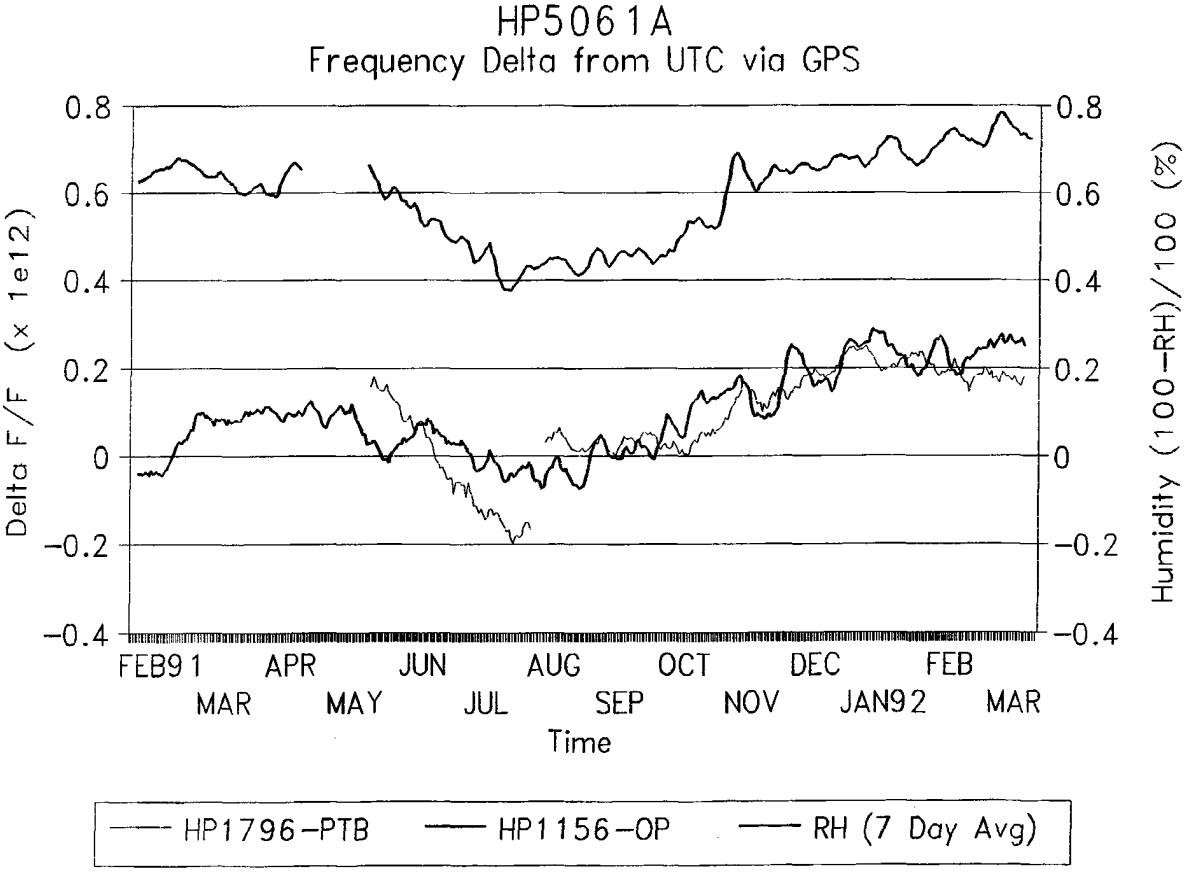


Figure 1: Upper trace - scaled relative humidity. Lower traces, - offset from UTC for two cesium standards as determined by common view GPS. (Data courtesy of HP Geneva Time & Frequency Laboratory)

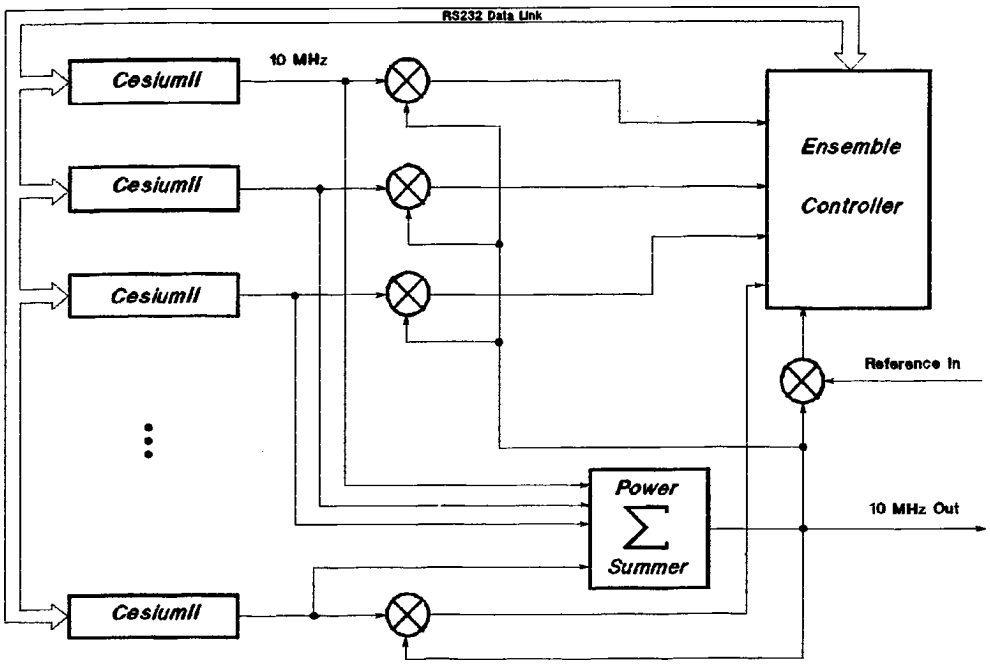


Figure 2: Block Diagram, real-time ensemble of cesium standards.

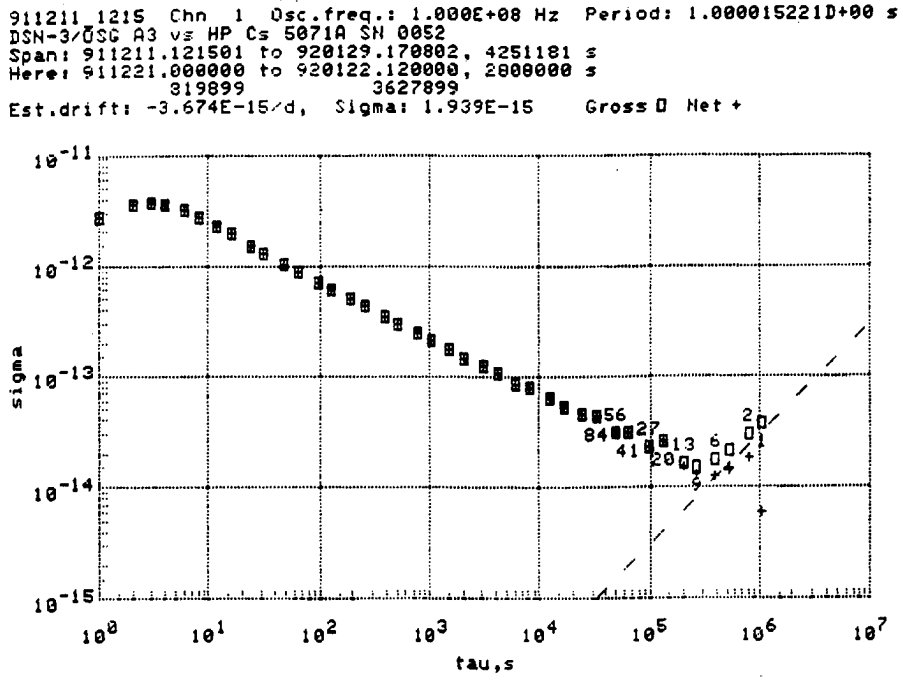


Figure 3: Root Allan Variance as determined by comparison to a hydrogen maser. Dashed line in lower right corner indicates expected instability in the hydrogen maser.
(Data courtesy of Jet Propulsion Lab, California Institute of Technology)

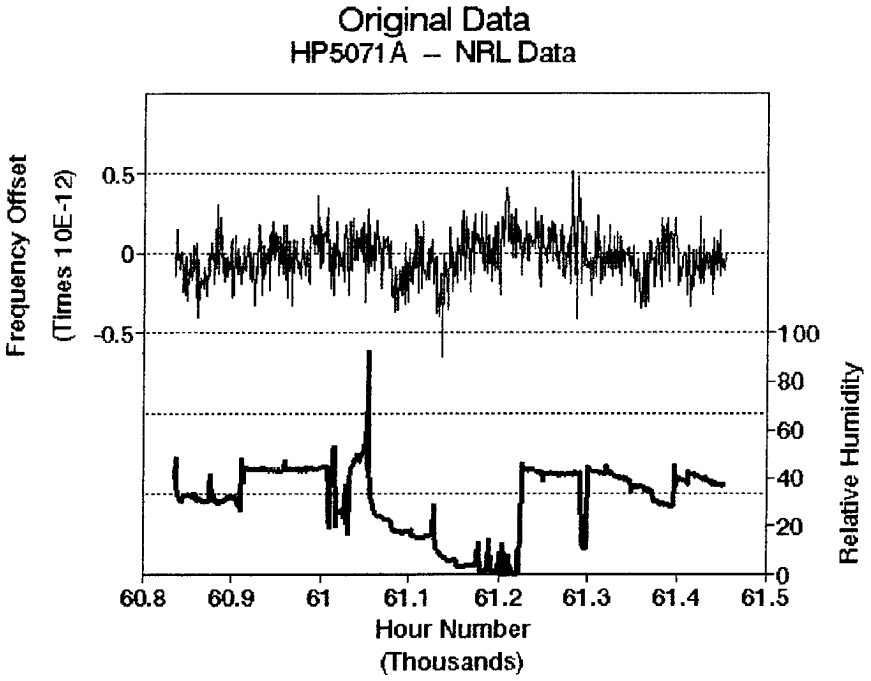


Figure 4: Upper trace - frequency offset from UTC.
Lower trace - measured relative humidity.
(Data courtesy of Naval Research Laboratory)

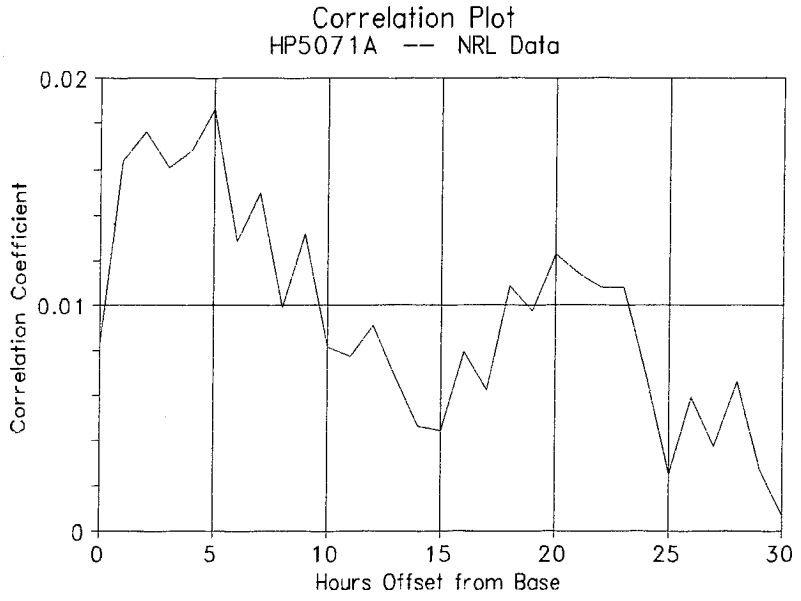


Figure 5: Correlation of data in Figure 4. Extremely weak peaks are observed at 5 and 20 hours offset.

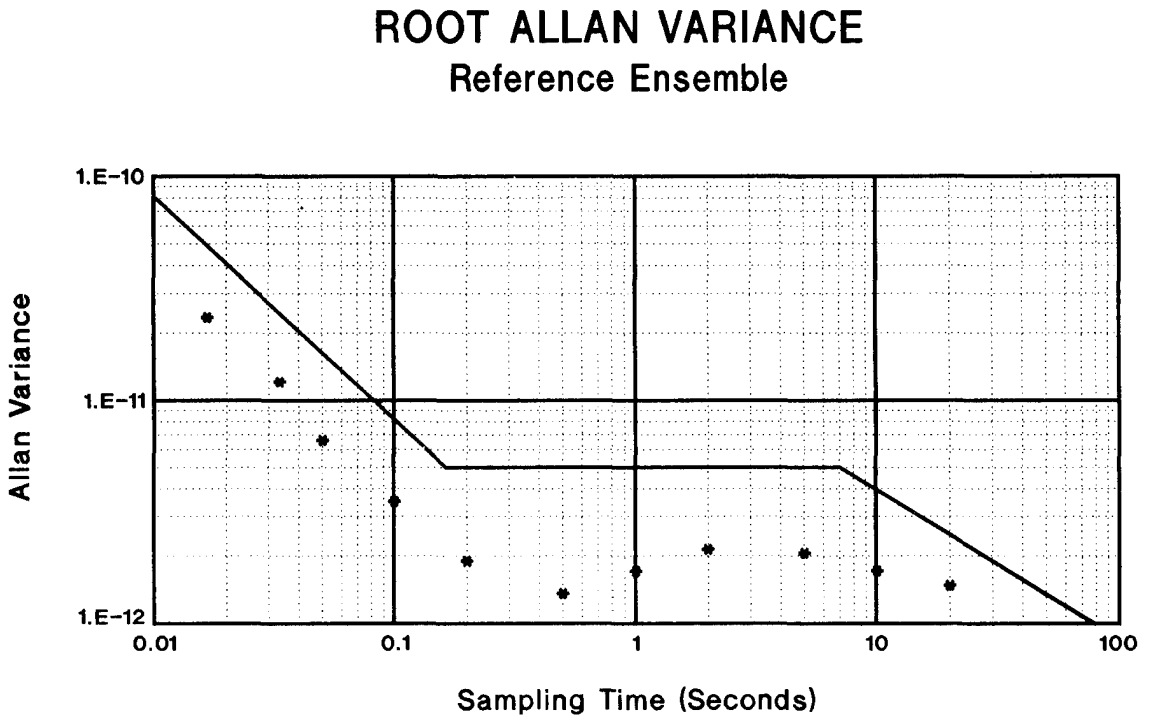


Figure 6: Root Allan Variance data for short sampling times.

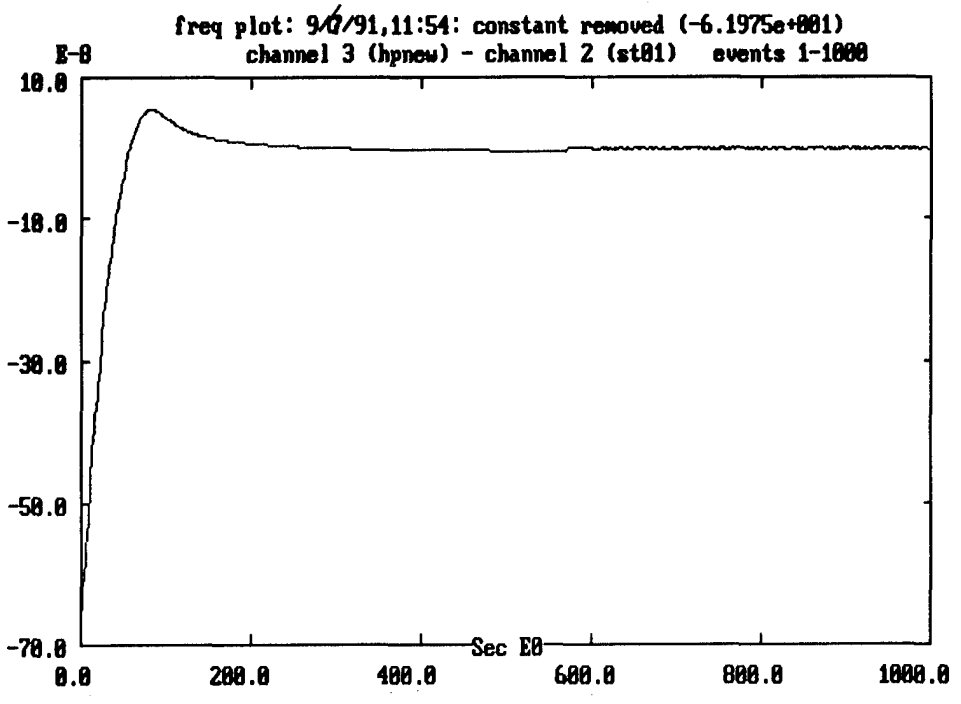


Figure 7: Measured frequency offset - from turn-on at NIST to 1000 seconds. Data indicates that frequency was stable at 616 seconds.

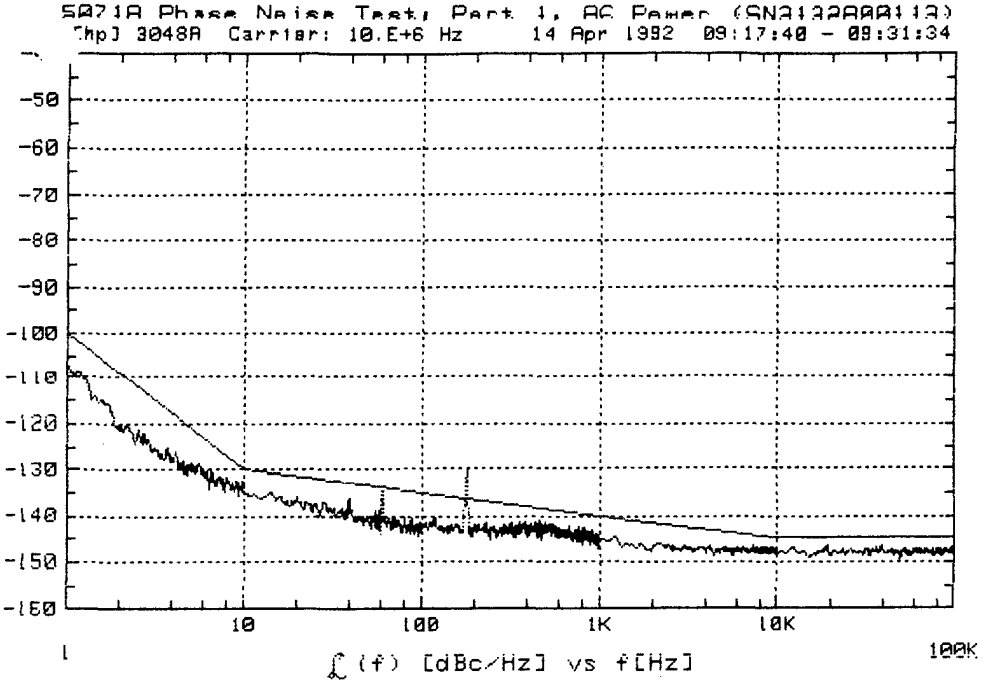


Figure 8: Phase Noise as measured on an HP3048A system

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A STUDY OF ATOMIC TIME SCALE STABILITY

Wei Guo

Shaanxi Astronomical Observatory, Academia Sinica
P. O. Box 18 Lintong, 710600 Shaanxi, China

Abstract

The principle of System Clock in time scale is presented in this paper. Two related propositions are proved, based on the analysis of atomic clock noise characteristics. The classical weighted average methodology is discussed in the light of System Clock principle and some drawbacks of that methodology are pointed out. Also, a new algorithm of time scales is proposed, using the dynamic model of the atomic clock. A simulation on computer is carried out and the results are presented in the paper.

Introduction

At present, the principle of weighted average methodology is generally used in computations of atomic time scales, the typical one is the so-called ALGOS algorithm used in BIPM[1]. The initial basis of the weighted average is the frequency stability theory which was developed in the late sixties[2]. After this period, with the development of statistics and digital signal processing, particularly the applications of time series analysis and state space analysis in atomic clock noise, we obtained more information about the characteristics of atomic clock noise, and some models different from the power-law model were developed. New developments also occurred in the study of time scale algorithms[3],[4]. But up to the present, there has been no reasonable theory which can describe the relationship between the noise model and the scale algorithm. For many years, the studies have been limited by the confines of the classical weighted average methodology. In this paper a new principle of time scale, *System Clock*, is presented, in order to analyze the relationship between atomic time scale and noise model, to establish a new frame for the description of scale algorithms and to provide a theoretical basis for algorithm selection and evaluation.

Noise Model of Atomic Clock

In general, the readings of atomic clock can be written as

$$T(t)=T_I(t)+T_D(t)+x(t) \tag{1}$$

where $T_I(t)$ is the ideal time as defined; $T_D(t)$ is the determined difference between the clock and ideal time; $x(t)$ represents all the uncertain factors for the difference.

The uncertain factor $x(t)$ arises mainly from the internal noise in the atomic clock. In addition, environmental factors, such as temperature, humidity, pressure and electromagnetic field, may also contribute to $x(t)$. Although there is some evidence for these effects[5], a suitable model has not yet been established. In usual discussions, noise model always means the internal noise of atomic clock. In this paper, too, $x(t)$ represents only internal noise of clock.

The basic factors of the model are the five recognized noise processes in atomic clocks,

$$x(t)=n_{-2}(t)+n_{-1}(t)+n_0(t)+n_1(t)+n_2(t) \tag{2}$$

The basic model is in the power-law form, which is power spectral density of fractional frequency fluctuations $y(t)=dx(t)/dt$[2],

$$S_y(f)=h_{-2}f^{-2}+h_{-1}f^{-1}+h_0+h_1f+h_2f^2 \tag{3}$$

By Eq.(3) and the relationship between $x(t)$ and $y(t)$, the spectral density of $x(t)$ can be written as[6]

$$S_x(f)=\frac{1}{(2\pi)^2}(h_{-2}f^{-4}+h_{-1}f^{-3}+h_0f^{-2}+h_1f^{-1}+h_2) \tag{4}$$

The power-law model is the characterization of frequency stability in frequency domain. The spectral coefficient h_α represents the magnitude of the noise processes. The more useful model for scale algorithm study is the form in time domain: dynamic model[7],[8]. The noise mechanism of the dynamic model is the same as in the power-law model, but the five noise processes are presented in three different layers: phase, frequency and frequency drift,

0-7803-0476-4/92 $3.00 © 1992 IEEE

$$\begin{aligned} x(t) &= \alpha(t) + n_2(t) + n_1(t) \\ \dot{\alpha}(t) &= \beta(t) + a_0(t) + n_{-1}(t) \\ \dot{\beta}(t) &= a_{-2}(t) \end{aligned} \tag{5}$$

where $a_0(t)$ and $a_{-2}(t)$ are white noise processes, corresponding to $n_0(t)$ and $n_{-2}(t)$ in Eq.(3), where they are color. This differences are merely the result of the different forms of models, and do not affect the noise properties themselves.

Another form of dynamic model is discrete,

$$\begin{aligned} x(k) &= \alpha(k) + r(k) \\ \alpha(k+1) &= \alpha(k) + \beta(k)\,T + q(k) \\ \beta(k+1) &= \beta(k) + f(k) \end{aligned} \tag{6}$$

where $r(k)$, $q(k)$ and $f(k)$ all are white noise processes, variance R, Q, and F. $q(k)$ and $f(k)$ are integrations of $a_0(t)$ and $a_{-2}(t)$ respectively in interval T. Because the color noise processes are not included in Eq.(6), the discrete dynamic form is a simplified model.

The Principle of System Clock

Since the final aim of a time scale is to give a time coordinate which is more stable than any individual clock, a time scale is equivalent to a clock in function. In addition, a time scale is a mathematical combination of readings of atomic clocks. The uncertainty of a time scale is regarded as one kind of noise process, which comes from the combination of noise processes in atomic clocks and fulfills a certain mathematical model. On the other hand, although the realization of a time scale is usually posterior, it can become real or quasi-real time by the techniques of computer control. So from the point of view of the physical process of producing time coordinate, there is almost no difference between a time scale and a real clock. According to the consideration above, we look upon the clock ensemble, which is involved in the calculation of a time scale, as a whole system. This system is the so-called *System Clock*, and the reading from such a clock is the calculated time scale. In a System Clock, each clock is one unit, and the relationship between the system and clocks is determined by the scale algorithm.

In a linear combination algorithm, the noise model of a System Clock is the same as one of a real clock and the noise coefficients are the linear combination of corresponding coefficients of real clocks. The purpose of studying scale algorithms is to find a mathematical combination to decrease the noise levels of the System Clock for certain clock ensembles.

Proposition I. Let $x_i(t)$ (i=1,2,...,N) be the noise model of N atomic clocks, fulfilling Eq.(2),(3),(4),(5), and noise coefficients be $\{h_{-2}(i), h_{-1}(i), h_0(i), h_1(i), h_2(i)\}$; if $x_s(t)$, the noise model of System Clock, is a linear combination of $x_i(t)$, then $\{h_{-2}(s), h_{-1}(s), h_0(s), h_1(s), h_2(s)\}$ must fulfil the inequality below, no matter what the algorithm is,

$$h_\alpha(s) \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{h_\alpha(i)}}, \quad \alpha=-2,-1,0,1,2 \tag{7}$$

Proof: Because $x_s(t)$ is the linear combination of $x_i(t)$, according to Eq.(3), $x_s(t)$ can always be written as the equation below, no matter what the algorithm is,

$$x_s(t) = \sum_{\alpha=-2}^{2} \sum_{i=1}^{N} C_\alpha(i)\, n_\alpha(t,i),$$

$$\sum_{i=1}^{N} C_\alpha(i) = 1, \quad \alpha=-2,-1,0,1,2$$

According to the assumption of independence between noise processes,

$$E n_\alpha(t,i)\, n_\beta(t',j) = 0, \quad \alpha \neq \beta \text{ or } i \neq j$$

$$\begin{aligned} R_{x_s}(\tau) &= E x_s(t)\, x_s(t+\tau) \\ &= \sum_{\alpha=-2}^{2} \sum_{i=1}^{N} C_\alpha(i)\, R_{n_\alpha}(\tau, i), \end{aligned}$$

$$\begin{aligned} S_{n_\alpha}(f) &= \int_{-\infty}^{+\infty} R_{x_\alpha}(\tau) \exp(-j2\pi f\tau)\, d\tau \\ &= \sum_{\alpha=-2}^{2} \sum_{i=1}^{N} C_\alpha(i)\, S_{n_\alpha}(f,i) \\ &= \frac{1}{(2\pi)^2} \sum_{\alpha=-2}^{2} h_\alpha(s)\, f^{\alpha-2} \end{aligned}$$

thus

$$h_\alpha(s) = \sum_{i=1}^{N} C_\alpha(i)\, h_\alpha(i), \quad \alpha=-2,-1,0,1,2$$

By Lagrange's method of multipliers, the minimum of $h_\alpha(s)$ can be solved as

$$h^*_\alpha(s) = \frac{1}{\sum_{i=1}^{N} \frac{1}{h_\alpha(i)}}, \quad \alpha=-2,-1,0,1,2$$

Because $h^*_\alpha(s)$ is the minimum of $h_\alpha(s)$, then

$$h_\alpha(s) \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{h_\alpha(i)}}, \quad \alpha=-2,-1,0,1,2$$

Quod erat demonstration.

In Proposition I, when the equality in Eq.(8) is valid for every α, all noise levels of a System Clock reach minimum, the scale is in the best stability and the algorithm is optimal. It is also indicated that there is a limitation for improvement of the scale stability for certain clock ensembles. The optimal point is just the limitation, which is not only the goal of the best algorithm but also the criterion of algorithm evaluation.

If the discrete noise model of Eq.(6) is substituted for the one in Proposition I, corresponding noise coefficients be $\{R_i, Q_i, F_i\}$ (i=1,2,...,N) and $\{R_s, Q_s, F_s\}$, then the inequality becomes

$$R_s \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{R_i}}, \quad Q_s \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{Q_i}}, \quad F_s \geq \frac{1}{\sum_{i=1}^{N} \frac{1}{F_i}} \tag{8}$$

Classical Weighted Average Methodology

The time scale algorithm based on the principle of classical weighted average can always be written as the following equation, no matter how the weights are calculated,

$$T_s(t) = \sum_{i=1}^{N} W_i T_i(t) \tag{9}$$

where W_i is the weight for clock $T_i(t)$.

The principle is generally valid, but for convenience we will consider only a two-clock algorithm here,

$$T_s(t) = W T_1(t) + (1-W) T_2(t) \tag{10}$$

The noise coefficients of System Clock $T_s(t)$ are

$$h_\alpha(s) = W^2 h_\alpha(1) + (1-W)^2 h_\alpha(2) \qquad \alpha = -2, -1, 0, 1, 2 \tag{11}$$

where $h_\alpha(1)$ and $h_\alpha(2)$ are noise coefficients of $T_1(t)$ and $T_2(t)$ respectively. According to the principle of the System Clock, the lower limitation of noise coefficients are

$$h_\alpha^*(s) = \frac{h_\alpha(1)\, h_\alpha(2)}{h_\alpha(1) + h_\alpha(2)} \qquad \alpha = -2, -1, 0, 1, 2 \tag{12}$$

Obviously, if the weight fulfills the equation below then the noise level will reach the lower limitation,

$$W_\alpha^* = \frac{h_\alpha(2)}{h_\alpha(1) + h_\alpha(2)} \tag{13}$$

and $h_\alpha(s) = h^*_\alpha(s)$ (α=-2,-1,0,1,2). Notice that the optimal weight in Eq.(13) is a function of α, it means that the optimal weights for each noise process are different from each other. Some weight may be optimal for certain noise and may not be optimal for another one. In general, there is almost no set of weights with which the whole noise level of the System Clock becomes minimum in classical weighted methodology, no matter how the weights are calculated, except in some special cases. It is not difficult to prove that if and only if

$$\frac{h_\alpha(1)}{h_\beta(1)} = \frac{h_\alpha(2)}{h_\beta(2)}, \quad \alpha \neq \beta \tag{14}$$

then $W^*_\alpha = W^*_\beta$ ($\alpha \neq \beta$), noise α and β are minimum simultaneously. If all five noise processes of two clocks fulfil the condition Eq.(14), there is certainly one set of weights with which the whole noise level of a System Clock becomes minimum. But this is merely an extremely special case.

One example is showed in Fig. 1. Two noise coefficients $h_{-2}(s)$ and $h_0(s)$ are the function of weight W. Because the condition of Eq.(14) is not fulfilled, two optimal weights are not coincident, two noise levels in System Clock will not be minimum simultaneously.

A New Algorithm

Noise characteristics of an atomic clock can not be reflected entirely in classical weighted average methodology, because the different noise processes can not be distinguished in the model upon which algorithm is deduced. This drawback can not be overcome by the algorithm itself. In contrast, the dynamic model has an advantage over the basic model of Eq.(2),(3),(4). The same noise processes are expressed in three different states, so the processes can be distinguished in at least three different layers. This important feature is expressed a in new algorithm.

A new algorithm based on the discrete dynamic model is

$$\begin{aligned} T_s(k) &= T_1(k) + \sum_{j=1}^{N} \theta_j [z_{j1}(k) - \alpha_{j1}(k)] \\ &\quad + \sum_{j=1}^{N} \lambda_j [\alpha_{j1}(k) + \alpha_{sj}(k)] \\ \alpha_{sj}(k+1) &= \alpha_{sj}(k) + [\sum_{i=1}^{N} \phi_i \beta_{i1}(k) - \beta_{j1}(k)] T \end{aligned} \tag{15}$$

$$\theta_j = \frac{\frac{1}{R_j}}{\sum_{i=1}^{N} \frac{1}{R_i}}, \quad \lambda_j = \frac{\frac{1}{Q_j}}{\sum_{i=1}^{N} \frac{1}{Q_i}}, \quad \phi_j = \frac{\frac{1}{F_j}}{\sum_{i=1}^{N} \frac{1}{F_i}}$$

where

$$z_{j1}(k) = T_j(k) - T_1(k), \quad \alpha_{j1}(k) = \alpha_j(k) - \alpha_1(k),$$

$$\beta_{j1}(k) = \beta_j(k) - \beta_1(k),$$

$$R_j = Dr_j(k), \quad Q_j(k) = Dq_j(k), \quad F_j(k) = Df_j(k)$$

$$j = 1, 2, \cdots, N$$

Proposition II. With the discrete dynamic model of Eq.(6), three noise levels of the System Clock are minimum simultaneously, provided the algorithm of Eq.(15) is used.

Proof: According to Eq.(6), noise model of clock j fulfills the equation below,

$$x_j(k) = \alpha_j(k) + r_j(k), \qquad Dr_j(k) = R_j$$

$$\alpha_j(k+1) = \alpha_j(k) + \beta_j(k) T + q_j(k), \qquad Dq_j(k) = Q_j$$

$$\beta_j(k+1) = \beta_j(k) + f_j(k), \qquad Df_j(k) = F_j$$

According to Eq.(1) and (15), System Clock $T_s(s)$ is

$$T_s(k) = T_I(k) + T_D(k) + x_s(k)$$

$$x_s(k) = \alpha_s(k) + r_s(k)$$

where

$$\alpha_s(k) = \sum_{j=1}^{N} \lambda_j [\alpha_j(k) + \alpha_{sj}(k)] \ ,$$

$$x_s(k) = \sum_{j=1}^{N} \theta_j x_j(k)$$

then

$$\alpha_s(k+1) = \sum_{j=1}^{N} \lambda_j [\alpha_j(k+1) + \alpha_{sj}(k+1)]$$

$$= \alpha_s(k) + \beta_s(k)\, T + q_s(k)$$

where

$$\beta_s(k) = \sum_{j=1}^{N} \phi_j \beta_j(k) \ , \quad q_s(k) = \sum_{j=1}^{N} \lambda_j q_j(k)$$

Similarly

$$\beta_s(k+1) = \sum_{j=1}^{N} \phi_j \beta_j(K+1)$$

$$= \beta_s(k) + f_s(k)$$

where

$$f_s(k) = \sum_{j=1}^{N} \phi_j f_j(k)$$

So $T_s(k)$ fulfills the equation below,

$$x_s(k) = \alpha_s(k) + x_s(k)$$

$$\alpha_s(k+1) = \alpha_s(k) + \beta_s(k)\, T + q_s(k)$$

$$\beta_s(k+1) = \beta_s(k+1) + f_s(k)$$

It is not difficult to see that

$$R_s = \frac{1}{\sum_{j=1}^{N} \frac{1}{R_j}} \ , \quad Q_s = \frac{1}{\sum_{j=1}^{N} \frac{1}{Q_j}} \ , \quad F_s = \frac{1}{\sum_{j=1}^{N} \frac{1}{F_j}}$$

Compare with Eq.(8), R_s, Q_s and F_s are minimum simultaneously.
Quod erat demonstration.

Phase difference $\alpha_{j1}(k)$ and frequency deviation $\beta_{j1}(k)$ are needed in computation of new algorithm. They can not be obtained directly from measurement of difference between clock readings, they should be substituted by suitable estimates.

Comparison of Two Algorithms by Simulation

The two algorithms, weighted average and new algorithm, are tested by simulant data. 11 clocks are simulated in different stabilities. The noise coefficients of these simulant clocks are adopted from the Ref.[4], the estimations of 11 real clocks in NIST.

The stabilities of each simulant clock and two scales, with old and new algorithm, are calculated in Allan variance, shown in Fig. 2, Fig. 3 and Table.

In classical weighted average methodology, one chooses weights merely to decrease one noise level of the System Clock in order to meet some special needs. Usually, the long term stability of a time scale is considered first. With the weighted average method, the stability of the scale is worse than some clocks of good short stability when the sample interval is less than 160 hours (about 7 days), even though the long stability of the scale is better than every individual clock. See the dashed line in Fig. 2.

The new algorithm takes advantages of each clock in long or/and short stabilities, so the short term stability of the new scale is good as well as its long term stability. So the new scale maintains the best stability in all sample intervals. See the dashed line in Fig. 3.

Discussions

(1) A new theoretical frame for the study of time scale algorithms has been established with the principle of the System Clock and the relevant propositions. By this frame, a time scale is expressed as the noise model of a real physical process, so the function of an algorithm becomes very clear: to decrease all noise levels in this real physical process.

(2) According to the principle of the System Clock, the noise level of a time scale is determined by the noise level of a clock ensemble and an algorithm. It is possible to evaluate the stability of a scale from the internal noise level of the system without any other external standard. This is similar to stability evaluation of a cesium base, where the uncertainty is evaluated by the uncertainties of units with which the base is constructed.

(3) In classical weighted average methodology, all noise level of a System Clock can not be decreased simultaneously, only one noise level can be controlled for some special purpose. Any change in weights is merely to adjust noise distributions in the System Clock. There will be a waste of resource, if the performance of each clock in the ensemble is very different from the others.

(4) The new algorithm presented in this paper is optimal for the discrete dynamic model. The algorithm will degenerate to quasi-optimal for all five noise processes. The value of the new algorithm is that it goes beyond the limitations of classical weighted average. If classical weighted average can be to a kind of plane algorithm, then the new algorithm is a solid one, because noise precesses are controlled in three different layers. It also suggests a possible way for finding an optimal algorithm for all five noise processes.

(5) The new algorithm is not a practical one, it is merely a principle, in a strict sense. Some techniques should be developed in order to put the algorithm into practical usage, such as estimations of clock parameters, phase and frequency, and adaptive corrections for slow changes of parameters. These problems are involved in digital signal processing and system control theory, and should not be very difficult to solve.

Reference

[1] B.Guinot, Annual Report for 1974, Bureau International de l'Heure, Paris

[2] J.A.Barnes, et al., "Characterization of Frequency Stability", IEEE Trans. Instrum. Meas., vol. IM-20, No. 2, 1971

[3] D.B.Percival, "The U. S. Naval Observatory Clock Time Scale", IEEE Trans. Instrum. Meas., vol. IM-27, No.4, 1978

[4] J.A.Barnes, "Time Scale Stability Based on Time and Frequency Kalman Filter", Proc. 39th Annual Frequency Control Symposium, 1985

[5] WU Gui-chen, WEI Guo, "Effect of Environmental Temperature and Humidity on the Rate of Cesium Clock", Proc. International Symposium of Electromagnetic Metrology, 1989

[6] P.Kartaschoff, Frequency and Time, Academic Press, London, 1978

[7] WEI Guo, "Characteristic Analysis of Clock Noise: A Dynamic Model", Proc. Third International Time Scale Algorithm Symposium, 1988

[8] S.R.Stein, "Kalman Ensemble Algorithm: Aiding Sources Approach", Proc. 43th Annual Frequency Control Symposium, 1989

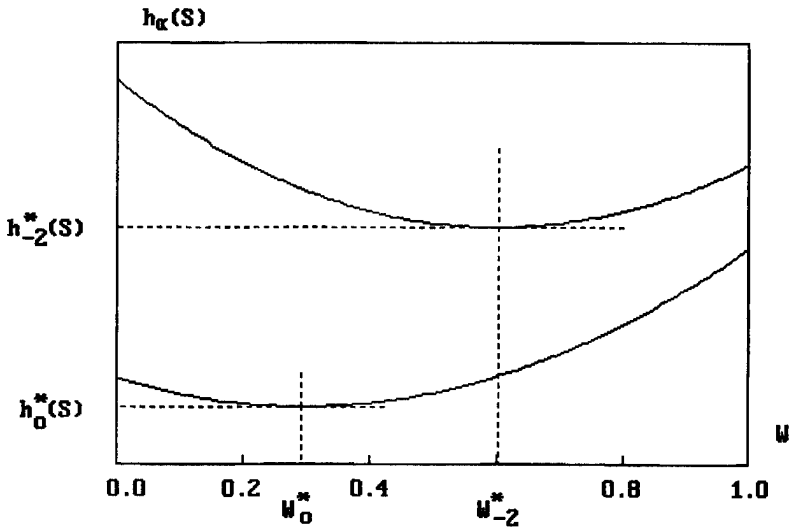


Fig. 1 Noise levels of System Clock vs. weight W

Tab. Stabilities of Time Scales with Old and New algorithm: $\sigma_y(\tau) \times 10^{-14}$

τ(hour)	1	2	4	9	18	38	78	162	336	696	1440
Old One	9.958	7.041	4.979	3.320	2.352	1.630	1.163	0.899	0.841	0.998	1.354
New One	2.115	1.496	1.060	0.713	0.520	0.405	0.392	0.477	0.652	0.927	1.329

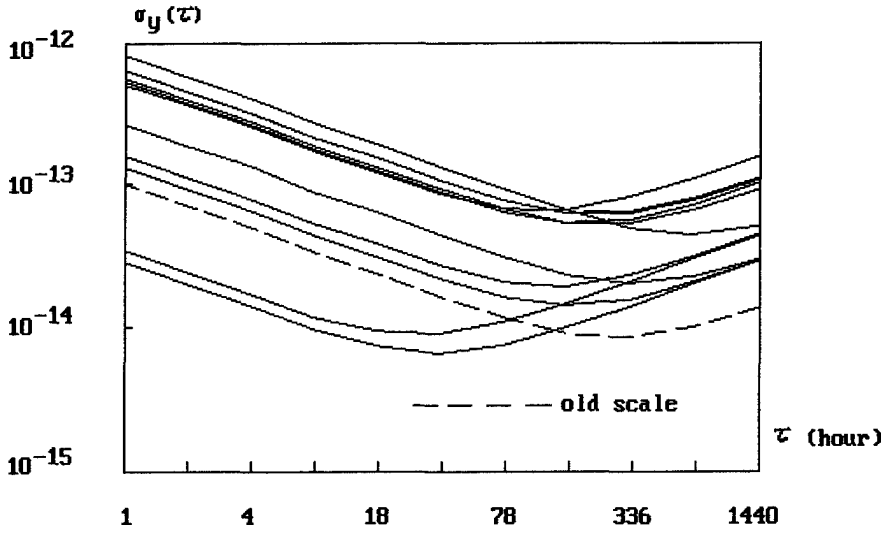


Fig. 2 Allan variance: clocks and old scale

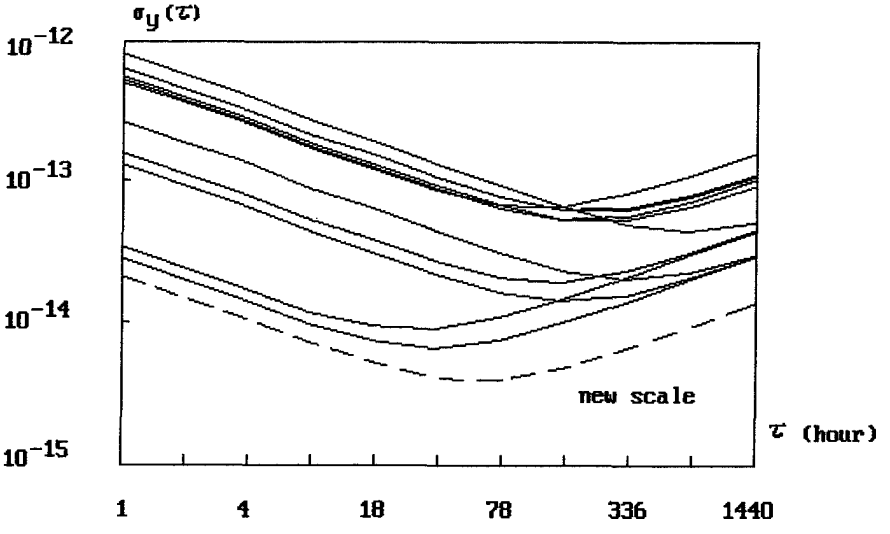


Fig. 3 Allan variance: clocks and new scale

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LOW NOISE, MICROWAVE SIGNAL GENERATION USING CRYOGENIC, SAPPHIRE DIELECTRIC RESONATORS: AN UPDATE

M. M. Driscoll* and R. W. Weinert**

*Westinghouse Electronics Systems Group, Baltimore, MD
**Westinghouse Science & Technology Center, Pittburgh, PA

Abstract

X-band Q factors of 10^5 to 10^9 have been obtained for sapphire dielectric resonators operated at temperatures in the range 4K to 77K. Based on considerations of exceptionally high Q and moderate signal handling capability, resonator use as a microwave signal generator reference element should allow realization of short-term frequency stabilities unattainable using alternative technologies.

In this paper, we will report on recent results obtained at Westinghouse using sapphire dielectric resonators for AFC stabilization of an X-band, SAWR VCO-based source. The resonators are operated on a low order (TE_{01}) mode and employ high temperature superconducting films in the metal enclosure covers in order to obtain unloaded Q-factors of $3X10^5$ to $5X10^5$. Signal source flicker-of-frequency noise characterized by $Sy(100Hz)=5X10^{-24}/f$ or L(100Hz)=-96dBc/Hz and noise floor levels of -165dBc/Hz have been measured at X=band.

While these results constitute state-of-the art performance, the measured values for flicker-of frequency noise is 10dB poorer than anticipated, based on the large value (72uV/Hz) of discriminator sensitivity obtained. Possible causes for current limitations in signal near-carrier spectral performance include resonator short-term frequency fluctuations resulting from environmentally-induced effects and/or frequency-drive (AM-to-FM) effects. Resonator vibration sensitivities (force applied parallel to enclosure covers) on the order of $1X10^{-8}$ per g have been measured.

This work was partially supported by DARPA, Contract No. 972-88-C-0050.

Introduction

In a low noise, microwave signal generator, output signal phase noise characteristics are largely determined by the generator reference oscillator resonator characteristics including loaded Q, operating frequency, drive level, insertion loss, short-term frequency stability, and frequency sensitivity to environmental stress. A spectral performance tradeoff usually exists with regard to obtaining low near-carrier noise vs noise floor levels. This tradeoff is a consequence of the fact that much higher resonator Qs are obtainable, for example, in a quartz crystal oscillator operating in the HF/VHF range, compared to those exhibit by conventional, microwave oscillator resonators [1-3].

It is possible, however, to obtain ultra-high resonator Q values directly at microwave frequencies (10^5 to 10^9 at X-band) using sapphire dielectric resonators (DRs) operated in the liquid nitrogen to liquid helium temperature range [3-5]. The use of this technology, therefore, offers the possibility of generating microwave signals exhibiting (phase noise) spectral performance previously unattainable using conventional resonator technologies. In a cooled, sapphire DR, achievement of ultra-high Q values are a consequence of the extremely low loss tangent exhibited by sapphire at low temperature. In these types of resonators, attainable Q can be limited by resonator enclosure metal wall losses, rather than by the sapphire itself.

At the 1991 Frequency Control Symposium, the authors described a concept for the utilization of a cooled, sapphire DR for noise reduction (via AFC stabilization) of a microwave frequency generator employing a VHF acoustic resonator-based oscillator and frequency multiplier [3]. Design parameters were described for a sapphire DR fabricated for X-band operation on a low order (TE_{01}) mode and employing high temperature, superconducting films in the

0-7803-0476-4/92 $3.00 © 1992 IEEE

resonature enclosure covers in order to obtain unloaded Q values of over $3X10^5$.

In this paper we describe (phase noise reduction) performance obtained using the DR in a high-sensitivity, frequency discriminator for AFC feedback loop stabilization of a surface-acoustic-wave resonator voltage-controlled oscillator (SAWR VCO)-based, X-band frequency generator.

Sapphire Dielectric Resonator Design

The TE_{01} dielectric resonator is comprised of a sapphire puck centered in a cylindrical copper enclosure and is described in detail in reference [3]. Low loss, ring supports are used to maintain the puck position, and loop-type coupling probes are used to excite and detect the resonance. The dimensions of the sapphire puck and metallic enclosure were chosen so that, at 77K, the TE_{01} mode frequency was 10GHz and the unloaded Q value was limited by the ohmic losses in the copper end planes (enclosure covers). Replacing the copper end planes with HTS films results in increased resonator Q without changing the resonant frequency. Using 1.5 inch diameter, $YBCO/LaAlO_3$ films, unloaded Q values of $3X10^5$ were realized at 77K. These films were of high quality, but were too thin (approx. 200nm) for proper shielding. For the work being reported here, 600nm thick, $TBCCO/LaAlO_3$, HTS films were used, and unloaded Q factors of $5X10^5$ were obtained [6]. These films had surface resistance values less than $1m\Omega$ at 77K, and unloaded Q values approaching $1X10^6$ should have been realized. The difference between predicted and actual Q values has been attributed to dielectric losses associated with the sapphire puck mounting rings. Currently, we are working on a lower loss mounting configuration.

Figure 1 shows the measured transmission and reflection responses of the (10.08GHz) resonator. The input and output coupling probes for the 10.08GHz and 10.24GHz resonators were adjusted for approximately 10 to 13dB input return loss and 0.5dB to 1dB output return loss. The 10.24GHz resonator transmission response symmetry and selectivity were somewhat degraded (compared to that of the 10.08GHz resonator) due to smaller frequency separation between the desired resonance and an undesired, hybrid mode.

In order to provide a stable-temperature, test environment and a rapid cool-down (approx. 1 hour), the resonators were mounted and sealed in a stainless steel can which was filled with hydrogen or helium gas, and

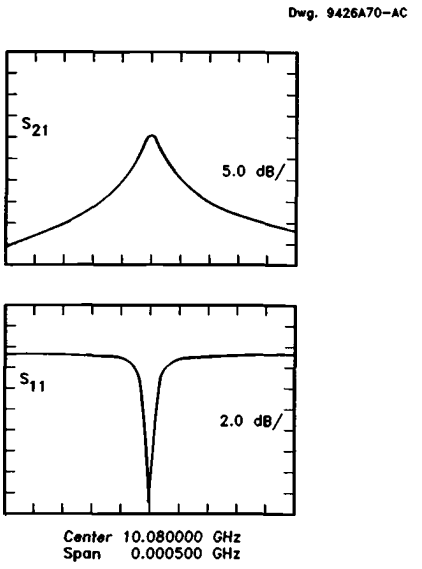


Figure 1. Measured Transmission and Reflection Responses for 10.08GHz DR: $Q_L = 3x10^5$ at 77K.

the entire can was submerged in a liquid nitrogen bath. Figure 2 shows the resonator mounted to the top flange of the stainless steel, outer enclosure. Sealing was accomplish using a copper gasket, and RF input/output connections were made via hermetic, bulkhead, SMA connectors soldered to the steel enclosure top flange.

Resonator-Based Frequency Discriminator

Two sapphire DRs were operated in a pair of reflection-mode type frequency discriminators [5, 7-9]. A block diagram of the AFC loop stabilization circuit including the DR-based, frequency discriminator is shown in Figure 3. Higher sensitivity is an advantage associated with utilization of resonator reflection response (compared to transmission response) in the discriminator circuit. For two-port resonator configuration, useful output signal power, subject to the narrowband resonator transmission response, can be extracted via the second (lightly-coupled) port. Utilizing resonators having unloaded Qs of approximately $5x10^5$ with input coupling adjusted for approximately -10dB return loss and relatively high (17 to 20dBm) drive

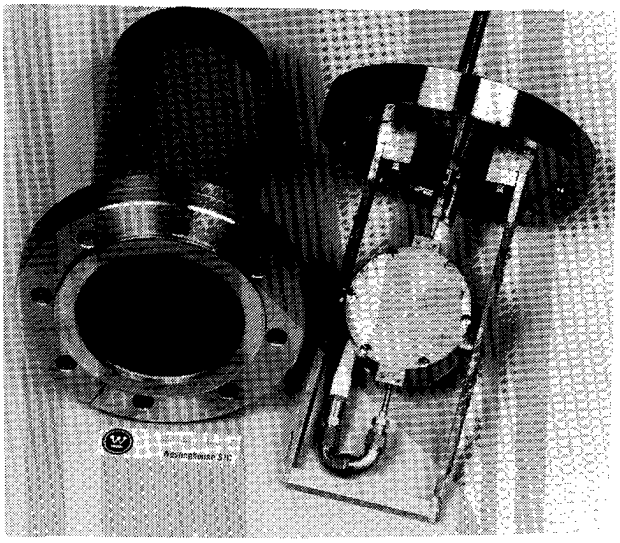


Figure 2. Photograph of DR mounted Top Flange of Stainless Steel, Outer Enclosure.

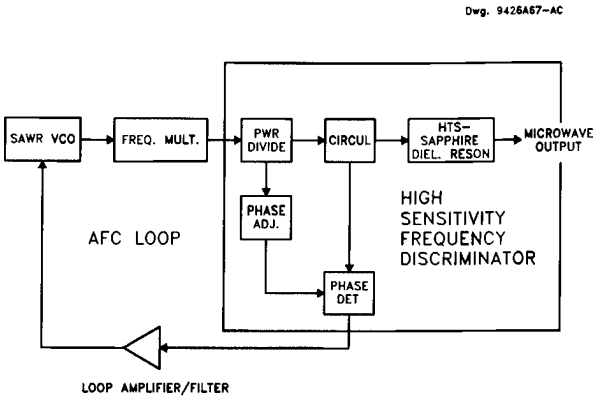


Figure 3. Block Diagram of AFC-Stabilized, X-Band Source Using Cooled, Sapphire DR-Based, Frequency Discriminator.

levels, discriminator sensitivities of approximately 72uvolt/Hz were obtained. A mechanically-adjustable phase shifter was used for coarse (open loop) and final (closed loop) resonator/discriminator alignment, as well as measurement of discriminator sensitivity. The alignment procedure simply consists of phase shifter adjustment for simultaneous attainment of a RF power null at the monitor output and zero DC voltage at the phase detector output.

Discriminator phase detector/AFC loop amplifier noise measurements were made with a broadband, 10dB return loss load (unterminated, 50-ohm pad) substituted for the DR. Referred to the phase detector (WJM79HC mixer) input, a 1/f noise level characterized by -136dBv/√Hz at 100Hz was measured. At 0.9volt/rad. detector sensitivity, this corresponds to relatively poor (-138dBc/Hz at 100Hz carrier offset frequency) mixer performance. Additional tests confirmed that generator (SAWR VCO+frequency multiplier) AM noise reduction via signal limiting was effective in reducing 1/f noise measured at the mixer output to a level of -146dBv/√Hz at 100Hz. Based on the measured value of 72uvolt/Hz discriminator sensitivity (with the DR installed), the expected, flicker-of-frequency noise level predicted for the AFC-stabilized source is characterized by $Sy(f)=5X10^{-25}/f$ or $\mathcal{L}(100Hz)=-106dBc/Hz$.

Measurement Test Set-Up

Figure 4 shows a block diagram of the hardware use to measure the (phase noise reduction) performance provide by the DR-based frequency discriminator/AFC loop. As shown in the figure, two, distinct DR-stabilized sources were constructed at X-band output frequencies separated by 160MHz. The Figure 4 measurement mechanization was chosen because it allowed the use of an on-hand, 160MHz, voltage-controlled crystal oscillator (VCXO) whose output signal phase noise sideband spectrum is superior (for $fm \leq 10kHz$) to that predicted for the X-band sources under test. The moderate-term stability of the DRs (and hence, locked, X-band sources) was such that the 1KHz/volt VCXO tuning sensitivity was more than adequate to maintain phase quadrature to the accuracy required in the HP11740A Phase Noise Measurement System during the measurement.

Each SAWR VCO was operated at 320MHz and employed a high level (Po=24dBm), 50-ohm, modular amplifier sustaining stage, schottky diode limiter, and a pair of cascaded SAWRs to achieve low near-carrier and noise floor levels. Similarly, the 10MHz crystal (offset) oscillator used in the second X-band source also employed multiple resonators and a high level sustaining stage in order to achieve excellent noise floor performance [10]. Figure 5 shows the measured phase noise performance for the unlocked (frequency discriminator disconnected), X-band sources. Phase noise peaking in the 10 to 20Hz region is a result of SAWR subassembly vibration isolator resonance. Overall, X-band, noise floor performance is limited by nearly equal contributions from the times 16, step-recovery diode multipliers, SAWR VCOs, and the 10MHz crystal oscillator.

SAWR VC0 tuning sensitivity (referred to X-band) was approximately 180KHz/volt. At

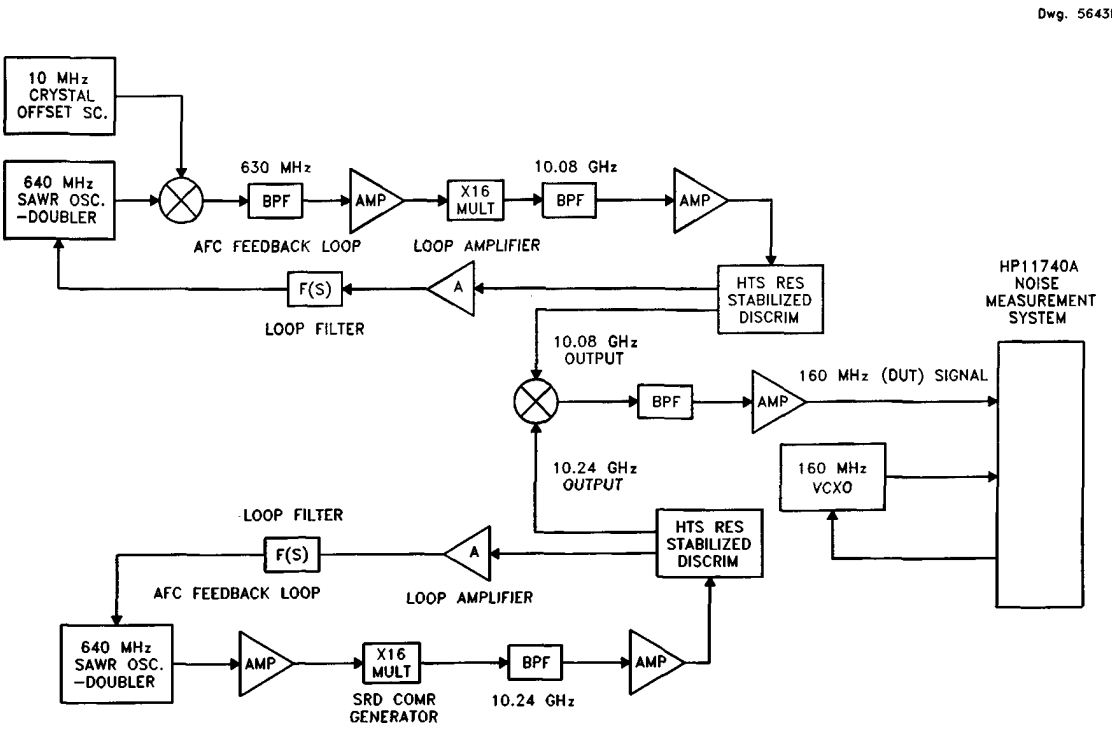


Figure 4. AFC-Stabilized, X-Band Source Phase Noise Measurement Test Set-Up.

Kd=72uvolt/Hz (discriminator sensitivity) and Kvco=180KHz/volt, a loop DC gain of Kd*Kvco=13 was achieved for the VCO/discriminator combination. The loop amplifier/filter was configured for approximately 40dB overall open loop gain at 1KHz and 100KHz (unity gain) bandwidth.

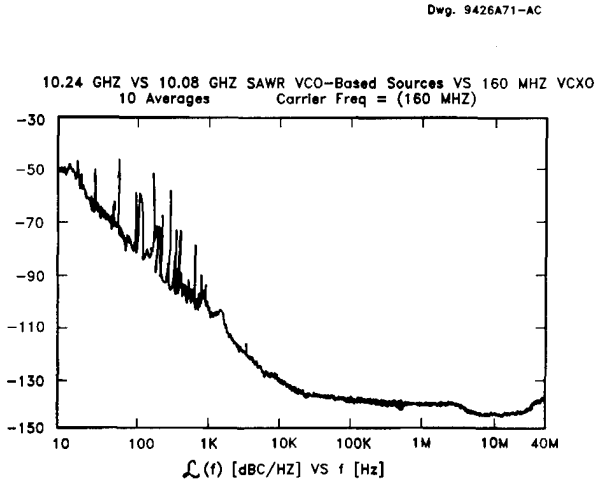


Figure 5. Phase Noise Measurement Results: Two, Un-stabilized, SAWR VCO-Based, X-Band Sources.

Figure 6 shows a photograph of the two, AFC-stabilized, X-band sources. In the figure, the sapphire DR assemblies (immersed in liquid nitrogen) are visible at the bottom of the photograph.

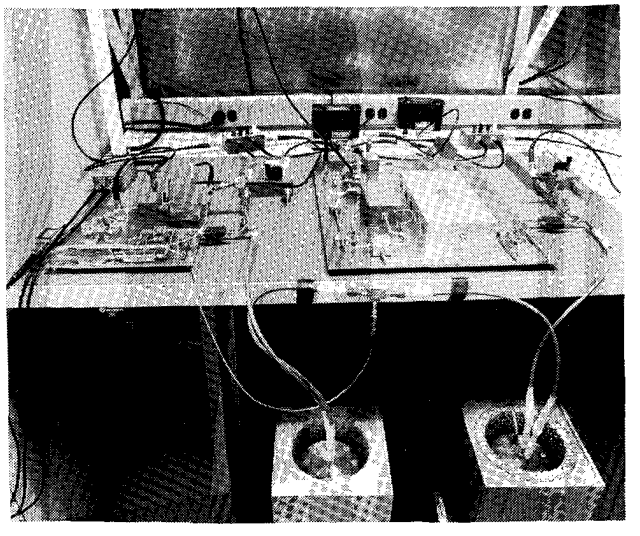


Figure 6. Photograph of AFC-Stabilized, X-Band Sources.

Measurement Results

Figure 7 shows the measured phase noise spectrum for the two, AFC-stabilized, X-band sources. As shown in the figure, the portion of the spectrum between 10Hz and 2KHz offset frequency has been 'corrupted' by what we believe to be vibration-induced instability in the DR. A separate set of noise measurements were made in the presence of variable-frequency mechanical stress applied to the DR assembly and (semi-rigid coaxial) DR input/output cables via a audio speaker-driven, polystyrene rod probe. Those measurements verified mechanical resonances at frequencies coincident with the Figure 7 noise peaks at 380Hz and 1.3KHz. In addition, probe excitation of the DR outer, sealed enclosure at non-mechanical resonance frequencies in a direction parallel to the plane of the DR covers indicated an acceleration sensitivity on the order of $1X10^{-8}$ per g. The discrete spurious at 9KHz in Figure 7 is a result of non-coherency between the two AFC-stabilized, X-band sources. In addition, the broadband noise peaking in the 1.5MHz to 20MHz frequency offset range is a direct result of spurious resonant mode responses exhibited by the 10.24GHz DR that result from utilization of a common DR design optimized for absence of nearby spurious modes at 10.08GHz operating frequency.

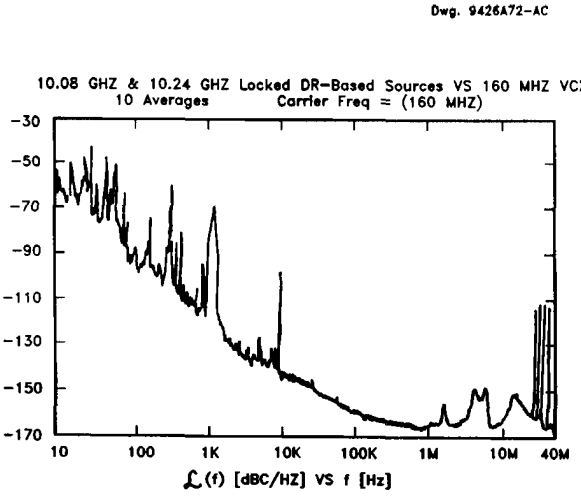


Figure 7. Phase Noise Measurement Results: Two, AFC-Stabilized, X-Band Sources.

In spite of the presence of vibration-induced spectral degradation, the data depicted in Figure 7 indicate, on a per-source basis, a flicker-of-frequency noise level characterized by $Sy(f)=5X10^{-24}/f$. This performance represents a 13dB reduction in (open loop) source flicker-of-frequency noise, but is 10dB poorer than the 23dB reduction predicted for the AFC-stabilized sources. At this time, we believe the discrepancy between predicted and measured data is due to short-term frequency instability in the DRs resulting from DR frequency sensitivity to environmental (temperature, vibration) stress and possible RF drive level effects (ie, AM to FM conversion). Nevertheless, the flicker-of frequency noise performance actually achieved is identical to state-of-the-art performance recently reported, for example, using UHF High-Overtone, Bulk Acoustic Resonator (HBAR) technology.

Conclusions

A cooled, sapphire dielectric resonator, designed to exhibit moderate X-band unloaded Q values on the order of $5X10^5$, can be used to achieve frequency discriminator sensitivities consistent with obtaining significant (10:1) improvements in signal generator flicker-of-frequency noise via AFC loop source stabilization.

In order for predicted levels of spectral performance improvement to be achieved, resonator short-term frequency stability, especially in the presence of environmental stress, needs to be improved.

References

[1] M. M. Driscoll, "Low Noise, Microwave Signal Generation Using Bulk- and Surface-Acoustic-Wave Resonator", IEEE Trans. UFFC, Vol. 35, No. 3, May, 1988, pp. 426-434.

[2] M. M. Driscoll R. A. Jelen, and N. Mathews, "Extremely Low Phase Noise UHF Oscillators Utilizing High-Overtone, Bulk-Acoustic Resonators", Proc. IEEE Ultras. Symp., Dec., 1990, pp. 513-519.

[3] M. M. Driscoll, et. al., "Cooled, Ultra-High Q, Sapphire Dielectric Resonators for Low Noise, Microwave Signal Generation", Proc. 45th Freq. Contr. Symp., May, 1991, pp. 700-706.

[4] A. Giles, et, al., "A High Stability Microwave Oscillator Based on a Sapphire-Loaded, Superconducting Cavity", Proc. 43rd Freg. Contr. Symp., May, 1989, pp. 89-93.

[5] J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Ring Resonator", Proc. 43rd Freq. Contr. Symp., May, 1989, pp. 107-114.

[6] TBCCO/$LaAlO_3$, HTS films supplied by Superconducting Technologies, Inc., Santa Barbara, CA.

[7] R. V. Pound, "Frequency Stabilization of Microwave Oscillators", Proc. IRE, Vol. 35, Dec., 1947, pp. 1405-1415.

[8] C. H. Grauling and D. J. Healey, "Instrumentation for Measurement of the Short-Term Frequency Stability of Microwave Sources", Proc. IEEE, Vol. 54, No. 2, Feb., 1966, pp. 249-257.

[9] F. L. Walls, et. al., "High Spectral Purity X-Band Source", Proc. 44th Freq. Contr. Symp., May, 1990, pp. 542-548.

[10] M. M. Driscoll, "Reduction of Quartz Crystal 0scillator Flicker-of-Frequency and White Phase Noise (Floor) Levels and Acceleration Sensitivity via Use of Multiple Resonators", 46th Freq. Contr. Symp., May, 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

HIGHER ORDER MODES, HIGH Q DIELECTRIC RESONATORS FOR OSCILLATOR APPLICATIONS

Kawthar A. Zaki & Xiao-Peng Liang
Electrical Engineering Department
University of Maryland, College Park, MD 20742

Abstract: Properties of the TE_{02} higher order mode of a sapphire dielectric resonator are investigated. This mode has significantly higher unloaded quality factor Q than that of TE_{01} mode. Significant reduction of the losses and corresponding increases in the unloaded Q of the resonator can be achieved by using superconductor enclosure covers. Detailed accurate quantitative analysis and results showing the achievable improvements are presented. This resonator can be advantageously used in very low phase noise oscillators applications.

1. Introduction

In Reference [1] it is demonstrated that unloaded Q's of the order of 500k can be achieved from cooled TE_{01}-mode dielectric sapphire resonators with high T_c superconductor enclosure covers at two ends. In this paper we investigate the possibility of achieving considerably higher unloaded Q's from similar configurations, but with higher order mode excitations of the sapphire resonators. Rigorous analyses of the loss mechanisms, field distributions for various modes and shapes of supporting structures are presented. The unloaded Q's of the modes are computed and compared for different conditions. It is shown that by careful optimization and appropriate selection of the resonant mode, unloaded Q's of the order of one Million are achievable at x-band frequencies, with HTc superconductor enclosure covers.

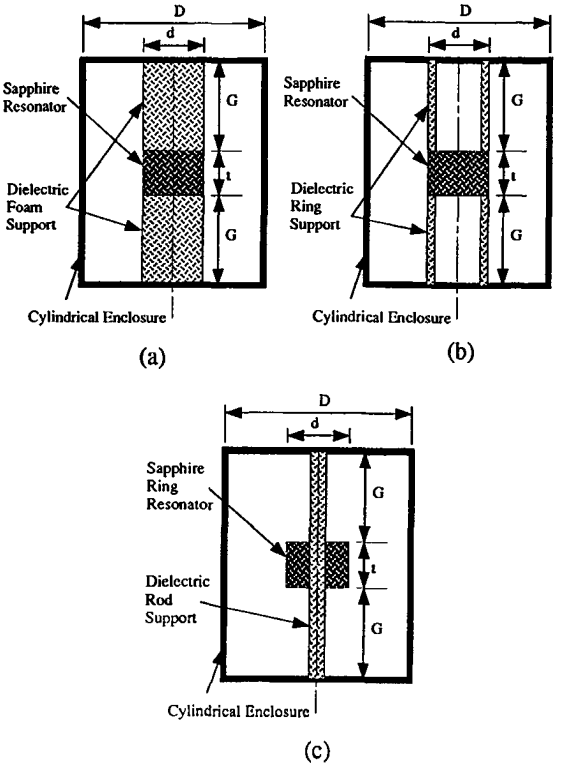


Fig. 1 Configurations of (a) sapphire resonator supported by solid foam discs, (b) sapphire resonator supported by cylindrical foam rings, and (c) sapphire ring resonator supported by a thin dielectric rod

2. Analysis

Typical structures of the resonators are shown in Fig. 1. Fig. 1-a consists of a cylindrical dielectric resonator of relative dielectric constant ϵ_{r_1}, diameter d, and length t, placed symmetrically inside a cylindrical con-

0-7803-0476-4/92 $3.00 © 1992 IEEE

ducting enclosure of diameter D and length $L(= 2G+t)$. The resonator is supported on its ends by two low loss foam discs of relative dielectric constant ϵ_{r_2}, diameter d and length G each. In order to minimize the losses in the foam support, hollow thin cylindrical rings can be used, instead of the solid foam cylinder, as shown in Fig. 1-b. A third configuration is shown in Fig. 1-c, in which the resonator has a small hole in its center that allows a thin dielectric rod extending from the top cover to the bottom cover through the resonator, to be used as a support.

Rigorous analysis methods for these types of structures have been recently developed [2]-[5]. These methods are based on the mode matching techniques and involve significant amounts of numerical computations, the results of which give the resonant frequencies of various modes that can exist on the structures, and electromagnetic field distributions of these resonant modes. From the knowledge of the fields, it is possible to compute the losses in various parts of the structure (i.e. the enclosure side wall, the enclosure covers, the resonator and the supporter) and also the total energy stored, and thus the unloaded Q's.

3. Results

A 10.3 GHz TE_{02} mode sapphire resonator mode chart is shown in Fig. 2. The dimensions of the resonator are changed so that the resonant frequency of the TE_{02} mode stays the same. Other resonant modes are close by, however, the aspect ratio (d/t) of the resonator can be chosen such that the TE_{02} mode is separated from the closest higher order and lower order modes by the same amount (in the case shown in Fig. 2 it is approximately 500 MHz). Fig. 3 shows the computed corresponding unloaded Q's of the various modes for a cooled structure (77°k), with copper enclosure. Clearly the TE_{02} mode has the highest unloaded Q of about 750K. Figs. 4-a and 4-b show the variations of unloaded Q's as function of

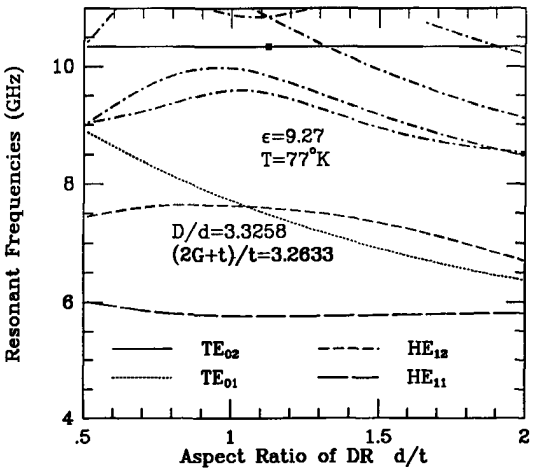


Fig. 2 A TE_{02} mode sapphire resonator mode chart

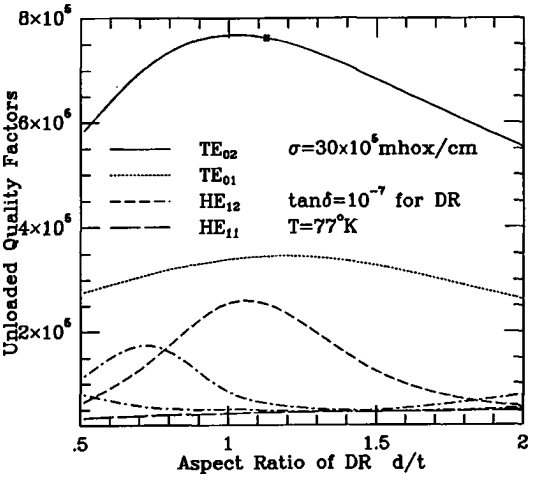
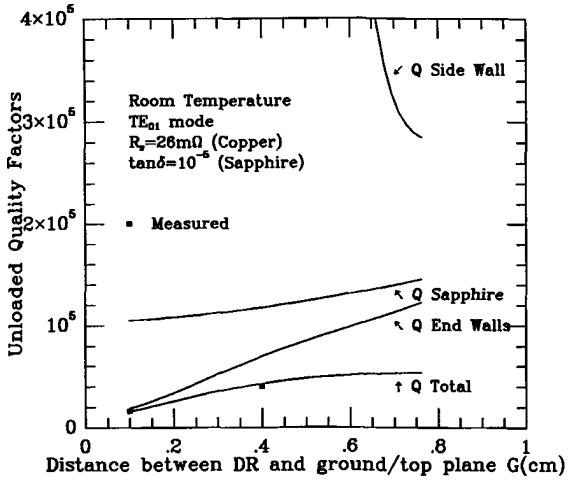


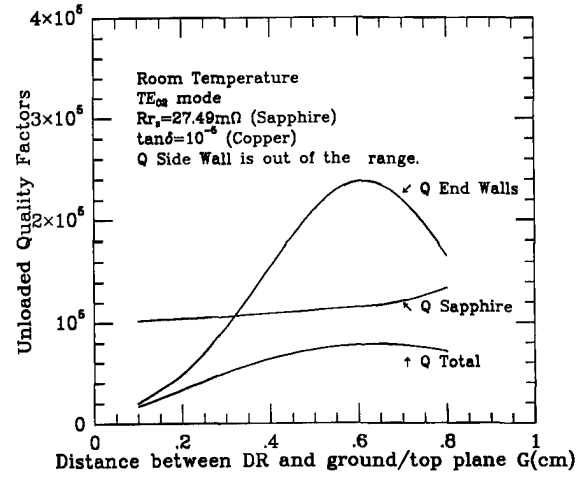
Fig. 3 The unloaded Q's of the various modes with copper enclosure at 77°k

the covers separation from the resonator G for TE_{01} and TE_{02} modes, respectively.

If high T_c superconductor covers are used for the enclosure, significant increase of the unloaded Q can be achieved as seen in Figs. 5-a and 5-b. These figures show the variations of Q_u as a function of the superconductor covers separation from the resonator G, for various values of (R_s/R_{copper}) for TE_{01} and TE_{02} modes, respectively.

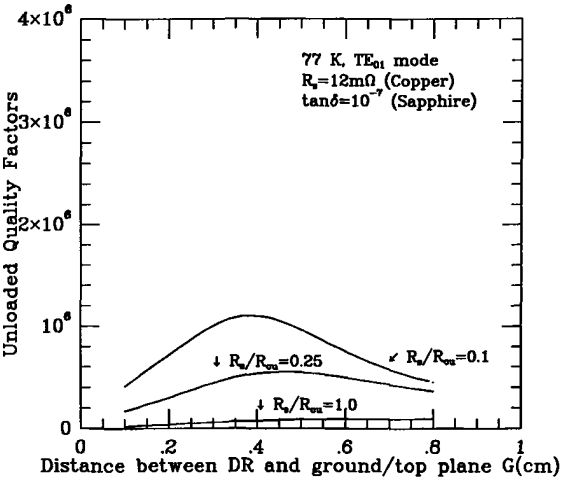


(a)

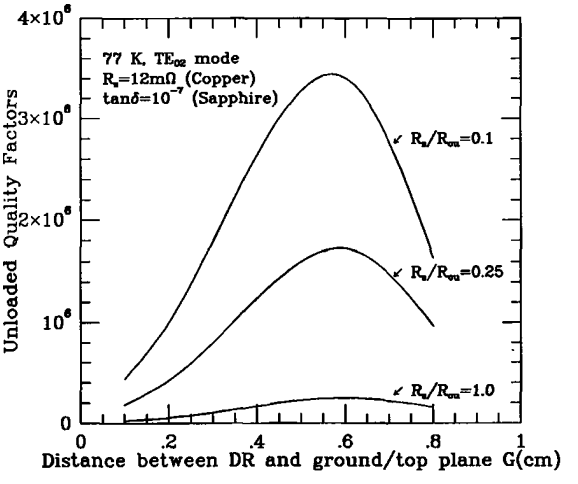


(b)

Fig. 4 The unloaded Q's as functions of G for (a) TE_{01} mode and (b) TE_{02} mode



(a)



(b)

Fig. 5 The unloaded Q's as functions of G with high Tc superconductor cover plates, (a) TE_{01} mode and (b) TE_{02} mode

It is clear that there is an optimum value of G to achieve the maximum Q's for both the copper covers and the superconductor covers.

Finally, Figs. 6-a and 6-b show the field distributions of the TE_{01} and TE_{02} modes, respectively.

4. Conclusions

The higher order TE_{02} mode sapphire resonator in a cooled copper enclosure offers significantly higher Q_u than a corresponding TE_{01} mode. Careful optimization of D.R. aspect ratio and the enclosure dimensions is needed to obtain the highest Q's and the biggest fre-

quency separations. The Q_u of the TE_{02} mode resonator can be increased further if superconductor cover plates are used. This type of resonator can be used as the frequency sensitive feedback element in very low phase noise oscillators.

References

[1] M.M. Driscoll et al, "Cooled, Ultrahigh Q, Sapphire Dielectric Resonators for Low Noise, Microwave Signal Generation," to appear in the *IEEE Trans. on MTT*, 1992. (Also presented at the *45th Annual Symposium on Frequency Control*, 1991).

[2] K.A. Zaki and C. Chen, "New Results in Dielectric Loaded Resonator," *IEEE Trans. on Microwave Theory and Techniques*, Vol. MTT-34, pp. 815-826, Dec. 1985.

[3] X.P. Liang, H.C. Chang and K.A. Zaki, "Hybrid Mode Dielectric Resonators on Suspended Substrates," *Proceedings of the 19th European Microwave Conference*, pp. 1141-1146, Sept. 1989.

[4] H.C. Chang and K.A. Zaki, "Unloaded Q's of Axially Asymmetric Modes of Dielectric Resonators," *IEEE MTT-S Int. Microwave Symposium Digest*, pp. 1231-1236, 1989.

[5] X.P. Liang, H.C. Chang, and K.A. Zaki, "Unloaded Q's of Hybrid Mode Dielectric Resonators on Grounded or Suspended Substrates," *Journal of Electromagnetic Waves and Applications*, Vol. 5, No. 3, 281-299, 1991.

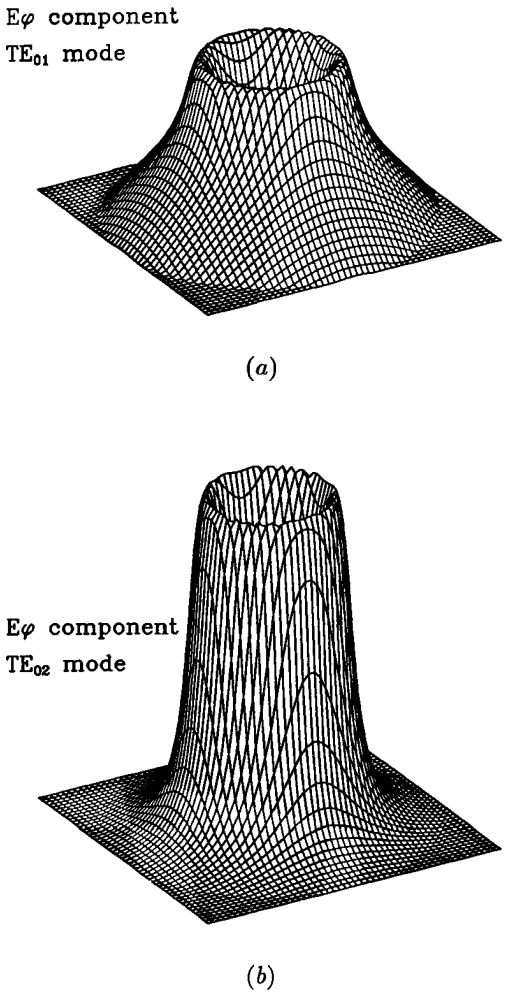


Fig. 6 Field distributions of (a) TE_{01} mode and (b) TE_{02} mode

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ULTRA-STABLE CRYOGENIC SAPPHIRE DIELECTRIC MICROWAVE RESONATORS

A. G. Mann, A. N. Luiten, D. G. Blair, M. J. Buckingham,

Department of Physics, University of Western Australia, Nedlands, W. A. 6009, Australia.

Abstract

This paper reports measurements of the properties of cryogenic sapphire resonators with various concentrations of paramagnetic ions and describes a new way of achieving self compensation of temperature fluctuations for such resonators mounted inside conducting shields. We have shown that frequency-temperature self-compensation, in which the T^4 temperature dependence of the sapphire dielectric constant compensates the effect of the T^{-1} Curie law of paramagnetic impurities, can be achieved at convenient operating temperatures around 6 K. Samples with a few ppm of chromium and iron impurity ions have displayed the expected strong influence on mode frequency and Q of nearby spin resonances at 11.45 GHz (Cr^{3+}) and 12.05 GHz (Fe^{3+}) and, also as expected, fail to compensate above the spin resonances. In samples with much reduced chromium and iron concentrations an ion with a very high frequency paramagnetic resonance (such as Ti^{3+}) has been exploited to obtain essentially frequency-independent self-compensation (up to the limit of our test equipment, 22 GHz), allowing freedom of choice of operating frequency. In such samples we have observed record unloaded Q's as high as 4×10^9 at 4.2 K. Previous self-compensated resonators achieved a relative frequency stability of better than 10^{-14} for integration times from 1 to 300 seconds; with the increased Q factors and reduced wall interactions we now expect substantial improvements in frequency stability.

Introduction

With the development of laser-cooled atomic frequency standards a requirement exists for new "flywheel" oscillators an order of magnitude more stable than state of the art quartz oscillators. Excellent short term stability is increasingly the province of cryogenic oscillators where resonator quality factors are in excess of 10^9. High purity monocrystalline sapphire at cryogenic temperatures is an excellent dielectric for high stability resonator applications owing to its high mechanical rigidity, low thermal expansion and very low dielectric loss. Ultra-stable microwave oscillators such as our Sapphire-Loaded Superconducting Cavity Oscillator (SLSCO) [1, 2] and JPL's Superconducting Cavity Maser (SCM) [3] both employ as the frequency determining element a very high Q sapphire resonator. Whereas the SCM operates in a low order mode the SLSCO utilizes a high order mode with much smaller fields at the shield walls and employs [1] a synthesis of loop oscillator and the form of Pound frequency stabilisation developed to a high degree by Stein and Turneaure [4]. A very important property of both oscillators is the existence of a maximum of mode frequency as a function of temperature and power which allows fractional frequency stabilities of 10^{-14} or better with quite modest temperature and power regulation. A fractional frequency stability of 9×10^{-15}, flat from 1 to 300 seconds integration time, has been reported with a 3 cm diameter sapphire resonator [1]. The frequency-temperature maximum of this resonator is due to the compensating effects of the Curie law temperature dependence of the ac susceptibility of residual paramagnetic impurities in the sapphire and the temperature-dependent surface reactance of the nearby superconducting walls [5, 6]. However the walls significantly limited the Q. To obtain higher Q factors and substantially increase the number of usable modes we have investigated more highly confined modes at X-band in a larger, 5 cm diameter resonator. This results, in contrast to the previously described maximum contributed by the temperature dependence of the superconducting penetration depth, in a contribution from the T^4

0-7803-0476-4/92 $3.00 © 1992 IEEE

temperature dependence of the dielectric constant which now compensates the effect of the T^{-1} Curie law of paramagnetic impurities.

The Resonator

The modes in a cylindrical dielectric resonator may be designated H-type (quasi-TE) or E-type (quasi-TM). Each mode is described by three mode numbers m, n and p characterizing the field variations in the azimuthal, radial and axial directions respectively. We have restricted our study to those modes with the highest field confinement to the dielectric, hence highest potential Q factor, namely those which have large azimuthal and small axial mode numbers; these are the H_{m11} ($TE_{m1\delta}$) and E_{m11} ($TM_{m1\delta}$) modes having m azimuthal field periods and only one for the radial and axial mode numbers. In a 5 cm diameter cylindrical resonator at X-band, m ranges from about 9 to 16 and mode spacings ($\delta m = 1$) are approximately 0.7 and 0.6 GHz for H_{m11} and E_{m11} modes respectively [7]. The two 5 cm diameter cylindrical resonator geometries we have investigated are shown in figure 1. Both are single crystals with their c-axis oriented along the geometric axis (to within 1 degree), optically polished surfaces and have integral (1.5 cm diameter) mounting spindles at each end. Figure 1a with the conventional rectangular cross-section refers to two samples of Union Carbide (UC) 1989 material while Figure 1b with the tapered ends refers to Crystal Systems (CS) "HEMEX" grade 1991 material.

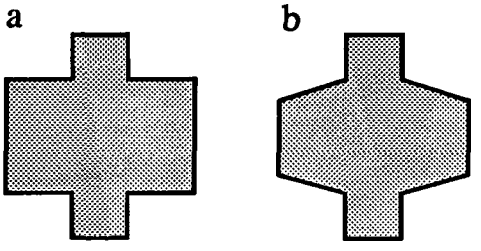


Figure 1 5cm diameter sapphire resonators: a Union Carbide, b Crystals Systems "HEMEX".

The sapphire resonator is mounted inside a cylindrical metallic shield (inside diameter 8cm and height 5 cm) which is housed in a permanently evacuated enclosure and cooled in a liquid helium dewar. Temperature control is provided by a carbon glass thermometer and a heater mounted in thermal contact with the shield. Both UC samples and one CS sample were tested inside copper shields while one CS sample was tested inside a niobium shield. The niobium shield was constructed with knife edges between the central cylinder and end plates to preserve the Q of cavity-like modes with currents crossing the walls so as to prevent Q degradation of nearby sapphire H or E modes.

Coupling to the mode fields was via fixed probes. Except for the resonator with the niobium shield, these were axial electric field probes protruding through holes in the end wall at a radius of about 2.2 cm. This coupling configuration necessarily favored E modes over H modes, but even E modes with the highest Q were undercoupled. For the resonator inside the niobium shield two probes were employed to enable loop oscillator operation on H_{m11} modes. One of these was an magnetic field loop probe in the end wall and the other a radial electric field probe in the center of the cylindrical wall.

The sapphire was cleaned in nitric acid and ethanol baths. It was given a final methanol rinse prior to assembly. The copper shield was mechanically polished and brightened with acidified copper sulphate solution. The niobium shield was cleaned in nitric acid. The assembled resonator was placed in a vacuum pot which was then evacuated and permanently sealed. The pot was attached to a cryostat for cooling to liquid nitrogen or helium temperatures. Mode families can be identified by the fractional frequency shift (ffs), between two reference temperatures (usually 4.2 K and 77 K) or at 4.2 K when an electron spin resonance is saturated [6]. For example, for both geometries of figure 1 the E_{m11} modes exhibit a ffs between 4.2 K and 77 K of 1.95×10^{-4} while the H_{m11} modes exhibit a ffs of 1.25×10^{-4}.

Q factors

We summarize here our measurements to be reported elsewhere [8] on the unloaded Q factors for the highest Q modes at 77 K and 4 K in UC samples with copper walls. The Qs at 77 K are in the range 1.7 to 5×10^7. Most modes are doublets with a splitting of about 1-2 kHz; an order of magnitude larger than their bandwidth at 77 K. The Q factors of the doublet components were equal at 77 K but at 4.2 K differed by a factor of up to 3 in the very highest Q modes. At 4.2 K the highest Q is 1.5×10^9 belonging to the E_{m11} mode family which, at a given frequency, has the largest value of m of any mode family. The Qs attained by the H_{m1p} modes ($p = 1, 2$) however are only a factor of 2 to 3 below this.

The new results for a CS sample with copper walls are represented by a plot of Q as a function of frequency for both E_{m11} and H_{m11} modes, shown in figure 2. We find maximum unloaded Qs (mainly in E modes) of 4×10^9 at 4.2 K compared to only $\sim 1.5 \times 10^9$ in the above UC results. We speculate that this may be attributable to the lower defect density expected from the Heat-Exchange Method compared to the Czochralski growth process. It is probable that the drop in some Q values below a

smooth curve, also observed in the UC resonators, is entirely due to nearby modes which have much lower Qs, for example cavity-like rather than quasi-TE or -TM modes. These have Qs ~ 10^6 determined by wall losses and need only be within a few MHz to pull the Q down to ~ 2 x 10^8.

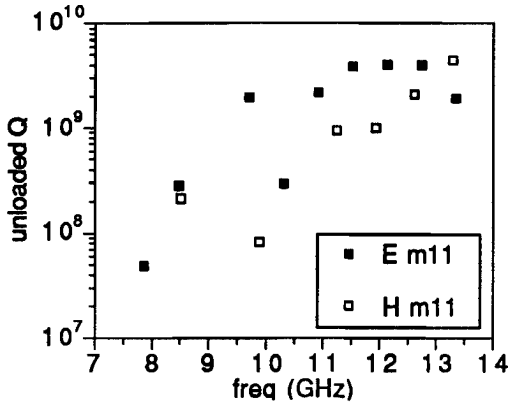


Figure 2 Unloaded Q at 4.2 K as a function of frequency for E_{m11} and H_{m11} mode families in a CS resonator with a copper shield.

The effect of the niobium walls is evident in figure 3 which displays mode Q values of a CS sample as a function of temperature for the H_{1011} and E_{1211} modes near 9 GHz. Across the niobium superconducting transition temperature, at 9.25 K, the Q drops from about 1.5 x 10^9 to 5 x 10^8 for the E mode and 2 x 10^8 for the H mode. Since the BCS surface resistance at 9 GHz drops by at least two orders of magnitude from 9.25 K to 5-6 K [9] we conclude that the superconducting wall losses are negligible at these latter operating temperatures. In this resonator only one mode, H_{1311} (11.25 GHz), is observed to have a Q as high as 4 x 10^9.

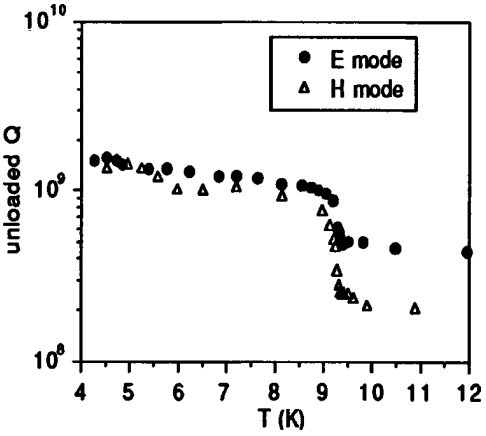


Figure 3 Unloaded Q of E_{1211} (9.1 GHz) and H_{1011} (9.2 GHz) modes in CS sample with a niobium shield.

Paramagnetic effects: In the UC samples the effect of chromium and iron paramagnetic impurities is pronounced. Near 4 K it is observed that modes between approximately 11.2 and 11.6 GHz and near 12.0 GHz are strongly influenced by the electron spin resonances (ESRs) at 11.45 GHz (Cr^{3+}) and 12.05 GHz (Fe^{3+}) respectively, even for paramagnetic ion concentrations of only a few ppm. Modes near these resonances are significantly degraded in Q at 4 K, but at 77 K this is only so if the mode lies inside the ESR linewidth (observed by applying a magnetic field along the sapphire c-axis to move the ESR onto the mode). This diminuition of effect is due to the 1/T dependence of the paramagnetic susceptibility. Furthermore, at 4.2 K these Q values are power dependent due to partial saturation of the ESRs because of the long spin-lattice relaxation time (~ 0.2 s). For example it is observed that the Q of an 11.60 GHz E mode reaches a limiting value of ~ 4 x 10^8 at high power (~ 1mW) but at power around 1μW drops to 2 x 10^7 [8]. In a separate series of experiments the real part of the ac susceptibility has been measured [6], and at constant frequency found to be proportional to a mode-dependent "filling factor", η. This factor is essentially unity for E modes with small axial mode numbers, but for H modes drops to ~ 0.03 for p=1 and 0.15 for p=2. This reflects the fact that H modes have a much smaller component of magnetic field perpendicular to the crystal c-axis which can couple appropriately to the spins. The fractional frequency shifts in E modes at 10 GHz obtained by saturating the chromium or iron ESRs are ~ 4 x 10^{-7} for UC compared to only ~ 10^{-9} for CS HEMEX, implying much lower ion concentrations in the latter of only parts per billion.

Frequency-Temperature Maximum

The temperature dependence of a given mode frequency, f, may be expressed as [6]

$$(f_0 - f) / f_0 = AT^4 + B\,y\,(T) + C/T \qquad (1)$$

where f_0 would be the frequency at absolute zero if no paramagnetic impurities were present. The first term represents the combined temperature dependence of the dielectric constant and the thermal expansion coefficient of the sapphire. A is somewhat mode dependent, being

largest for E_{m11} modes and smallest for H_{m11} modes, the ratio of the two extremes being about 1.5. The second term represents the temperature dependence of the surface reactance (X_S) of the metallic shield, normally only appreciable for superconducting walls. In the latter case [5, 6] the function y(T) may be approximated by the two-fluid model expression $\{1 - (T/T_c)^4\}^{-1/2}$ above half the superconducting transition temperature, T_c. In the third term, the T^{-1} factor represents the Curie law temperature dependence of the susceptibility of paramagnetic impurities in the sapphire. The coefficient C is actually a strong function of frequency, C(ω), and is related to the real part of the ac susceptibility, χ', by [6]:

$$2C(\omega)/[T\eta] = \chi'(\omega) = \sum_{j=1}^{k} \chi_{0j}\, g(\omega, \omega_j) \qquad (2)$$

where η is the mode filling factor and χ_{0j} is a partial dc susceptibility, proportional to the concentration of ions responsible for the jth, of a total of k ESRs, at frequency ω_j. Except for $\omega/2\pi$ within a few tens of MHz of $\omega_j/2\pi$, the frequency dependence is given by $g(\omega, \omega_j) \sim \omega_j/(\omega_j - \omega)$. It is to be noted that g changes sign and therefore frequency-temperature compensation does not occur above the ESR of the dominant paramagnetic species.

If the field confinement is sufficiently high that the shield has negligible influence then frequency-temperature "self-compensation" can occur. That is, the temperature dependence of the sapphire dielectric constant (proportional to T^4) compensates the Curie law of paramagnetic impurities, provided the mode frequency is below the ESR of the dominant paramagnetic species. The temperature at the frequency maximum, T_m, is then equal to $(C/4A)^{1/5}$. It is thus proportional to the fifth root of the real part of the ac susceptibility (proportional to ion concentration and filling factor), whereas the second derivative of frequency with respect to temperature is proportional to the two-fifths power of the same. As illustrated in figure 4, T_m is a factor of about 1.5 larger in E modes than in H modes, mainly reflecting the large ratio of filling factors (about 20:1) between these modes.

In UC samples with ~ 1 ppm of chromium we have found that frequency-temperature "self-compensation" can be achieved below 11.2 GHz and at convenient operating temperatures ~ 6 K in H modes [6].

In samples with much reduced chromium and iron concentrations an ion with a very high frequency paramagnetic resonance (such as Ti^{3+}) can be used to obtain essentially frequency independent self-compensation ($g(\omega, \omega_j) \rightarrow 1$), allowing freedom of choice of operating frequency. T_m then only depends on the mode type through the filling factor. Such a situation occurs in both our CS samples, in which we have a Ti^{3+} concentration of about 1 ppm. We observe that self-compensation occurs for all microwave modes up to the limit of our test equipment (22 GHz). This is illustrated in figure 5 for E modes between 4.7 and 22 GHz, where the temperature maximum is observed around 9-10 K.

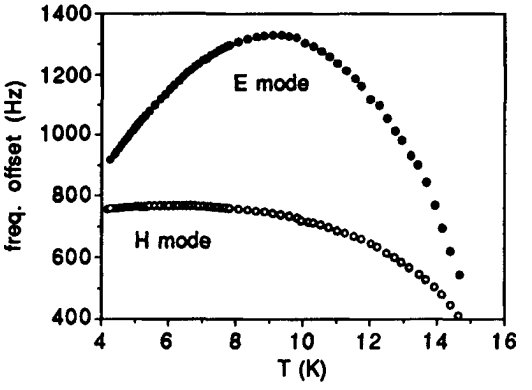


Figure 4 Temperature dependence of mode frequency for E_{1311} (9.7 GHz) and H_{1011} (9.2 GHz) modes in a CS resonator with a copper shield.

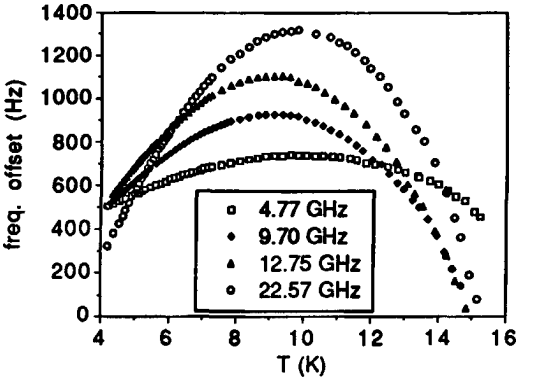


Figure 5 Temperature dependence of mode frequency for E modes in a CS resonator with a copper shield.

Frequency-Power Dependence

In the original 3cm resonator a maximum in H_{611} mode frequency versus power was observed at a few mW and can be explained by two or more competing effects: partial (self-) saturation of the ESRs by microwave power at the mode frequency (9.7GHz) which dominates in the low

power regime, and effects linear in the power that dominate at high power. The latter include differential heating between the sapphire and the niobium walls and radiation pressure-induced permittivity changes [10]. The self-saturation can be modelled [11] as a factor slightly less than unity multiplying $g(\omega, \omega_j)$ in equation (2) which changes by only a few percent between zero and maximum mode power. The slope ($\sim 10^{-8}$ W^{-1}) of relative frequency shift at high power is consistent with a thermal boundary resistance between sapphire spindles and niobium end walls of a few K per Watt which is quite reasonable and probably well above the radiation pressure term. In practice one operates at a power near the turning point, which has a second derivative of 3×10^{-8} W^{-2}.

In a 5cm sapphire resonator with much smaller field at the walls (estimated at $< 10^{-1}$ that of the 3 cm resonator) the differential heating term should be correspondingly smaller and probably less than the radiation pressure term. In HEMEX samples, the nearby ESRs are very weak and easily saturated while the very high frequency ESR is so far from the frequencies of interest that it cannot be saturated, hence self-saturation effects are expected to be very small.

Conclusions

We have shown that in a 5 cm resonator Q values of 4×10^9 at 4.2 K, not limited by wall losses, are possible. Frequency-temperature self-compensation, in which the Curie law temperature dependence balances that due to the dielectric constant of the sapphire, can be achieved at convenient operating temperatures. In samples with much reduced chromium and iron concentrations an ion with a very high frequency ESR such as Ti^{3+} can be used to obtain frequency-independent self-compensation.

Acknowledgements

The authors would like to thank Adrian Giles, Marco Costa and Mike Tobar for their general assistance. This research 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] 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, pp. 145-146, 1990.

[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.Ultrason. Ferroelec. Freq. Contr., vol. 37, pp 339-346, Sept. 1990.

[4] S. R. Stein and J. P. Turneaure, "The Development of the Superconducting Cavity Stabilized Oscillator", Proc. 27th Annual Frequency Control Symposium, 1973, p414 .

[5] S.K. Jones, D.G. Blair and M.J. Buckingham, "Effect of paramagnetic impurities on frequency of sapphire-loaded superconducting cavity resonator", Electron Lett. vol 24 , pp 346-347, 1988.

[6] 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., vol 25, 1992, in press.

[7] Michael E. Tobar and Anthony G. Mann, "Resonant frequencies of higher order modes in cylindrical anisotropic dielectric resonators", IEEE Trans.Microwave Theory and Techn., vol. 39, pp 2077-2082, Dec. 1991.

[8]A.G.Mann, A.G.Truscott, D.G. Blair and E.N.Ivanov, "Ultra-low loss cryogenic sapphire loaded copper cavity resonators", in preparation.

[9] V.B. Braginskii, V.P. Mitrafanov and V.I. Panov Systems with Small Dissipation. Chicago:University of Chicago, 1985, p 78.

[10] C.M. Lyneis Experimental Studies of the Microwave Properties of 8.6 GHz Superconducting Niobium, Niobium-Tantalum and Tantalum Cavities, Ph.D. Thesis, Stanford University, Stanford, CA, 1974, unpublished, pp 39-42.

[11] A.J. Giles Electron Spin Resonance Phenomena in a Sapphire Loaded Superconducting Cavity Oscillator, Ph.D. Thesis, Physics Dept., University of Western Australia, 1992, unpublished.

ENVIRONMENTAL SENSITIVITY OF THE SUPERCONDUCTING CAVITY MASER OSCILLATOR*

R. T. Wang and G. J. Dick
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

The Superconducting Cavity Maser Oscillator (SCMO) was tested for environmental sensitivity in the controlled environment of JPL's Frequency Standards Test Facility (FSTF). Measured sensitivities to temperature, magnetic field, pressure and humidity were than used to predict SCMO performance in various laboratory environments. A comparison to SCMO frequency stability tests in these same environments shows good results, and indicates a need to reduce sensitivities to temperature and magnetic field. Characterization provides continual evaluation of SCMO for the next generation of frequency standards for NASA's Deep Space Network.

Introduction

The Superconducting Cavity Maser Oscillator (SCMO) has demonstrated a short term stability of $1 \cdot 10^{-14}$ at one second[1,2,3], a factor of ten better than a hydrogen maser. The SCMO is an all-cryogenic, helium cooled oscillator with a lead-plated sapphire cavity and an ultra-low noise cryogenic ruby maser. A second unit is under construction to demonstrate a calculated performance of $3 \cdot 10^{-15}$ at one second.

Environmental tests were performed to necessitate a second SCMO with reduced sensitivities. This test integrated four parameters: temperature, magnetic field, barometric pressure and relative humidity. Our dual purposes are to interrogate the intrinsic behavior of the whole system and to evaluate critical subsystems to improve overall performance.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

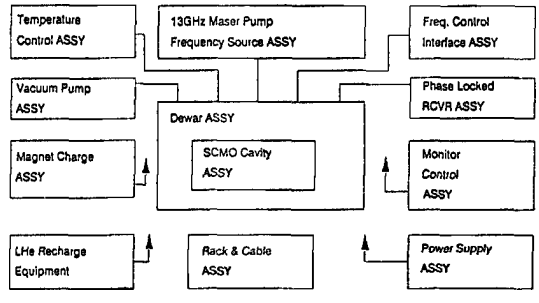


Figure 1: SCMO subsystems.

In this paper we report test procedures and results. Figure 1 shows SCMO major subsystems, in particular, physics unit and electronics. The physics unit consists of a three-cavity oscillator operated at 1.5 Kelvin in a liquid helium dewar which has a built-in magnetic shield. Electronics are made up of a conductance bridge, a 13 GHz microwave synthesizer and other supporting parts. SCMO characterization was performed at Jet Propulsion Laboratory (JPL) Frequency Standard Test Facility in an environmentally controlled chamber. These tests provide insight on the long term stability of SCMO and on variations of operational frequency which may be expected under different environmental conditions.

Test Plans

Sensitivities were measured from frequency difference of SCMO and a hydrogen maser at 100MHz. The hydrogen maser was chosen as reference because of a much lower drift rate, approximately 100

0-7803-0476-4/92 $3.00 © 1992 IEEE

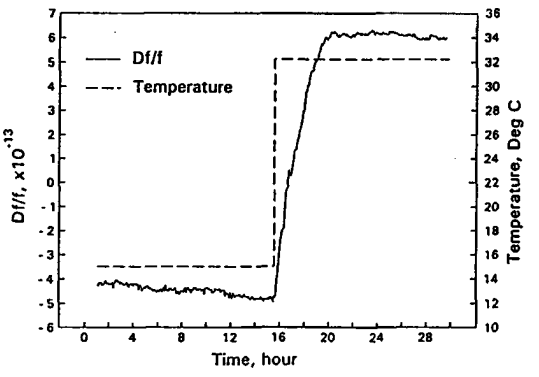


Figure 2: Example of temperature test. Data shown frequency difference of SCMO and a hydrogen maser at 100MHz.

times lower than SCMO. The frequency difference was monitored and recorded on a HP 1000 computer for analysis. Detail description of test facility and measurement system can be found in reference [4]. In each test, only one parameter was varied and others were kept constant. Tests were done on the whole system and major subsystems. During subsystem tests, a particular subsystem was isolated in the chamber while the remaining subsystems were operated outside of the chamber. Preliminary tests, test procedures and some precautions are discussed in the following.

1. Temperature: Although the chamber temperature can be varied between 15 and 35 degrees C with step size as small as 0.01 deg C, we noticed for temperature above 33 deg C the microwave power of the 13 GHz source dropped below threshold. Chamber temperature was then cycled between 15 and 32.5 deg C every 24 hours; for each fixed temperature we also measured the temperature standard deviation to be 11 millidegree C. Figure 2 shows an example of temperature sensitivity test of the whole system. A time constant of four hours was identified and it was used to calculate our test cycle of 24 hours.

2. Magnetic field: A helmholtz coil is used to generated a static DC magnetic field. Tests were done both vertically and horizontally. No change was observed for from horizontal magnetic field variations. For simplicity the rest of our discussion will be on vertical direction only. The increment steps are 50, 100 and 200 mG, both positive and negative direction alternatively. There are two reasons to choose such increment steps. First, the step will give a clear frequency shift which is above SCMO daily drift rate. Second, the steps are much large than daily magnetic field variation ($\approx$ 1mG) to keep data from contamination. Magnetic field was always set back to zero to check for background before each increment and, at end of each increment, the system was allowed to stabilize for one hour before next change.

3. Pressure: Chamber control range is 48 inches of water which is equivalent to 12 percent of atmosphere change or 760 $\pm$ 45 mmHg. Test pressure was cycled between +10" and -10" of water every 12 hours. According to a JPL barometric record, the maximum variation for the past 10 years is 12" of water a value which is smaller than our test range of 20" of water.

4. Relative Humidity: Humidity was stepped from 10% to 30%, 50% and 70% each with a duration of 8 hours and temperature was fixed at 20 degree C. We did not go above 70% humidity to avoid possible water condensation which might cause permanent damage to the electronics.

Test Results and Analysis

Test results are tabulated in Table I and II. Table I lists major subsystem sensitivities in comparison with the whole unit. Clearly humidity has no obserable effect. Magnetic field only seems to impact the physics unit which is no surprise because the ruby maser is operated with a DC magnetic field of 500 Gauss. Both temperature and pressure sensitivity are seeing some competing effect from different electronics. It's interesting to note that the sum of subsystems sensitivities is approximately equal to the overall system response. This gives us confidence that no critical subsystem has been omitted.

Table II compares sensitivities of SCMO and a hydrogen maser[5]. Clearly temperature and magnetic field sensitivities are the two areas for improvement.

To understand the long term impact of any particular environmental parameter, we took a different approach. There are three steps in this analysis. (1) We measured the time dependence of temperature, magnetic field and pressure at three different locations. (2) Allan Deviation of these parameters were calculated from measurements[6]. Since this may be a new approach to readers, it deserves a word of explanation. For example, one can generate an Allan Deviation σ_T (τ) from a data set T(t) for the time dependence of temperature T in the same way that we conventionally generate the Allan Deviation of

fractional frequency σ_y (τ) from a data set of measurements y(t) where y≡δf/f. Instead of Allan Deviation of frequency, one would obtain Allan Deviation of temperature. (3) Multiplied by the appropriate SCMO sensitivity from Table I, we obtain predicted SCMO frequency deviation contributed from different environmental parameter for various measuring times. One word of caution, due to various time constants of each system, this exercise can only be applied at longer measuring times or else the interpretation will not be valid.

Figure 3 shows a comparison of predictions and measured SCMO Allan Deviations of fractional frequency at three locations. SCMO measurements represent the collective environmental effects on SCMO. At location No.1, long term stability clearly limited by temperature instability. Not only so, the frequency stability curve turns up for longer times at the same slope as temperature. Location No.2 has better temperature control and the SCMO long term stability slightly improved. Location No.3, our environmentally controlled chamber, has the best temperature control but with more magnetic disturbance due to equipment transfer in an adjacent chamber during the course of the measurement. Stability didn't seem to be better than location No.2 even though this would be expected from the curves of B,T and P. This may indicated an increased value for magnetic field sensitivity compared to the value given in Table I.

While this approach is in a preliminary stage of development, we believe it has important advantages for understanding and predicting the long-term stability of ultra-stable oscillators in any environment.

Conclusions

We have performed environmental sensitivities tests on temperature, magnetic field, pressure and humidity. Temperature and magnetic field are two major contributors to long term instability and their effects can be limited with modest modifications.

For temperature sensitivity, one immediate improvement is to add a temperature control unit to the electronics. Temperature stability improves by a factor of 2. Also a 13 GHz synthesizer is being built to provide a 10 times less temperature sensitivity. On the magnetic field sensitivity, a newly designed magnetic shield will provide a 20 times more shielding.

Acknowledgements

Special thanks to W. A. Diener for his assistance in sensitivity measurements, R. E. Taylor for easy access to the capabilities of the Test Facility and Dr. C. A. Greenhall for software enhancements meeting our special requirements and calculation of parameter Allan deviation. The authors would like to thank Dr. L. Maleki for many helpful suggestions.

References

[1] S. Thakoor, D. M. Strayer, G. J. Dick and J. E. Mercereau, "A Lead–on–Sapphire Superconducting Cavity of Superior Quality," *J. Appl. Phys.*, **59**, 854–858 (1986).

[2] R. T. Wang and G. J. Dick, "Improved Performance of the Superconducting Cavity Maser at Short Measuring Time," in *Proceedings of the 44th Annual Frequency Control Symposium*, 89–93 (1990).

[3] R. T. Wang and G. J. Dick, "Combined Stability of the Hydrogen Maser and Superconducting Cavity Maser Oscillator," in *Proceedings of the 45th Annual Frequency Control Symposium*, 491–494 (1991).

[4] R. Sydnor, "Environmental Testing at the Jet Propulsion Laboratory's Frequency Standard Laboratory," in *Proceedings of the 43th Annual Frequency Control Symposium*, 289–295 (1989).

[5] A. Kirk, P. Kuhnle and R. Sydnor,"Evaluation of Hydrogen Maser," in *Proceedings of the 14th Annual Precise Time and Time interval (PTTI) Planning and Applications Meeting*, 359–392 (1982).

[6] Private communication with Dr. C. A. Greenhall. Calculation performed by Dr. C. A. Greenhall.

Table 1: Environmental sensitivity of SCMO subsystems.

Sensitivity	Temperature (per deg C)	Magnetic Field (per mG)	Pressure (per inches of water)	Humidity (per %)
Whole System	$6.3 \cdot 10^{-14}$	$1.2 \cdot 10^{-15}$	$2.2 \cdot 10^{-14}$	No Change
Physics Unit	No Change	$1.2 \cdot 10^{-15}$	$3.2 \cdot 10^{-14}$	No Change
Electronics	$8.3 \cdot 10^{-14}$	No Change	$-1.0 \cdot 10^{-14}$	No Change
Synthesizer	$3.8 \cdot 10^{-14}$	No Change	$-0.5 \cdot 10^{-14}$	No Change
Bridge	$-4.1 \cdot 10^{-15}$	No Change	No Change	No Change

Table 2: Comparison of SCMO and hydrogen maser.

Parameter	SCMO	Comparison with Hydrogen Maser
Temperature	$6.3 \cdot 10^{-14}/degC$	x6 worse
Magnetic Field	$1.2 \cdot 10^{-15}/mG$	x5 worse
Pressure	$2.2 \cdot 10^{-14}/"water$	x4 better
Humidity	No change	Same

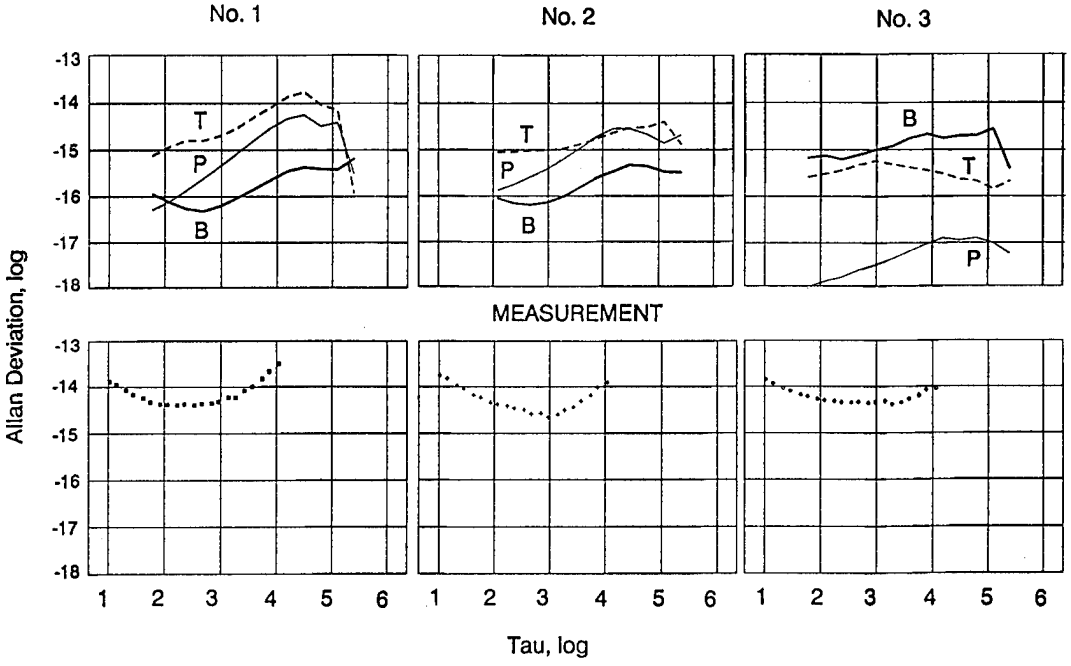


Figure 3: Comparison of measured SCMO Allan deviations and the calculated Allan deviation from environmental parameters at three locations. Parameters stability are varied as much as 100 time. Allan deviations are clearly limited by the worse case of the three parameters.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

MICROWAVE FREQUENCY DISCRIMINATOR WITH A COOLED SAPPHIRE RESONATOR FOR ULTRA-LOW PHASE NOISE*

David G. Santiago and G. John Dick
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Pasadena, California 91009, U.S.A.

ABSTRACT

First results are presented for an X-band frequency discriminator using a cooled sapphire microwave resonator. These results show a lower close-in (1 Hz - 1 kHz offset) phase noise measurement floor than any oscillator presently available. This performance is made possible by a sapphire "whispering gallery" mode resonator which shows the highest quality factor (with Q's up to 30 million) of any rf, microwave, or acoustic resonator at temperatures down to 77K. Performance is further increased by novel phase detection circuitry. The sapphire discriminator was used to characterize the phase noise of a single crystal quartz oscillator of the highest quality, without the use of a second similar oscillator as reference.

INTRODUCTION

The fundamental instrument for spacecraft radio science experiments is the ultra-stable oscillator (USO). The USO provides signals with spectral purity and stability to obtain data on fluctuations in phase, frequency, and intensity of radio signals as they propagate through intervening media before Earth reception[1]. The data yields information on planetary atmospheres and ionospheres, as well as information on planetary interiors by observing the perturbation of signals due to gravity fields. The spectral purity of the signal derived from the USO determines the resolution and accuracy of measured parameters in such radio science investigations.

Use of a sapphire phase stabilizer (SPS) operating at X-band or higher corrects phase fluctuations in the multiplied frequency of the USO[2]. The SPS employs a critically coupled sapphire resonator operating with a Q of 15 million at 80 Kelvin to potentially provide a 30 dB or greater improvement in phase noise over the best conventional sources at offset frequencies of 1 Hz and higher. For comparison, the Q of the best quartz crystal oscillators is about 2 million.

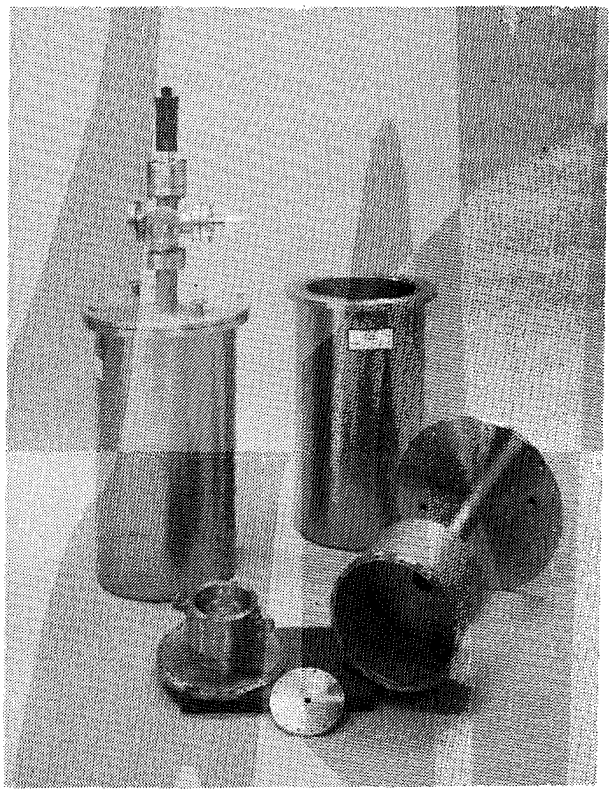


Figure 1. Cryogenic components for the two sapphire resonator subsystems.

The SPS is based on a novel technology consisting of a cooled sapphire "whispering gallery" mode X-band resonator operating in the temperature range from 70 Kelvin to 300 Kelvin - values achievable by means of radiative and thermoelectric cooling.

This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

We demonstrate for the first time a measurement of the phase noise of the quietest available quartz oscillator with a noise floor reduced by 10 to 45 dB for offset frequencies from 1 Hz to 1000 Hz. For operation at 80 Kelvin, this noise floor is almost exactly as predicted for the measured Q of 6 million; it would be reduced by another 8 dB at the design Q of 15 million.

Also significant are room-temperature tests using new rf circuitry[3]. These tests show noise further reduced by approximately 10 dB compared to the conventional configuration. This circuitry is also applicable to the 80 Kelvin SPS.

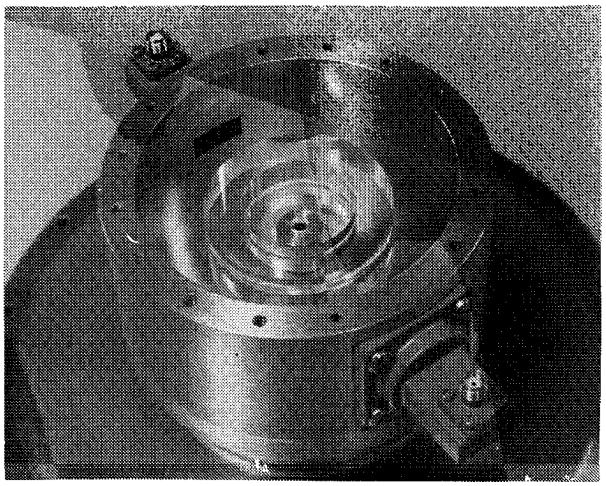


Figure 2. Close-up view of the sapphire resonator with cover removed.

Resonator Characteristics

Figure 2 shows the sapphire element whhich is 5 cm in diameter and 2 cm high in the approximately 7.6 cm diameter copper containing can. The $WGH_{10,0,0}$ mode of the sapphire wheel has a resonant frequency of 7.85 GHz at room temperature and 7.95 GHz at 77K.

Dielectric losses in sapphire are rapidly reduced as its temperature is cooled below ambient, showing an approximate T^{-5} dependence down to 60 Kelvin[4-7]. While metallic losses limit the achievable Q values in most microwave resonators, recently developed "whispering gallery" mode sapphire resonators confine the resonant microwave energy in the sapphire, and so allow the inherent Q value of the sapphire to be realized[4-9]. Unloaded Q values at 8-10 GHz (X-band) for these resonators are approximately $2 \cdot 10^5$ at room temperature, increasing to $3 \cdot 10^7$ at 80K.

The thermal coefficient of frequency of the sapphire resonators has also been studied and has a value of $5 \cdot 10^{-5}$ at room temperature reducing to $3 \cdot 10^{-6}$ at 80K[7]. Even this reduced coefficient at 80K limits the resonator's applicability as a frequency stabilizing element. However, application as a phase stabilizer is not adversely effected for the relatively higher fluctuation frequencies above 1 Hz, where the thermal mass of the sapphire resonator effectively stabilizes its operating frequency.

Phase Noise Considerations

Phase noise performance for rf and microwave cavity and DRO oscillators at small offsets (f < 1000 Hz) is determined by resonator Q and the 1/f noise level in the active device[10-13]. For slow phase fluctuations ($f \ll \nu/Q$), the characteristic phase slope of the resonator $\partial\phi/\partial\nu = 2Q/\nu$ implies a corresponding slow fluctuation in the frequency of the oscillator. Here f represents the fluctuation frequency, ν the microwave frequency, Q the quality factor of the resonator and ϕ the phase of the microwave signal. In this way, the power spectral density of phase fluctuations for the active device $S_\phi(f)|_a$ results in an oscillator output frequency noise spectral density of

$$S_y(f)|_{out} = \frac{S_\phi(f)|_a}{(2Q)^2} \qquad (1)$$

or the mathematically equivalent output phase noise

$$S_\phi(f)|_{out} = \frac{\nu^2}{f^2}\frac{S_\phi(f)|_a}{(2Q)^2} \qquad (2)$$

where $y \equiv \delta\nu/\nu$ is the fractional frequency deviation.

The use of a stabilized local oscillator (STALO) configuration allows a passive mixer to take the place of the rf amplifier or transistor used in simpler oscillator configurations. While somewhat more complicated due to the use of feedback at baseband (frequencies nominally zero) rather than at rf, the low noise of the mixer ($S_\phi(1\text{ Hz})|_a \approx$ -135dB) at X-band equals that of active devices at lower frequencies and essentially eliminates any performance disadvantage for high frequency operation[10-13].

Further improvement of the phase noise performance of sapphire stabilized oscillators (at room and cryogenic temperatures) is made possible by

carrier suppression circuitry. We have developed oscillator configurations which reduce the effect of semiconductor 1/f phase noise sources for stabilized local oscillator circuits. A small signal is produced which suffers no loss of loop phase information or signal-to-noise ratio. This small signal can then be amplified without degradation due to multiplicative amplifier noise and detected without saturation of the detector. This circuitry can give a 20 dB or greater reduction in phase noise.

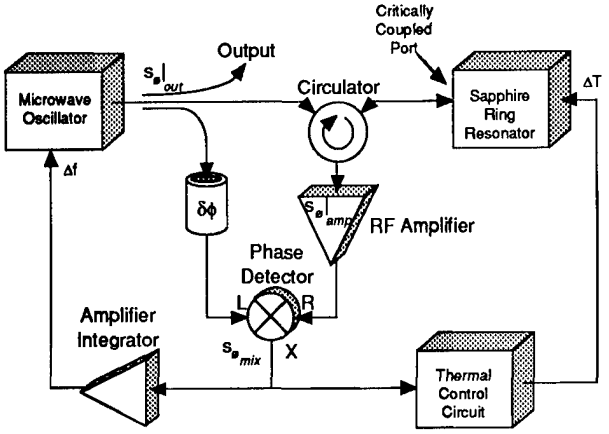


Figure 3. SPS configuration for reduced phase noise. Critical coupling to resonator and operation very near ν_0 allow insensitivity to phase noise of amplifier. Amplifier gain allows reduction of phase detector noise.

SPS Design using Suppressed Carrier

Figure 3 shows a version of the STALO design which can result in improved performance[3]. A small (nominally zero) signal returned from the resonator is amplified before it enters the mixer. Two effects of this addition are easy to understand. The loop gain is increased by the added gain and may be compensated for in the design of the base-band amplifier. Secondly, the gain of the amplifier increases the sensitivity of the mixer output to phase error in the resonator without significantly affecting mixer noise. Thus the effective mixer phase noise is reduced by the amount of amplifier gain. This can be a very substantial improvement.

A third effect of this modification is a little more complicated. Since flicker noise in X-band amplifiers is somewhat greater than for mixers (-120 dB vs -135 dB at 1 Hz offset), a crucial point is the proper analysis of the contribution of amplifier noise. The kind of noise under discussion is not additive noise, which is independent of any large signal also present, but instead is multiplicative noise which transforms a large signal by slightly modifying its amplitude and phase. (Additive noise is typically insignificant except at offset frequencies f > ≈ 10 kHz.)

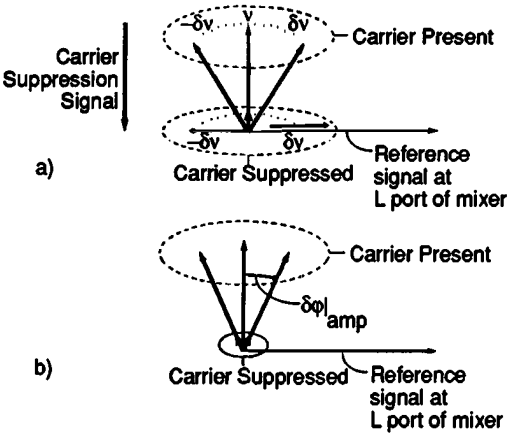


Figure 4. (a) Phasor diagram showing effect of LO frequency error on rf resonator signal with and without carrier suppression. Also shown are (constant) carrier suppression signal and mixer reference signal. Component of resonator signal in phase with reference is unchanged by 20 dB carrier suppression.

(b) Phasor diagram showing effect of amplifier phase noise on resonator signal. Reduction of carrier by 20 dB results in similar reduction in signal error in phase with the mixer reference. (Horizontal scale for a) and b) both greatly enhanced)

Figure 4(a) shows the cavity signals in phasor form, with and without carrier suppression. The added signal due to small frequency variations in the local oscillator (LO) is also shown. These added signals detected by the mixer allow feedback circuitry to cancel the frequency variations. The effect of multiplicative phase noise in the amplifier for both cases is shown in Figure 4(b). It is clear from the figure that multiplicative noise generates signals indistinguishable from those caused by actual frequency variations, but are only generated in the presence of a large carrier. Thus there is a reduction in amplifier noise proportional to the degree of carrier suppression of the microwave signal.

Thus, oscillator phase noise is determined by a combination of mixer noise (reduced by amplifier gain) and amplifier noise (reduced by the degree of carrier suppression). For example, if mixer noise of -135 dB/f(/Hz) is reduced by 25 dB of amplifier gain to -160 dB/f(/Hz), and amplifier noise of -120 dB/f(/Hz) is reduced by 40 dB of carrier suppression to the same value, the combined noise of $S_\phi(f)|_a$ = -157 dB/f(/Hz) would determine oscillator performance as given by Equation 2.

EXPERIMENTAL ASPECTS

Here we only briefly describe significant experimental elements, since design of the setup has been completely reported elsewhere[2]. A small commercial Dewar was chosen for reliability and low cost. It includes a stainless steel vacuum insert with a copper bottom for resonator cooling. The 80K operating point is held above the 77K liquid nitrogen boiling point by a combination of rf and dc heating with detected rf feedback and thermal fluctuations are reduced by a 400 second thermal time constant. The waveguide coupling ports are below cutoff (non-propagating) at the 8.1 GHz operating frequency, but the desired coupling is achieved by easily machining teflon (ϵ = 2) inserts to reduce the cutoff frequency.

We have verified all aspects of component performance including vacuum integrity, thermal time constants, mixer noise levels, and resonator frequency, Q, coupling constants and mode purity. The sapphire resonator element was purchased slightly larger than the design value to allow tuning to frequency by grinding. The initial frequency is approximately 7.94 GHz at 80 Kelvin. All tests to date have been done at this frequency. After tuning to within 0.1% by grinding, a thin (0.1 cm) sapphire tuning disk will be installed above the resonator to bring the frequency within the required 0.001% for operation. Final tuning to the 8.1 GHz design frequency can then be accomplished by variation of the operating temperature by $\approx$ 1 Kelvin.

The waveguide coupling ports had approximately 2% of critical coupling before installation of their teflon inserts, and under this circumstance an unloaded resonator Q of 30 million was demonstrated. A first try at insert design created an overcoupled condition, with a loaded Q of 6 million for two ports. The phase-noise measurements described below were made under these conditions. A single-port configuration is probably more advantageous, allowing a critically coupled Q of 15 million. Subsequent tests will thus be done with only one teflon insert, the second port being only weakly coupled.

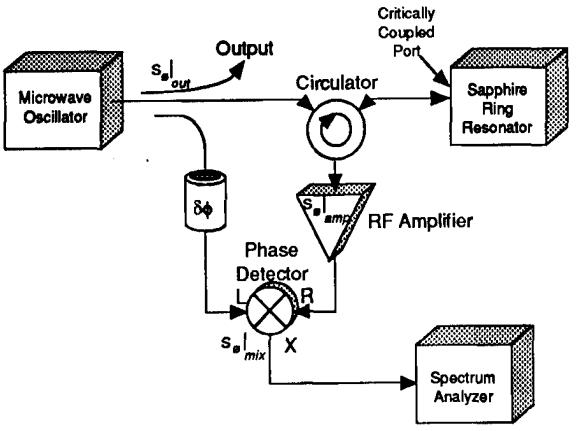


Figure 5. Measurement setup for open-loop test of sapphire phase stabilizer, configured as an ultra-low noise frequency discriminator.

Phase Noise Measurements

Open-loop tests of the SPS as a frequency discriminator were made at both 80 Kelvin and 300 Kelvin using the circuit shown in Fig. 5 with the circulator replaced by a 3 dB hybrid coupler. The hybrid allowed lower noise operation at the expense of signal strength. These tests show (at 80K) lower phase noise than the quietest available quartz crystal oscillator, and agree with the expected mixer noise of -135 dB/f(/Hz) and the measured Q of 6 million.

Figure 6 shows a measurement of phase noise using the cryogenic discriminator. The ultra-low noise quartz crystal oscillator[14] under test is incorporated in a clean-up loop (CUL) for the NASA VLG-10 hydrogen masers. Previous tests of the VLG-10 CUL required comparing two units, and dividing the resulting measured noise by 2 (-3dB) to compensate for the noise of the second oscillator. The SPS discriminator allowed the noise of a single unit to be characterized for the first time, and gave results in excellent agreement with previous tests of the oscillator.

These measurements were made in spite of the relatively low Q, poor mode purity, overcoupled ports, and unadjusted frequency as described above, and yet already exhibit a noise floor 10 dB to 70 dB below the VLG-10 CUL. When the resonator frequency is properly adjusted and critically coupled, we expect an improvement of approximately 10 dB.

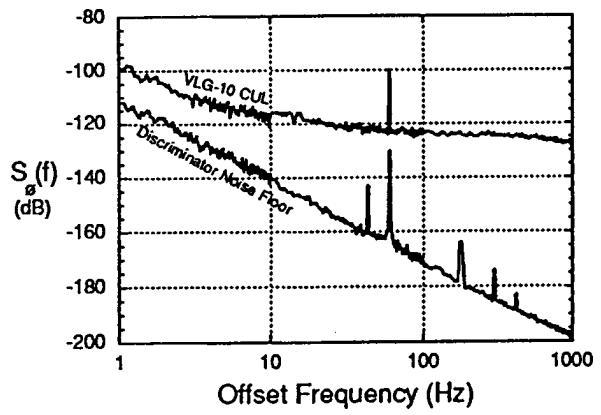


Figure 6. Phase noise characterization at 100MHz of an ultra-low noise quartz crystal oscillator using the frequency discriminator shown in Fig. 5 operating at 80 Kelvin with a loaded Q of 6 million. This represents the first time that this phase noise could be characterized without subtracting a similar noise contribution due to the reference oscillator.

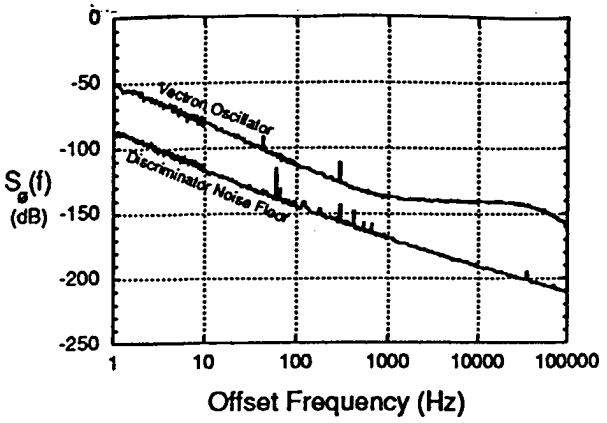


Figure 7. Phase noise characterization of a 100 MHz Vectron oscillator using a frequency discriminator shown in Fig. 5 operating at 300 Kelvin with a loaded Q of 150,000.

While cryogenic results have yet to show an effective reduction of mixer noise due to carrier suppression techniques, room temperature tests demonstrated substantial improvement. Greater freedom of access at room temperature allowed the circuitry to be more easily optimized, resulting in a 10 dB reduction in mixer noise. Figure 7 shows the results of these tests using a critically coupled resonator with a Q of 150,000. We were able to tune the resonator to 7.800 GHz in this case by the addition of two sapphire tuning disks, and thereby eliminate a synthesizer which may have degraded the low temperature measurements. We could also precisely adjust to the critically coupled condition, allowing performance optimization not possible in the cryogenic tests. The Vectron 100 MHz oscillator, with its low noise at high offset frequencies, passes a relatively small noise signal through the rf amplifier shown in Figure 5. In contrast, measurements using the cryogenic resonator involved noise levels which severely taxed the amplifier's dynamic range, resulting in uncompensated high-frequency noise that may have degraded the system.

The room temperature noise floor corresponds to an active device noise of $S_\phi(1\text{ Hz})|_a \approx$ -145 dB in Eq. (2) - a value 10 dB lower than expected for mixer noise, and so shows the efficacy of the carrier suppression circuitry.

Extrapolating the noise floor of the room temperature measurements to the loaded Q of 15 million expected at 80 Kelvin suggests an improvement of almost 20 dB below the floor shown in Figure 6; this noise floor would be 30 dB and more below that of the ultra-quiet quartz oscillator incorporated in the VLG-10 clean up loop.

Mode Purity

The $WGH_{10,0,0}$ whispering gallery modes occur in degenerate mode pairs common to circularly-symmetric waveguide structures. At 80 Kelvin we observe that geometrical imperfections in the sapphire geometry lift the degeneracy and produce a mode pair frequency split by several kilohertz (about a part in 10^6 difference). Because the sapphire alone determines the modes, its angular orientation determines mode alignment with respect to the coupling port. Variation in sapphire mounting from run to run gives rise to a variable coupling coefficient. Under optimal conditions, only the strongly coupled mode shows a reduced, loaded Q, while the other mode should show the intrinsic Q if it is of detectable amplitude. In contrast, our observations showed the "uncoupled" mode had coupling typically reduced by only $\approx$50%.

Mounting the sapphire in the proper angular position would require a difficult trial and error method. Therefore, to remove this dependence on details of the installation, a perturbing element which dominates the imperfections in the sapphire was installed in the SPS as in Figure 8. The mode splitter was designed to provide relatively large frequency splitting ($\approx$54 kHz compared to only $\approx$3 kHz due to

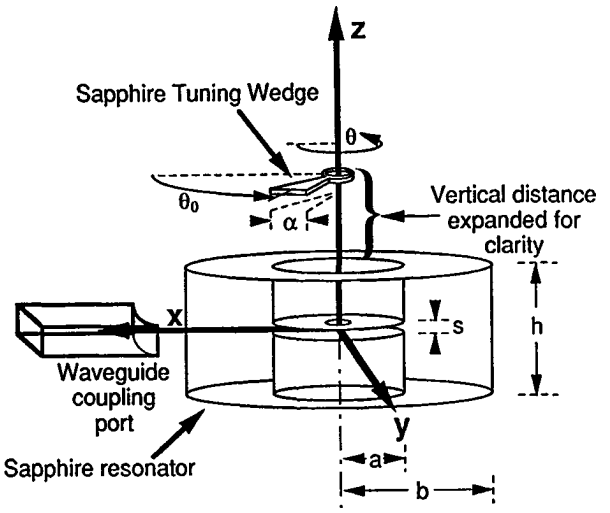


Figure 8. Sapphire resonator geometry. The sapphire tuning wedge controls mode alignment and resonant frequency.

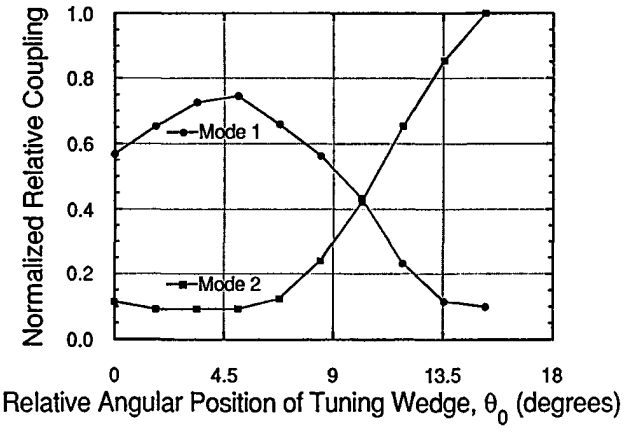


Figure 9. Mode coupling strength as a function of angular position of the tuning wedge.

the imperfections) required for mode control. Variable angle control was used to align one of the modes with the coupling port; the other mode, rotated by 90 degrees, would have nominally zero coupling.

As can be seen in Figure 9, adjusting the angular position of the mode splitter in cryogenic tests controlled mode alignment with the coupling port. The adjustment gave critical coupling to the chosen mode and, importantly, did not degrade resonator Q. This should allow us to reproducibly critically couple to the resonator for the design Q of 15 million at 80K.

ANALYSIS

These results clearly establish the credibility of liquid nitrogen temperature sapphire technology. With open loop operation showing excellent noise floor performance, systematics involved in closed loop operation must now be addressed. Because these phase noise levels are below those obtainable any other way, one unit must act as a reference for a second, and phase comparisons will probably need to be done at X-band rather than at 100 MHz as was done so far.

Some problem areas are bypassed by the open loop measurements. For example, various sources of phase fluctuation which are below the VLG-10 noise level may still be above that of the noise floor. Effects due to uncorrected noise in the VLG-10 at high offset frequencies may be eliminated by phase compensation by means of an X-band phase shifter. Other problems include AM noise in the signal applied to the SPS, AM noise in the harmonics of the 8.1 GHz carrier, circulator non-linearities and noise, and PM-AM conversion in the X-band phase shifter.

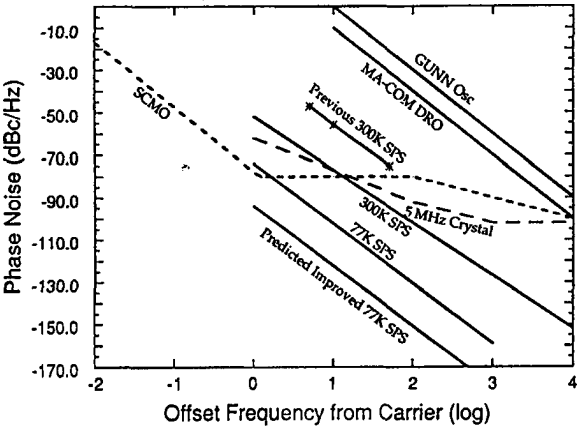


Figure 10. Phase noise, referred to 10 GHz, for various frequency sources including the results from Figs. 6 and 7 as well as our prediction for performance of the 77K SPS.

Figure 10 compares the data shown in Figs. 6 and 7 with phase noise data at X-band for various low noise sources, and with predicted performance for the 80K SPS. Our previously reported results are also shown[2]. The cryogenic sources show significant advantages over the entire range of offset frequencies ($0.01\ \text{Hz} \leq f \leq 10{,}000\ \text{Hz}$), with SPS performance taking up where superconducting cavity maser

oscillator (SCMO) performance flattens out at $f \approx 1$ Hz[15].

It seems likely that a compromise between the expense and complication of the liquid helium-cooled SCMO and the relatively simple SPS may also be an attractive option. Operating at several milliwatts rf power, compared to $\approx$ 1 nanowatt for the SCMO, the SPS has a much lower noise floor. The SCMO, on the other hand, has a nominally zero temperature coefficient of frequency and so shows much greater stability at the lowest offset frequencies. Because thermal effects "freeze out" with an approximate T^3 dependence, an SPS cooled by a closed-cycle refrigerator to 10 Kelvin might be an attractive compromise, showing both low noise and higher stability.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to R. T. Wang for cryogenic and rf expertise, A. Kirk for assistance with phase noise tests, and T. K. Tucker for assistance with circuit design and component selection.

REFERENCES

[1] L. Maleki, "Frequency and Timing Requirements for Space Exploration," *Proc. 45th Ann. Symposium on Frequency Control,* 707-711 (1991).

[2] J. Dick, J. Saunders and T. Tucker, "Ultra-Low Noise Microwave Phase Stabilizer using Sapphire Ring Resonator," *Proc. 44th Ann. Symposium on Frequency Control,* 577-84 (1990).

[3] G. J. Dick and J. Saunders, "Method and Apparatus for Reducing Microwave Oscillator Output Noise", U.S. Patent #5,036,299, July 30, 1991.

[4] 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).

[5] V. B. Braginsky, V.P. Mitrofanov and V.I. Panov, *Systems with Small Dissipation,* (Univ. of Chicago Press, Chicago, 1985), 85-89.

[6] V. I. Panov and P. R. Stankov, "Frequency Stabilization of oscillators with high-Q leucosapphire dielectric resonators", *Radiotekhnika i Electronika* **31**, 213, (1986), (In Russian).

[7] D. G. Blair and I. N. Evans, "High-Q Microwave Properties of a Sapphire Ring Resonator," *J. Phys. D: Appl. Phys.*, **15**, 1651-1656, (1982).

[8] A. Giles, S. Jones and D. Blair, "A High Stability Microwave Oscillator Based on a Sapphire Loaded Superconducting Cavity", *Proc. 43rd Ann. Symposium on Frequency Control,* 89-93, (1989).

[9] X. H. Jiao, P. Guillon, and L. A. Bermudez, "Resonant frequencies of whispering-gallery dielectric resonator modes", *IEE Proceedings* **134**, Pt. H, 497 (1987).

[10] 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," *Proc. 42nd Ann. Symposium on Frequency Control,* 432-441 (1988).

[11] F. L. Walls and C. M. Felton, "High Spectral Purity X-band Source," *Proc. 44th Ann. Symposium on Frequency Control,* 542-548 (1990).

[12] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", *Proc. 41st Ann. Symposium on Frequency Control,* 99-110, (1987).

[13] G. K. Montress, T. E. Parker, M. J. Loboda, and J. A. Greer, "Extremely Low Phase Noise SAW Resonators and Oscillator: Design and Performance," IEEE *Trans. Ultrasonics, Ferroelectrics, and Frequency Control* **UFFC-35**, #6, 657-667 (1988).

[14] Quartz oscillator model 8600.03 by Oscilloquartz SA, CH-2002 Neuchâtel 2, Switzerland.

[15] R. T. Wang and G. J. Dick, "Improved Performance of the Superconducting Cavity Maser at Short Measuring Time," in *Proc. 44th Ann. Symposium on Frequency Control,* 89-93 (1990).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ULTRA-HIGH STABILITY SYNTHESIZER FOR DIODE LASER PUMPED RUBIDIUM

John P. Lowe, F. L. Walls, and R. E. Drullinger

Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80303

Abstract

We describe the design of a synthesized local oscillator for a rubidium [Rb] passive frequency standard pumped by radiation from a diode laser. The design goals for this new oscillator are:

1) To operate at room temperature,
2) To have the ability to step quickly to different frequencies for measuring parameters such as magnetic fields,
3) To contribute less than 2 x 10^{-14} $\tau^{-1/2}$ to the frequency stability of the Rb standard for measurement times between 30 and 200 seconds.

Potential limits for obtaining these goals, as well as solutions to such problems, are discussed.

Introduction

A number of concepts for new atomic frequency standards have the potential for achieving a frequency stability better than 1 x 10^{-14} $\tau^{-1/2}$ and inaccuracies less than 1 x 10^{-16} [1-7]. One of the difficulties of realizing this performance in passive atomic standards, however, is obtaining a local oscillator with the necessary frequency stability for interrogating the atomic resonance. Several of these new frequency standards require an interrogation cycle that range from 30 s to 200 s. To achieve the full capability of the new atomic standards, the local oscillator must maintain a frequency stability that is roughly comparable to the atomic performance for a time at least as long as the interrogation cycle time. Dick, et al. have investigated in detail the frequency instability in a standard that is caused by noise in the local oscillator [8]. In situations where there are two or more atomic samples, it is possible to shape the interrogation cycle to significantly reduce the contribution of the local oscillator to the white frequency noise level [8]. However, for those systems where only one atomic sample is used (currently the majority of the systems), the noise in the local oscillator is still an overriding limitation.

Contribution of the U. S. Government, not subject to copyright.

Since none of the present room temperature rf oscillators possesses a frequency stability of 1 x 10^{-14} $\tau^{-1/2}$ for 30 s to 200 s, it is of interest to pursue special frequency standards which might achieve this level of stability. One such candidate is the passive, diode-laser-pumped, rubidium-gas-cell standard analyzed by Camparo, et al. [9]. They project that with an optimum choice of parameters, a rubidium frequency standard could be built that has a frequency stability of less than 1 x 10^{-14} $\tau^{-1/2}$. Although this passive standard also needs a local oscillator with stability near 1 x 10^{-13} at 10 to 30 ms, the loop attack time could be as short as 0.03 s since the linewidth of the resonance is expected to be on the order of 100 Hz and the modulation frequency about 37.5 Hz. A good quartz oscillator can meet this requirement.

For a passive standard using sinusoidal phase modulation, the phase noise at the second harmonic and fourth of the modulation frequency appear as additional white frequency terms which limit the ultimate short-term frequency stability [10-15]. The most thorough treatment of this problem is given by Audoin, et al. [14,15]. They have shown that this limit is approximately given by

$$\sigma_y(\tau) = (f_m/\nu_o)\,[0.723\,S_\phi(2f_m) + (C)\,S_\phi(4f_m)]^{1/2}\,\tau^{-1/2} \qquad (1)$$

where $S_\phi(2f_m)$ and $S_\phi(4f_m)$ is the power spectral density of phase noise in the oscillator at twice and four times the modulation frequency f_m, respectively, and ν_o is the carrier frequency. Modulation

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

parameters such as modulation depth determine the value of C (typically C $\simeq$ .1) [15]. A typical quartz oscillator and a modulation frequency of 37.5 Hz would produce a standard with a stability of about 1 x 10^{-13} $\tau^{-1/2}$. In this paper we describe a method using currently available technologies to reduce this effect to below 2 x 10^{-14} $\tau^{-1/2}$.

Design of a New Passive Rubidium Servo System

Since one of the limitations to the frequency stability is the noise at the second harmonic of the modulation frequency and not significantly from the noise at other Fourier frequencies [10-15], filtering the noise from the local oscillator signal only in the region about the second harmonic of the modulation is sufficient to reduce this effect. Figure 1 shows a block diagram of one approach to achieve this filtering of a signal suitable for interrogating the Rb resonance at 6.83 GHz. The output of our low noise 5 MHz oscillator [S_ϕ(75 Hz) = 6.3 x 10^{-17} (rad^2/Hz),($\mathcal{L}$(75 Hz) = -165 dBc/Hz)] is doubled to 10 MHz, phase modulated at f_m, filtered at $\pm$ $2f_m$, frequency multiplied to 500 MHz, and mixed with a tunable synthesizer (1/16 μHz resolution) to produce a submultiple of the Rb resonance near 488.14 MHz. The 488.14 MHz signal is multiplied by 14 in a step recovery diode (SRD) to produce a 6.834 xx GHz signal which can be used to interrogate the Rb resonance. The error signal from the detector is used in the traditional way to steer the 5 MHz quartz oscillator.

The key element is the filter following the phase modulator. This filter (Figure 2) has notches that are spaced symmetrically about the carrier and located at twice the modulation frequency from the carrier (ν_o $\pm$ 75 Hz). Calculations by Smythe [16] show that using crystal resonators with an unloaded Q-factor of 1 x 10^6 should yield notch filters with an attenuation at the center of the notch of order 20 dB and an insertion loss at the carrier frequency of less than 1 dB. Placing the notch filter after the modulator also allows the filter to attenuate the frequency error associated with second harmonic distortion within the modulator.

A notch filter has significant advantages over a passband filter for this application. Since the crystal resonators in the notch filters do not absorb significant power from the carrier, we can transmit power more efficiently through this type of notch filter than through a passband filter. Also, the phase of the carrier signal is affected much less by small variations in the frequency of the notch filter than for a passband filter. Calculations indicate [16] that an equivalent noise resistance of order 10 Ω is possible. This corresponds to an added noise of 0.4 nV/$\sqrt{\text{Hz}}$ or S_ϕ(f) = 8 x 10^{-20} (rad^2/Hz), ($\mathcal{L}$(f) = -194 dBc/Hz) for a carrier level of + 13 dBm. At this noise level, the stability limit calculated from Eq.(1) is approximately 1 x 10^{-15}.

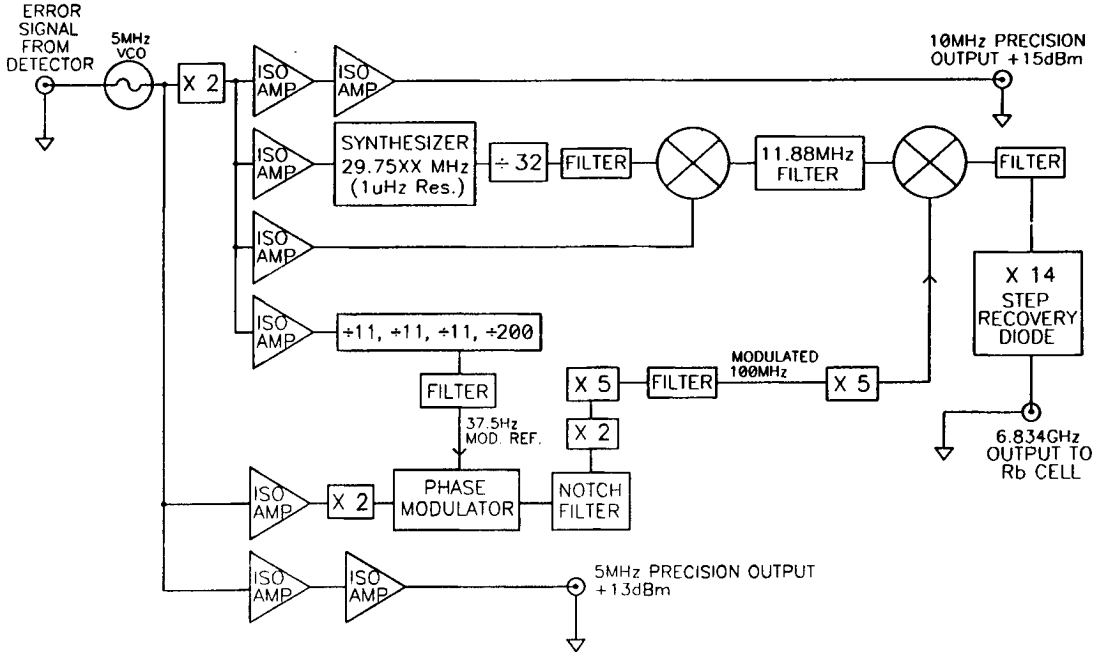


Figure 1. Block diagram of Rb synthesizer.

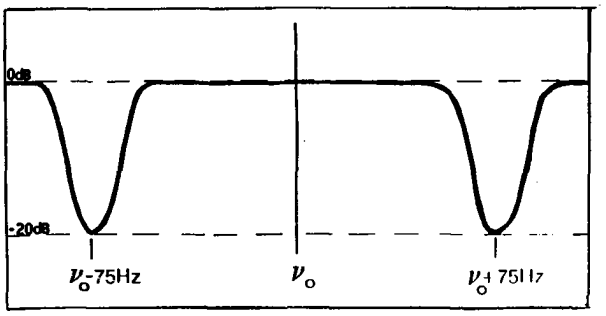


Figure 2. Notch filter at $\nu_o \pm 2f_m$.

A potential limit to the overall stability of the servo system is the noise added by the frequency multipliers. Typical doublers have added phase noise at approximately $S_\phi(75\ Hz) = 6.3 \times 10^{-17}$ (rad^2/Hz), and $S_\phi(150\ Hz) = 4.0 \times 10^{-17}$ (rad^2/Hz) which would limit the stability to $2.4 \times 10^{-14}\ \tau^{-1/2}$. Using the work of Felton [17], we have developed a series of ultra-low noise doublers. These multipliers feature an equivalent input phase noise of $S_\phi(75\ Hz) = 4.2 \times 10^{-18}$ (rad^2/Hz), and $S_\phi(150) = 2.6 \times 10^{-18}$ (rad^2/Hz) for doubling 10 MHz to 20 MHz (Figure 3). Based on this level of performance, the stability limitation would be 6.7×10^{-15}. Preliminary results show the phase noise of the 20 MHz to 100 MHz multipliers will contribute approximately 1 dB to the overall phase noise.

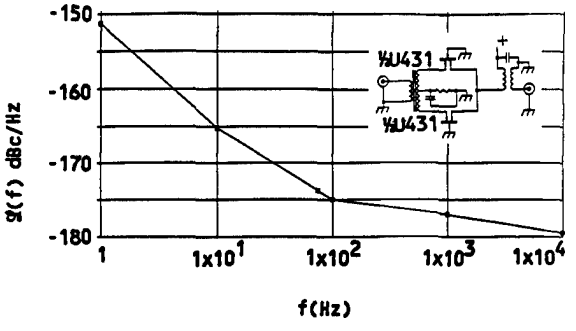


Figure 3. Residual phase noise of the 10 to 20 MHz doubler circuit.

The 100 MHz distribution amplifiers and the 100 MHz to 6.83 GHz (through 488 MHz) synthesis is completed. Results on similar 100 MHz to 10.6 GHz synthesizers show that phase variations within the synthesizers correspond to a frequency stability of 2×10^{-17} at 1000 s, which is well below the requirements [18].

The phase modulator is a linear, small-angle crystal modulator [19]. This phase modulator has very low second harmonic distortion and would enable one to find the center of the resonance to an accuracy of 0.1 ppm of the resonance linewidth. However, the added phase noise of the modulator is $S_\phi(75\ Hz) = 8 \times 10^{-16}$ (rad^2/Hz), ($\mathcal{L}(75\ Hz) = -154$ dBc/Hz) (Figure 4a). Figure 4b shows the anticipated improvement of the phase noise at 75 Hz due to the incorporation of

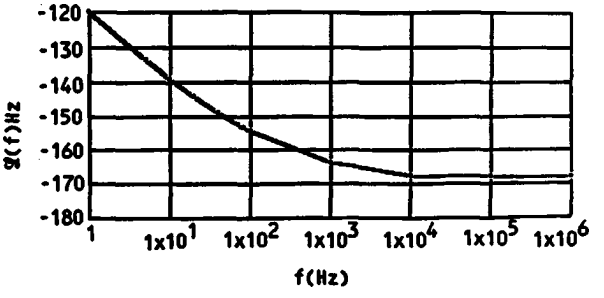


Figure 4a. The phase noise of the small-angle crystal modulator.

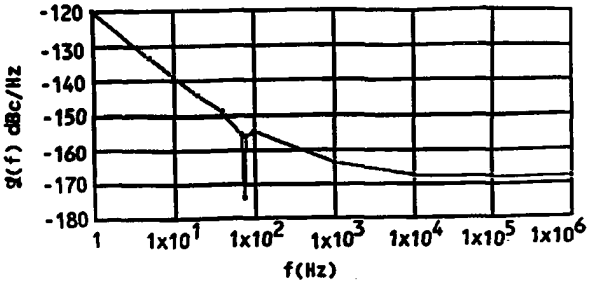


Figure 4b. The phase noise of the small-angle crystal modulator and notch filter.

the notch filter. Figure 5 shows the contribution of the electronics to $\sigma_y(\tau)$ versus τ for the complete standard, calculated with and without the noise in the present small-angle modulator and a bandwidth of 100 Hz. The phase lock loop was assumed to be second order with a unity-gain attack time of 0.03 s and an integrator attack time of 0.12 s. Clearly the noise of the modulator dominates the frequency stability. Reducing the level of random walk FM has no effect on the stability nor does increasing the measurement bandwidth.

Although the white frequency noise limit is within specifications, $\sigma_y(\tau) = 1.4 \times 10^{-14}\ \tau^{-1/2}$, further improvements are possible by reducing the phase noise of the small-angle phase modulator. One possibility is to use a lower noise phase modulator in

place of the small-angle modulator at the expense of second harmonic distortion. The notch filter following the modulator would help reduce the effect of higher distortion. If necessary, two filters in series could be used.

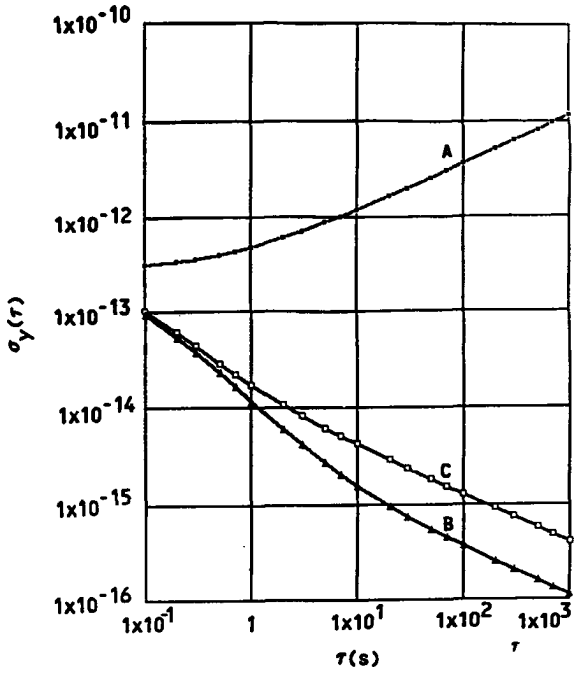


Figure 5. Stability calculations for the complete standard. Curve A shows the stability of the open loop oscillator, curve B shows the stability of the closed loop oscillator without the noise of the modulator and curve C shows the stability with the noise of the modulator added.

Conclusion

We have shown that it is possible to develop a synthesizer for a laser-pumped rubidium standard with the noise performance necessary to support a stability of 1.4 x 10^{-14} $\tau^{-1/2}$ up to approximately 200 s. This was accomplished by incorporating notch filters about the carrier at the second harmonic of the modulation frequency. Also, lower noise multipliers were incorporated to help reduce the phase noise present. The resulting rubidium standard may be used as the local oscillator for other standards which need high stability during their interrogation cycle. Other approaches are being pursued to increase the stability even more.

Acknowledgement

The authors would like to thank Rolland Barillet from the Laboratoire de l'Horloge Atomique for his help.

References

[1] D.J. Wineland, J.C. Bergquist, J.J. Bollinger, W.M. Itano, D.J. Heinzen, S.L. Gilbert, C.H. Manney, and M.G. Raizen, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions", IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, vol. 37, No. 6, pp. 515-523, Nov. 1990.

[2] J.D. Prestage, G.J. Dick, and L. Maleki, "JPL Trapped Ion Frequency Standard Development", IEEE 41st Annual Frequency Control Symposium, pp. 20-24, 1987.

[3] M.G. Raizen, J.M. Gilligan, J.C. Bergquist, W.M. Itano, D.J. Wineland, "Ionic Crystals in a Linear Paul Trap", Physical Review, vol. 45, Number 9, pp. 6493-6501, May 1, 1992.

[4] A. de Marchi."The Optically Pumped Caesium Fountain: 10^{-15} Frequency Accuracy?", Metrologia, vol. 18, pp. 103-116, 1982.

[5] A. Clairon, C. Salomon, S. Guellati, and W.D. Phillips, "Ramsey Resonance in a Zacharias Fountain", Europhysics Letters vol. 16(2), pp. 165-170, September, 1991.

[6] S. Chu "Laser Manipulation of Atoms and Particals", Science, vol. 253, pp. 861-866, 1991.

[7] J.L. Hall, M. Zhu, and p. Buch, "Prospects for Using Laser-Prepared Atomic Fountains for Optical Frequency Standards Applications", J. Opt. Soc. Am. B, vol 6, pp. 2194-2205, 1989.

[8] G.J. Dick, J.D. Prestage, C.A. Greenhall, and L. Maleki, "Local Oscillator Induced Degradation of Medium-Term Stability in Passive Atomic Frequency Standards", 22nd PTTI, NASA CP 3116, pp. 487-508, 1990.

[9] J. C. Camparo, R. E. Drullinger, C. Szekely, "Strategies for Developing a Diode Laser Pumped Gas Cell Atomic Clock", these proceedings.

[10] F.L. Walls and S.R. Stein, "Servo Techniques in Oscillators and Measurement Systems", NBS Technical Note 692, 1976.

[11] G. Kramer, "Noise in Passive Frequency Standards", CPEM 1974 Digest, IEEE Conference Publication 113, IEEE Catalogue Ch 0770-8 IM.

[12] F.L. Walls, "Stability of Frequency Locked Loops", in Proc. Fourth Symposium on Frequency Standards and Metrology, A. de Marchi, editor, Springer-Verlag, pp. 145-149, 1988.

[13] L.S. Cutler, Hewlett-Packard Co., 3500 Deercreek, Palo Alto, Ca. 94304, Private communication, 1988.

[14] C. Audoin, V. Candelier, and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards", Conference on Precision Electromagnetic Measurements, 1990.

[15] R. Barillet, V. Giordano, J. Viennet, C. Audoin, "Microwave Interrogation Frequency Noise and Clock Frequency Stability: Experimental Results", Proc. 6th EFTF, 1992.

[16] R.C. Smythe, Piezo Technology Inc. 2525 Shader Rd. Orlando, Fl. 32842, Private Communication, 1992.

[17] C.M. Felton, "Superimposing Low-Phase-Noise, Low-Drift Instrumentation Techniques on RF Design", RF Design, pp. 65-74, October 1990.

[18] C.W. Nelson, F.L. Walls, F.G. Ascarrunz, P.A. Ponds, "Progress on Prototype Synthesizer Electronics for $^{199}Hg+$", these proceedings.

[19] John P. Lowe and F. L. Walls, "Ultralinear Small-Angle Phase Modulator", Proc. 45th SFC, pp. 645-648, 1991.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A LOW-FREQUENCY, HIGH RESOLUTION DIGITAL SYNTHESIZER

Robin P. Giffard and Leonard S. Cutler
Hewlett Packard Laboratories
P.O. Box 10350, Palo Alto, CA 94303-0867

Abstract-A high resolution digital frequency synthesizer was designed to provide modulation and fine frequency interpolation in a new Cesium frequency standard. Requirements to be satisfied include resolution, spectral purity, phase-continuous frequency modulation, simplicity and low power. In the following paper we describe how these requirements were satisfied with a Direct Digital Frequency Synthesizer (DDFS). This type of frequency synthesizer was originally described by Tierney, Rader, and Gold [1], and has since been widely used and discussed [2,6]. The application to a hydrogen maser has been described by Matisson and Coyle [7].

Design Requirements

In the new Cesium standard, the microwave probe frequency has to be tuned to both the F=3, Mf=0 to F=4, Mf=0 "main" transition near 9192.632 MHz, and the F=3, Mf=1 to F=4, Mf=1 "Zeeman" line. At the chosen magnetic field, this requires the synthesizer to have a range of about 50 kHz. The necessary resolution is fixed at better than 92 microhertz by the requirement that the standard be steerable with a settability of 1 part in 10^{14} or less. To carry out effective square-wave frequency modulation, the synthesis chain must change frequency and settle to an accuracy of 1 part in 10^{14} in a time of about 2 ms, the duration of the transient generated by the tube when the frequency changes. For frequency changes of less than 1 kHz, the frequency should change without any phase discontinuity. Finally, to avoid offensive frequency pulling, the synthesizer should be free of sidebands close to the resonance line to a level of about -70 dBc. The synthesizer output has to be free of sidebands coherent with the modulation process and its harmonics to about -120 dBc. Because of the need for fast phase-continuous frequency switching, the requirements for this synthesizer are somewhat different from those relevant to hydrogen maser design [7].

A Direct Digital Frequency synthesizer was chosen to obtain the necessary resolution. In order not to degrade phase noise, the output frequency of the synthesizer is translated without multiplication to the probe frequency. To minimize power consumption and cost, the DDFS operates at the lowest satisfactory clock frequency. The overall synthesis in the cesium standard is defined by the following equation:

$$Fp = 29 \times 320 \text{ MHz} - ((35/32) \times 80 \text{ MHz} - Fs), \quad (1)$$

where Fp is the probe frequency, and Fs is the frequency of the digital synthesizer. The required range of the synthesizer is thus 125 kHz to 175 kHz. The architecture and overall synthesis scheme of the new Cesium standard are discussed in other papers [8,9].

Direct Digital Synthesis

The principal of the DDFS is well-known [1-3]. Its performance depends on the operation of the digital phase accumulator shown in Fig 1. The adder A causes the contents of the phase register R to increase by the value Na of the addend at each clock cycle. If the adder is binary and the phase register contains P bits, after N clock cycles the phase register will contain the number $N \cdot Na$ MODULO 2^P. The output of the phase accumulator is a sampled saw-tooth wave with a repetition period of $2^P/(Fc \cdot Na)$. It is well known that if the exact regularly spaced phase samples are used to calculate the inputs Vi to an ideal digital-to-analog converter (DAC) using the equation:

$$Vi = Const \times \sin\left[\frac{2\pi Ai}{2^P}\right], \quad (2)$$

the DAC output signal after an ideal anti-alias filter will be a pure sinusoid at a frequency $Fs = Fc \cdot Na/2^P$, completely free of non-harmonic distortion. The Nyquist condition: $Fs < Fc/2$, must be satisfied.

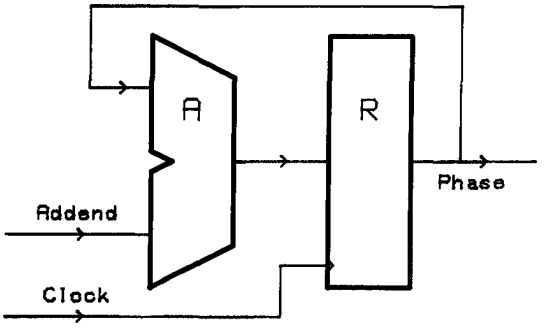


Fig 1. Digital phase accumulator consisting of a binary full-adder A and a clocked register R.

0-7803-0476-4/92 $3.00 © 1992 IEEE

In practice, the conversion of linear phase to sinusoidal DAC input values is usually carried out by a "look-up table" or "waveform-map" consisting of a static read-only memory (ROM), avoiding the need for real-time calculation. Since the amount of memory required to encode the entire width of the phase accumulator would usually be prohibitive, only the Q most significant bits of the accumulator output are generally used to calculate the sine-wave samples. The truncation of P-Q bits of the phase leads to errors in the computation, and the output waveform cannot be a pure sine-wave, although its average frequency is not changed. The errors cause spurious frequencies known as "phase truncation sidebands" to appear [4,5].

It is well known in the field of digital signal processing that the finite resolution of the DAC used to construct the waveform introduces sidebands due to quantization error. Non-ideal static and dynamic DAC effects lead to further degradation.

We will now consider how a synthesizer was designed to meet the requirements of the frequency standard.

Design

It is shown above that the DDFS resolution df is related to the clock frequency and the phase register length by $df=Fc/2^P$, so that register length and adder width are minimized by choosing the lowest clock frequency that will allow satisfactory anti-alias filtering. The requirements of the instrument are satisfied by a 34-bit register and a 1 MHz clock frequency. The clock frequency is more than 4 times the highest output frequency required, allowing the two highest bits of the adder to be implemented as a synchronous counter, and making the anti-alias filter less critical.

The choice of the DAC resolution is dictated by the -70 dB spectral purity requirement. The quantization noise power in an ideal uniform DAC is easily shown to be of the order of $S^2/12$, where S is the step size. For a DAC having D bits, the maximum sine-wave output power is $2^{(2D-3)}S^2$, so that the total noise power is theoretically -1.76-6.02 D dBc [10]. In practice this power is distributed over a number of frequencies which are m/n intermodulation products of the clock and synthesizer frequencies. The power in any given sideband can be minimized to some extent by choice of the ROM program [5]. A 12-bit DAC was chosen, and a computer program was used to calculate the level of the dominant spurious sideband and choose the ROM coding. Allowing for static errors, an ideal 11-bit DAC was simulated, and it was found that the level of the unwanted sidebands should be below -90 dBc.

In order to obtain sideband levels approaching these values in real DACs, it is normally necessary to suppress glitches either by following the DAC with a sample-and-hold circuit, or by blanking its output while it settles after each data transition. Large-scale non-linearity in the DAC transfer function is expected to cause harmonic distortion, which is not important in this application.

The magnitude and spectral distribution of truncation sidebands has been explored theoretically by Nicholas and Samueli [4] and by Nicolas, Samueli and Kim [5]. A very large number of sidebands is produced in general, with a complex distribution of frequencies. The amplitude of the sidebands is found to depend on the width of the phase accumulator P, and the number of bits Q input to the waveform map. A small modification of

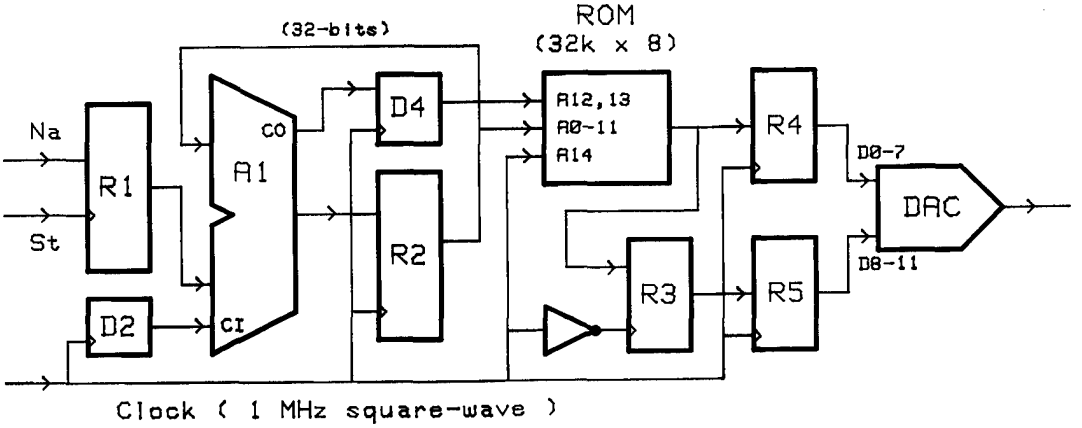


Fig 2. Block diagram of the Direct Digital Frequency synthesizer. The clock input is a 1 MHz square wave. The operation of the circuit is explained in the text. The DAC de-glitcher, and the anti-alias filter are not shown.

the accumulator reduces the worst possible sideband amplitude by about 4 dB. It was decided to use an address width of 14 bits in the synthesizer since this was a convenient match to a 256-k ROM with two bytes for each phase value. With an effective P value of 35, and $Q=14$, worst-case truncation sidebands are calculated to be below -84 dBc. We expect that as a result of the frequency switching sequence used in our application, the chance of a spurious sideband coherent with the modulation process is negligible.

Hardware Design

The hardware design of the synthesizer is shown in fig. 2. The 32 bit addend word *Na* is transferred to the register R1 by a hardware-derived timing strobe. The adder A1 is a 32-bit full binary adder whose output is latched into the 32-bit phase register R2 at each clock cycle. The carry output of the adder forms the enable of a 2-bit synchronous binary counter D4 whose output forms the two most significant bits of phase. This configuration is equivalent to a 34-bit phase accumulator with a 32 bit input word, limiting the output frequency to *Fc*/4. As suggested by Nicholas and Samueli, the carry input of the adder is at half the clock frequency. The upper 14 bits of the phase word are fed to address bits 0 through 13 of the ROM, and bit 14 alternates at the clock frequency. Registers R3-R5 align the upper and lower byte outputs of the ROM in time, and ensure that the DAC inputs all change simultaneously. To obtain good phase noise it is essential that the clock input to R4 and R5 has very low jitter. The DAC is a 12-bit, bipolar device with symmetric current outputs. De-glitching is implemented by gating off the DAC output for 200 nS after each data change with a balanced current switch. A

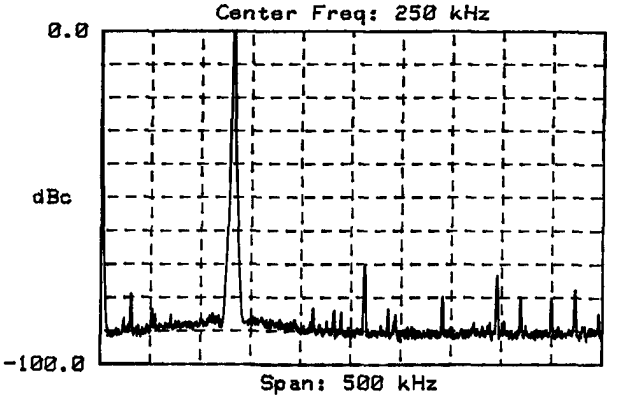


Fig 3. Output spectrum of the DDFS up to the Nyquist limit for a programmed output frequency of 131770 Hz. The lines in the spectrum are listed in Table I.

Table I

Frequency (Hz)	Level (dBc)	Origin
263540	-72.5	$2Fs$
395310	-74.1	$3Fs$
472920	-78.2	Fc-$4Fs$
30710	-78.4	$23Fs$-$3Fc$
341150	-79.2	Fc-$5Fs$
449470	-79.7	$11Fs$-Fc
418760	-80.1	$2Fc$-$12Fs$
51450	-83.0	$(2^{14}+1)Fs$-$2159Fc$

three-pole anti-alias filter with a cut-off of about 300 kHz is used. The power consumption of the complete synthesizer, implemented in the HCMOS logic family, is 760 mW. With no large-scale integration, the circuit is made on a single-sided surface-mount circuit board 10 x 19 cm in size.

Performance

Figure 3 shows the output spectrum of the synthesizer, programmed for a frequency of 131770 Hz. The plot is from zero frequency to 500kHz, the Nyquist limit. The output is taken before the anti-alias filter. Table I shows the amplitudes and numerical structure of the major spurious lines in the spectrum. The most prominent lines apart from the desired output are the second and third harmonics. A few m/n intermodulation products are visible at a level of about -80 dBc and below. An unexpected line which is an extremely high order mixing product with $m=1+2^{14}$ is seen at -83 dBc.

In our application, the critical frequency range for spurious signals is the region of the main resonance line, corresponding to frequencies between 120 and 145 kHz. The required synthesizer output frequencies are between 131.0 and 132.5 kHz. A computer program was used to find which orders of intermodulation would generate significant spurious sidebands. The output spectrum was then examined in detail to find the amplitudes of these products. Figure 4 shows how the sideband amplitude varies with order n up to 100. The lowest significant order is 37, at which the amplitude has fallen to about -90 dBc.

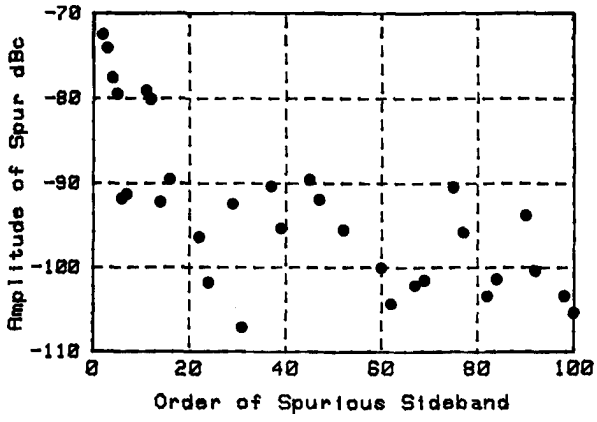


Fig 4. The observed variation of the amplitude of spurious products with their intermodulation order *n*. The frequency of these products is given by ABS(*m* • *Fc*-*n* • *Fs*) where *m* and *n* are integers. *m* runs form 0 to 13.

Figure 5 shows the frequency region around a 131770 Hz output signal with greater resolution. A pair of spurious sidebands at $Fs \pm (38Fs\text{-}5Fc)$ is present at a level of about -90 dBc. At the level of about -93 dBc no truncation sidebands are seen for this addend value.

Figure 6 shows, for comparison, the same frequency range as Figure 5 but with the DAC de-glitching disabled. A large number of sidebands have appeared with amplitudes up to -66 dBc. Careful examination shows that these are *m*/*n* products. Clearly the de-glitching is useful.

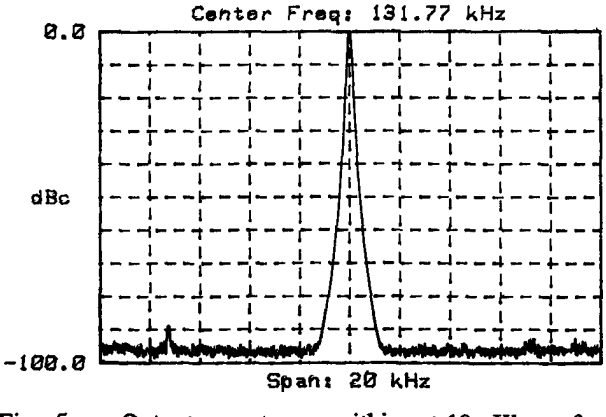


Fig 5. Output spectrum within ±10 Khz of a programmed output of 131770 Hz. The effective noise bandwidth of the spectrum analyzer is 100 Hz. The frequencies of the spurious sidebands are given by $5Fc\text{-}37Fs$ and $38Fs\text{-}5Fc$.

Summary

A DDFS consisting of a 34-bit phase accumulator feeding a 14-bit wide address map and a 12-bit de-glitched DAC has been evaluated. For output frequencies in the range 131.0 to 132.5 kHz, which are required to excite the main Cesium resonance line, the level of spurious sidebands is less than -85 dBc. This is compatible with the expected performance of the DAC. No truncation sidebands have been seen down to -90 dBc. A set of sidebands with frequencies given by $2154Fc\text{-}(2^{14}\text{-}1)Fs$ and $(2^{14}+1)Fs\text{-}2154Fc$ are detectable with an amplitude of -83 dBc. The origin of these sidebands is not understood, and they do not occur at significant output frequencies.

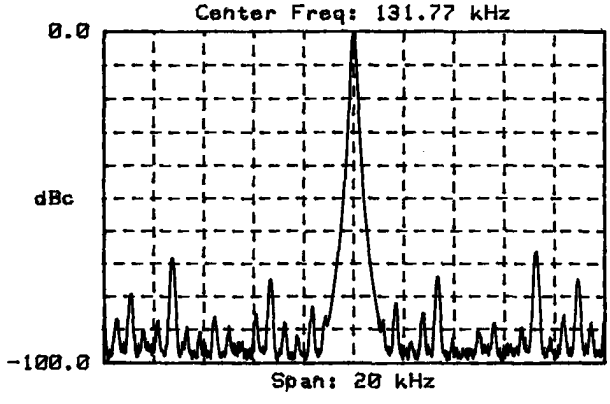


Fig 6. Same as Fig 5, with DAC de-glitcher disabled. The spurious sidebands visible are *m*/*n* intermodulation products with values of *m* and *n* between 37 and 260.

References

[1] J. Tierney, C.M.Rader, and B. Gold, "A digital frequency synthesizer" IEEE Trans. Audio Electroacoust., Vol AU-19, pp. 48-56, Mar. 1971.

[2] A.L. Bramble, "Direct digital frequency synthesis", in Proceedings of the 35th Annual Symposium on Frequency control, pp 406-414, IEEE, 1981.

[3] V. Reinhardt, K. Gould, K. McNab, and M. Bustamante, " A short survey of frequency synthesizer techniques", in Proceedings of the 40th Annual Frequency Control Symposium, pp 355-365, IEEE 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", in Proceedings of the 41st Annual Frequency Control Symposium, pp 495-502, IEEE 1987.

[5] H.T. Nicholas, III, H. Samueli, and B. Kim, "The optimization of direct digital frequency synthesizer performance in the presence of finite word length effects", in Proceedings of the 42nd Annual Frequency Control Symposium, pp 357-363, IEEE, 1988.

[6] J.F. Garvey, and Daniel Babitch, "An exact spectral analysis of a number controlled oscillator based synthesizer", in Proceedings of the 44th Annual Frequency Control Symposium, pp 511-521, IEEE, 1990.

[7] E.M.Matisson and L.M.Coyle, "Phase noise in direct digital synthesizers", in Proceedings of the 42nd Annual Frequency Control Symposium, pp 352-356, IEEE, 1988.

[8] L.S. Cutler and R.P. Giffard, "Architecture and algorithms for a new cesium beam frequency standard electronics", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[9] R. Karlquist, "A new RF architecture for Cesium frequency standards", to be published in Proceedings of the 1992 IEEE Frequency Control Symposium, IEEE 1992.

[10] "Digital Signal Processing", by A.V. Oppenheim and R.W. Shafer, Prentice-Hall, Englewood Cliffs, 1974.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

SPECTRAL PERFORMANCE OF FREQUENCY MULTIPLIERS AND DIVIDERS

M. M. DRISCOLL and T. D. MERRELL
Westinghouse Electronic Systems Group, Baltimore, MD

ABSTRACT

This paper constitutes a review of phase noise performance data accumulated by the authors for commonly used, frequency multiplier (transistor and step-recovery diode impulse generator, and diode and transistor full-wave rectifier doubler) circuits and frequency divider (digital logic-level, analog regenerative and varactor diode) circuits.

The data reveals large variations in the spectral performance exhibited by the various circuit types, and can be used as an important, database guide in the design of low noise, frequency synthesizer hardware.

INTRODUCTION

Low noise signal generator near-carrier spectral performance is usually limited by the corresponding spectral performance achieved in the oscillator(s) incorporated in the hardware. However, as oscillator/resonator short-term frequency stability performance has improved along with increased signal generator capability and circuit complexity, the net, additive effect of necessary use of large numbers of individual signal amplification, frequency addition and subtraction (mixer), and frequency multiplication and division stages often limits attainable output signal spectral performance.

With regard to frequency multiplication and division circuitry, it is usually necessary to tradeoff frequency synthesis circuit spectral performance with other circuit performance parameters including bandwidth, cost, complexity, efficiency, reliability, etc. Intelligent design choices require knowledge of the spectral performance characteristics attainable using alternative frequency multiplier and divider circuitry and an understanding of the mechanisms involved that determine device phase noise performance.

IMPULSE GENERATION-TYPE FREQUENCY MULTIPLIERS

1. Step Recovery Diode Frequency Multipliers

Step recovery diodes (SRD) are usually used for single-step, high-order (multiplication factor) signal frequency multiplication. The SRD operates as a charge-controlled switch. A forward diode bias stores charge, and a reverse bias depletes the stored charge. In the SRD, current conduction turn-off occurs fast enough to produce an extremely narrow, impulse-like, output waveform. For periodic (sinusoidal) diode excitation, the output signal frequency "comb" spectrum consists of harmonics of the input signal extending from the input frequency to the reciprocal of the output waveform, impulse width. Used as a frequency multiplier, the desired, input signal harmonic is extracted via a bandpass filter. Coaxial-connectorized, self-contained, SRD comb generator modules are available from several suppliers, and standard input/output frequency ranges extend from 100 MHz (input) to 25 GHz (output). Device input impedance matching bandwidths are relatively narrow, and recommended input drive and per-harmonic output power levels lie in the range 20 to 27 dBm and 0 to -20 dBm, respectively [1-3].

Experiments conducted at Westinghouse indicate that lossless (direct) connection of the harmonic selection, bandpass filter to the SRD output, as well as attempts to increase multiplier efficiency by optimizing the match at the diode output-bandpass filter interface, can result in significant increase in multiplier phase noise. Use of isolators and/or circulators at the SRD output do not always constitute a solution, owing to device bandwidth limitations. Use of a small amount of broadband, resistive attenuation at the interface for controlling the effective diode load impedance at all output signal spectral component frequencies results in uniform, improved, phase noise performance. Figures 1 and 2 show the results of measurement of (times ten multiplication) SRD multipliers, operating at 100 MHz and 1 GHz input frequency. Figure 1(b) shows the phase noise performance of the 100 MHz–1.1 GHz multiplier with SRD direct (lossless connection between the diode and output bandpass filter. As shown in the figure, the result is a 50 dB increase in white phase noise (floor) level. For the 1–10 GHz multiplier phase noise measurements (figure 2), the X-Band, GaAs, output amplifiers constitute the dominant source of additive, flicker-of-phase noise. In contrast to figure 1, figure 3 shows the results of additive phase noise measurements made a number of years ago by Hewlett-Packard for a pair of times twenty (100 MHz–2 GHz), SRD multipliers [4].

2. Narrow Pulse Generation, Transistor Frequency Multipliers

High-order multiplication can also be accomplished (at lower frequency ranges than those used with the SRD) using transistor circuits. Figure 4 shows a simplified schematic diagram for a 10 MHz to 80 MHz multiplier fabricated and tested specifically for the purpose of the additive phase noise performance for this class of multiplier circuit. As shown in

0-7803-0476-4/92 $3.00 © 1992 IEEE

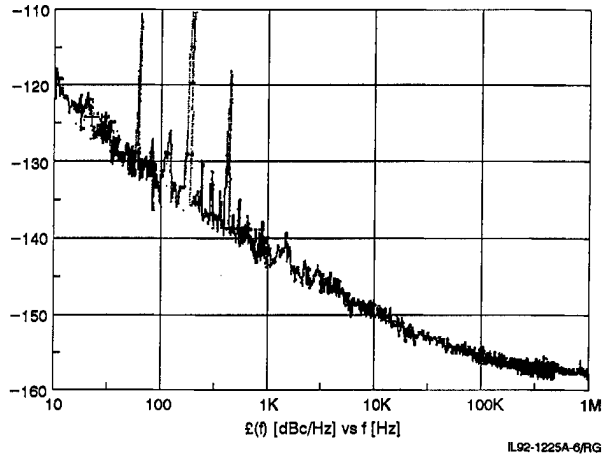


Figure 1(a). Phase Noise Measurement Results for a Pair of Times Eleven (100 MHz – 1.1 GHz) Step Recovery Diode Multipliers

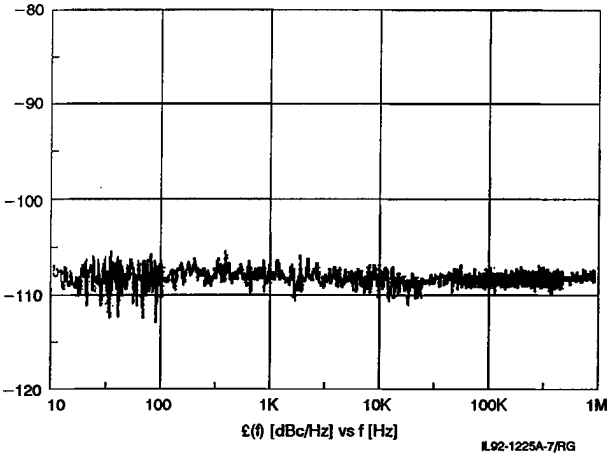


Figure 1(b). Phase Noise Measurement Results for the Figure 1(a) Multipliers with Lossless Connection Between Diode and Output Filter

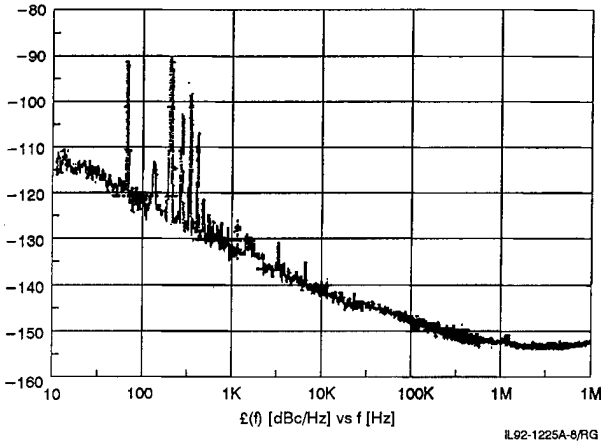


Figure 2. Phase Noise Measurement Results for a Pair of Times Ten (1 – 10 GHz) Step Recovery Diode Multipliers

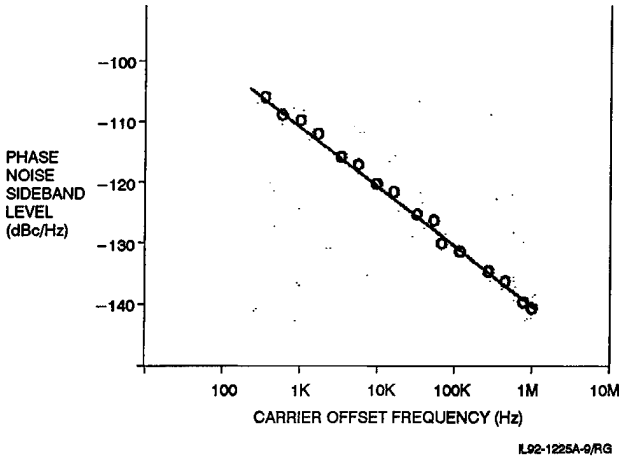


Figure 3. Phase Noise (Hewlett-Packard) Measurement Results for a Pair of Times Twenty (100 MHz – 2 GHz) Step Recovery Diode Multipliers

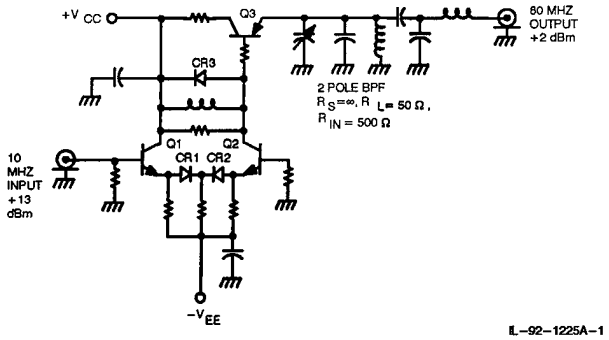


Figure 4. Simplified Schematic Diagram for Transistor "Impulse" Generation Type Frequency Multiplier

the figure, the impulse-generating portion of the circuit consists of two, emitter-coupled, switching transistors. In order to increase switching speed, the circuit incorporates a pair of schottky diodes and is arranged so that the transistors conduct during the entire waveform cycle. A single-ended, "impulse" waveform is created by differentiating the transistor, "square wave" collector current using a inductor/resistor/diode combination. A class-C, output amplifier with a multipole, bandpass output matching structure is used for waveform amplification and desired signal harmonic extraction. The circuit is similar in nature to one first described by Dick Baugh in 1972 [5]. Figure 5 shows the measured phase noise performance for a pair of the figure 4, test circuits. As shown in the figure, the per-device, phase noise floor level, referred to the input frequency, is -166 dBc/Hz and is likely a result of un-optimized, relatively small, class-C amplifier drive level. To the extent that the output amplifier of the figure 4 circuit conducts over a small portion of the input signal period, one might expect to obtain similar spectral performance from frequency multipliers designed using sinusoidal excitation, small conduction angle, transistor amplifiers. With regard to flicker-of-phase noise, the results shown in figure 5 are 15 dB

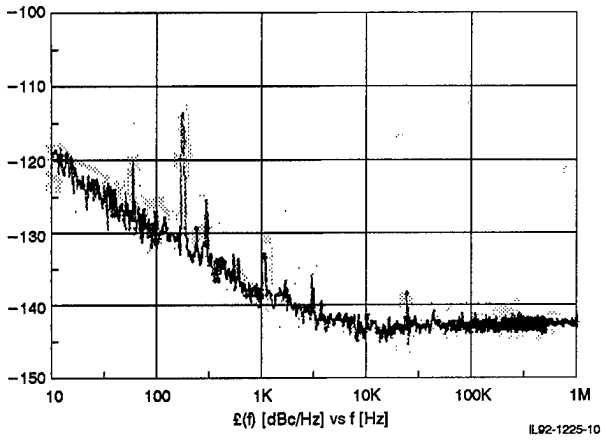


Figure 5. Phase Noise Measurement Results for a Pair of Times Eight (10 – 80 MHz) Transistor Impulse Generation Type Multipliers

poorer, compared to the (1 MHz input frequency) Baugh multiplier, which additionally incorporated an emitter-coupled, zero-crossing detector-type output limiter.

FULL WAVE RECTIFIER-TYPE, FREQUENCY DOUBLERS

1. Diode Doublers

Multiple-decade bandwidth, schottky diode-type, frequency doublers are available as standard, vendor components that operate over the HF to microwave frequency region. Doubler conversion loss is approximately 10 dB, and units are available for optimized efficiency at moderate (10 dBm) and high (23 dBm) input drive levels. Similar performance can and has been obtained by splitting the input signal and driving two ports of a double-balanced mixer. Doubler spectral performance, in terms of flicker-of-phase noise, is excellent, but varies with the type of diodes used and, to a much lesser extent, with device operating frequency [6].

In general, use of individual, cascaded stages of schottky diode-type, frequency doublers and low, flicker-of-phase noise, interstage amplifiers provides best obtainable, frequency multiplier spectral performance. Figures 6 to 8 show the phase noise performance obtained for low drive level diode doublers, high drive level diode doublers, and for a pair of times-sixteen, frequency multipliers incorporating a cascade of four similar type doublers. For the times-sixteen multipliers, the 2.56 GHz, silicon bipolar, doubler interstage amplifier used constituted the dominant source of multiplier flicker-of-phase noise [7].

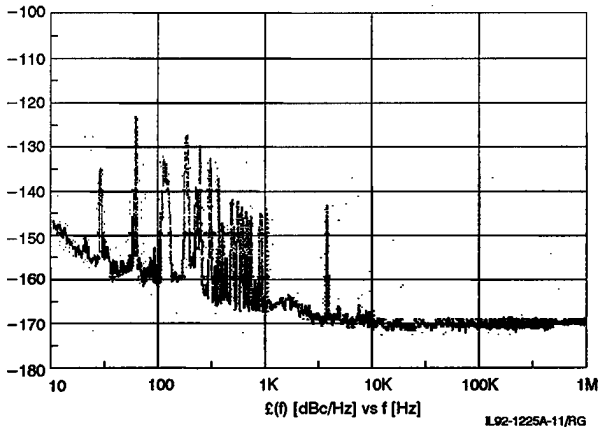


Figure 6. Phase Noise Measurement Results for a Pair of (13 dBm Drive Level, 100 – 200 MHz) Schottky Diode Frequency Doublers with Output Amplifiers

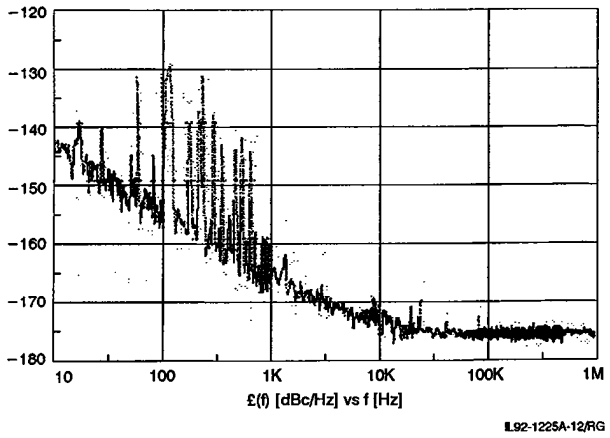


Figure 7. Phase Noise Measurement Results for a Pair of (23 dBm Drive Level, 80 – 160 MHz) Schottky Diode Frequency Doublers

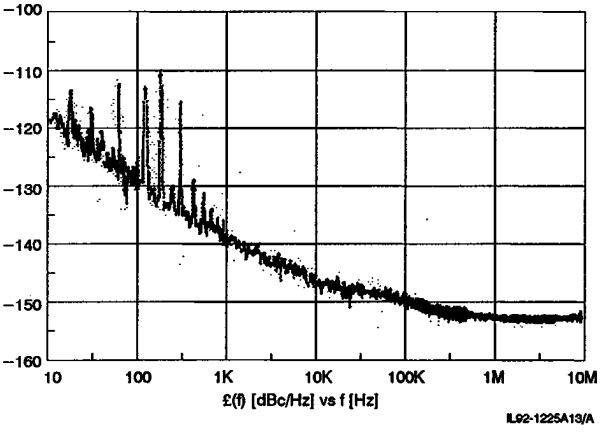


Figure 8. Phase Noise Measurement Results for a Pair of Times Sixteen (640 MHz – 10.24 GHz) Frequency Multipliers Incorporating Cascaded Stages of Schottky Diode Doublers with Interstage Amplifiers and Bandpass Filters

2. Transistor Doublers

Class C, transistor amplifiers can be arranged in a "push-push" circuit configuration to perform frequency doubling in much the same manner as the diode doubler. In the case of the transistor doubler, there is a net conversion gain of typically 6 to 10 dB. Figure 9 shows a simplified schematic diagram for the transistor doubler, and figure 10 shows additive phase noise measurement results for a pair of 80 MHz–160 MHz, transistor doublers. As shown in the figure, both the flicker-of-phase and white phase noise performance of the transistor doubler are poorer than those exhibited by the schottky diode doublers. In the case of the transistor doublers, phase noise performance is influenced by transistor type, drive level, operating frequency, and emitter degeneration [6,8].

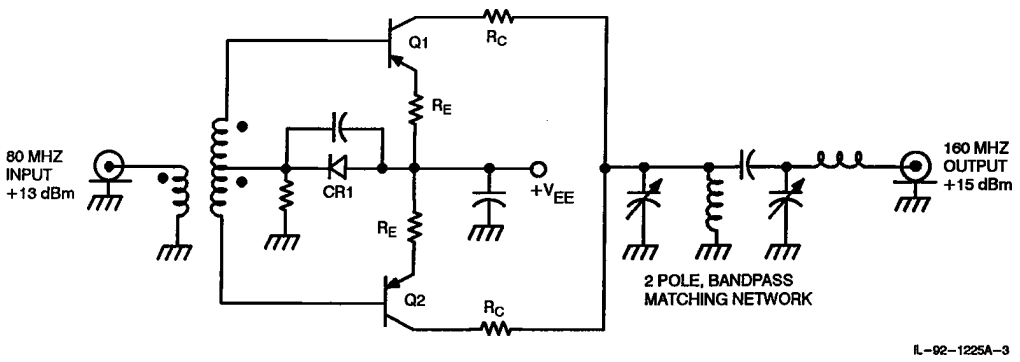


Figure 9. Simplified Schematic Diagram for a Transistor "Push-Push" Type Frequency Doubler

Figure 11 shows a summary of the measurement results obtained for various types of frequency multipliers. As can be seen from the figure, there is a wide variation in device type near-carrier and noise floor performance, with the schottky diode doublers exhibiting lowest noise performance, but also lowest multiplication factor. Multiplier circuit spectral performance must therefore be traded off against other circuit performance parameters such as overall circuit cost, complexity, and power consumption.

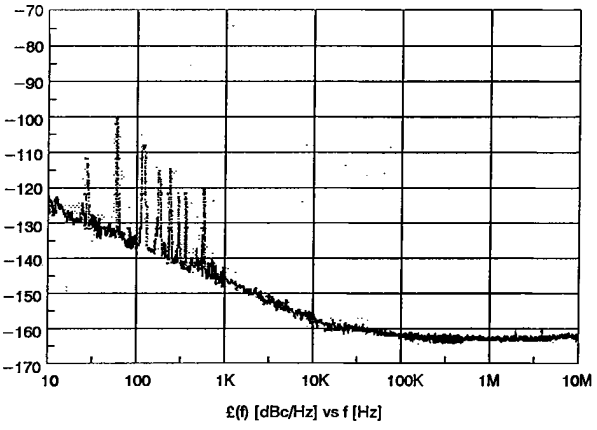


Figure 10. Phase Noise Measurement Results for a Pair of (13 dBm Drive Level, 80 – 160 MHz) Transistor "Push-Push" Type Frequency Doublers

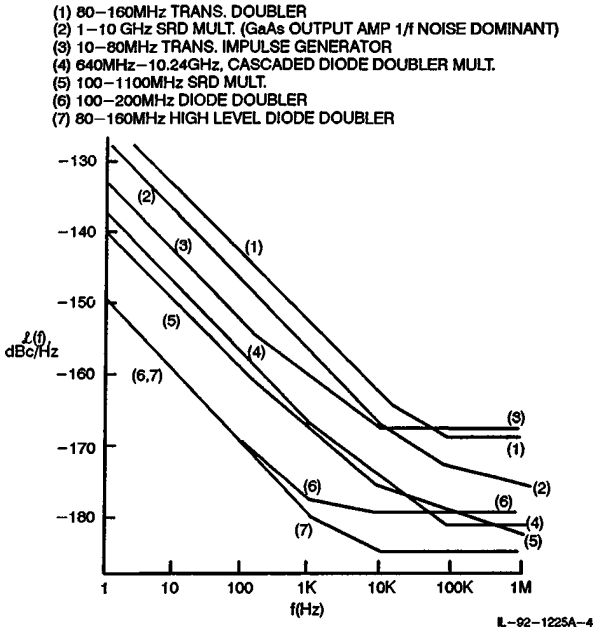


Figure 11. Summary of Frequency Multiplier Phase Noise (Referred to Device Input) Measurement Results

VARACTOR DIODE FREQUENCY DIVIDERS

Varactor diode or parametric, frequency dividers (and multipliers) are well-known [1,9–10]. Although various analyses of circuit operation have been described in detail in the literature, not much information is available with regard to circuit phase noise performance. The divide-by-two, varactor frequency divider operates by generating negative conductance at half the input (diode pump) frequency. A generalized block diagram is shown in figure 12. Experiments performed at Westinghouse using a straightforward, 'cookbook' design for a relatively low frequency (40 MHz to 20 MHz) divider indicated: (1) optimum efficiency is typically

obtained for moderately high (20 dBm) input drive levels, (2) input and output matching networks are relatively narrow-band, owing to the large ratio between diode and generator/load resistances, (3) in spite of matching network losses, divider efficiency is quite good, and (4) compared to digital, logic-level dividers, varactor divider flicker-of-phase noise level is similar, and white phase noise (floor) level is much lower [9]. Figure 13 shows the additive phase noise spectra measured for the (figure 12) divider circuits.

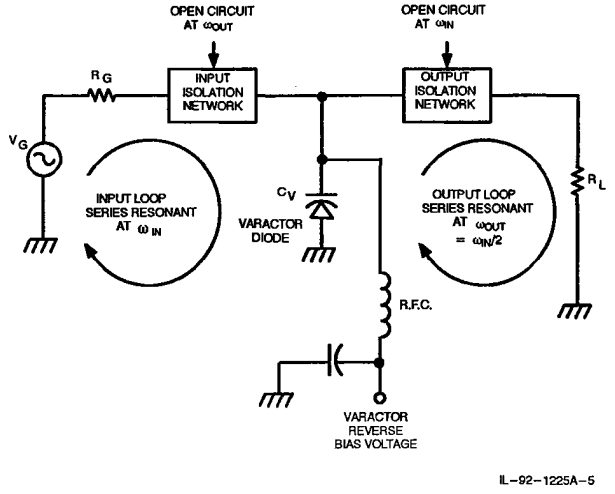


Figure 12. Simplified Schematic Diagram for a Varactor Diode Type (Divide-by-Two) Frequency Divider

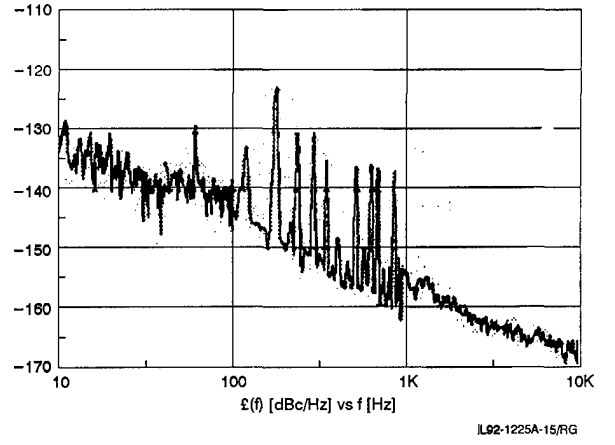


Figure 13. Phase Noise Measurement Results for a Pair of Varactor Diode Type (40–20 MHz) Frequency Dividers

REGENERATIVE FREQUENCY DIVIDERS

Regenerative frequency (divide-by-two) division, while not providing the (divide ratio) versatility of digital, logic level devices, offers several important circuit performance advantages [1,11–13]. As shown in figure 14, the regenerative divider acts as an oscillator operating at the output frequency and incorporates a frequency conversion (double-balanced

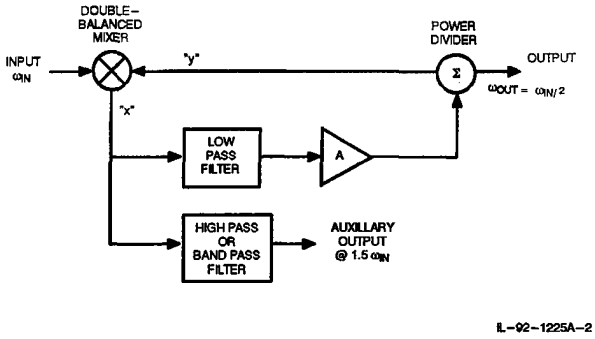


Figure 14. Simplified Schematic Diagram for a Regenerative Type (Divide-by-Two) Frequency Divider

mixer) in the oscillator positive feedback loop signal path. Steady-state, loop (unity) gain and ($2N\pi$ radians) phase requirements are the same as for a conventional, feedback loop oscillator, except that the requisite signal phase shift across the mixer (from nodes x to y in figure 14) occurs when the oscillating output signal assumes the corresponding, requisite time relationship with that of the input signal. Although conditions for stable circuit operation have been defined differently, depending on circuit analysis approach, a common point of agreement is that widest, stable operating bandwidth requires use of minimal loop signal propagation/group delay. Analyses of divider signal phase relationships show that the loop amplifier and mixer open-loop, flicker-of-phase noise sideband level is reduced by 6 dB at the divider output [13]. The circuit can be configured so that mixer conversion loss is feedback signal level-dependent. The mixer can therefore provide the requisite loop AGC mechanism so that steady-state operation is obtained for signal levels below those driving the loop amplifier into gain compression. The advantages of the regenerative divider include: (1) very low flicker-of-phase noise and noise floor performance (coincident with use of high compression point mixers and loop amplifiers exhibiting low flicker-of-phase noise, (2) typical octave, stable operating bandwidth, and (3) straightforward design using 50-ohm modular components. In addition, a potentially useful, second output signal at 1.5 times the input signal frequency is available at the mixer output.

Figure 15 shows the (typical) measured phase noise performance for a pair of VHF regenerative dividers designed at Westinghouse [11].

DIGITAL, LOGIC-LEVEL DIVIDERS

Due to low cost and design flexibility, logic dividers play an important role in both commercial and military frequency synthesis applications. Bipolar, fixed-modulus dividers are available with input frequencies up to 2.5 GHz, dual-modulus dividers with input frequencies up to 1 GHz, and fully programmable counters and dividers with input frequencies up to 600 MHz.

Logic divider phase noise characteristics are normally not vendor-specified. As is the case with many types of analog frequency synthesis components such as RF amplifiers, mixers, frequency doublers, it is necessary, then, to accumulate a "library" of phase noise performance data on a variety of

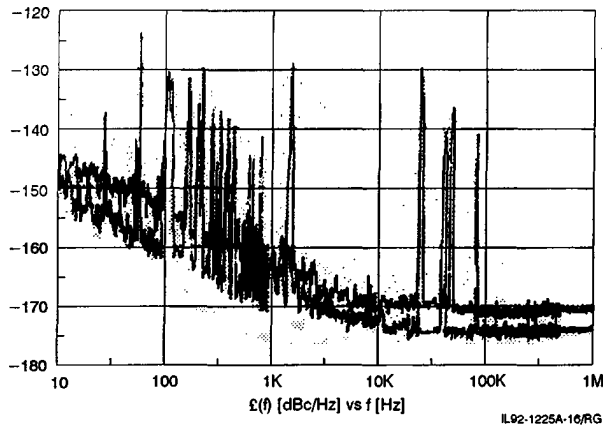


Figure 15. Phase Noise Measurement Results for a Pair of Regenerative (160 – 80 MHz) Frequency Dividers. Top Curve: Measured Performance for Two Dividers. Bottom Curve: Measured Instrumentation Noise Level

candidate devices. The following portion of this paper contains a brief analysis of phase noise characteristics followed by measurement results.

1. Analysis

In a logic divider output, transitions occur in response to a clock input signal. Phase noise is a measure of unpredictable, random variations in transition times at the output for a well defined input and can be modelled in the following way. In figure 16, the solid-lined waveform represents an ideal output response, and the dashed-lined waveform represents an actual response. The difference is a sequence of narrow, random-amplitude pulses which characterize the noise. The delay variations in figure 16 are exaggerated. In practice, delay variations are on the order of a few picoseconds and the rise and fall times on the order of a few nanoseconds, so that, to a good approximation, the width of a particular noise pulse is equal to the rise or fall time of the output signal and the amplitude $A_n = \Delta\tau_n\, dV/dt$ is proportional to the random variations in delay. It is convenient to suppose that the noise pulse train repeats every second so that a Fourier series representation of the nth pulse consists of discrete frequencies at 1 Hz intervals. The power spectral density of this pulse is:

$$S_{nk} = 2A_n^2 t_r^2 \frac{\sin^2(\pi f_k t_r)}{(\pi f_k t_r)^2} \qquad (1)$$

If the risetime t_r is small compared to the period of the output waveform, the spectral density of the frequency components close to the carrier can be approximated by:

$$S_{nk} = 2(A_n t_r)^2 = 2(\Delta\tau_n \frac{dV}{dt} t_r)^2 = 2(\Delta\tau_n V)^2 \qquad (2)$$

Also, if the random variations in the output are uncorrelated pulse-to-pulse, the power spectral density of the entire noise pulse train is the sum of the spectral densities of the individual pulses,

$$S_k = \sum_{n=1}^{2f_c} 2\Delta\tau_n^2 V^2 = 4f_C(\Delta\tau_{RMS})^2 V^2 \qquad (3)$$

where f_c is the frequency of the output waveform. Finally, if the ideal output response is approximated by a square wave, the ratio of the noise power spectral density to the fundamental component of the output signal is

$$\mathcal{L}(f) = 10\ log[2\pi^2 f_c(\Delta\tau_{RMS})^2] \qquad (4)$$

This result characterizes single-sideband, wideband phase noise in logic devices in terms of the frequency of the output signal and the RMS jitter; it is independent of risetime and, in general, insensitive to the transition characteristics of the output signal.

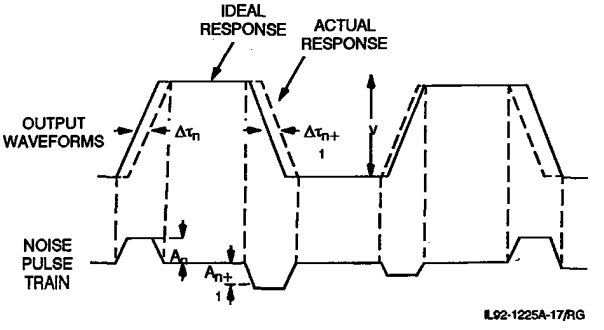


Figure 16. Representation of Divider Wideband Phase Noise as a Narrow Pulsewidth, Pulse Train

If the input to a divider is corrupted with wideband phase noise, the effect of the divider is to reduce the frequency f_c so that a noiseless divider would reduce phase noise at the output by 10 logN where N is the divide ratio. If the delay variations in figure 16 are correlated pulse-to-pulse then the noise is characterized as narrowband. In this case, the effect of a noiseless divider is to reduce the phase noise by 20 logN. From a frequency point of view, phase noise improvement in the wideband case is degraded by folding of the frequency components about zero frequency.

2. Measurement Results

Table 1 summarizes additive phase noise obtained for a number of different logic dividers. The table includes measurements on comparators and translators because it is often convenient to use these devices in practical designs. Figures 17 and 18 show actual measurement data obtained for ACT CMOS and ECL dividers. These are two source measurements. Assuming equal noise contributions from each device, per-device phase noise levels are 3 dB lower than indicated by the plots. The table 1 results are consistent with results obtained by others and available in published literature [6, 14–15].

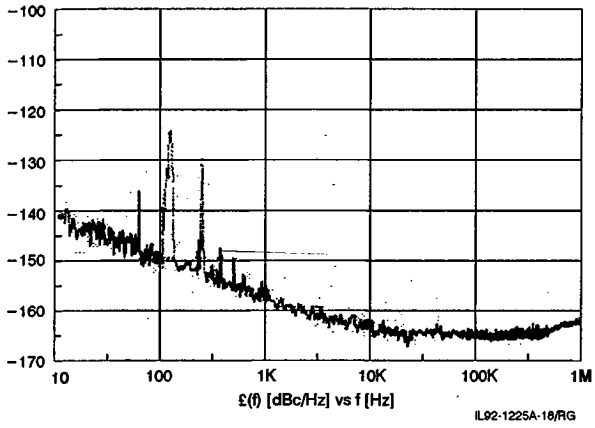


Figure 17. Phase Noise Measurement Results for a Pair of CMOS ACT (Divide-by-Four, 20 – 5 MHz) Frequency Dividers

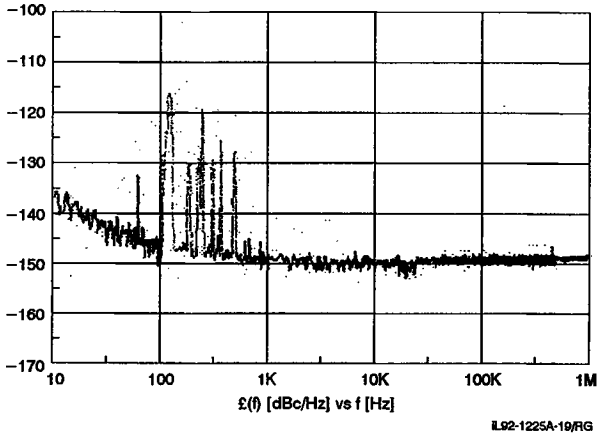


Figure 18. Phase Noise Measurement Results for a Pair of ECL (Divide-by-Eight, 400 – 50 MHz) Frequency Dividers

Table 1. Summary of Phase Noise Measurement Results (Referred to Device Output) for Logic Dividers

DEVICE	DESCRIPTION	OUTPUT LOGIC LEVEL	OUTPUT FREQUENCY (MHz)	£(f) FLOOR (dBc/Hz)	£(f) @100 Hz (dBc/Hz)
10H125	TRANSLATOR	TTL	20	161	153
LT1016	COMPARITOR	TTL	12	152	150
74ACT161	DIVIDE-BY-4	TTL	5	169	153
74HC161	DIVIDE-BY-4	TTL	5	166	150
74AS161	DIVIDE-BY-4	TTL	5	162	152
74F161	DIVIDE-BY-4	TTL	5	163	152
74LS161	DIVIDE-BY-4	TTL	5	157	151
74161	DIVIDE-BY-4	TTL	5	157	151
10115	RECEIVER	ECL	12	159	156
MC12011	DIVIDE-BY-8	ECL	50	153	150
SP8828	DIVIDE-BY-8	ECL	60	149	145
SP8828	DIVIDE-BY-8	ECL	230	141	134
SP8818	DIVIDE-BY-8	ECL	60	154	148
SP8818	DIVIDE-BY-8	ECL	230	149	141

CONCLUSIONS

For frequency multiplication circuitry, schottky diode doublers exhibit lowest overall phase noise performance over the entire HF through microwave frequency range. Unfortunately, per-device multiplication factor is lowest. In contrast, SRD multipliers offer single-step, high-order multiplication but poorer noise performance, compared to the diode doublers. When consideration is given to necessary use of interstage amplifiers in high-order multipliers using cascaded doublers, the SRD approach is superior from a DC power consumption standpoint.

Ironically, a similar situation exists for frequency division circuitry, with analog regenerative (divide-by-two) dividers providing superior noise performance, compared to (higher division ratio) logic dividers.

REFERENCES

[1] V. Manassewitsch, "Frequency Synthesizers Theory and Design," John Wiley and Sons, New York, 1976, Chapter 6.

[2] "Step Recovery Diode Frequency Multiplier Design," Hewlett-Packard Application Note 913.

[3] Step Recovery Diode Component Catalogues, Hewlett-Packard, Inc., and Herotek, Inc..

[4] "Harmonic Generation Using Step Recovery Diodes and SRD Modules," Hewlett-Packard Application Note 920.

[5] R. A. Baugh, "Low Noise Frequency Multiplication," Proc. 26th Freq. Contr. Symp., June, 1972, pp. 50–54.

[6] F. L. Walls, "Low Noise Frequency Synthesis," Proc. 41st Freq. Contr. Symp., May, 1987, pp. 512–518.

[7] M. M. Driscoll, et.al., "Extremely Low Phase Noise UHF Oscillators Utilizing High-Overtone, Bulk Acoustic Resonators," Proc. 1990 IEEE Ultras. Symp., Dec., 1990, pp. 513–518.

[8] D. Halford, et.al., "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure," Proc. 22nd Freq. Contr. Symp., April, 1968, pp. 340–341.

[9] J. Hillibrand, et.al., "Semiconductor Diodes in Parametric Subharmonic Oscillators," RCA Review, June, 1959, pp. 229–253.

[10] P. Penfield and R. P. Rafuse, "Varactor Applications," M.I.T. Press, 1962, Chapter 9.

[11] M. M. Driscoll, "Phase Noise Performance of Analog Frequency Dividers," IEEE Trans. UFFC, Vol. 37, No. 4, July, 1990, pp.295–301.

[12] R. G. Harrison, "Theory of Regenerative Dividers Using Double Balanced Mixers," 1989 IEEE MTT-S Symp. Digest, Vol. 1, June, 1989, pp.459–462.

[13] V. F. Kroupa, "Frequency Synthesis," John Wiley and Sons, New York, 1973, Chapter 3.

[14] W. F. Egan, "Modeling Phase Noise in Frequency Dividers," IEEE TRans. UFFC, Vol. 37, No. 4, July, 1990, pp. 307–315. 15.

[15] D. E. Phillips, "Random Noise in Digital Gates and Dividers," Proc. 41st Freq. Contr. Symp., May, 1987, pp. 507–511.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

FREQUENCY INSTABILITIES IN PHASE-LOCKED SYNTHESIZERS INDUCED BY TIME DELAYS

W. Wischert[1,2] M. Olivier[1] and J. Groslambert[1]

[1] Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté, 32, av. de l'Observatoire, F-25000 Besançon, France

[2] Institut für Theoretische Physik und Synergetik,
Universität Stuttgart, Pfaffenwaldring 57/4, W-7000 Stuttgart 80, Fed. Rep. of Germany

Abstract

This paper deals with the influence of time delays on the dynamical behaviour of phase-locked synthesizers. We have performed experiments where analog and/or numerical delay lines were introduced in a phase-locked loop (PLL), the main building block of such devices.

We show that even a first order PLL can exhibit a variety of instabilities leading to a chaotic behaviour. These instabilities which can lead to a loss of synchronization of the loop are the origin of phase- and frequency instabilities of the whole device. A theoretical analysis is performed in order to compare the experimental results with the theory.

Introduction

Frequency Synthesizers based on the phase-locked principle are nowadays in common use. Their main building blocks consist of PLL's and frequency dividers. Dynamic instabilities of such loops are the origin of phase- and frequency instabilities in the synthesizer [1].

To implement PLL's with a large locking range one has to choose a high open loop gain since there is a direct proportionality between these two quantities. As will be shown below a high open loop gain can cause retardation effects of the signal to play an important role of phase- and frequency instabilities in the synthesizer.

Time delays are likely to occur in engineering sciences. In radar systems they can become quite large. In a synthesizer they have their physical origin mainly in the propagation time of the signal through frequency dividers.

From the point of view of dynamical system theory a nonlinear control system where delay effects play a role can be characterized by the following diagram

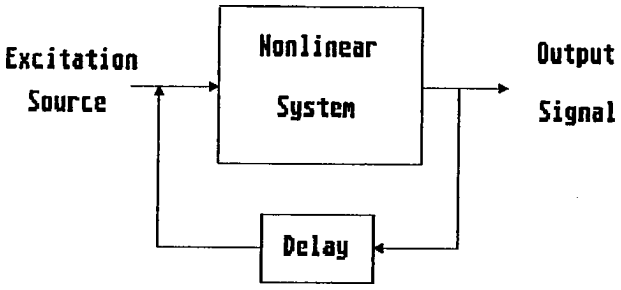


Fig.1. Simplified Diagram of a Control System with delay

A mathematical description of such a system leads in the simplest case to an evolution equation of the following general form

$$\dot{\mathbf{q}}(t) = \mathbf{N}(\mathbf{q}(t),\mathbf{q}(t-\tau)), \qquad (1)$$

$\mathbf{q}(t)$ stands thereby for the dynamic variable of interest, $\mathbf{N}$ is a vector-valued nonlinear function of its arguments and τ represents the time delay. Eq. (1) is a so-called delay-differential equation.

We would like to remark that such an equation always occur if delay effects are taken into account. The reason why its mathematical treatment is so difficult lies in the fact that Eq. (1) has a functional character, that is, represents an infinite dimensional system and as a consequence, can exhibit a rich dynamical behaviour.

0-7803-0476-4/92 $3.00 © 1992 IEEE

In what follows we present first the experimental apparatus and derive a model equation for it. We then proceed to a linear and nonlinear stability analysis of this model equation in order to understand the effects produced by varying the time delay. Finally the experimental results are presented and compared with the theoretical results.

Derivation of a model equation

To examine delay effects in synthesizers we have focused our interest to instabilities in their main element namely PLL's. For the sake of simplicity we examine a first order PLL where a delay line is introduced. This is done because such a PLL, after transients have died out, doesn't show any dynamical behaviour in the absence of delay. So we can be sure that all the instabilities we observe are induced by retardation effects. The experimental apparatus with the analog delay line is shown in Fig. 2.

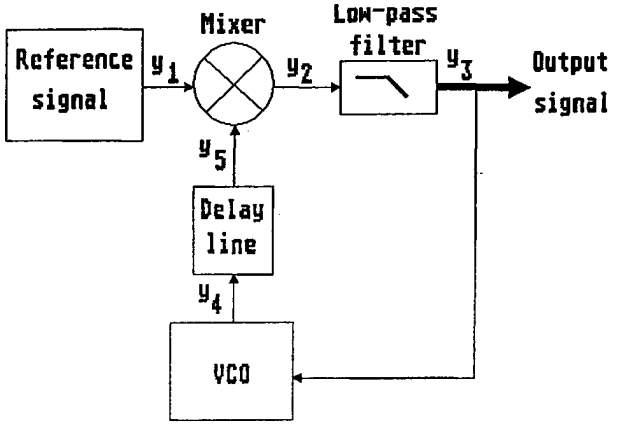


Fig. 2. First order phase-locked loop with delay

The low-pass filter shown in Fig. 2. is assumed to have a sufficiently high bandwith as not to influence the loop properties and is only used to eliminate higher frequency components at the output of the mixer.
It is only the analog line which can be placed at the output of the VCO. If we work with the numerical delay line we have to place it directly after the low-pass filter because its sampling time is too low to support the high frequency oscillations at the output of the VCO. It can easily be shown, that the theoretical results are independent of the positionnement of the delay line.

The reference signal is supposed to be an oscillation with angular frequency ω_0, that is

$$\begin{aligned} y_1(t) &= A \sin \Theta(t) \\ \dot{\Theta}(t) &= \omega_0 \ . \end{aligned} \tag{2}$$

The voltage controlled Oscillator (VCO) produces a signal whose angular frequency depends linearly on the incoming tension y_3

$$\begin{aligned} y_4(t) &= B \cos \Psi(t) \\ \dot{\Psi}(t) &= \omega + K_1 y_3(t) \ , \end{aligned} \tag{3}$$

ω is the central angular frequency of the VCO and K_1 its sensitivity.
The delay line produces a temporal retardation of the output signal of the VCO, so that at the input of the mixer we have

$$y_5(t) = y_4(t-\tau) \ . \tag{4}$$

The mixer working as a phase detector is the only nonlinear element in the circuit. If it has a sinusoidal characteristic its output signal is proportional to the product of the two input signals

$$\begin{aligned} y_2(t) &= \beta y_1(t) y_5(t) \\ &= \frac{\beta AB}{2} \left(\sin[\Theta(t)-\Psi(t-\tau)] + \sin[\Theta(t)+\Psi(t-\tau)] \right), \end{aligned} \tag{5}$$

$\frac{\beta AB}{2} = \mu$ is the sensitivity of the mixer.
Higher frequency components of $y_2(t)$ i.e. the second term of the right hand side of Eq. (5) are eliminated by the low-pass filter which leads to

$$y_3(t) = \mu \sin [\Theta(t)-\Psi(t-\tau)] \ . \tag{6}$$

We introduce the phase difference $\varphi(t)$ between the reference signal and the output signal of the VCO, the so-called phase error as the dynamic variable

$$\varphi(t) = \Theta(t) - \Psi(t-\tau) \ , \tag{7}$$

If Eq. (7) is differentiated with respect to time t we obtain the final equation describing the dynamics of a first order PLL with time delay

$$\dot{\varphi}(t) + K \sin [\varphi(t-\tau)] = \omega - \omega_0 \ , \tag{8}$$

where $K = K_1 \mu$ is the open loop gain of the PLL.

Eq. (8) has the same form as the general Eq. (1) and represents a first order scalar delay differential equation for the phase error signal.

In what follows we already assume that the PLL is synchronized to the reference signal, so that the right hand side of Eq. (8) can be omitted. The steady state solutions $\bar{\varphi}$ of Eq. (8) are then given by

$$\bar{\varphi} = 0 \pmod{2\pi} \ . \tag{9}$$

While these steady states are always stable in a PLL without delay and therefore synchronization always take place, the stability of the steady state solutions in the presence of a time delay can be influenced. This is why the dynamics of delayed systems can be drastically different from the ones without any retardation effect.

Linear Stability Analysis

To perform a linear stability analysis we disturb the system from its steady state

$$\varphi(t) = \bar{\varphi} + \delta\varphi(t) \quad , \tag{10}$$

with $\delta\varphi(t)$ a small quantity. The hypothesis

$$\delta\varphi(t) \sim e^{\lambda t} \tag{11}$$

leads, keeping only linear terms, to a transcendental equation for the eigenvalues λ

$$\lambda + K e^{-\lambda\tau} = 0 \; . \tag{12}$$

Eq. (12) tells us, that the system has an infinity of (complex) eigenvalues and an infinity of corresponding eigenvectors. Its dynamical behaviour can only be correctly described in an infinite dimensional phase space [2].

We also see, that in the case without delay (τ=0), there is only one solution of Eq. (12), namely λ=-K, which is always real and negative, expressing the fact that the steady state solutions never become unstable.

If we split the complex eigenvalue λ into its real and imaginary part ($\lambda=\gamma+j\Omega$), Eq. (12) is equivalent to

$$\begin{aligned} \gamma + K e^{-\gamma\tau} \cos\Omega\tau &= 0 \\ \Omega - K e^{-\gamma\tau} \sin\Omega\tau &= 0 \; . \end{aligned} \tag{13}$$

The steady state solution $\bar{\varphi}$ loses its stability if the real part of one of the eigenvalues becomes positive. At the threshold ($\gamma\equiv 0$) Eq. (13) leads immediately to

$$\Omega = K \; . \tag{14}$$

The first of the Eqs. (13) has the solutions

$$\Omega\tau = \left(n + \frac{1}{2}\right)\pi, \quad n \in \mathbb{N}_0 \; , \tag{15}$$

so that the condition for the first instability (n=0) is

$$K\tau \geq \frac{\pi}{2} \; . \tag{16}$$

With Eq. (16) Eq. (14) can be rewritten as

$$\Omega = \frac{\pi}{2\tau} \; . \tag{17}$$

The angular frequency of the oscillation is therefore inversely proportional to the time delay.

Eq. (16) shows clearly that delay induced instabilities in PLL's can occur at very low retardation times if the open loop gain is sufficiently high.

Nonlinear Stability Analysis

The linear stability analysis provides a method to determine the frequency of the oscillation at threshold. The corresponding amplitude can only be obtained by a nonlinear stability analysis.

The disturbed solution of Eq. (8) can be written as

$$\varphi(t) = \bar{\varphi} + M \sin\Omega t + \varepsilon R(t) \tag{18}$$

with M the amplitude of the oscillation. R(t) contains all higher harmonic terms of the solution and the perturbation parameter ε expresses the fact that these terms are supposed to be small. Insertion of Eq. (18) into Eq. (8) gives, after some tedious calculations, in lowest order of the perturbation parameter, an implicit equation for the amplitude of the oscillation

$$M = \frac{4K\tau}{\pi} J_1(M) \, , \tag{19}$$

J_1 is the first Besselfunction of the first kind.

The first oscillatory instability, a so-called Hopf bifurcation is now completely described. We derived analytical expressions for the angular frequency and the amplitude of the oscillation both as a function of the open loop gain and the time delay.

We would like to remark that if the nonlinear function in Eq. (8) would have a monotonic form, the instability found would be the only one since such functions can't show other instabilities than an oscillatory one [3]. In the case of a mixed nonlinear function, that is, a function which is neither strictly decreasing nor strictly increasing, as it is the case in a PLL, other instabilities can occur if K or τ are varied.

Unfortunately there is no theory available until now which could describe these higher bifurcations. This is mainly due to the fact that no complete Floquet theory exists for delay differential equations. What we can only say is that numerical calculations showed us that the oscillating solution bifurcates further to a chaotic state as the control parameters are varied.

It should be mentioned that in physical sciences a lot of work was done over the past ten years in order to understand the dynamical behaviour of systems influenced by time delays and in particular the chaotic phenomena which arise. An important study was consacreted to acousto- and electrooptical devices in the field of optical bistability (see e.g. [4,5]). The underlying model equations have practically the same form as the one we derived. This indicates that chaotic phenomena in PLL's are also likely to occur.

Experimental Results

The experimental setup of a first order PLL with time delay is shown in Fig. 2. The experiments were done at a reference frequency of 135 MHz with analog and numerical delay lines which were introduced into the loop. In the case of the analog delay line, delays could be fixed at 15 μsec and 30 μsec. The characteristic of such a delay line is shown in Fig. 3.

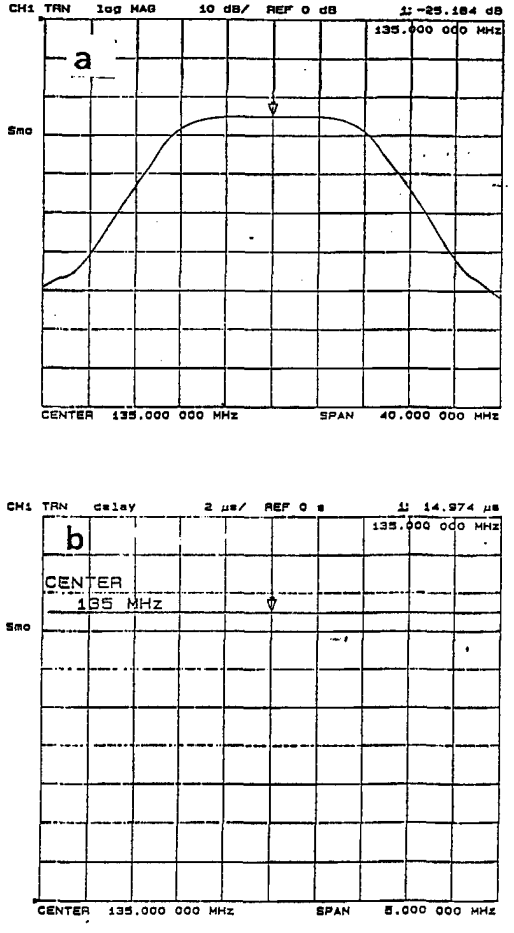


Fig. 3. Characteristic of an analog delay line
a) Transmission characteristic $S_{21}(f)$
b) Delay versus frequency

Fig. 3 shows clearly that in the frequency domain considered the analog delay line is quasi dispersionless. So our theoretical assumption of a fixed value of τ (see Eq. (4)) is in accordance with the experiments.

In the case of the numerical delay line the retardation could be varied between 70 μsec and 500 msec. Fig. 4. represents a schematic diagram of a numerical delay line.

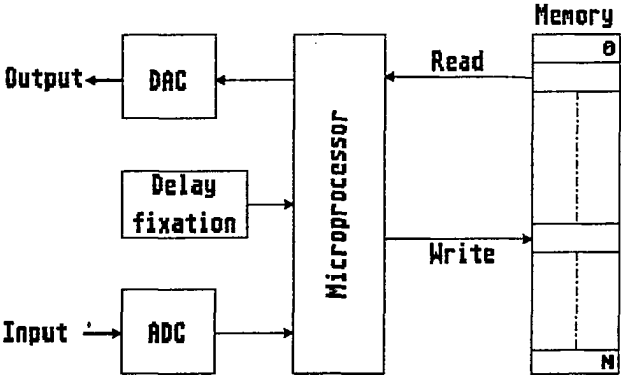


Fig. 4. Schematic diagram of a numerical delay line. The minimum delay T_s is fixed by the sampling time of the two converters and the read and write procedure to and from the memory. The delay τ is determined by the reading of the information which is written at a position $N=\tau/T_s$. The maximum delay is limited by the capacity of the memory.

Due to its relatively low sampling rate this delay line had to be introduced directly after the mixer, as already mentioned above.

In a first step we examined the locking behaviour at fixed delay τ when the open loop gain K was slowly varied. Assuring that synchronization takes place at a given open loop gain, we increased the gain by varying either the amplitude of the reference signal or the sensitivity of the VCO. At low gain the locking behaviour didn't change. In increasing K further we arrived at a threshold where the phase error signal began to oscillate. This phase instability, however, didn't destroy the locking of the PLL. We repeated the same procedure with other delay values in order to show that the instability is uniquely due to the delay line. The experimental result of this first oscillatory instability can be seen in Fig. 5., where as a comparaison the theoretical curve (Eq. (17)) is also included.

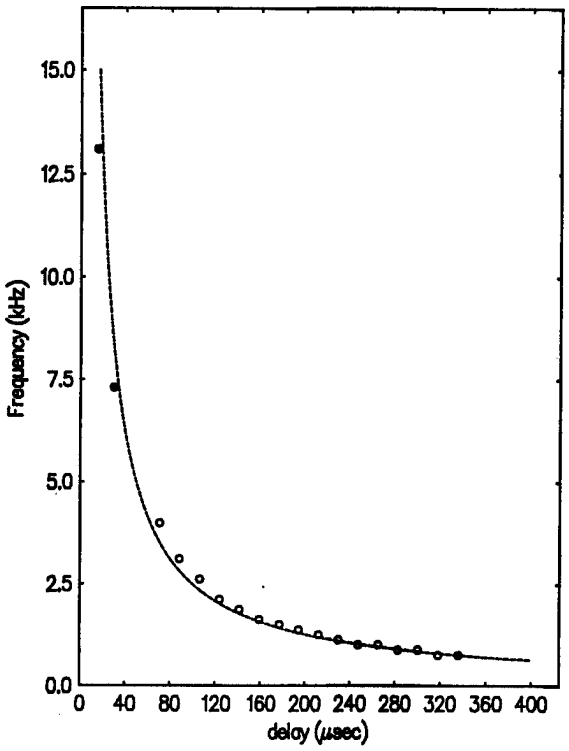


Fig. 5. First oscillatory instability of a PLL with delay
o numerical delay line
• analog delay line
— theoretical result

At higher values of the gain other instabilities appeared indicating the begining of a chaotic behaviour of the loop. The study of chaotic phenomena in PLL's became more popular in recent years [6-8]. However, all the experiments which were done so far consist of second order PLL's driven by an external force. Their dynamics are, as a consequence, very similar to the ones of the driven pendulum, a well studied system.

With our experiment we present another way to study chaotic behaviour in PLL's. To generate at least three-dimensional phase spaces one doesn't have to refer to experiments like the former ones, where the external forcing is necessary to provide the third degree of freedom. One can instead simply use the fact that a time delay, if it is sufficiently high can generate in many nonlinear systems a phase space dimension which can be as high as one wishes. This is why our experimental studies about the chaotic behaviour in PLL's can be done with first order PLL's without any external forcing term. We are not aware of any study of higher instabilities with such a system, although it seems to be the simplest one to be realizable.

In increasing the open loop gain further while keeping the delay at a fixed value, we observed effectively higher instabilities of the phase error signal leading to a chaotic behaviour. This was visible by a broadband power spectrum which replaced the discrete line spectrum in the regular regime. In Fig. 6. the power spectrum for different open loop gains is presented, clearly showing a period doubling to chaos, one of the fundamental and frequently observed transitions.

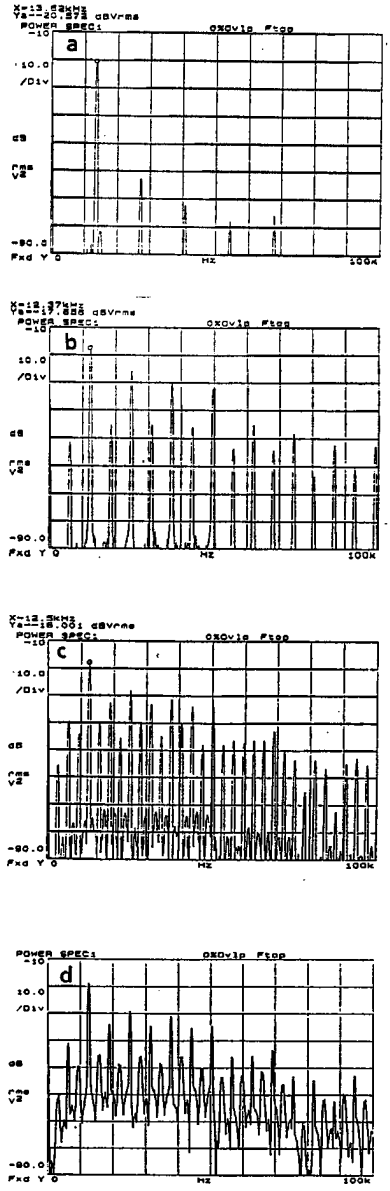


Fig. 6. Power spectrum of a first order PLL with delay at the transition to chaos for different values of the open loop gain.
a) oscillation
b) and c) period doubling
d) chaotic regime

We would like to remark that an experimental study and a theoretical explanation of the observed phenomena is extremely difficult since we are now operating the PLL in a fully nonlinear domain, where the assumptions about the dynamical behaviour of the components of the loop will be certainly modified.

Conclusion

We have analyzed the dynamics of a first order phase-locked loop where a variable delay line was introduced. Theoretical calculations performed in the time domain showed us that an instability of the phase error signal can be induced by time delays. Experimental measurements have confirmed this instability and revealed also a chaotic behaviour of the phase error signal at very high open loop gain. Thus, the requirement of a high locking range makes delay induced instabilities an important instability mechanism in frequency synthesizers.

Acknowledgments

One of us (W.W.) would like to thank J.J. Gagnepain for his hospitality at the LPMO and the Robert Bosch foundation for the support of this work. All of us would like to thank C. Ferrandez for the fabrication of the numerical delay line.

References

[1] W.F. Egan, *Frequency Synthesis by Phase Lock*, (Wiley&Sons, New York 1981)

[2] J.K. Hale, *Functional Differential Equations*, (Springer, New York 1977)

[3] U. an der Heiden and M.C. Mackey, "Mixed Feedback: A Paradigm for Regular and Irregular Oscillations," in *Temporal Disorder in Human Oscillatory Systems*, eds. L. Rensing, U. an der Heiden, M.C. Mackey, (Springer, New York 1987)

[4] P. Nardone, P. Mandel, and R. Kapral, " Analysis of a delay-differential equation in optical bistability," Phys. Rev. **A33**, 2465 (1986)

[5] R. Vallée and C. Delisle, "Route to chaos in an acousto-optic bistable device," Phys. Rev. **A31**, 2390 (1985)

[6] D. D'Humieres , M.R. Beasley, B.A. Huberman, and A. Libchaber, "Chaotic states and routes to chaos in the forced pendulum," Phys. Rev. **A26**, 3483 (1982)

[7] M. Olivier and J.J. Gagnepain, "Chaotic States and Anormalous Noise in Resonators," Proc. 3rd International Conference *on 1/f* Noise, Montpellier (1983)

[8] T. Endo and L.O. Chua, " Chaos From Phase-Locked Loops I & II," IEEE Trans. Circuits Syst., vol. 35, 987 (1988) and 255 (1989)

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

STABILITY ANALYSIS OF AN 18 GHz MICROWAVE FREQUENCY SYNTHESIZER

Stephen D. Turner, P.E., MAXTECH, Inc.
2820 E. College Ave. Suite G. State College, PA. 16803
Ali Behagi, Associate Professor, Penn State Harrisburg
Middletown, PA. 17057

Abstract

The control loop stability design is presented for a single loop 18 GHz frequency synthesizer. The Root Locus is used as the primary stability analysis tool. Using the Root Locus along with the Routh Hurwitz stability criterion, the control loop can be made stable over a wide range of loop gain. It is shown that a great deal of stability data can be obtained from the root locus while giving insight to the closed loop settling time. All data from the root locus has exhibited excellent correlation with the Bode analysis techniques.

Introduction

A classical control system analysis method developed by W.R. Evans in 1948 is applied to phase locked loop systems. The root locus technique is a graphical approach to determining the relative stability and transient performance of a closed loop control system. This stability analysis is presented for an efficient class of 18 GHz frequency synthesizers. It is used as a local oscillator in an 18 GHz microwave radio system. These systems are used in Local Area Communication Networks and are often found in remote and isolated locations. As a result there are stringent performance requirements placed on reliability, size, and DC power consumption.

The synthesizer is required to cover a frequency range of 18.5 GHz to 19.2 GHz in 1 MHz steps. A DC input power requirement of 3.5 Watts maximum is placed on this synthesizer along with a 37.5 in^3 maximum package size. The switching speed was specified at 55 milliseconds. For these reasons it was decided that this synthesizer should consist of a single loop topology. Table 1 summarizes the synthesizer specifications.

Performance Parameter	Specification
Frequency Range	18.5 to 19.2 GHz
Step Size	1 MHz
Reference Frequency	120 MHz
Control Word	16 bit parallel CMOS
Output Power	+10 dBm into 50 Ω
Stability	± 10 PPM over temp.
Temperature Range	-30 °C to +85 °C
Switching Time	55 msec.
Phase Noise	-64 dBc/Hz @ 10KHz
	-84 dBc/Hz @ 100 KHz
	-104 dBc/Hz @ 1 MHz
DC Input Power	+12 vx,290 mA(3.5W)
Physical Size	10" x 3.75"1x 1"

Table 1. Synthesizer Specifications

Hardware Realization

The entire synthesizer was fabricated with discrete devices and teflon based substrates. An X-band VCO was designed and fabricated that could meet the phase noise specification when multiplied to 18 GHz [1]. The VCO utilized a silicon bipolar transistor and a resonator comprised of a microstrip circuit and a gallium arsenide hyperabrupt varactor. The VCO tunes a bandwidth of 9.00 to 9.80 GHz which provides sufficient margin to account for frequency drift over temperature. The VCO output is then multiplied to 18 GHz using an FET frequency doubler.

A coupled sample from the VCO is mixed with a multiplied reference signal of 9.720 GHz that produces an IF frequency in the VHF range. This IF is prescaled using commonly available silicon prescalers which is important to the overall DC efficiency of the synthesizer.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The synthesizer integrated circuit chosen for this design is the NJ8820 CMOS device made by Plessey Semiconductor. This device has the advantage of an integral sample and hold phase detector which increases the attenuation of phase detector reference frequency output [2]. The channel data is decoded to the variable modulus registers using a 27C64 EPROM. A block diagram of the synthesizer is shown in figure 1.

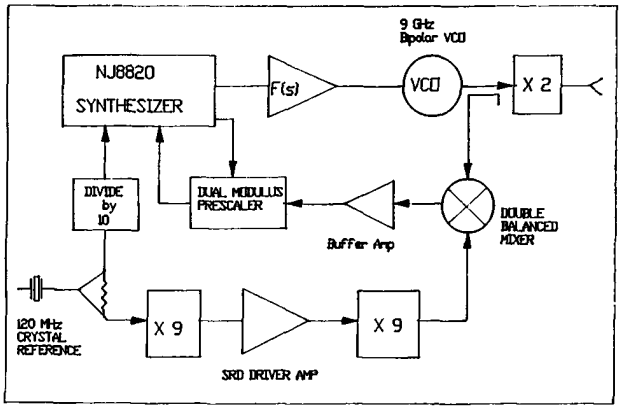


Figure 1. 18 GHz Synthesizer Block Diagram

Control Loop Design

The use of a single loop can present some difficult challenges in the compromise of loop bandwidth, optimum phase noise, and loop stability. To meet the performance specifications of this synthesizer it was determined that the loop bandwidth would have to be positioned between 500 Hz and 3 KHz. With the prescaling constraints of the NJ8820 a phase detector reference frequency of 50 KHz had to be used. This resulted in a divider modulus range of K_N=2400 to 9400 to achieve the required bandwidth of the synthesizer. A block diagram of the control loop is given in figure 2.

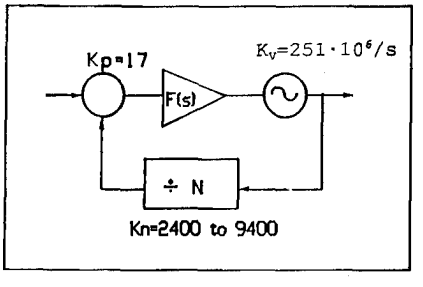


Figure 2. Control Loop Block Diagram

In determining the open loop transfer function the VCO is treated as a simple integrator, neglecting the effects of its modulation bandwidth. The value of the VCO's gain is fixed at K_v=251·10⁶ rad/s/v. The phase detector is a variable gain element with its initial value set at K_p=17 v/rad. The effect of phase detector delay is included by multiplying the open loop transfer function by an exponential with a time constant value of $T=1/f_{ref}$ [3]. The loop filter consists of a basic fourth order active low pass filter with its schematic given in Figure 3. The effect of the Op Amp's voltage gain and gain-bandwidth product is included and results in the following transfer function in pole-zero form [4].

$$F(s)=\frac{(1+T_2 s)}{\frac{s}{K1}(1+T_3 s)(1+T_4 s)+\left[\frac{(1+T_0 s)(1+T_4 s)}{A_0}\cdot\frac{1}{\left(\frac{s}{2\cdot K1}(1+T_3 s)+(1+T_2 s)\right)}\right]} \quad (1)$$

where:

$K1=1/(2R1\cdot C2)$
$T_2=R2\cdot C2$
$T_3=(R1\cdot C1)/2$
$T_4=R3\cdot C3$
A_o=Op Amp Voltage Gain
T_o=Op Amp Dominant Pole

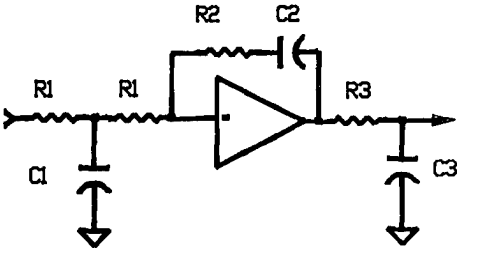


Figure 3. Loop Filter Schematic

The resulting open loop transfer function for the control loop is of the form shown in equation (2).

$$GH(s)=\frac{K_v}{s}\frac{K_p\cdot F(s)}{K_N} \quad (2)$$

Observation of the open loop transfer function shows the variable parameters of the control loop. Note that along with the zero and the poles of the loop filter, the loop filter gain, K1, and the phase detector gain, K_p, exist as variables.

As a starting point in the control loop design, the loop filter zero shall be set approximately 10 dB above the loop bandwidth of 500 Hz. The pole, 1/T3, shall be set at ten times the loop bandwidth and pole, 1/T4, set at fifty times the loop bandwidth. With the divider modulus initially set at K_N=9400, the loop filter gain, K1, is adjusted to set the loop bandwidth near 500 Hz.

Root Locus Design

Because the closed loop roots of the transfer function characteristic equation determine the control loop stability and transient response, the root locus is an effective means of analyzing the stability of the frequency synthesizer. The root locus is simply a graphical plot of these roots on the s-plane. When any parameter changes in the loop transfer function the location of the closed loop roots on the s-plane also changes. A frequency synthesizer differs from a fixed frequency phase locked loop in that the loop gain varies as the divider modulus,K_N, is varied. It is then important to guarantee control loop stability over the entire range of K_N, the operational bandwidth of the frequency synthesizer. For the stability analysis presented in this paper, the parameter of variation on the root locus is always the loop gain.

Consider the basic control system given in figure 4. The characteristic for this system is given by equation (3) [5]. This equation can be rearranged so that the loop gain appears as a variable, K_T. In this case $K_T = K_p \cdot K1 \cdot K_v \cdot 1/K_N$. The remaining zero and poles of the loop filter are denoted as P(s) and equation (3) is rewritten in the form of equation (4).

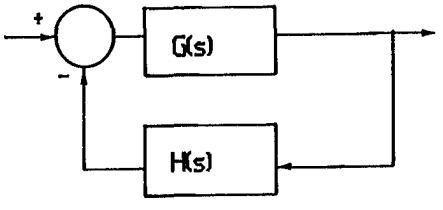


Figure 4. Basic Single Loop Control System

$$1 + GH(s) = 0 \tag{3}$$

$$1 + K_T P(s) = 0 \tag{4}$$

The root locus then consists of a graphical plot of the roots of equation (4) as K_T is varied from 0 to some specified maximum value. The roots of equation (4) must lie on the left hand side of the s-plane for conditional stability.

The initial open loop transfer function for this control loop is given in equation (5) and the resulting root locus is shown in figure 5. This is a classical root locus for a fifth order control loop containing a zero. It is obvious from this plot that a value of loop gain exists where the root locus crosses into the right hand plane and the system becomes unstable. This could result in a synthesizer that would be usable only over a very narrow bandwidth.

$$GH(s) = \frac{2.44 \cdot 10^{20}(s+2506)}{s(s^3+5.03 \cdot 10^7 s^2+9.48 \cdot 10^{11} s+5.93 \cdot 10^7)} \cdot \frac{e^{-.00002s}}{(s+113636)} \tag{5}$$

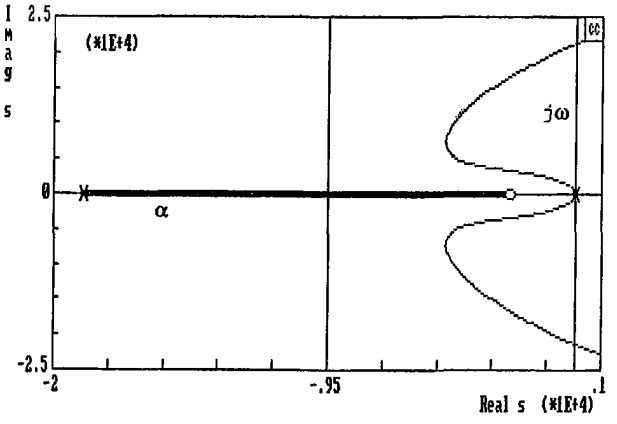


Figure 5. Root Locus of Control Loop

There are many fine references that describe the procedures for drawing the root locus [5-7]. Basically the locus of roots of the characteristic equation begin at the poles of P(s) and end at the zeros of P(s) as K_T increases from zero to infinity. An interactive CAD program is recommended to assist in the calculations involved and the iterative process of optimizing a control system. All of the root locus plots in this

paper were constructed using Program CC by Systems Technology,Inc. Note that the locus of figure 5 does not show one remaining pole at $s=-5\cdot10^7$. This pole is due to the dominant pole of the Op Amp and is far enough removed from the loop bandwidth that it presents negligible phase degradation and is not in the resolution of this root locus plot.

When examining the root locus, the area where the locus is near the jω axis is the most important. For each complex root on the locus its impact on stability is determined by its real part and the magnitude of its residue. A larger real part represents a larger decay of the transient response due to that root. The larger the residue of a particular root, in comparison to the residues of the other roots in the system, the more that root tends to dominate the overall system response. The root locus of figure 5 shows that the complex conjugate pair of roots dominate this fifth order system. Because these dominate complex roots are conjugates, they are symmetric with the real axis. The root locus crosses the positive jω axis and the negative jω axis at the same value of K_T. For this reason the root locus can be magnified to show only the upper left quadrant of the s-plane and expanded around the area of the dominant complex conjugate roots. Figure 6 shows such a root locus plot.

A great deal of stability information can be obtained from the root locus. As previously mentioned the limiting value of loop gain, K_T, can be found where the root locus crosses the jω axis. A related parameter, the gain margin, can also be obtained from the root locus. The gain margin is defined by equation (6). A system is generally considered stable for gain margins of at least 10 dB [6].

$$Gain\ Margin = 20\cdot\log\frac{K\ at\ j\omega\ crossover}{K\ at\ operating\ point} \qquad (6)$$

The system phase margin can be obtained from the root locus and is defined by equation (7).

$$Phase\ Margin = 180^\circ + arg\ GH(j\omega') \qquad (7)$$

Here jω' is the point on the jω axis at the operating value of loop gain. By substituting this value of ω into the open loop transfer function the angle of the complex transfer function, argGH(jω'), can be obtained. Using Program CC this is easily obtained by moving the graphics cursor to the desired value of loop gain on the jω axis. The phase margin is then read directly from the cursor readout as the argument of the loop gain exponential.

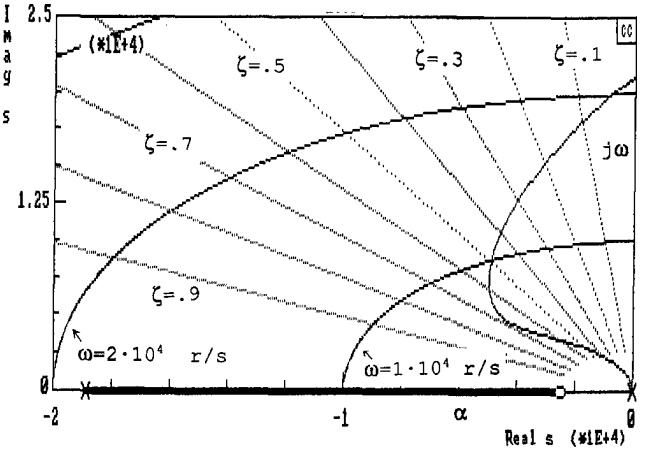


Figure 6. Root Locus-Upper Left Quadrant displaying damping factor,ζ ,and ω lines

The closed loop damping ratio for the dominate complex roots can also be found on the root locus. Damping ratio lines are placed on the s-plane by drawing lines from the origin at a angle of $\pm\ \cos^{-1}\zeta$ with the negative real axis. Additionally the closed loop bandwidth can be obtained for the dominate roots by displaying the ω lines on the s-plane. Both damping and omega lines are shown on the root locus in figure 6.

Routh-Hurwitz Criterion

The limiting value of loop gain where the root locus crosses the jω axis can be evaluated using the Routh-Hurwitz technique [5]. The Routh-Hurwitz technique utilizes the characteristic equation of the loop transfer function of equation (5). Consider equation (5) with the gain value of $2.44\cdot10^{20}$ replaced by the variable K. The transfer function is then factored into its characteristic equation (8). Note that the exponential multiplier has been excluded for simplicity.

$$p(s) = s^5 + 5.04\cdot10^7 s^4 + 6.65\cdot10^{12} s^3 + 1.08\cdot10^{17} s^2 + (6.74\cdot10^{12} + K)\, s + 2506K \qquad (8)$$

The coefficients of s, a_5 through a_0, in the characteristic equation are arranged into an array as described by reference [5]. These coefficients are also used to define the remaining elements, b through e, of the array shown in figure 7.

s^5	a_5	a_3	a_1
s^4	a_4	a_2	a_0
s^3	b_1	b_2	0
s^2	c_1	c_2	0
s^1	d_1	d_2	0
s^0	e_1	e_2	0

Figure 7. Routh-Hurwitz Array for a fifth order system

Moving down the first column of coefficients in the array, the last three terms exist as functions of K. These three terms are each set equal to zero and solved for K. The minimum non zero value of K is that normalized value where the root locus crosses the jω axis. It is then obvious that the Routh-Hurwitz criteria becomes an integral part of the root locus analysis. In fact most CAD programs allow the K value to be displayed by setting a graphics cursor on the locus at the jω crossover. Using the Routh-Hurwitz criteria, the system gain margin can be calculated using equation (6).

Control Loop Compensation

Reconsider the open loop transfer function of equation (2) having a multiplier, K', representing a variable gain multiplier from one to infinity. This enables the CAD program to construct the root locus with the fixed loop gain of the user defined transfer function such as equation (5). The root locus path can then be traced over the synthesizer's modulus range. In this case the open loop transfer function of equation (5) was defined with K_N=9400. Therefore when the synthesizer modulus is changed to K_N=2400 the loop gain is increased by a factor of 3.92 . On the root locus the synthesizer bandwidth can be tracked from the point where K'=1 to K'=3.92. The designer can then optimize the parameters of the loop transfer function for the desired operating characteristic over the entire synthesizer bandwidth. This can be very helpful in difficult designs requiring tradeoffs between the location of the loop bandwidth, loop damping, and placement of reference suppression poles. The root locus allows the designer to view the relative damping of the loop over the entire operating range as well as obtain an approximation of the closed loop bandwidth via the bandwidth of the dominant roots. At the same time, the loop gain margin and phase margin are readily available at any synthesizer channel. In a high order control system containing dominant complex conjugate roots, as is typical in frequency synthesizers, the parameters associated with these roots very closely predict the performance of the loop.

Consider the root locus of the initial loop design shown in figure 6. Over the range from K'=1 to K'=3.92, the loop bandwidth is close to the desired range of 500 Hz to 3 KHz but the system has poor damping over most of the operating bandwidth. These low values of loop damping contribute to closed loop peaking, poor gain and phase margin, and longer settling times. The first loop change is the movement of the loop filter zero closer to the jω axis. This causes the dominant root locus to move towards encircling the zero on the negative real axis. This is because, for a small range of K', the system is beginning to approximate a pure second order system for which the dominant complex roots form a circle around the zero. The poles, $1/T_3$ and $1/T_4$, can then be moved closer to the jω axis for increased reference sideband suppression. The design then becomes an iterative process of optimizing the pole-zero locations for the desired damping and loop bandwidth. At this step in the design process the transfer function loop gain can be modified to move the modulus range of K' to curve fit the root locus at the desired range. Typically the band extremes, K'=1 and K'=3.92, are set at the minimum values of damping that the system can tolerate. In this synthesizer both the loop filter gain, K1, as well as the phase detector gain, K_P, can be varied. Figure 8 shows the optimized root locus for the synthesizer's control loop. Note that the locus has moved into an area of improved damping while the range of loop bandwidth has remained relatively unchanged. The locus of the dominant roots now move from the origin to encircle the zero and then move along the negative real axis before breaking away again as a conjugate pair. As figure 8 shows, the band extremes have been limited to a minimum damping value of 0.6 for reasonable settling. At K'=1,which corresponds to K_N=9400, the phase margin equals 50°, the gain margin is 22 dB, and the loop bandwidth is 430 Hz. At K'=3.92, which corresponds to K_N=2400, the phase margin equals 42°, the gain margin is 10.5 dB, and the loop bandwidth is 2.6 KHz.

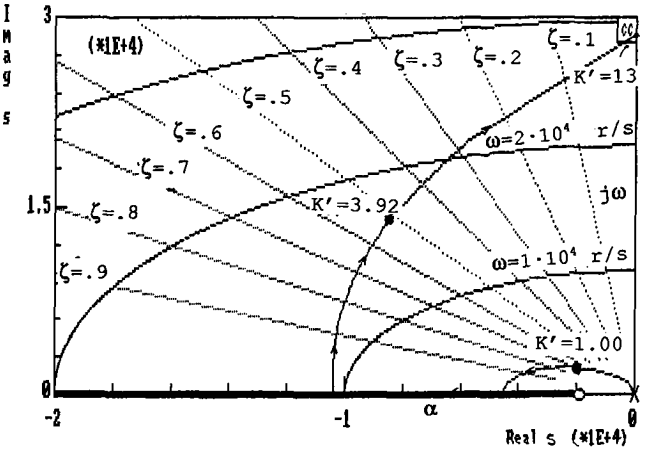


Figure 8. Optimized Root Locus

Conclusions

The examination of the closed loop roots of a control system's characteristic equation has been used in the design of a frequency synthesizer. The root locus is a graphical tool that enables the analysis and design of a control system based on its characteristic equation. In the design of wideband frequency synthesizers it is helpful to have a means of displaying a variety of stability parameters simultaneously over the entire operating bandwidth. The design information available from the root locus can be summarized as:

1) Limiting value of loop gain at jω crossover
2) Gain Margin
3) Phase Margin
4) Control Loop Damping Ratio,ζ
5) Closed Loop Bandwidth

Figure 9 shows the closed loop Bode plots for the control loop at each band extreme. The closed loop bandwidth when K_N=9400 was measured to be 550 Hz. At K_N=2400 the loop bandwidth measured 2.5 KHz. Each case shows very good correlation with the closed loop bandwidth for the dominant poles of the root locus. In each case, 60 to 70 dB suppression of the reference sidebands has been achieved from the control loop with very little loop peaking. An analysis of the open loop Bode plots showed that the phase margin was 51° with K_N=9400 and 44° with K_N=2400. Again realizing very good correlation with the root locus dominant poles. Using the root locus, all of this information can be obtained from one plot and for the entire bandwidth of the synthesizer. The root locus should not always be the preferred analysis method particularly with fixed frequency loops and simple order loops. However high order control loops and wideband synthesizers can benefit from the root locus analysis. A good working knowledge of the root locus as well as other methods can enable the frequency control engineer to develop a gestalt of control loop design techniques.

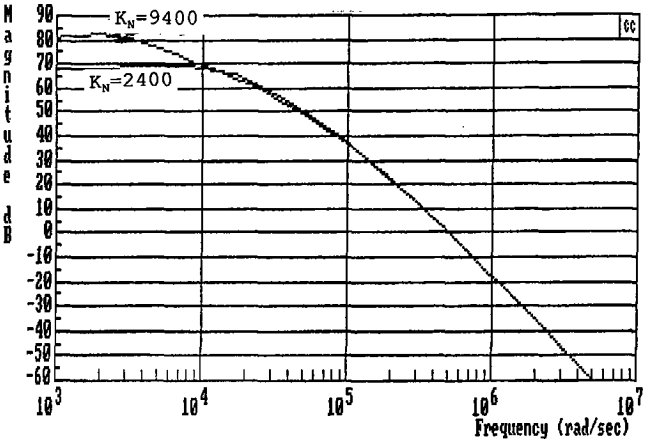


Figure 9. Closed Loop Bode Plot for K_N=9400 and K_N=2400

References

[1] Stephen D. Turner,"An Analysis of Microwave Varactor Tuned Transistor Oscillators Based on Circuit Q Factor",Master of Engineering Thesis, The Pennsylvania State University,1990.

[2] Ulrich L. Rhode, Digital PLL Frequency Synthesizers. Englewood Cliffs,NJ:Prentise-Hall,Inc., 1983.

[3] Vadim Manassewitsch, Frequency Synthesizers, New York: John Wiley & Sons,1987.

[4] Andrzej B. Przedpelski, "Phase Locked Loops", R.F.Design , pp. 24-30,Sept.1979.

[5] Richard C.Dorf,Modern Control Systems,Reading, MA.:Addison-Wesley Publishing Company, 1980.

[6] Joseph J. DiStefano, Allen R. Stubberud, Ivan J. Williams,Feedback and Control Systems,New York: McGraw-Hill Book Company,1967.

[7] Motorola Application Note AN-535,"Phase-Locked Loop Design Fundamentals", Motorola Semiconductor Products,Inc.,Phoenix,AZ.,1973.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

SENSITIVITY OF FAST SETTLING PLLs TO DIFFERENTIAL LOOP FILTER COMPONENT VARIATIONS

Thomas J. Endres and James B. Kirkpatrick

HUGHES AIRCRAFT COMPANY
Space and Communications Group
Bldg R10, M/S S325
P.O. Box 92919
Los Angeles, CA 90009

ABSTRACT

An op-amp configured as a combined differential amplifier and integrator may exhibit a transient response which can be catastrophic to a phase-locked loop (PLL) intended for fast phase settling. A practical mathematical model for the loop filter is derived and corresponding effects are shown in a frequency synthesizer via computer simulations and empirical data.

FAST SETTLING PLL INTRODUCTION

A phase-locked loop (PLL) architecture can be employed for fast hopping frequency synthesis (see Figure 1). The assumption of linearity allows the loop to be treated as a classical control system and modeled via Laplace techniques. The following notation and dimensions will be used throughout this text:

MODULE	GAIN	UNITS
Phase Detector	K_d	volts/rad
VCO	K_v	rad/volt/sec
Divide-by-N	1/N	--
Loop Filter	F(s)	--

When the loop is locked, the phase detector compares the divide-by-N output frequency with the reference frequency such that the two are equal. The phase detector's output is a voltage or current based on the phase difference between these two signals. This phase-sensitive signal is then integrated by the loop filter, which in turn slews the VCO to the proper frequency. In this manner, the loop remains locked to the reference frequency.

Typically, a second order, type 2 (two integrators in the loop path) loop is chosen to permit a small static phase error and non-zero tuning voltage for the VCO [4], [7], [16]. The loop filter for this implementation can be ideally modeled as

$$F(s) = \frac{s\tau_2+1}{s\tau_1}$$

Second order system parameters can be defined

$$\omega_n = \sqrt{\frac{K_d K_v}{N\tau_1}}$$

$$\zeta = \frac{\omega_n \tau_2}{2}$$

and the closed loop transfer function is then

$$\frac{\emptyset_{out}(s)}{\emptyset_{in}(s)} = \frac{N(2\zeta\omega_n s+\omega_n^2)}{s^2+2\zeta\omega_n s+\omega_n^2}.$$

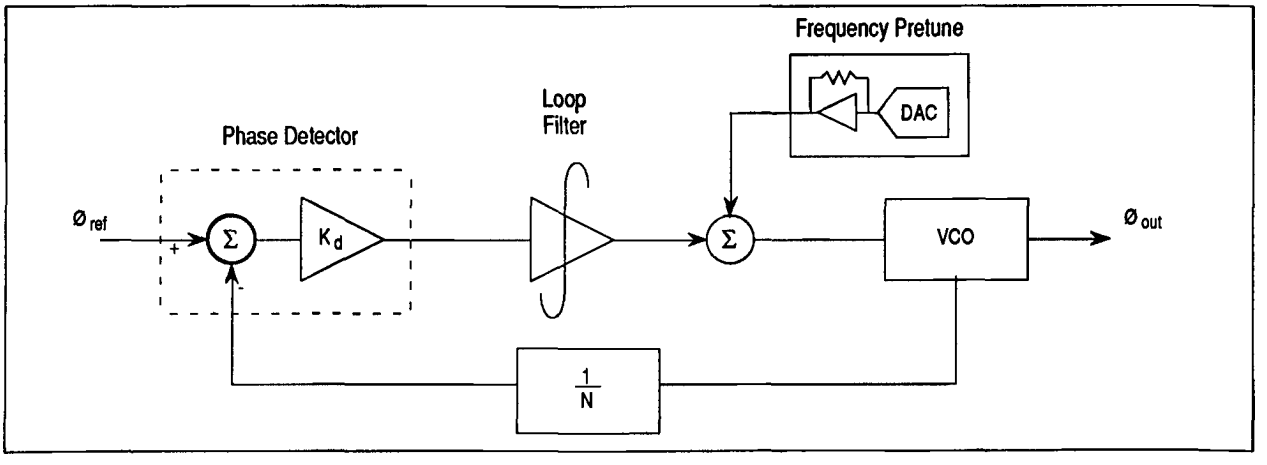


Figure 1. A PLL as a Frequency Synthesizer

0-7803-0476-4/92 $3.00 © 1992 IEEE

The loop dynamics can be chosen to reflect the system specifications. Consider the input as a step in frequency; a critically damped (ζ=1) system settles fastest to zero phase error. However, practical systems are usually required to settle to within a finite error window in a given time, e.g., ±6 degrees in 50 µsec. Therefore, a critically damped system is often not the best choice; an underdamped system will "settle" faster because of the nature of the specification. The ideal time response due to a reference frequency step for an underdamped system is given as

$$\phi_{error}(t) = \frac{\Delta\omega}{\omega_n\sqrt{1-\zeta^2}} \sin\left(\sqrt{1-\zeta^2}\omega_n t\right) e^{-\zeta\omega_n t}$$

and plotted in Figure 2 for ζ=0.7 and ζ=0.1 [7].

The phase error present at the VCO output is then N times that in Figure 2.

PRETUNE CIRCUIT

Typically, the loop requires some type of steering or help circuitry for fast phase settling. Figure 1 includes a pretune circuit that receives digital commands to output a stepped dc voltage. This voltage is superimposed with the loop filter output voltage to tune the VCO. When the loop is hopped, the pretune is synchronously commanded with the divide-by-N and coarsely tunes the VCO to a frequency close to $N*f_{REF}$. The remaining voltage needed to tune the VCO properly is referred to as loop stress error and must be made up by the loop filter. Hence, the pretune effectively decreases the magnitude of the frequency step by decreasing the loop stress error.

DUAL FUNCTION LOOP FILTER

A typical loop filter implementation that accommodates differential phase detector outputs is shown in Figure 3.

The time response given in Figure 2, based on the ideal loop filter model, fails to be consistently satisfied. The ideal model assumes a perfectly balanced op-amp, i.e., all like components are exactly equal. This model does not account for a transient term associated with the non-inverting leg of the op-amp due to mismatches in like components. A more rigorous loop filter model is needed.

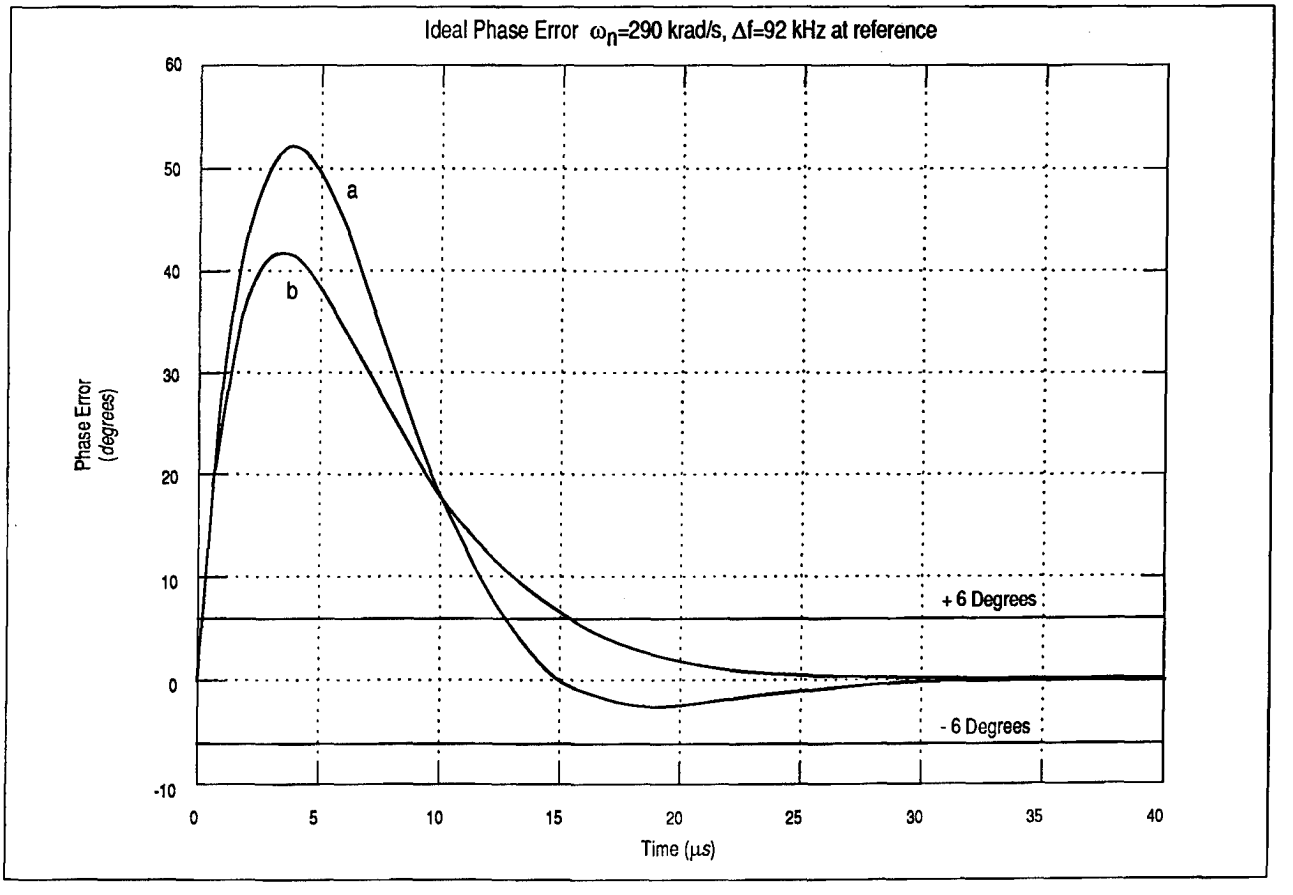


Figure 2. Phase Error Versus Time, a) ζ=0.7 and b) ζ=1 for Reference Frequency Step

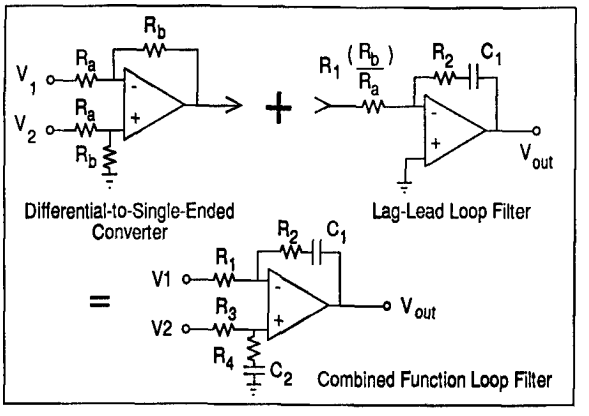


Figure 3. Dual Function Loop Filter

Let $Z_2 = R_2 + \frac{1}{sC_1}$

and $Z_4 = R_4 + \frac{1}{sC_2}$

By definition,

$$V_{out} = A_v(V^+ - V^-)$$

Then

$$V_{out} = A_v\left\{V_2\left(\frac{Z_4}{Z_4+R_3}\right) - V_1\left(\frac{Z_2}{Z_2+R_1}\right) - V_{out}\left(\frac{R_1}{Z_2+R_1}\right)\right\}$$

And after some algebra,

$$V_{out} = A_v\left\{V_2\left(\frac{Z_4}{Z_4+R_3}\right)\left(\frac{Z_2+R_1}{Z_2+R_1(1+A_v)}\right) - V_1\left(\frac{Z_2}{Z_2+R_1(1+A_v)}\right)\right\}$$

Now assume that $A_V >> 1$ such that

$$1 + A_V \approx A_V \text{ and}$$
$$Z_2 + R_1(1+A_V) \approx R_1A_V.$$

Then

$$V_{out} \approx V_2\left(\frac{Z_4}{R_1}\right)\left(\frac{R_1+Z_2}{R_3+Z_4}\right) - V_1\left(\frac{Z_2}{R_1}\right).$$

Now substituting for Z_2 and Z_4 and with the following change of variables,

$$a = C_1(R_1 + R_2)$$
$$b = C_2(R_3 + R_4)$$
$$c = R_4C_2$$
$$d = R_1C_1$$
$$e = R_2C_1$$

yields a Laplace solution

$$V_{out}(s) = V_2(s)\left\{\left(\frac{sc+1}{sd}\right)\left(\frac{sa+1}{sb+1}\right)\right\} - V_1(s)\left\{\frac{se+1}{sd}\right\}$$

For the ideal model, the output is zero for equal inputs V_1 and V_2. For the practical model presented above with equal inputs V_1 and V_2, the output can be non-zero. This fact is more easily seen in the time domain. Let the inputs equal a unit step, i.e., $V_1=V_2=1/s$. Then

$$V_{out}(t) = \left\{1 - \frac{R_3C_2}{R_1C_1}\right\} + \frac{R_3}{R_1}\left\{\frac{C_2}{C_1} - \frac{R_1+R_2}{R_3+R_4}\right\} e^{-\frac{t}{C_2(R_3+R_4)}}$$

Inspection shows the output is zero for equal inputs if and only if the loop filter components are exactly matched. That is, for $R_1=R_3$, $R_2=R_4$, and $C_1=C_2$, the output is zero. Else, an exponential term is introduced at the loop filter's output (see Figure 4).

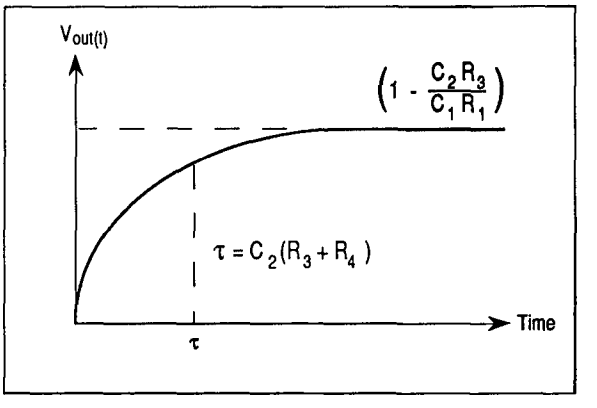


Figure 4. Loop Filter Exponential for Two Equal Step Inputs

The direction and magnitude of this exponential are dependent on the offset in the loop filter components. The time constant is dependent on the loop filter selects, i.e., loop dynamics. Note that this exponential term is associated with the non-inverting leg of the op-amp. For any non-zero V_2, some undesired transient will be introduced at the loop filter's output. If $V_2=0$, the practical model reduces to the ideal model.

This transient may appear deceivingly negligible. However, for sufficiently high loop gain, the transient is greatly amplified for component mismatches well within the device tolerances, such that the effect on phase error must be considered.

Note that for a frequency step within the loop the input to the loop filter is a saturating ramp. However, if sufficiently fast, this ramp input can be effectively modeled as a step function. Indeed, the slope of this ramp can be calculated for a frequency step in a high gain loop and be shown to approximate a step. Suppose the VCO is stepped as from a pretune voltage. The VCO phase output

in time is $V_{step}K_vt$. The divide-by-N divides the VCO output by N and the phase detector multiplies this function by K_d. The loop filter sees the input $(V_{step}K_dK_v/N)t$. Thus, for sufficiently high loop gain the slope of this ramp is indeed fast. Figure 5 shows the output of the loop filter from a PSpice loop filter model with a given mismatch for two different inputs, a unit step and a saturating ramp with slope of approximately 4.5 V/μsec (V_{step} = 5V, K_d = 0.636 V/rad, K_v = 150 MHz/V, and N = 654).

The loop filter outputs are indistinguishable so that a step input to the loop filter is a good approximation for a frequency step within a high gain loop. Note that a step input to the loop filter is also exact for a phase step within the loop.

Practically, it may be unreasonable to drive an op-amp with these functions because of their high slew rate edges. Thus, it is typical to have a lowpass filter between the phase detector and loop filter. For a reasonably high reference rate, the reference period T_{REF} is usually much faster that the loop filter time constants. Since the lowpass filter time constant may be on the same order as the reference, the loop filter output is affected little by the lowpass network. Hence, a step function is a reasonable mathematical approximation to the loop filter input.

PERCENT MISMATCH

Inspection of the time-domain loop filter transient, derived from the practical model, shows that the worst case exponential occurs when all three components on one leg are varied in one direction relative to the components on the other leg. Define a positive X percent mismatch such that R_1>R_3 by X percent of their nominal value, R_2>R_4 by X percent of their nominal value, and C_1>C_2 by X percent of their nominal value. Conversely, a negative percent mismatch implies the components on the non-inverting leg are greater than the inverting leg components. Note that it is only the relative difference of

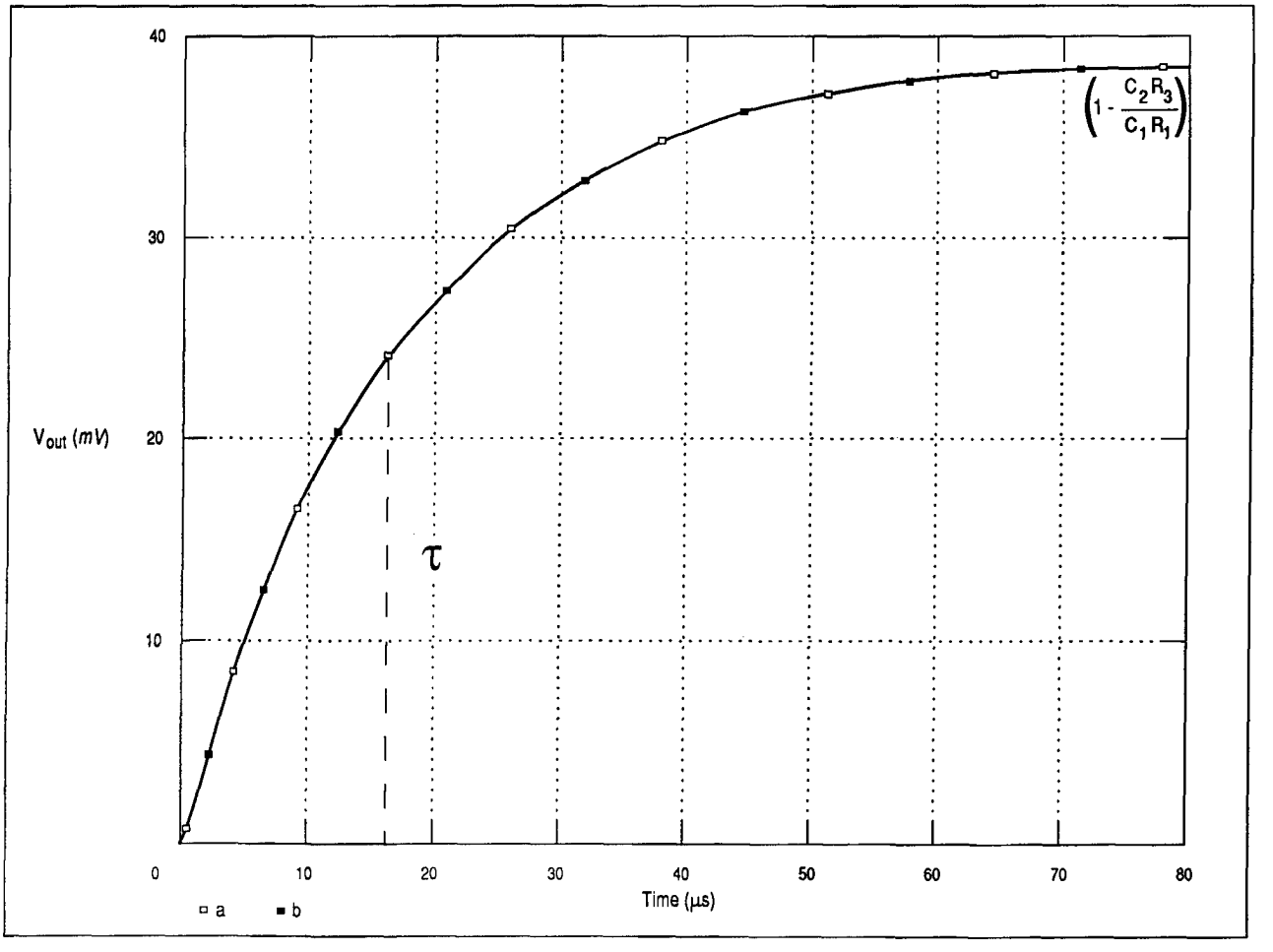


Figure 5. PSpice Loop Filter Output for a) Unit Step Input and b) Saturating Ramp Input (R_1=11,118Ω, R_3=10,900Ω, R_2=6324Ω, R_4=6200Ω, C_1=1020pF, C_2=1000pF)

the components that drives the transient term. Thus, if the nominal value capacitance is 1000 pF, a set of components C_1=990 pF and C_2=1010 pF has essentially the same effect as the set of components C_1=1000 pF and C_2=1020 pF. Given either set of these capacitors and resistors offset in the same direction by 2 percent of their nominal value, the effect is a -2 percent mismatch.

THEORETICAL PREDICTIONS

By assuming a linear system and employing superposition principals, a convenient way to theoretically model the loop filter transient is as a disturbance introduced at the loop filter's output (see Figure 6).

Suppose the divide-by-N ratio is changed such that a new output frequency is desired. Egan has shown this frequency step to be analogous to stepping the reference frequency [4]. The output phase error is then approximately equal to the superposition of the ideal second order response and a term introduced from the loop filter mismatch. By approximating the inputs to the loop filter as stepped voltages, it is found that

$$\varnothing_{out}(s) \approx \frac{\Delta\omega N(2\zeta\omega_n s+\omega_n^2)}{s^2(s^2+2\zeta\omega_n s+\omega_n^2)} + \frac{AK_v}{d}\left(\frac{s(ac-be)+(a-b+c-e)}{(sb+1)(s^2+2\zeta\omega_n s+\omega_n^2)}\right)$$

where A = equivalent amplitude of step into op-amp
= $V_{PD,MAX}$ *[Duty cycle of PD]
$\Delta\omega$ = reference frequency step.

Assume a typical loop with ω_n=290 krad/sec, ζ=0.9, and +2 percent mismatch in the loop filter components. Now transforming to the time domain reveals a dominant exponentially decaying transient of time constant τ = b:

$$\varnothing_{out}(t) \approx \frac{AK_v}{d}\left(\frac{(a-b)(b-c)}{(1-2\zeta\omega_n b+\omega_n^2 b^2)}\right)e^{-\frac{t}{b}}$$

Using loop filter components equal to those in Figure 5, and the K_d, K_v, and N values mentioned earlier, this time constant τ is much longer than the ideal PLL transient response time constant:

$$\tau \equiv b = C_2(R_3+R_4) \approx 17\ \mu\text{sec} \approx 5\left(\frac{1}{\zeta\omega_n}\right)$$

Thus, the transient response due to loop filter mismatch will swamp the ideal response term and cause settling times much longer than that predicted by the ideal equation. Figure 7 shows the $\varnothing_{error}$ at the VCO output for (a) the ideal second-order system with a reference frequency step, (b) the dominant exponential transient above, and (c) the LaPlace system using the practical instead of ideal loop filter model with a reference frequency step.

The dominant exponentially decaying transient closely approximates the response of the system employing the practical loop filter model for phase error approaching final value. Hence, the dominant decaying exponential should be a reasonable predictor of settling time.

Figure 8 shows the same data as Figure 7 except for a -2 percent mismatch in the loop filter components.

The superposition of the ideal second-order response and loop filter transient will alter the shape of the resultant phase settling response. The mathematical model predicts that the direction of the loop filter exponential depends on the loop filter mismatch. More precisely, a negative percent mismatch may add destructively with the ideal second-order-system phase error, while a positive percent mismatch may add constructively.

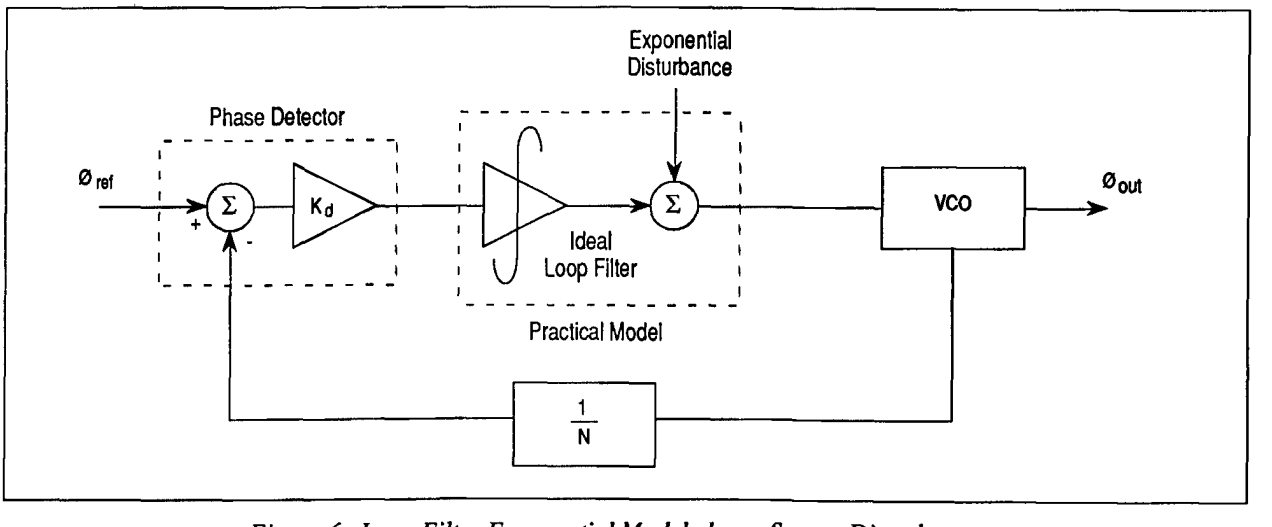


Figure 6. Loop Filter Exponential Modeled as a System Disturbance

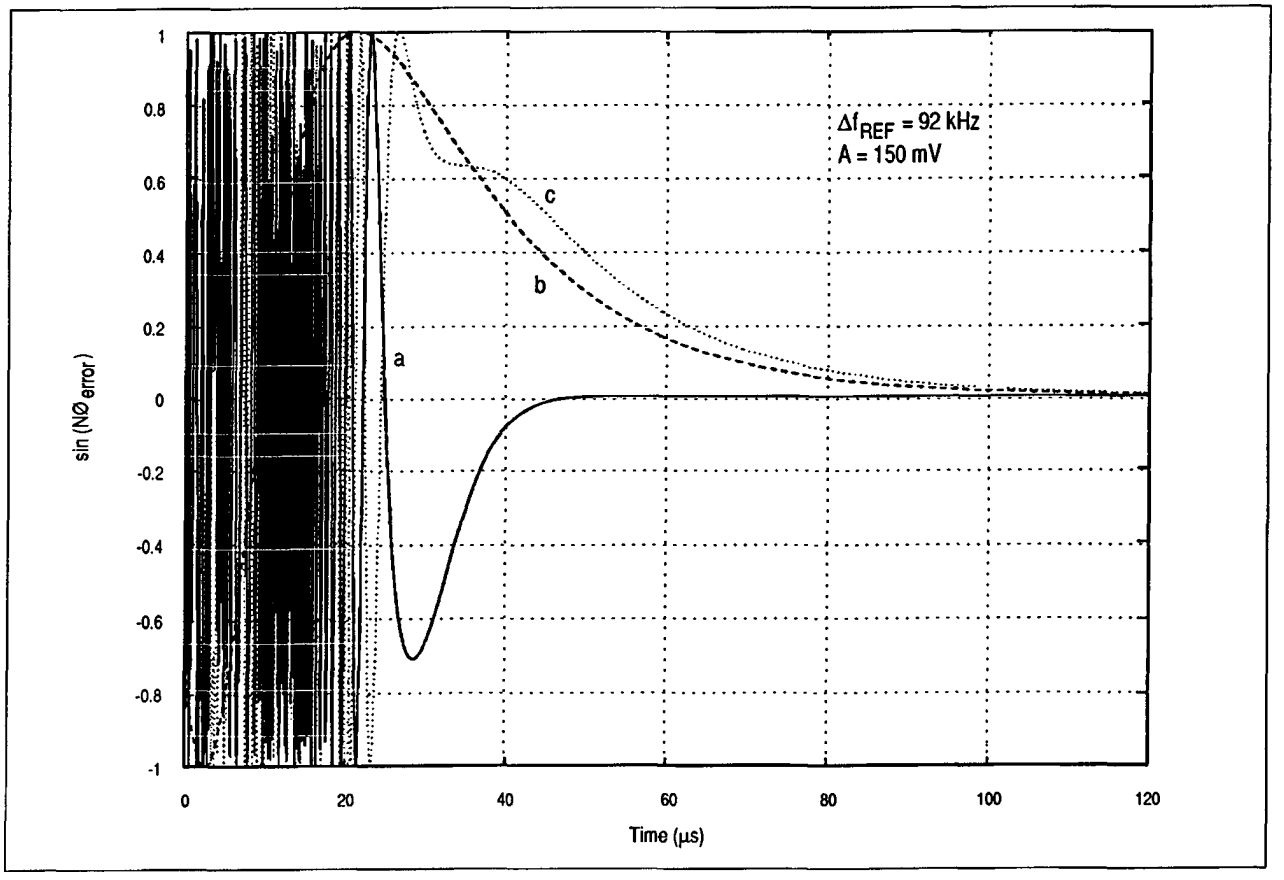


Figure 7. Output Phase Error for a) Reference Frequency Step for Ideal Second-Order System, b) Loop Filter Transient From +2 Percent Mismatch, and c) Reference Frequency Step for LaPlace System Employing Practical Loop Filter Model with +2 Percent Mismatch

PSPICE MODEL

A loop model is developed in PSpice for comparison (see Figure 9). This simple model accurately models the phenomenon, while allowing a more elegant model to be easily developed.

The reference spur filter is a notch filter used to attenuate reference spurs and harmonics. The filter may be modeled explicitly in PSpice (using R's, L's and C's) but can be thought of as roughly a constant gain phase shift for reference frequencies sufficiently outside the loop bandwidth.

The differential phase detector model is accomplished via ideal voltage-controlled voltage sources (VCVS) with positive and negative gains. This implementation is reasonable since the loop does not slip cycles.

The VCO gain, K_v, is determined by choice of the feedback capacitor. The loop filter components are easily changed to accommodate any percent mismatch.

The input is a voltage step to the VCO as from a pretune, causing a frequency step out of the VCO. This situation is analogous (within a factor of N) to a frequency step at the reference. $\emptyset_{out}$ of the VCO can then be viewed. For similar loop dynamics relative to Figures 7 and 8, consider $\emptyset_{out}$ for +2 percent and -2 percent mismatch in the loop filter components (Figures 10 and 11, respectively). The PSpice results evidence the loop filter transient and agree well with theoretical predictions.

EMPIRICAL DATA

Final comparison is done with empirical data. The actual loop contains a pretune circuit and reference spur filter. The loop dynamics are similar to the above data in Figures 7 through 11. Phase settling for +2 percent and -2 percent loop filter mismatch is shown in Figures 12 and 13, respectively.

The loop filter exponential resulting from a negative percent mismatch, when superimposed with the ideal second-order-system response for simulations and

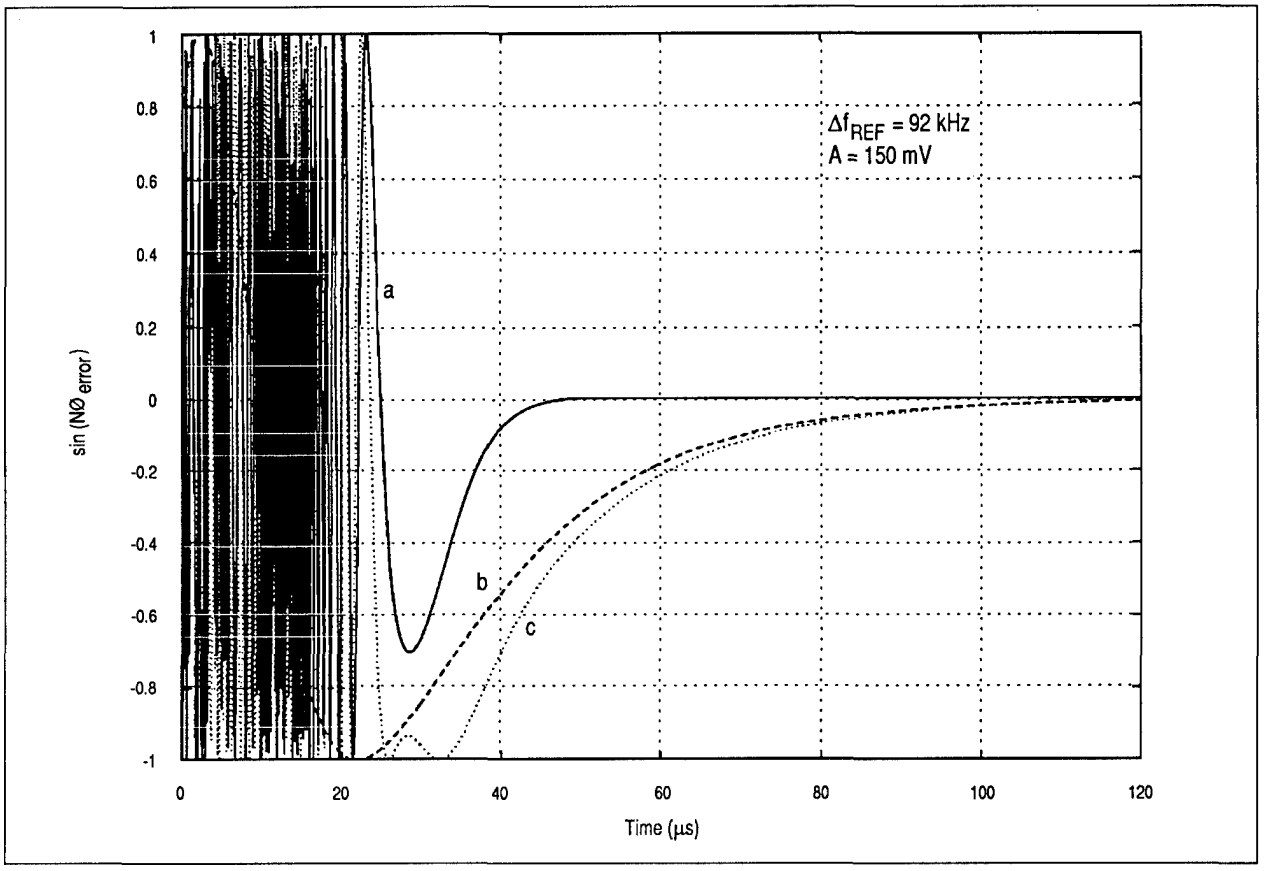


Figure 8. Output Phase Error for a) Reference Frequency Step for Ideal Second-Order System, b) Loop Filter Transient From -2 Percent Mismatch, and c) Reference Frequency Step for LaPlace System Employing Practical Loop Filter Model with -2 Percent Mismatch

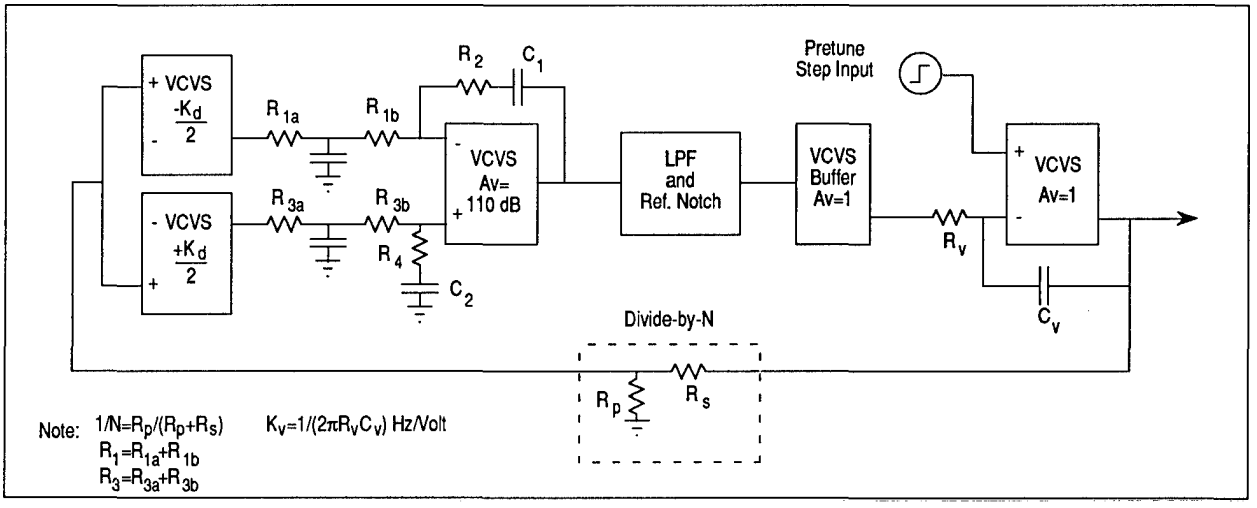


Figure 9. PSpice PLL Model for Time and Frequency Response Prediction

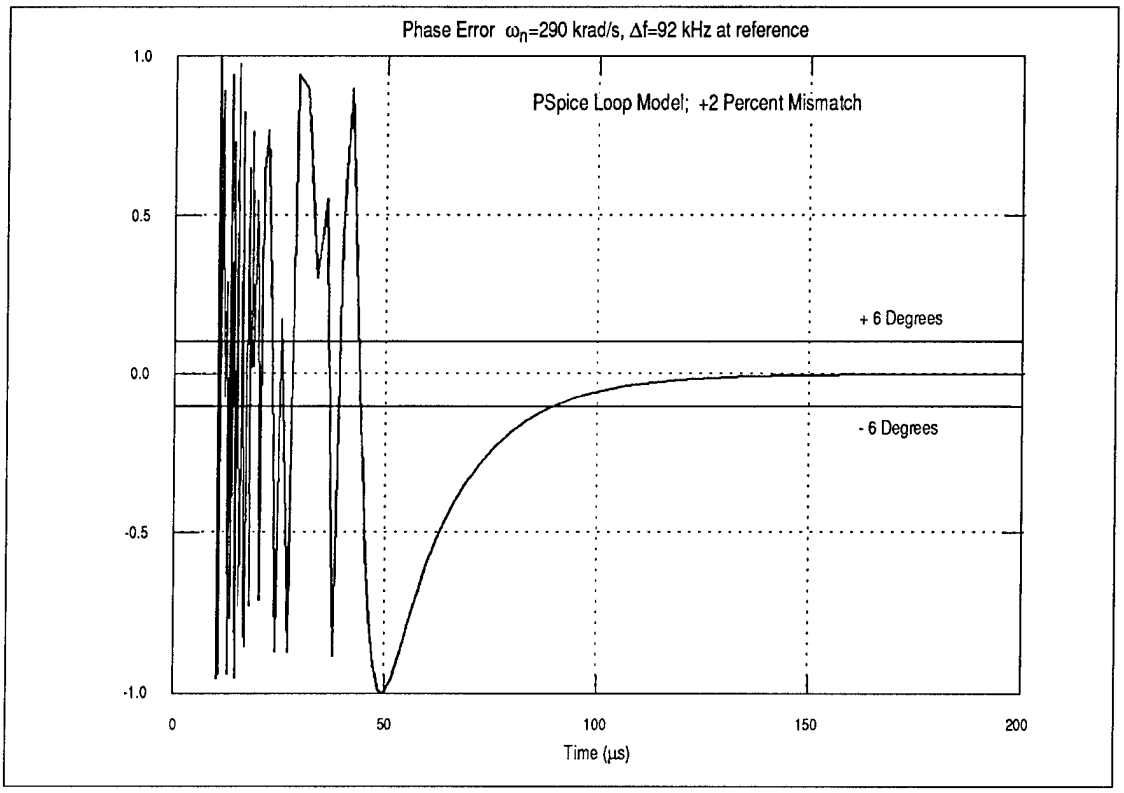


Figure 10. PSpice Phase Settling for ζ=0.9, and ω_n=290krad/s, with +2 Percent Loop Filter Mismatch

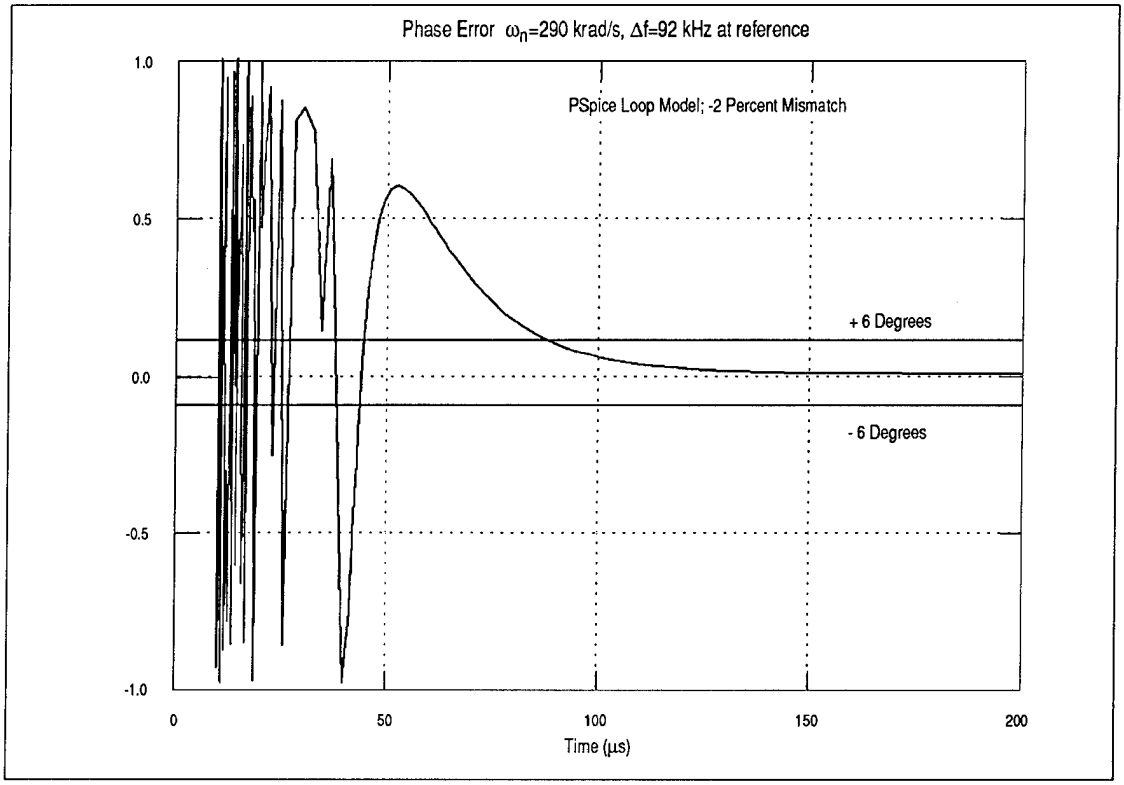


Figure 11. PSpice Phase Settling for ζ=0.9, and ω_n=290krad/s, with -2 Percent Loop Filter Mismatch

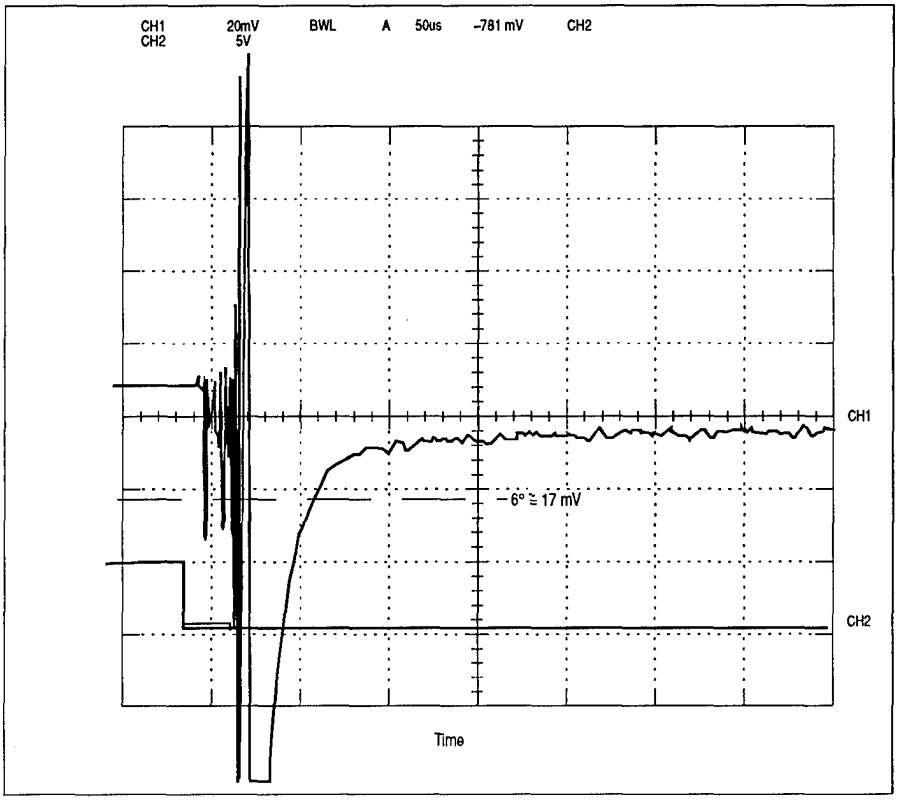


Figure 12. Empirical Phase Settling for +2 Percent Loop Filter Mismatch

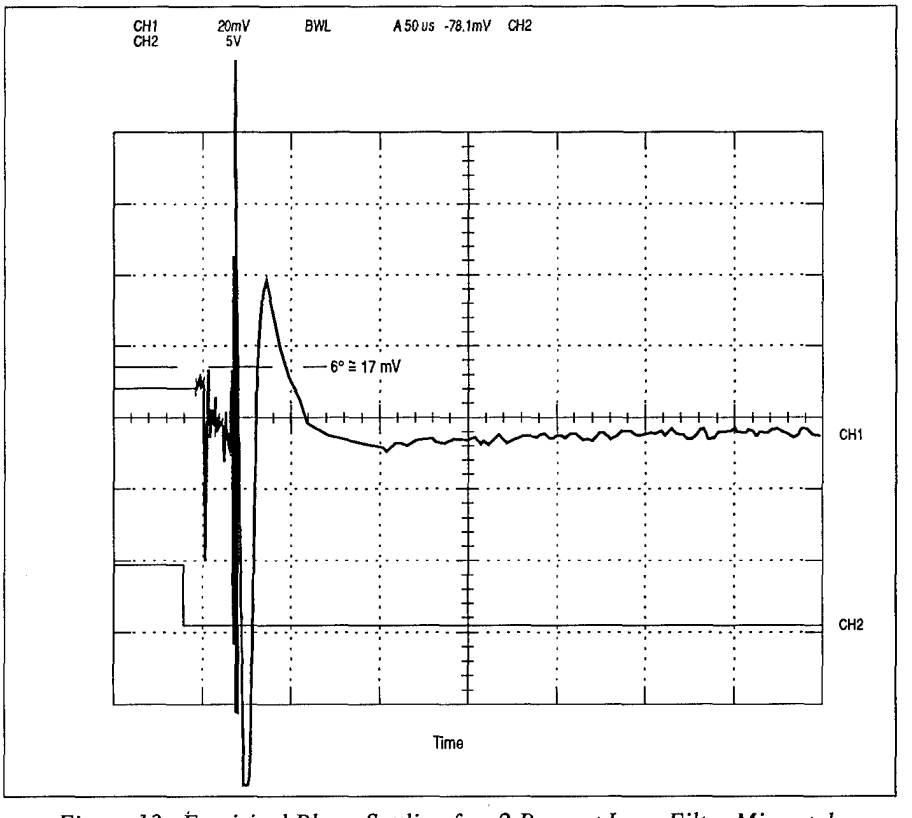


Figure 13. Empirical Phase Settling for -2 Percent Loop Filter Mismatch

empirical data, causes the phase error to overshoot the zero mark. Suppose the specification is ±6 degrees, i.e., the loop is required to settle within ±6 degrees of final phase in a given time. If the overshoot is just small enough that it misses the 6 degree mark, settling time will be optimized. Hence, because of the nature of the specification, a slight negative percent mismatch should minimize the settling time.

Consider the settling time to ±6 degrees, as the loop filter components are varied, for the three loops: PSpice loop model, actual loop, and theoretical disturbance model. These results are plotted below in Figure 14.

There exists a minimum for the PSpice and actual loops which is offset from 0 percent due to the nature of the settling time specification, i.e., ±6 degrees. The theoretical disturbance model has non-zero output only when the loop filter components are mismatched, hence the null at 0 percent.

Most practical systems do not require knowledge concerning the exact values of loop filter components. For 1 percent tolerance parts, a ±2 percent mismatch is feasible. These small variations, well within device tolerances, can degrade phase settling time. Since omponents may not age at the same rate and the loop stress error may age in an unknown direction, it is unlikely that the operating point of Figure 14 is known at end-of-life. Hence, a system that behaves well at beginning-of-life is not guaranteed satisfactory end-of-life performance.

SOLUTIONS

Three possible solutions that minimize or eliminate this phenomenon are presented.

COMPONENT MATCHING

Since the undesired transient term is driven solely by the relative difference in loop filter components, the option exists to handpick these parts. They can be chosen at beginning-of-life such that no undesired loop filter transient is introduced (that is, exactly match the components to 0 percent). Figure 14 curves suggest an alternative for an underdamped system: optimally mismatch the components toward a negative percent mismatch, thus minimizing settling time. Careful selection of parts is necessary to ensure aging in the same direction.

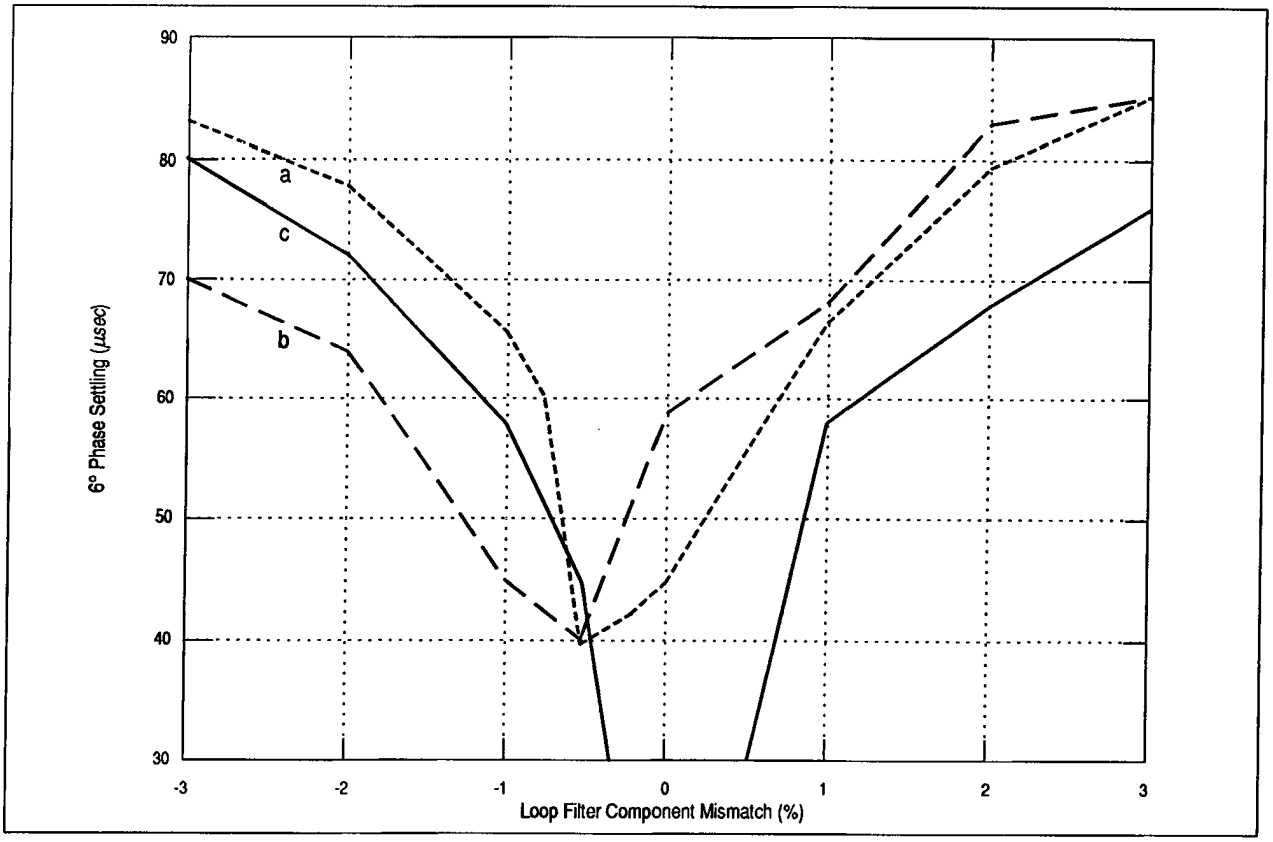


Figure 14. Phase Settling Versus Percent Mismatch for a) PSpice Loop Model , b) Actual Loop and c) Theoretical Loop Filter Disturbance Model

SINGLE-ENDED SCHEME

Another option is to employ a purely single-ended phase detector and loop filter scheme. Since the undesired transient results from a non-zero input to the non-inverting leg of the op-amp, a single-ended scheme eliminates the undesired term by forcing the non-inverting input to zero. However, this idea may force a difficult trade-off between K_d and phase detector range if a digital flip-flop type phase/frequency detector is used. That is, a single-ended output may provide $\pm\pi$ or $0\text{-}2\pi$ linear range with full output swing (maximum K_d), or $\pm 2\pi$ range with half-swing ($K_d/2$). Thus, noise contribution increases due to lower K_d must be weighed against loop characteristics such as pull-in range and acquisition time. Also, if a 50 percent nominal duty cycle output pulse is chosen for the single-ended output with full K_d, reference spurious will be increased considerably.

TWO-STAGE DUAL-FUNCTION LOOP FILTER

To accommodate differential phase detector outputs and maintain the associated advantages, it is possible to separate the two functions of the loop filter via two op-amps, the first as a difference amplifier and the second as an integrator (see Figure 15).

Note that for equal inputs (V_1=V_2), the difference amplifier's transfer function is identical to the combined loop filter's transfer function with $C_1 = C_2 = \infty$; that is, the exponential term reduces to a simple gain offset so that it is not critical to match the difference amplifier's components.

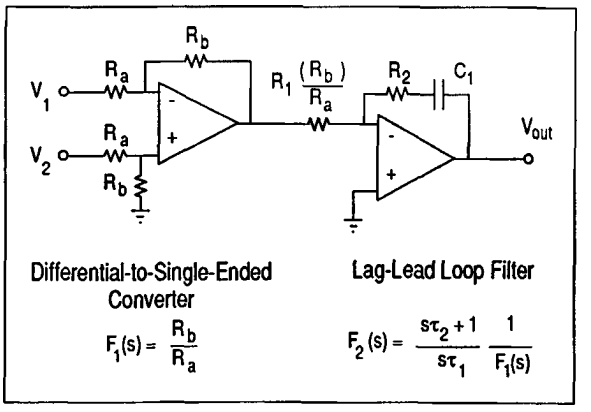


Figure 15. Two-Stage Dual-Function Loop Filter

The additional op-amp imposes a possible noise impact on the loop. However, with careful component selection, the two-stage loop filter can achieve roughly equal or even better noise performance than the combined loop filter.

CONCLUSION

We have presented a typical loop filter configuration and shown its transient effects in a fast hopping PLL. Phase settling is shown to be sensitive to small variations in the loop filter components, well within the device tolerances. We have derived a practical mathematical model for the loop filter that correlates very well with PSpice simulations and empirical data. Finally, we have offered three alternative solutions to the phenomenon, one of which requires no design change.

ACKNOWLEDGMENTS

This work was performed under subcontract with TRW Incorporated, Electronic Systems Group, Space Communications Division. The authors wish to acknowledge the support of the following individuals: John Chernekoff, Mike Lenzo, Paul Sako, and everyone else who helped to further this investigation.

BIBLIOGRAPHY

[1] J. Blake, "Design of Wideband Frequency Synthesizers," RF Design, May, 1988

[2] J. A. Crawford, "Extending sampling to Type II phase-locked loops," Microwaves and RF, September, 1984

[3] J. A. Crawford, "Understanding the specifics of sampling in synthesis," Microwaves and RF, August, 1984

[4] W. F. Egan, *Frequency Synthesis by Phase Lock* (John Wiley and Sons, Inc., 1981)

[5] G. F. Franklin, J. D. Powell, and A. Emami-Naeini, *Feedback Control of Dynamic Systems* (Addison-Wesley, 1986)

[6] F. M. Gardner, "Charge-Pump Phase-Lock Loops," IEEE Trans. on Comm., Vol. Com-28, No. 11, November, 1980

[7] F. M. Gardner, *Phaselock Techniques*, 2nd edition (John Wiley and Sons, Inc., 1979)

[8] R. W. Hall, "Synthesizer analysis clarifies PLL circuits," Microwaves and RF, May, 1985

[9] R. G. Irvine, *Operational Amplifier Characteristics and Design* (Prentice Hall, Englewood Cliffs, NJ, 1981)

[10] V. Manassewitsch, *Frequency Synthesizers* (John Wiley and Sons, Inc., 1987)

[11] A. V. Oppenheim, A. S. Willsky, and I. T. Young, *Signals and Systems* (Prentice Hall, Englewood Cliffs, NJ, 1983)

[12] C. W. Price II, "Modeling Discrete-Time Phase Detection in the Phase-Locked Loop," RF Design, May 1990

[13] A. B. Przedpelski, "PLL Primer-Part IV, Description of Loop Integrator Circuit," RF Design, November, 1987

[14] B. Sklar, *Digital Communications* (Prentice Hall, Englewood Cliffs, N.J., 1988)

[15] J. Smith, *Modern Communication Circuits* (McGraw-Hill, Inc., 1986)

[16] D. H. Wolaver, *Phase-Locked Loop Circuit Design* (Prentice Hall, Englewood Cliffs, NJ, 1991)

IMPROVING NOISE IN MEASUREMENT SYSTEMS*

R. L. Hamell, P. F. Kuhnle, R. L. Sydnor,
G.J. Dick, A. Kirk, and W.A. Diener
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, California 91109

Abstract

The Frequency Standards Laboratory (FSL) at the Jet Propulsion Laboratory has the task of evaluating a wide variety of products used by the frequency and timing community. Supporting these activities requires ever increasing improvements in our measurement equipment and techniques as better frequency standards evolve. Several specialized items of equipment have been developed recently at JPL for measuring Allan Deviation and power spectral density of phase and amplitude on a variety of equipment types, covering frequency ranges to 10 GHz. These include improved zero crossing detector and offset generator electronics, low noise mixers and frequency multipliers used for frequency translation, high stability bandpass filters, temperature control methods, and a fiber optics signal transportation link to the data processing computer that is immune to spurious signal contamination. The design and use of these items, along with measurement data will be presented.

INTRODUCTION

The NASA/JPL Frequency Standard Laboratory was established in 1977 in a 700 square foot facility to test and maintain the Hydrogen Maser Frequency Standards for the Deep Space Network. In 1986, this facility was replaced with a Test Laboratory of 2400 square feet in a new 14,000 square foot building to house all the NASA/JPL Frequency and Timing Research and Development and Implementation offices and laboratory. The Test Laboratory supports Hydrogen Maser work, and testing of any DSN equipment requiring stability and spectrum measurements.

Measurement and stimulus equipment is continuously being upgraded and requirements are increased. In the past 3 years we have designed replacement equipment and new capabilities as described below.

ALLAN DEVIATION MEASUREMENT

These measurements are usually performed between two atomic frequency standards that cannot easily be offset by 1 Hz at the output frequency of 100 MHz. The normal test configuration utilizes a 1 Hz offset generator at 100 MHz cascaded with one of the frequency standards.[1] These two signals are then connected to a double balanced mixer, Zero Crossing detector and subsequently a period counter. The counter is connected to a computer which computes the Allan deviation data. This test configuration provides data with a successive number of approximately one second period measurements. The general configuration is shown in Figure 1.

ZERO CROSSING DETECTOR

The existing Zero Crossing Detector (ZCD) utilized for Allan Deviation Measurement has been in use since 1977. The new design exhibits an order of magnitude improvement in time jitter. The existing units in use for the past 15 years have been adequate for Allan Deviation measurement of active Hydrogen Maser Frequency Standards. Mercury Trapped Ion Frequency Standards and Superconducting Cavity Maser Oscillator development requires improved Allan Deviation measurement capability.

The ZCD's function is to transform a 1 Hz beat frequency between two frequency standards into a train of approximately one second pulses to be characterized by a conventional period counter.

The present Zero Crossing Detectors exhibit time jitter of approximately an order of magnitude better (10^{-7} versus 10^{-6} second) performance.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The new design uses a one stage amplifier and a three stage limiter after the RF mixer. Each successive limiter has a wider bandwidth to process the increasing slew rate. Figure 2 is a block diagram of this ZCD.

The present ZCD has a 50 ohm coaxial output to the counter. This results in 100 Ma current pulses in the cable with each zero crossing. With multiple channels operating concurrently, there have been crosstalk problems, which sometimes contaminate the data. The new devices use a multimode fiberoptic transmitter mounted directly in the ZCD circuit board and chassis. A fiberoptic receiver is mounted in the counter. This eliminates the multiple channel current pulses in the cables. The approximate 30 Ma transmitter current is contained within the chassis. This has eliminated the crosstalk problems when multiple channels are concurrently in use.

Reference 2 contains a detailed description of this new ZCD design including analysis of noise, crosstalk and the design.

ALLAN DEVIATION OFFSET GENERATOR

A 1 Hz offset generator is used to generate a 1 Hz beat frequency between 100 MHz sources in an Allan Deviation test[1]. The offset generator becomes necessary when the sources cannot be offset in frequency to develop a 1 Hz beat, or when a 2-port device is the item in test.

A recent JPL redesign of the offset generator has improved phase noise performance more than five times compared to the previous design.

The output, offset by 1 Hz from 100 MHz, is developed in two stages, using direct frequency synthesis. The first stage generates an output offset by 10 kHz from 100 MHz. The second stage generates the required output, offset by 1 Hz from 100 MHz.

The first stage divides the input signal frequency by 10,000, and then mixes the divider output with the input signal in a single-sideband generator. The lower sideband output at 99.99 MHz is then fed to the second stage. The second stage duplicates this process, except that a 5 MHz crystal oscillator and X20 frequency multiplier are operated in a phase locked loop to generate the output signal at 100 MHz minus 1 Hz. A more detailed description of this design is given in reference 3.

The Allan Deviation test results of this design are as follows:

tau, seconds	**sigma**
1	3.2×10^{-15}
10	5.0×10^{-16}
100	9.0×10^{-17}
1000	2.0×10^{-17}
10,000	4.0×10^{-18}

PHASE NOISE MEASUREMENTS

Phase noise is measured today using low frequency (0.1 Hz to 100 KHz) Fast Fourier Transform Spectrum Analyzers. Low noise crystal oscillators are generally used as the common frequency source to reduce possible noise group delay dephasing of the two channels under test. Usually, these measurements on atomic standards and crystal oscillators are performed at 100 MHz. In this case, a 5 to 100 MHz frequency multiplier has been designed and implemented. In addition, a "multiple amplifier" low noise preamplifier between the phase detector and the analyzer was designed and is being evaluated. Figure 3 is a block diagram of two typical test configurations.

LOW PHASE NOISE SOURCE

The FSL has been using selected 5 MHz Oscilloquartz 8600 and Frequency Electronics FE150L Crystal Voltage Controlled oscillators as sources for state of the art phase noise performance.

The single sideband phase noise at 1 Hz from the carrier is -127dBc and -134 dBc respectively on the best units currently in the FSL. The high frequency floor is approximately -155 dBc for both types.

Concurrently both models exhibit Allan Deviation performance of 1 to 3 X 10^{-13} over the 1 to 100 second sampling range.

As a comment, JPL is currently using selected OSA 8600 selected VCO's in all the Hydrogen Maser Frequency Standards in the NASA/JPL Deep Space Network. These units exhibit single sideband phase noise performance of better than -123 dBc at 1 Hz from the carrier. The typical performance is -125 dBc.[4]

FREQUENCY MULTIPLIERS

A series of frequency multipliers have been developed at JPL to generate reference signals at integer multiples of the low noise 5 MHz crystal oscillator

reference source. Additive phase noise contribution of the multipliers is sufficiently below the source noise that source noise establishes the measurement performance.

One design commonly used for testing has outputs at 10 MHz, 20 MHz, and 100 MHz. Frequency multiplication is obtained from diode type push-push frequency doublers from Mini-Circuits Laboratories, followed by temperature compensated bandpass filters, and high isolation, low noise amplifiers.

The 5 MHz input reference signal is multiplied to 80 MHz, with outputs taken at 10 MHz, and 20 MHz, then 80 MHz and 20 MHz are mixed to obtain a 100 MHz output. The adjacent, unwanted harmonic outputs (f_1 and f_3) from the frequency doublers prior to any filtering are typically 30 dB below the desired output at f_2, easing the requirements on the bandpass filters.

The bandpass filters are narrowband circuit elements by definition, while all other circuit elements in the frequency multiplier are broadband. For this reason, the filters become the most critical elements in maintaining good phase stability over temperature. For this reason, special effort is applied in the design and construction of bandpass filters.

The bandpass filters in use are a 5-pole Butterworth design, with a typical 3dB bandwidth of 10%. The filters are constructed with NPO capacitors and 30 PPM temperature stability inductors, and use selected negative temperature compensating capacitors to further improve temperature stability to better than 2PPM/°C.

Specifications for this multiplier are as follows:

Input Frequency: 5 MHz

Output frequencies: 10, 20, and 100 MHz

Spurious Outputs: 85 dB below carrier

Harmonic Output: 40 dB below carrier

Output Level: 13 dBm

Stability/ 5°C step: 10^{-13}

Phase Noise $L(f)$ at 5MHz

1 Hz	-141 dBc
10 Hz	-150 dBc
1 kHz to 100 kHz	-155 dBc

Allan Deviation

1 second	1.5×10^{-14}
10 second	4.0×10^{-15}
100 second	1.0×10^{-15}
1k second	3.0×10^{-16}
10k second	1.5×10^{-16}

Reference signals above the 1 GHz range are obtained from commercially available frequency multipliers that phase lock a microwave cavity oscillator to a multiple of the 5 MHz low noise reference oscillator.

LOW NOISE PREAMPLIFIER

Phase noise measurement of low noise devices will usually require a low noise preamplifier at the input to the FFT analyzer to override the input noise of the analyzer.

One ultra-low noise preamplifier design in use at JPL has 16 parallel connected IC amplifiers in the input stage[5]. Since the noise generated by each device is uncorrelated, noise contributed by n stages increases by $\sqrt{n}$, while signal output increases by n. With 16 devices used, noise reduction is ideally 12 dB compared to a single stage amplifier (measured with zero source impedance). In an operating system with a 50 ohm source impedance driving the amplifier, noise reduction measures 5 dB. Specifications for this preamplifier are given below.

Equivalent input noise:

1 Hz	1.6 nanovolt
10 to 10^5 Hz	0.8 nanovolt

Preamplifier gain: 60 dB

POWERLINE SPURIOUS

For noise measurements on low noise equipment, the measurement equipment as well as the equipment in test may be responsive to powerline magnetic fields found in the test area, or to conducted power line frequencies present on DC power inputs to sensitive electronics. Responses that can be established as measurement equipment caused can be discounted.

To isolate and differentiate responses of the test instruments from equipment in test, a variable frequency power inverter is used to drive a Helmholtz coil for magnetic field tests, or to supply 115 VAC as input power to various elements in the test setup for conducted response tests. An inverter frequency close to the measured response is chosen so that response sensitivity is equivalent, yet

allows differentiation from the response in question.

Circuit elements particularly sensitive to powerline frequency magnetic fields include inductors of all types, mixers, phase detectors, and RF power dividers which usually contain ferrite core transformers.

MAGNETIC FIELD SENSITIVITY TESTING

To measure magnetic field sensitivity of equipment, several Helmholtz coils are available at JPL to develop either AC or DC magnetic fields with a known, controllable amplitude.

A 4 foot diameter coil is used to develop single axis AC magnetic fields with field intensities up to 2 Gauss. This coil is used to measure magnetic sensitivity of small assemblies with dimensions up to 2 to 4 cubic feet.

A larger, 6 foot diameter three-axis coil is used to test larger assemblies, and is used for either AC or DC fields.

MAGNETIC SHIELDS

Magnetic shield enclosures are used as a diagnostic tool and measurement aid by locating sensitive elements and isolating sensitive electronics from powerline frequency magnetic fields in the testing area. Particularly sensitive electronic assemblies are permanently installed in magnetic shields. The shielding material presently used is fully annealed Co-Netic type AA.

To shield small components such as mixers and power splitters, several layers of .004 inch thick foil material can be wrapped around the part. To shield larger assemblies, magnetic shield enclosures are usually fabricated from .025 or .050 inch thick material. Magnetic field attenuation of 20 to 30 dB is usually achieved at powerline frequencies for the shield enclosure, which is usually adequate to reduce responses below the measurement floor.

ACCELERATION SENSITIVITY

Characterizing the acceleration sensitivity of electronics is one task performed in the FSL. Equipment installed in the DSN antennas may be subjected to acceleration resonances as low as 1Hz due to the very large mass of these antennas.

A long stroke, horizontal shaker with air bearings is utilized at the FSL to determine acceleration sensitivity at these low frequencies. The shaker is effective at frequencies between 0.1 Hz and 200 Hz, and will develop acceleration forces up to 1.5 g. The air bearings result in an acceleration force that accurately reproduces the shaker excitation input, so that equipment responses can be accurately calibrated.

SUMMARY

Equipment developed today is ready for measurement of the next generation of frequency standards currently being developed by JPL for NASA. New environmental stimulus equipment to evaluate and test state of the art equipment has been developed or purchased.

REFERENCES

[1] Richard L. Sydnor, "Environmental Testing at the Jet Propulsion Laboratory's Frequency Standards Laboratory", *43rd Annual Symposium on Frequency Control,* 1989.

[2] G.J.Dick, P.F. Kuhnle, and R.L. Sydnor, "Zero-Crossing Detector with Sub-Microsecond Jitter and Crosstalk", *Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting,* 1990.

[3] R.L. Hamell, P.K. Kuhnle, R.L. Sydnor, An improved offset generator for Allan Deviation Measurement of ultra stable frequencies, *Proceedings of the 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting,* 1991

[4] Paul F. Kuhnle, "NASA/JPL Deep Space Network Frequency and Timing", *Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting,* 1989.

[5] *Ultra-Low Noise "N" Stage Parallel Amplifier,* Burr-Brown Integrated circuits Data Book, Volume 33

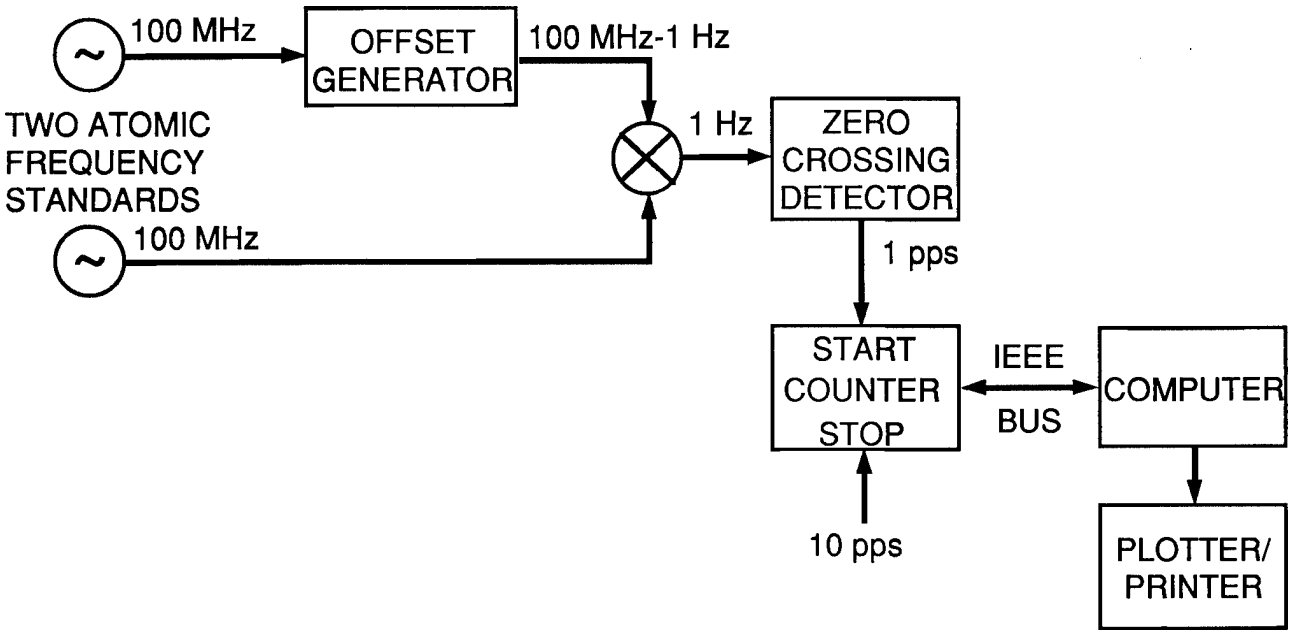


Figure 1. Typical Allan deviation test configuration.

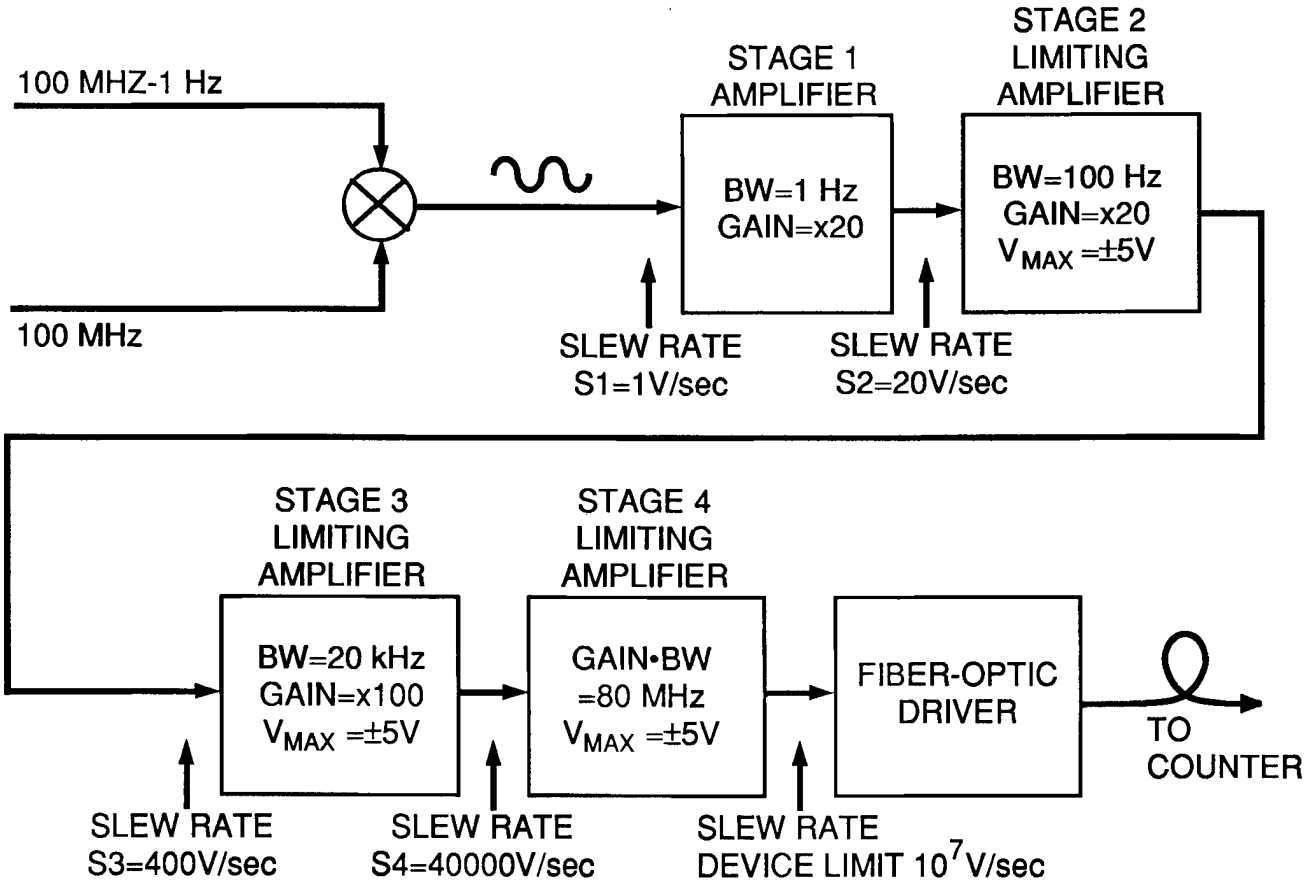
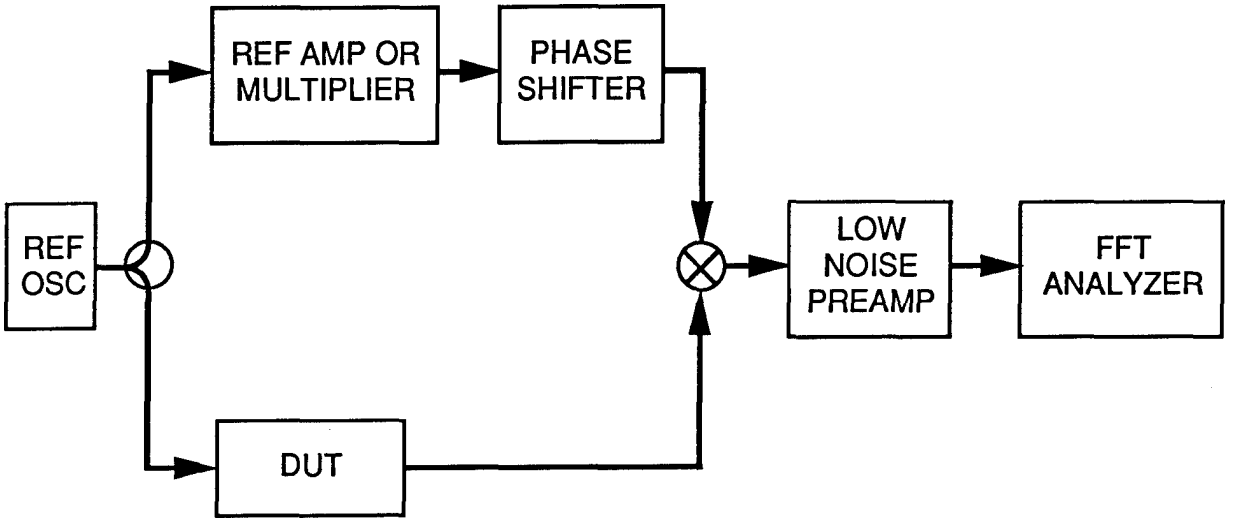
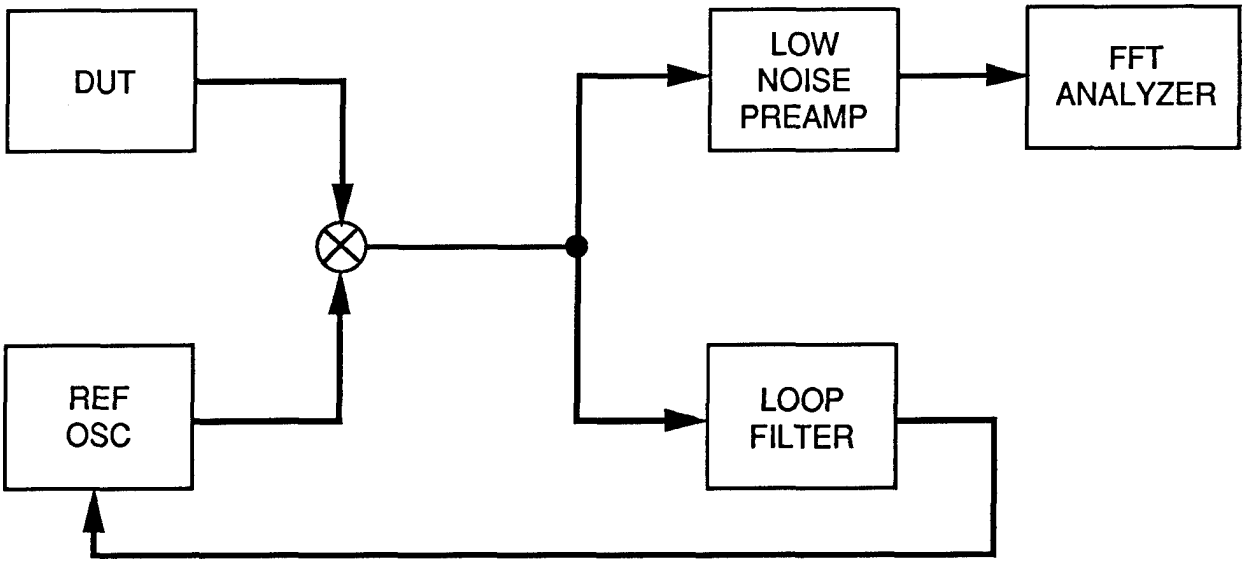


Figure 2. Block diagram of the new ZCD design.



AMPLIFIER / MULTIPLIER TEST SETUP



OSCILLATOR TEST SETUP

Figure 3. Phase noise test configuration.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF KALMAN FILTERING TECHNIQUES TO THE PRECISION CLOCK WITH NON-CONSTANT AGING

Wei Su
Geo-Centers, Inc.
615 Hope Road
Eatontown, NJ 07724

Raymond L. Filler
US Army Research Laboratory, SLCET-EF
Fort Monmouth, NJ 07703-5601

Abstract

The availability of onboard intelligence has created the possibility of compensating for systematic effects in time and frequency subsystems. To that end, a Kalman filter approach to compensate for aging in precision oscillators will be presented. The approach predicts the frequency based upon past measurements taken at arbitrary intervals. In an earlier algorithm, the aging rate was assumed to be constant after the most recent measurement so the predicted frequency change was always linear with time. This assumption is not generally satisfied in quartz oscillators. In the new approach, the oscillation frequency is described by a weighted sum of logarithmic functions. The weighting parameters are random variables estimated by the Kalman algorithm. Quartz oscillator examples are presented to show that the new algorithm is superior than the earlier one.

Introduction

The frequency compensation technique, utilizing the Kalman algorithm[1], presented here is an intelligent algorithm to be used in the Modular Intelligent Frequency, Time, and Time Interval (MIFTTI)[2] subsystem to enhance frequency standard performance. In many systems, a time reference is not continuously available. Thus, each user must intelligently estimate the time error between his clock and the system time (reference clock). The idea of an intelligent clock is illustrated in Figure 1. When the reference clock is available (switch closed) the Kalman algorithm learns the behavior of the crystal oscillator. When the reference clock is not available (switch open) the Kalman algorithm predicts and compensates for any frequency error due to aging. In reference [3] an algorithm was proposed to compensate for frequency aging which was assumed to have a constant aging rate during the prediction interval. The constant aging rate assumption may be suitable for Rubidium frequency standards but not for quartz oscillators. In this paper the algorithm is extended to improve the performance of quartz crystal oscillator based frequency standards.

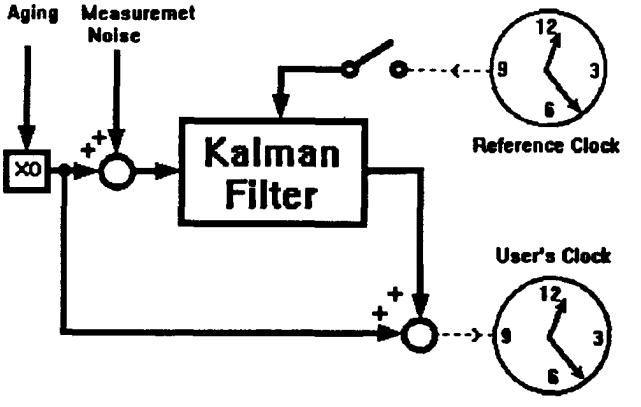


Figure 1 Intelligent Clock.

Oscillator Model Formulation

The behavior of an oscillator, or a clock, can be described by

$$\frac{dy(t)}{dt} = a(t) + \xi(t)$$

$$\frac{da(t)}{dt} = \zeta(t) \qquad (1)$$

$$y(t) = \frac{f(t) - f(t_0)}{f(t_0)}.$$

where, f(t) is the oscillation frequency at time t, a(t) is the aging rate of the oscillator, and $\xi(t)$ and $\zeta(t)$ are random variables. When the frequency is observed at unequally spaced times $t_0, t_1, \ldots, t_k, t_{k+1}, \ldots$, Eq.(1) can be written as a discrete-time linear state equation with noises ϕ_k and φ_k as

$$y_{k+1} = y_k + (t_{k+1} - t_k)a_k + \phi_k$$
$$a_{k+1} = a_k + \varphi_k. \quad (k = 0,1,2,\ldots) \qquad (2)$$

0-7803-0476-4/92 $3.00 © 1992 IEEE

Reference [3] estimates y_k and a_k by using the Kalman algorithm[1], as shown in Figure 2, and predicts y_{k+1} by ignoring the noises ϕ_k and φ_k. That is

$$\hat{y}_{k+1}^{-} = \hat{y}_k + (t_{k+1} - t_k)\hat{a}_k$$
$$\hat{a}_{k+1}^{-} = \hat{a}_k \ , \qquad (3)$$

where "^-" denotes a prediction and "^" denotes an estimate. Letting p=k+1 and c=k, Eq.(3) implies that the frequency at time t_p can be predicted, based on the knowledge of $\hat{y}_0$ and $\hat{a}_c$, via a linear discrete-time equation

$$\hat{y}_p^{-} = \hat{a}_c t_p + \hat{y}_0 \ , \qquad (4)$$

where $\hat{y}_0=\hat{y}_c-\hat{a}_c t_c$. The estimated aging rate $\hat{a}_c$ is updated after every measurement but is maintained at that constant value determined at the most recent calibration time, t_c, after which the reference is not available. However, references [4] and [5] indicate that the aging rate for a quartz crystal oscillator generally decreases with time. This implies $a_{k+1}\neq a_k$, so the frequency is not a linear time function. In this case, Eq.(3) will not give a satisfactory prediction because the expectation value of φ_k is not zero so the aging noise φ_k in Eq.(2) cannot be ignored.

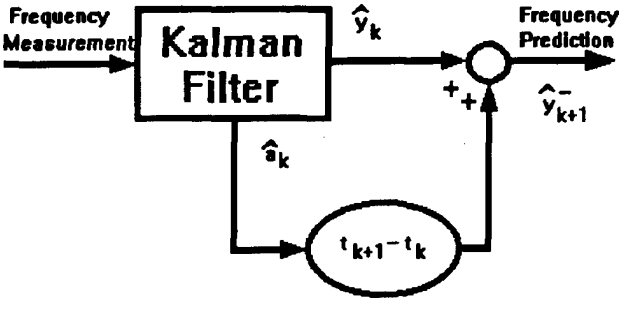


Figure 2 "Constant" Aging Kalman Algorithm.

A valid description[6] of quartz crystal oscillators may be to fit the frequency measurements to a non-linear time function such as

$$\eta(t) = A \cdot \log(Bt + 1) + C \ , \qquad (5)$$

where A, B, and C are parameters which are properly chosen to yield the least squares fitting errors

$$\sum_k (\eta_k - y_{m,k})^2 \rightarrow \min, \qquad (k = 0,1,2,\ldots) \qquad (6)$$

and where $\eta_k=\eta(t_k)$ is a discrete-time sample of $\eta(t)$, and $y_{m,k}$ is defined as follows

$$y_{m,k} = \frac{f_{m,k} - f_{m,0}}{f_{m,0}} . \qquad (k = 0,1,2,\ldots) \qquad (7)$$

Here, we have denoted $f_{m,k}=f_m(t_k)$ as the frequency measured at time t_k.

A discrete form of Eq.(5) is obtained by sampling $\eta(t)$ at $t=t_k$, which gives

$$\eta_k = \eta(t_k) = A \cdot \log(Bt_k + 1) + C. \qquad (8)$$

Note that the frequency prediction, Eq.(4), with constant aging assumption[3] is a special case of Eq.(8). To show this we choose a small B such that $|Bt_k| << 1$ and $\log(Bt_k+1)\rightarrow Bt_k$. Then, Eq.(8) becomes

$$\eta_k = ABt_k + C \ , \qquad |Bt_k| << 1 \qquad (9)$$

which yields Eq.(4).

Furthermore, the logarithmic model can be extended to a general form

$$\eta_k = \sum_{j=1}^{M} A_j \cdot \log(B_j t_k + 1) + C \ . \qquad (k = 0,1,2,\ldots) \qquad (10)$$

This is a weighted sum of logarithmic functions which can describe multiple processes.

However, the logarithmic model, in some cases, doesn't satisfactorily fit the long-term data if the data has unpredictable spikes, jumps, or sudden slope changes as shown in the examples. To address this problem, state noises are considered in the following discussion.

State Description of The Logarithmic Model

A recursive form of Eq.(10) can be obtained by subtracting η_k from η_{k+1} to form

$$\eta_{k+1} = \eta_k + \sum_{j=1}^{M} A_{j,k} \cdot \log\left[\frac{B_j t_{k+1} + 1}{B_j t_k + 1}\right], \qquad (11)$$

where $A_{i,k}$ is the parameter A_i at time t_k. Then, we assume a random disturbance $\omega_{i,k}$ (i=1,2, ... ,M) and update the parameter $A_{i,k}$ by

$$A_{j,k+1} = A_{j,k} + \omega_{j,k}. \qquad (j = 1, 2, \ldots, M) \qquad (12)$$

The combination of Eqs.(11) and (12) gives a state variable representation

$$\begin{bmatrix} \eta_{k+1} \\ A_{1,k+1} \\ \vdots \\ A_{M,k+1} \end{bmatrix} = \begin{bmatrix} 1 & \log\frac{B_1 t_{k+1}+1}{B_1 t_k+1} & \cdots & \log\frac{B_M t_{k+1}+1}{B_M t_k+1} \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} \eta_k \\ A_{1,k} \\ \vdots \\ A_{M,k} \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_{1,k} \\ \vdots \\ \omega_{M,k} \end{bmatrix} \qquad (13)$$

and the measurement is then described by

$$y_{m,k} = \eta_k + \mu_k$$

$$= [\,1\ 0\ \bullet\bullet\ 0] \begin{bmatrix} \eta_k \\ A_{1,k} \\ \bullet \\ \bullet \\ \bullet \\ A_{M,k} \end{bmatrix} + \mu_k, \tag{14}$$

where μ_k is a random variable. Eqs.(13) and (14) is a state variable representation of an oscillator which does not need any of the information from before the kth sampling step. If the B_i parameter can be determined in advance, Eq.(13) will be a linear state variable equation and the frequency will be predicted, as shown in Figure 3, by using the Kalman algorithm. Unfortunately, the determination of the B_i's are, usually, formidable and memory-consuming. Thus, it is necessary to modify Eq.(13) to get a simpler model.

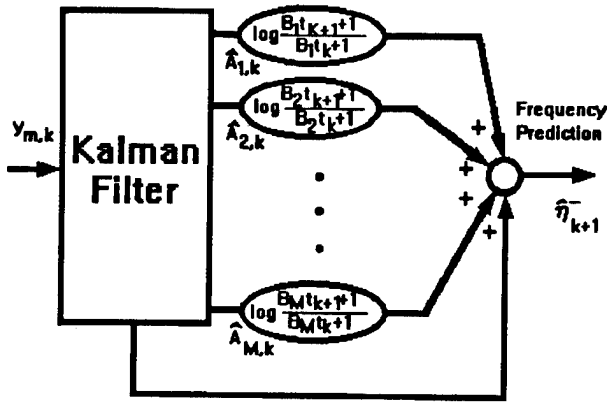


Figure 3 "Log" Aging Kalman Algorithm.

To obtain a logarithmic model without being bothered by the B_i's is to assume $B_i t_k >> 1$, thus $(B_i t_{k+1}+1)/(B_i t_k+1) \rightarrow t_{k+1}/t_k$. In this case, Eq.(11) yields

$$\eta_{k+1} = \eta_k + \log\left[\frac{t_{k+1}}{t_k}\right] \cdot \sum_{j=1}^{M} A_{j,k}\ ; \qquad B_j\, t_k >> 1 \tag{15}$$

$$= \eta_k + \log\left[\frac{t_{k+1}}{t_k}\right] \cdot \alpha_k\ ,$$

where

$$\alpha_k = \sum_{j=1}^{M} A_{j,k} \tag{16}$$

is updated by

$$\alpha_{k+1} = \alpha_k + \omega_k\ , \tag{17}$$

and where

$$\omega_k = \sum_{j=1}^{M} \omega_{j,k}\ . \tag{18}$$

A simplified state representation of the logarithmic aging model is then

$$\begin{bmatrix} \eta_{k+1} \\ \alpha_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & \log\frac{t_{k+1}}{t_k} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \eta_k \\ \alpha_k \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_k \end{bmatrix} \tag{19}$$

and

$$y_{m,k} = [\,1\ \ 0\,] \begin{bmatrix} \eta_k \\ \alpha_k \end{bmatrix} + \mu_k\ . \tag{20}$$

Assuming ω_k and μ_k are white, uncorrelated, and have known covariance, we can optimally estimate the values of η_k and α_k by using the Kalman algorithm, Figure 4, based on the measurements before and at time t_k.

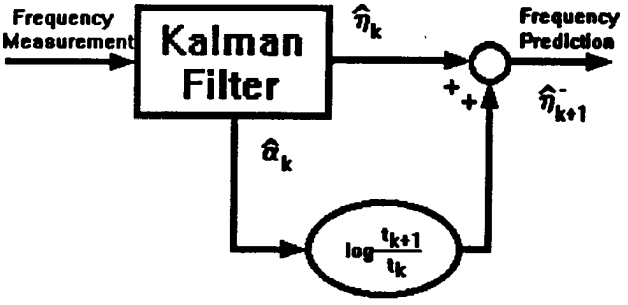


Figure 4 Simplified "Log" Algorithm.

Kalman Algorithm

The Kalman algorithm[1,7] is a recursive computation for obtaining the optimum estimate of a random process in which both the state of the process and the measurements are affected by noises.

The discrete-time model of a random process is described as

$$x_{k+1} = \Phi_{k+1,k} x_k + \rho_k \tag{21}$$

and

$$z_k = H_k x_k + \mu_k\ . \tag{22}$$

where x_k is the n dimensional process state vector at time t_k, z_k is the m dimensional measurement vector at time t_k, ρ_k is assumed to be an n dimensional white noise with

known covariance, μ_k is assumed to be an m dimensional white noise with known covariance and uncorrelated with the ρ_k, $\Phi_{k+1,k}$ is the transition matrix to translate the state x_k into x_{k+1}, and H_k is the measurement matrix which relates the state to the measurements.

The covariance matrices used in the algorithm are

$$Q_k = E[\rho_k \rho_k]$$

$$R_k = E[\mu_k \mu_k]$$

$$P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \quad (23)$$

$$P_k^- = E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] .$$

$$(k = 0,1,2,\ldots)$$

The optimal state estimate $\hat{x}_k$ and the error covariance matrix P_k are calculated recursively. At time t_k, we have

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - H_k\hat{x}_k^-) \quad (24)$$

and

$$P_k = (I - K_kH_k)P_k^- , \quad (25)$$

where, K_k is the optimal blending factor and is obtained by

$$K_k = P_k^-H_k^T(H_kP_k^-H_k^T + R_k)^{-1} . \quad (26)$$

Then, the prior state (one-step state prediction) and error covariance matrices are updated for the next step by

$$\hat{x}_{k+1}^- = \Phi_{k+1,k}\hat{x}_k \quad (27)$$

and

$$P_{k+1}^- = \Phi_{k+1,k}P_k\Phi_{k+1,k}^T + Q_k . \quad (28)$$

Frequency Compensation

Assume the reference clock is available until time t_c. We desire to know the frequency at time t_p ($t_p > t_c$) in order to eliminate the frequency error of a user's clock without the reference information. When $t_k < t_c$, the frequency measurement is fed to the Kalman algorithm at each discrete time t_k and the Kalman algorithm learns the system behavior by searching for the "best" estimate of α_k and η_k if the logarithmic model is considered. After t_c, the Kalman algorithm receives no further information but predicts the frequency at time t_p based on the knowledge gathered up until time t_c. To show this, we define

$$x_k = \begin{bmatrix} \eta_k \\ \alpha_k \end{bmatrix}$$

$$\Phi_{k+1,k} = \begin{bmatrix} 1 & \log\frac{t_{k+1}}{t_k} \\ 0 & 1 \end{bmatrix} \quad (29)$$

$$\rho_k = \begin{bmatrix} 0 \\ \omega_k \end{bmatrix}$$

$$H = [1 \quad 0]$$

$$z_k = y_{m,k} .$$

With c=k and p=k+1, we can rewrite Eq.(27) to

$$\hat{\eta}_p^- = \hat{\eta}_c + \hat{\alpha}_c\log\left[\frac{t_p}{t_c}\right]$$

$$= \hat{\alpha}_c\log(t_p) + \hat{c}_c , \quad (30)$$

where $\hat{c}_c = \hat{\eta}_c - \log(t_c)$.

The values of $\hat{\eta}_c$ and $\hat{\alpha}_c$ are estimated at time t_c by using the Kalman algorithm, Eq.(24), and the frequency prediction $\hat{\eta}_p^-$ at time t_p is obtained by using the logarithmic equation, Eq.(30).

Examples

Data for three quartz crystal oscillators are chosen to illustrate the aging compensation. The first example shows the superiority of using the logarithmic model over the constant aging rate model[3], and the second and third examples present the advantage in using the Kalman algorithm rather than using least squares fitting.

Example 1

Figure 5 illustrates the frequency prediction of Oscillator #1 utilizing the constant aging algorithm. Where the squares are the measured data and the solid line is the prediction. When $t > t_c$ = 35 days, the frequencies are predicted by a straight line. However, this linear prediction curve doesn't match the frequencies observed after the time t_c since the actual aging rate of Oscillator #1 is not constant. Thus, the prediction error increases with time. To show the improvement, Figure 6 presents the prediction result by using the Kalman filter algorithm with a logarithmic model. As we see, the prediction error is significantly reduced.

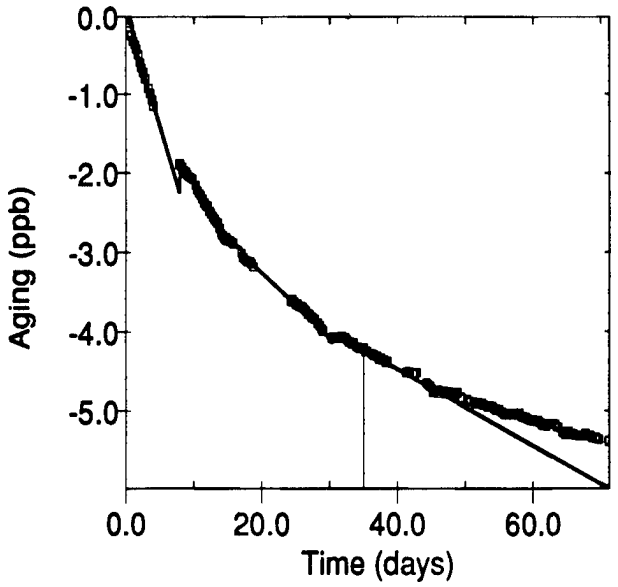


Figure 5 Kalman Algorithm Updated with Measurements Until t = 35 days. Aging Rate Assumed Constant After 35 days.

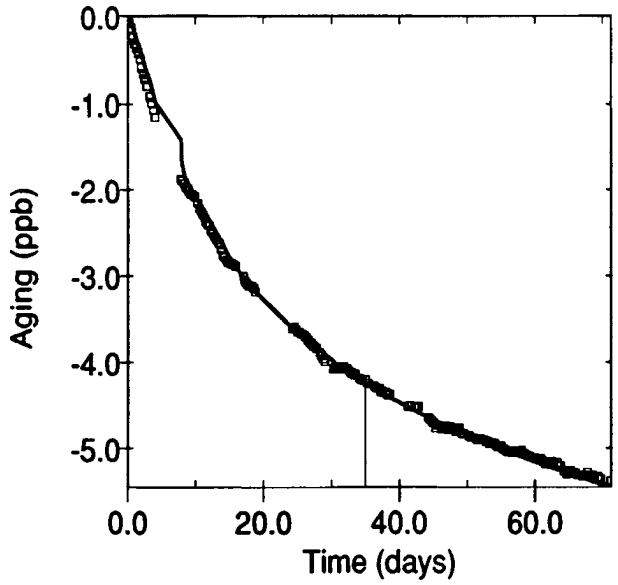


Figure 6 Kalman Algorithm Updated with Measurements Until t = 35 days. Logarithmic Aging Assumed After 35 Days.

Example 2

This example use both least squares fitting and Kalman algorithm to predict the frequency of Oscillator #2. First, Let us define a fitting function

$$\eta_p = A \cdot \log(t_p) + C \qquad (31)$$

so it has the same form as in Eq.(30). The parameters A and C can be calculated by using the well known least squares fitting formula based on the measurements $y_{m,k}$'s and the sampling steps t_k's. Unfortunately, the measurements in Figure 7 (squares) are not fit well by a single logarithmic function. Rather the data is approximately a piece-wise logarithmic function with a discontinuity at 600 days due to low temperature storage while in "off" state[4]. Although the least squares fitting can be implemented by using the data after the discontinuity, as shown in Figure 7 (the solid line between 650-878 days), the least squares fitting method is not straightforward in predicting the oscillation frequency since the discontinuities may occur randomly. However, the Kalman algorithm does not require the knowledge of discontinuities. As shown in Figure 8, the frequency prediction (solid line) switches from one logarithmic function to another at 600 days to fit the measured data (squares) until t_c = 900 days. Then, it gives predictions up to 1150 days without the reference time information.

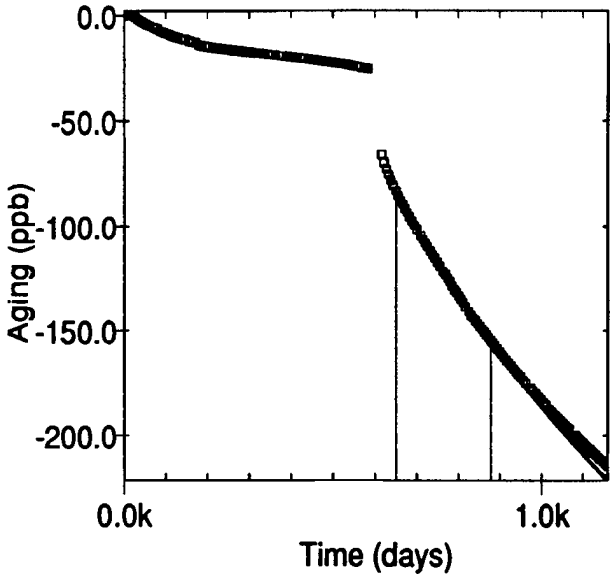


Figure 7 Least Squares Log Fit Using Data Between 650 - 878 days. Logarithmic Aging Assumed After 878 days.

Example 3

The Oscillator #3 is a difficult example since, first, it does not match a single logarithmic function, second, there is a rapid slope change at t = 25 days and an anomalous spike observed at t = 120 days. If the least squares fitting, as shown in Figure 9, is used, only the data after 120 days will be acceptable. However, in the

Kalman algorithm, we do not need the knowledge of slope changes and spikes since it affects only a few iterations of data estimation. As we see in Figure 10, the prediction (solid line) agrees with the measurements (squares) when time $t_k \leq t_c = 300$ days, and gives a satisfactory prediction while $t_k > t_c$.

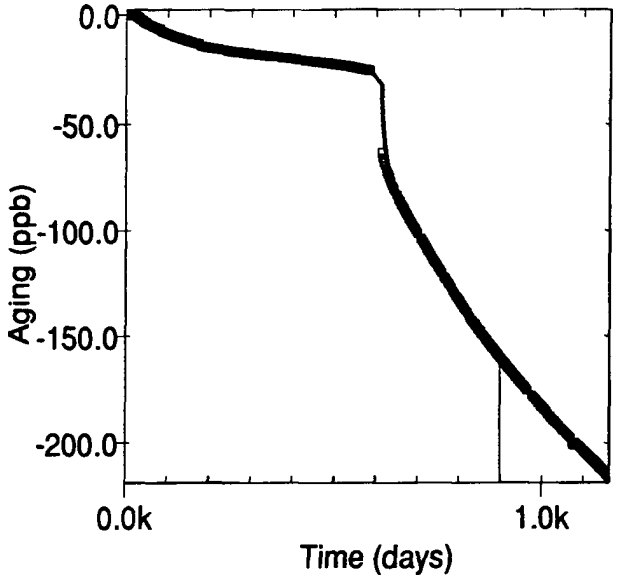


Figure 8 Kalman Algorithm Updated with Measurements Until t = 900 days. Logarithmic Aging Assumed After 900 days.

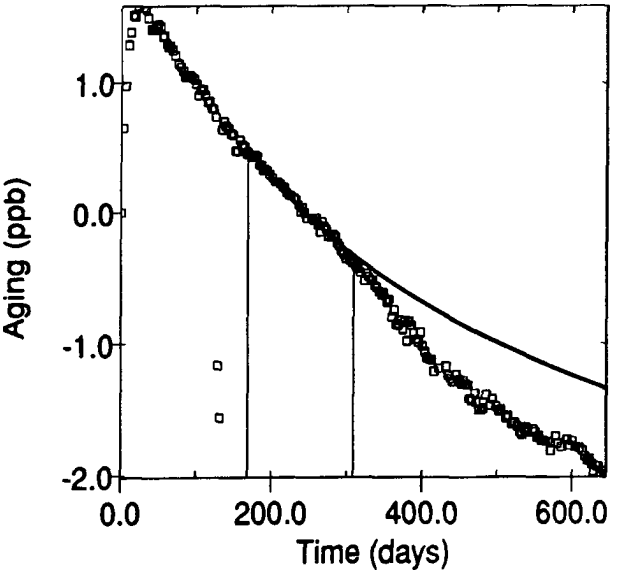


Figure 9 Least Squares Log Fit Using Data Between 168 - 309 days. Logarithmic Aging Assumed After 309 days.

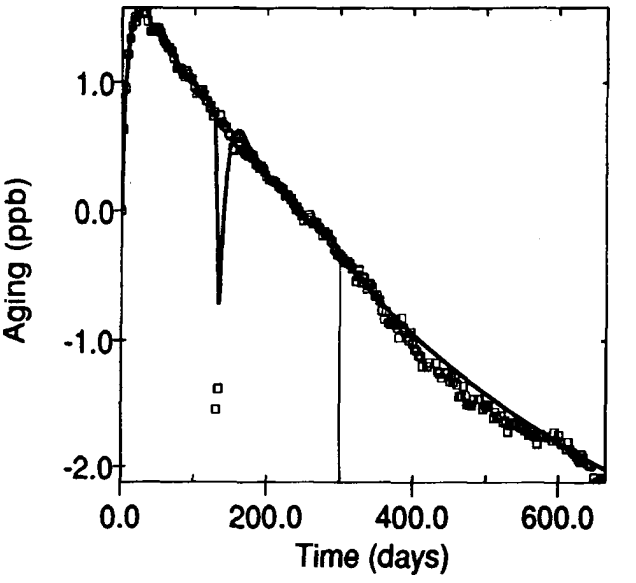


Figure 10 Kalman Algorithm Updated with Measurements Until t = 300 days. Logarithmic Aging Assumed After 300 days.

Conclusions

The frequency of a quartz oscillator can be modelled by a logarithmic function and can be compensated by using the Kalman algorithm. The logarithmic model is more suited to crystal oscillators than the constant aging model proposed previously, and the Kalman algorithm is better than least squares fitting method and it requires much less data storage.

References

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Trans. of the ASME-J. of Basic Engr., pp.35-45, March 1960.

[2] K. D. Lyon, W. J. Riley, and J. R. Vaccaro, "The Modular Intelligent Frequency, Time, and Time Interval (MIFTTI) Subsystem Program," Proc. 43rd Annual Symposium on Frequency Control, pp.225-231, 1989.

[3] R. L. Filler and S. R. Stein, "Kalman Filter Analysis for Real Time Applications of Clocks and Oscillators," Proc. 42nd Annual Symposium of Frequency Control, pp.447-452, June 1988.

[4] R. L. Filler and J. R. Vig, "Long Term Aging of Oscillators," Proc. 1992 IEEE Frequency Control Symposium, May 1992.

[5] J. R. Vig, "Introduction to Quartz Frequency Standards," Proc. 23rd Annual Precise Time and Time Interval(PTTI) Applications and Planning Meeting, December 1991.

[6] J. R. Vig and T. R. Meeker, "The aging of Bulk Acoustic Wave Resonators, Filters and Oscillators," Proc. 45th Annual Symposium of Frequency Control, pp.77-101, May 1991.

[7] R. G. Brown, Introduction to Random Signal Analysis and Kalman Filtering, New York: John Wiley & Sons, 1983.

PERFORMANCE OF A PID PHASE LOCK LOOP WITH KALMAN FILTERED INPUT DATA

Robert Hardin and Mark Yankowski

Ball Corporation, Efratom Division, 3 Parker, Irvine, CA 92718-1605

Abstract- **In this paper, the performance of a rubidium oscillator disciplined to the Global Positioning System (GPS) satellite time transmissions is presented. Also presented is the phase lock loop technique used. The performance data were obtained after the level of Selective Availability (SA) was turned up in November of 1991. The results show an offset frequency accuracy of 1.0E-11 while maintaining a phase accuracy of 300 nanoseconds versus UTC. Typically the offset frequency error was ≤2.0E-12. Also presented are results of analyzing recently acquired GPS data. The fractional frequency stability of least squares estimates performed on the data demonstrate that with SA active, one can at best expect 2.0E-11 frequency accuracy with an OCXO and can expect accuracies of 2.0E-12 (typically) with rubidium oscillators.**

The disciplining technique used performs least squares filtering on the phase error data from the receiver. The phase at the end of the least squares line ($z_m(k+1)$) is used as the measurement at time index 'k+1'. The $z_m(k+1)$ measurements are filtered by an adaptive one state Kalman filter to obtain the estimates ($\hat{x}(k+1|k+1)$).

The $\hat{x}(k+1|k+1)$ phase error estimates are processed by a Proportional-Integral-Differential (PID) control loop to provide the control input to the rubidium oscillator.

The phase comparisons were referenced to a GPS disciplined rubidium oscillator in which only Block 1 GPS satellites were tracked by the GPS receiver in use. The frequency comparisons were referenced to the Block 1 system and to a Loran-C station's cesium standard. The receivers used were Trimble Advanced Navigation Sensor Timing modules from Trimble Navigation.

I. Introduction

Ever since the United States Department of Defense (DOD) announced their decision to intentionally degrade the accuracy of GPS via Selective Availability (SA) being implemented on Block 2 satellites, many users have been concerned about the time and frequency accuracies obtainable under SA. The DOD formally implemented SA on March 25, 1990 and put an end to "a long period of uncertainty, apprehension and heated debate," as the May/June 1990 issue of GPS World put it [13] [1]. However, during the Desert Storm conflict SA was turned off for strategic reasons. Because of this users only got a glimpse of the effects of SA. Selective Availability was turned back up in November of 1991.

This paper presents the phase and frequency performance results obtained by using an adaptive one state Kalman filter with prefiltering. The results show an offset frequency accuracy of 1.0E-11 while maintaining a phase accuracy of 300 nanoseconds. Typically the offset frequency error was ≤2.0E-12.

Also presented are results of analyzing recently acquired GPS data. The fractional frequency stability of least squares estimates performed on the data demonstrate that with SA active, one can at best expect 2.0E-11 frequency accuracy with an OCXO and can expect accuracies of 2.0E-12 (typically) with rubidium oscillators.

Theoretical formulations are included in Appendix A.

[1]References are indicated by numbers enclosed in square brackets and listed in Appendix B.

0-7803-0476-4/92 $3.00 © 1992 IEEE

II. Problem Statement

Figure 1.0 represents the control problem to be solved and the system configuration employed in controlling the oscillator. (See Figure 1.0 on the following page for reference.)

The problem statement was to use a microprocessor based controller to control the oscillator and maintain a phase accuracy of 500 nanoseconds and an offset frequency of 1.0E-11 in the presence of Selective Availability (SA). Assume that GPS satellites are continuously available.

In the environment under consideration the pressure changes and vibration are insignificant and the temperature of the baseplate of the oscillator varies only ±2.5 °C over any 24 hour period. If wider temperature range operation is required a temperature stabilized oscillator will be employed.

III. Problem Analysis

The state variables of the rubidium oscillator are the phase, offset frequency, and drift. Note that the offset frequency and drift can be estimated from the phase measurements. With a full GPS constellation no drift estimation and correction is required. The state of the MRK is perturbed by environmental variables such as temperature, pressure, vibration, etc. Since the temperature variations are so benign no temperature control technique is needed (other than that employed within the MRK). The MRK temperature coefficient is typically 2.0E-12 per °C. The MRK phase error (process state to be estimated) is further perturbed by MRK system noise. We shall assume that the MRK noise components produce zero mean Gaussian phase jitter. The total peak-to-peak jitter is much less than 1.0 nanosecond which is quite small compared to GPS system noise.

If the Trimble receiver were used with an on-board Temperature Compensated Crystal Oscillator (TCXO) it would provide a 1.0 millisecond accuracy on its 1 PPS output signal. The 1 PPS output would be 1.51 to 0.49 milliseconds early. The Trimble receiver also contains bias (phase error between its 1 PPS output and GPS) on its RS-422 data port. The bias data provides error information accurate to tens of nanoseconds in the absent of SA. Furthermore, if the TCXO is removed and a synthesized RF input from the MRK is provided the receiver's 1 PPS output moves synchronously with the MDRC generated 1 PPS output. Therefore, at start-up the MDRC 1 PPS output is synchronized with the receiver's 1 PPS output. After this initial synchronization is performed the bias data is a direct indication of the phase error of the MRK oscillator. These bias data are perturbed by receiver noise and GPS system noise. These additive noise components combine to degrade the measurement capability of the GPS receiver by inducing a phase jitter and dither on the reference data. We shall assume that the GPS observation noise is Gaussian.

An important property of Gaussian random signals is that they remain Gaussian after passing through a linear system. Assuming independent Gaussian primary random sources, if the observed random signal is also Gaussian, we may assume that the dynamic system between the observer and the primary source is linear. This conclusion may be forced on us also because of lack of detailed knowledge of the statistical properties of the observed random signal. Given any random process with known first and second-order averages, we can find a Gaussian random process with the same properties. Thus Gaussian distributions and linear dynamics are natural, mutually plausible assumptions particularly when the statistical data are scant [3].

Prior to the activation of SA we were using a least squares filter (LSF) to estimate the phase error of the MRK oscillator and using these phase estimates in a Proportional-Differential-Integral (PID) type controller. A two hour integration period had been decided upon based on results obtained from limited testing during the time when SA was activated just prior to Desert Shield.

We felt that with SA active the PID controller could provide suitable control of the MRK as long as the least squares phase estimates were not excessively noisy and the PID terms were suitably chosen.

We did not want to increase the integration period beyond the existing two hour period. This was so that slowly varying environmental temperatures could be compensated for by the PLL and a temperature compensation scheme avoided. Also we did not want the phase performance to suffer by extending the integration time. It was reasoned that the output of the LSF could be further filtered by a

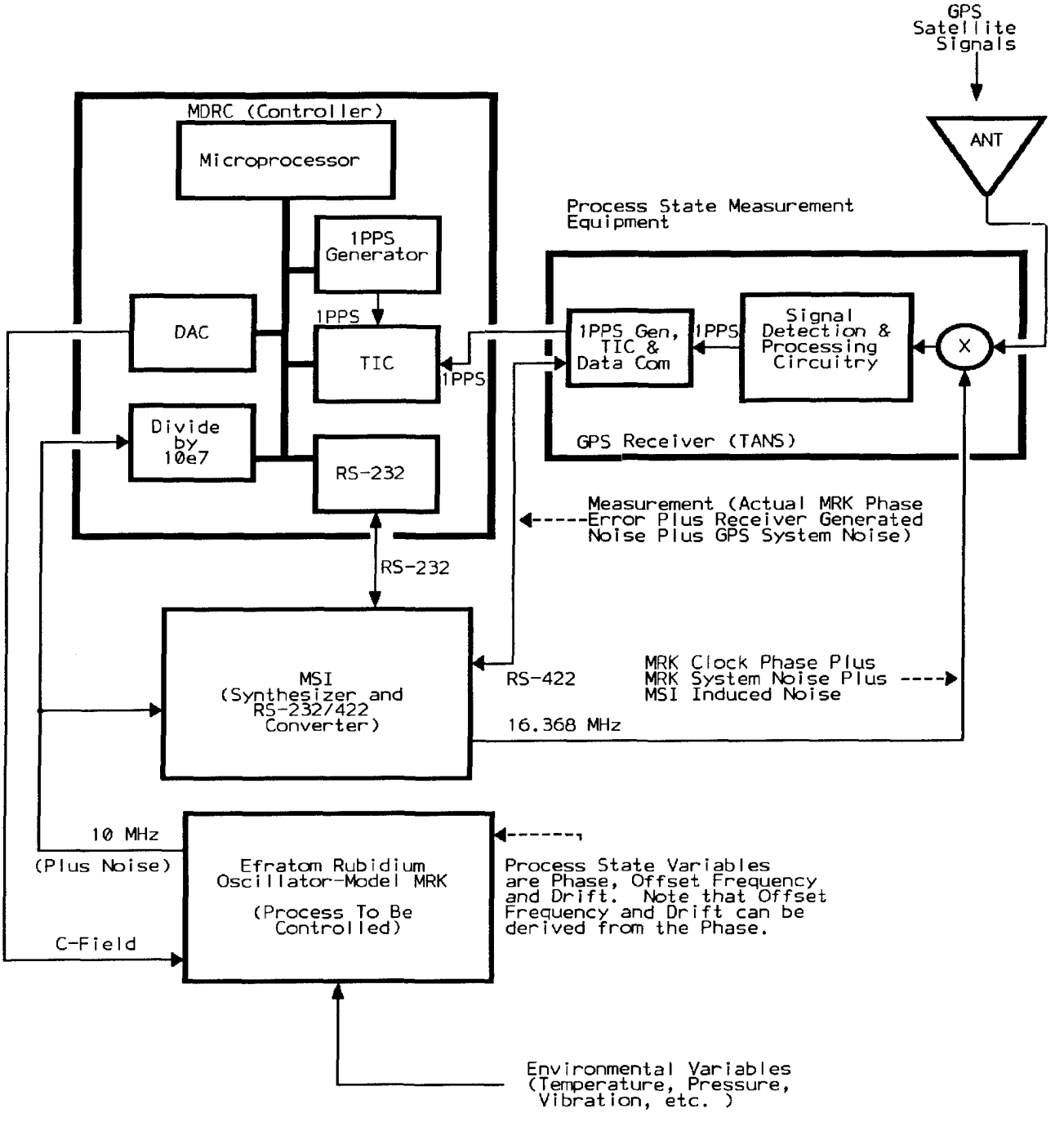
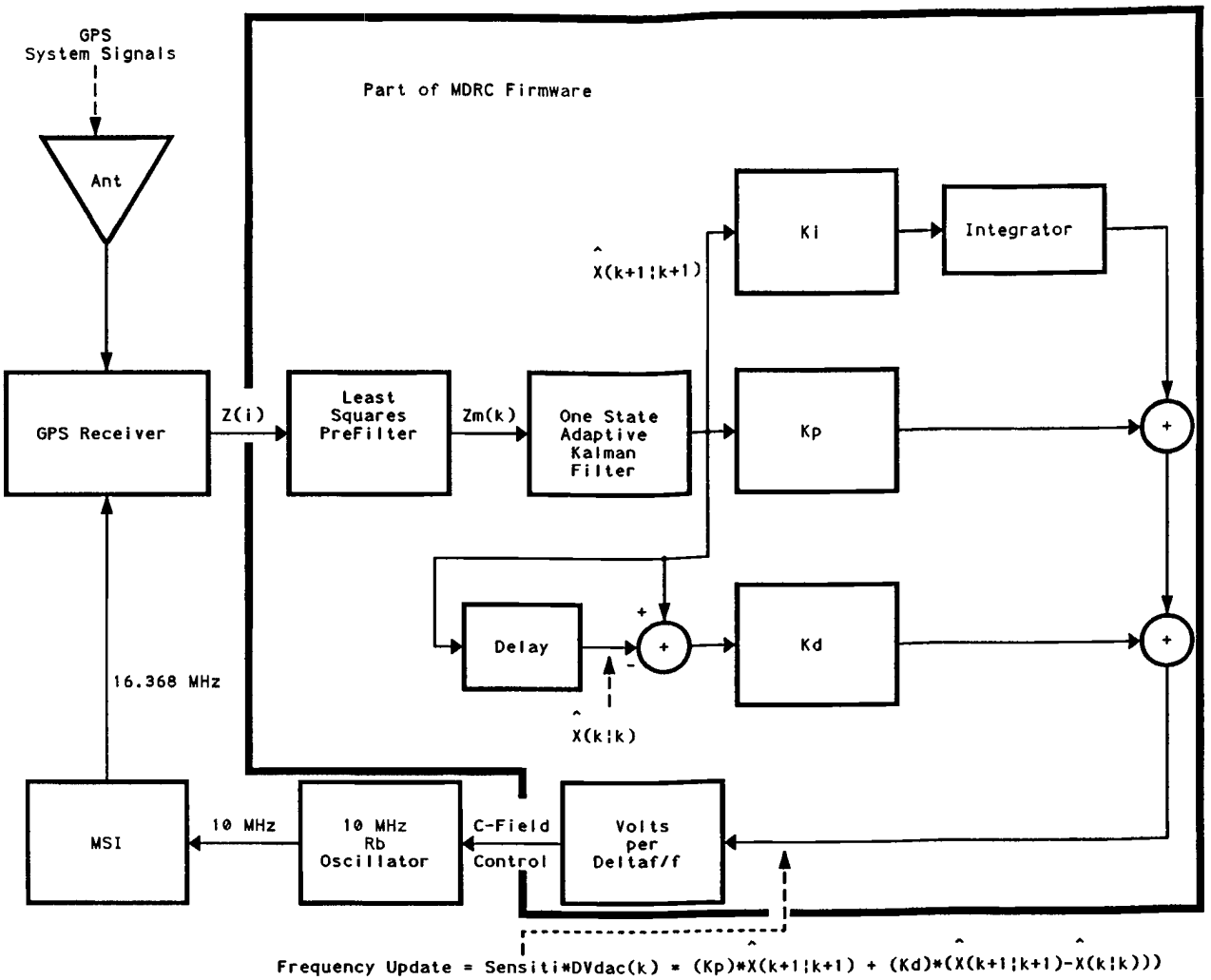


Figure 1. MRK Control Problem Representation



$$\text{Frequency Update} = \text{Sensiti}*DVdac(k) * (Kp)*\hat{X}(k+1|k+1) + (Kd)*(\hat{X}(k+1|k+1)-\hat{X}(k|k))) + (Ki)*\sum_{k=1}^{M}(\hat{X}(k+1|k+1)) + \text{InitDeltaf}$$

DVdac(k) = Vdac(k) - 2048

DVdac(k) = Change (in counts) from initial setting of 12 bit DAC at update k

Vdac(k) = Count value of 12 Bit DAC at update k

Z(i) = Phase error data from GPS Receiver
Zm(k) = Phase error measurement from prefilter

$\hat{X}(k+1|k+1)$ = Phase error estimate from Kalman filter
Sensiti = MRK oscillator offset frequency to C-Field voltage sensitivity
InitDeltaf = Estimate of Initial Oscillator Offset Freq.

Kp = Gain of proportional phase lock loop term
Ki = Gain of integral phase lock loop term
Kd = Gain of differential phase lock loop term

Figure 2. Phase Lock Loop Representation

Kalman filter to reduce the noise on the LSF phase estimates.

The literature is quite rich on Kalman filtering [2-4,7, 9,10-12]. Appendix A.1 summarizes the Kalman filter equations and assumptions. The notation that we have adopted is that of Meditch [7].

The use of a prefilter in Kalman filtering has been extensively discussed in the literature [10]. Appendix A.2 has a derivation of the effect of the LSF prefilter on the phase observations from the GPS receiver. It is seen that the observation noise, v_i is smoothed; however, there is some additional noise introduced by smoothing the state. As long the process noise is small compared to the measurement noise and the integration time is adequate the LSF prefilter will provide an improvement. The observation noise is much greater than the noise on the Efratom MRK oscillator.

Further note that by using the output of a least squares prefilter as the measurement (z_m) the effect of the control update is contained in the measurement. Therefore, $z_m(k+1)$ is used directly as input to the Kalman filter equation. Figure 2.0 is a block diagram of the phase lock loop employed.

IV. Simulation Results

At the time of development of the control algorithm, sufficient receiver data with SA active did not exist. Therefore, SA was simulated and the entire control loop was simulated by programs written in the C and Pascal programming languages. During the simulations the PID loop terms were tweaked to provide the desired performance. Figures 3.0 and 4.0 show the simulation results. In the figures, the 'LSF' plots are the results of using only a least squares filter on the observation data. The 'Kalman' makes use of a least squares prefilter which is followed by an adaptive Kalman filter.

From the figures it is seen that the Kalman filter greatly improves the phase lock loop performance and indicates that the design goal could be met. The 'LSF' data show that the least squares filter only solution (with a 7200 second integration period) performs satisfactorily in phase but does not meet the frequency accuracy requirements. For more than 50% of the time the least squares filter solution had an offset frequency of more than 1.0E-11 (actually got as high as 3.2E-11).

As expected, the Kalman filter simulation results indicate slightly better performance than the experimental results shown in figures 7.0 and 8.0. However, the simulation results show that the simulation can be used to predict actual performance and therefore serves as a valuable design tool.

V. Experimental Results

Figure 5.0 is a plot of bias data experimentally obtained between 3/18/92 and 4/18/92. The fact that SA is active is evident from this plot. Figure 6.0 is a histogram done on the data in figure 5.0. The data represents 9552 samples spaced 5.0 minutes apart. The histogram shows that the bias data is between -400 nanoseconds and +400 nanoseconds 95% of the time with a distribution that is very close to Gaussian. The GPS position accuracy distributions were expected to be non-Gaussian in the 95% region [1]. The Standard Position Service (SPS) timing accuracy with respect to UTC was expected to be 363 nanoseconds 95% of the time [1]. Our findings support our expectations.

Figure 7.0 is a histogram plot of phase data recorded to disk between 3/18/92 and 5/6/92 (see the test setup in figure 10.0). The data show that the phase error between the Block 1 reference and the unit under test (UUT) was < 300 nanoseconds 92% of the time. Therefore, we expect the typical phase maintenance of MDRC disciplined MRK oscillators referenced to GPS to degrade by 300 nanoseconds when comparing SA active performance with SA inactive performance.

Figures 8.0 and 9.0 show the absolute and 1 day averaged offset frequency performance respectively. Figure 8.0 shows that only 1% of the time did the offset frequency exceed 1.0E-11 (by a few parts in ten to the twelfth). Figure 9.0 shows that 65% of the time the 1 day averaged offset frequency is $\leq$2.0E-12 and at no time does it exceed 1.0E-11.

Figure 10.0 shows the test setup used.

VI. Recently Acquired GPS Data

Figure 11.0 is a plot of observation data (receiver bias data) acquired on 3/18/92. Figure 12.0 is an

autocorrelation plot obtained from the Figure 11.0 data. The autocorrelation damps out rather quickly.

The expectation value is -86.79 nanoseconds indicating that the mean of the GPS noise is nearly zero but biased slightly negatively. The autocorrelation calculation was repeated with a shift in time origin of 30 days without significant change. Therefore, we shall assume that if sufficient samples were analyzed the autocorrelation function would be independent of time origin and the mean value would be constant (definition of wide-sense stationary) [6]. Furthermore, a property of stationary processes is that their autocorrelation functions die out quickly [5]. Thus we shall assume that the GPS system noise is a Gaussian wide-sense stationary process. A Gaussian stationary process is completely characterized by its mean, variance, and autocorrelation function [5].

Figures 13.0 and 14.0 are spectral density plots derived from the autocorrelation function of figure 12.0. It is seen that most of the variance of the series is concentrated at low frequencies.

The autocorrelation plot was obtained by making use the following equations [5]:

$$(1)\quad r_k = c_k/c_0$$

$$(2)\quad c_k = (1/N)\sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}),\ k = 0,1,2,....,K$$

The spectral density function plot was obtained by using the following equation [5]:

$$(3)\quad g(f) = p(f) / \sigma_z^2$$

$$= 2\{1 + 2\sum_{k=1}^{\infty} p_k \cos 2\pi fk\},\ 0 < f < \tfrac{1}{2}$$

The spectrum p(f) shows how the variance of a stochastic process is distributed between a continuous range of frequencies [5].

From figure 14.0 we see that spectral components below 1.25 millihertz are relatively strong. Using the Nyquist maximum interval for sampling (T = 1/2B; B is the maximum frequency desired for recovery) we see that a sampling interval of 1/(2*1.25E-3) or 400 seconds would not provide filtering of the effects of SA. Therefore, to remove or attenuate the effects of SA, the integration period must be considerably longer than 400 seconds.

Figure 15.0 is a plot of the fractional frequency stability $\sigma_y(\tau)$ of offset frequency estimates of a rubidium oscillator versus integration time. The estimates were performed using actual GPS observation data spaced five minutes apart. One hundred estimates were performed at each integration value in arriving at each $\sigma_y(\tau)$. The data were obtained by calculating $\sigma_y(\tau)$ on the data used to generate the plot in figure 6.0. Figure 15.0 shows that $\sigma_y(\tau)$ of the estimates decreases rapidly with integration time for integration times less than 5,000 seconds. For integration times between 5,000 seconds and 27,000 seconds (our data record at this time does not allow calculation beyond 27,000 seconds) $\sigma_y(\tau)$ of the estimates approximately decreases by a factor of two (from 2.0E-11 to about 3.0E-12) as τ doubles. Therefore, one can expect good performance of a control loop using a least squares filter alone (no Kalman filtering) provided the integration period is $\geq$ 4.0 hours.

Figure 16.0 shows the fractional frequency stability of several precision commercial and laboratory frequency standards [15]. Some of the figure 15.0 data points have been plotted on figure 16.0 to readily compare the fractional frequency stability of least squares estimates from GPS data against precision commercial oscillators.

The flicker floor of an unperturbed rubidium oscillator is well below 5.0E-13. Because of environmental influences, namely temperature we can expect a flicker floor of 3.0E-12 for fielded rubidium standards. For a 3.0E-12 noise floor an averaging time of 27,000 seconds provides an optimal crossover.

The OCXO curve shows that at 100 seconds the OCXO starts to drift and has a $\sigma_y(\tau)$ of about 2.0E-11 at 5,000 seconds which is the value of τ for which the estimate has an equivalent $\sigma_y(\tau)$. This demonstrates that with SA active one can at best expect 2.0E-11 (at the worst averaging time of 5000 seconds) frequency accuracy with an OCXO.

Figure 16.0 provides a good overview of the expected performance of different system configurations.

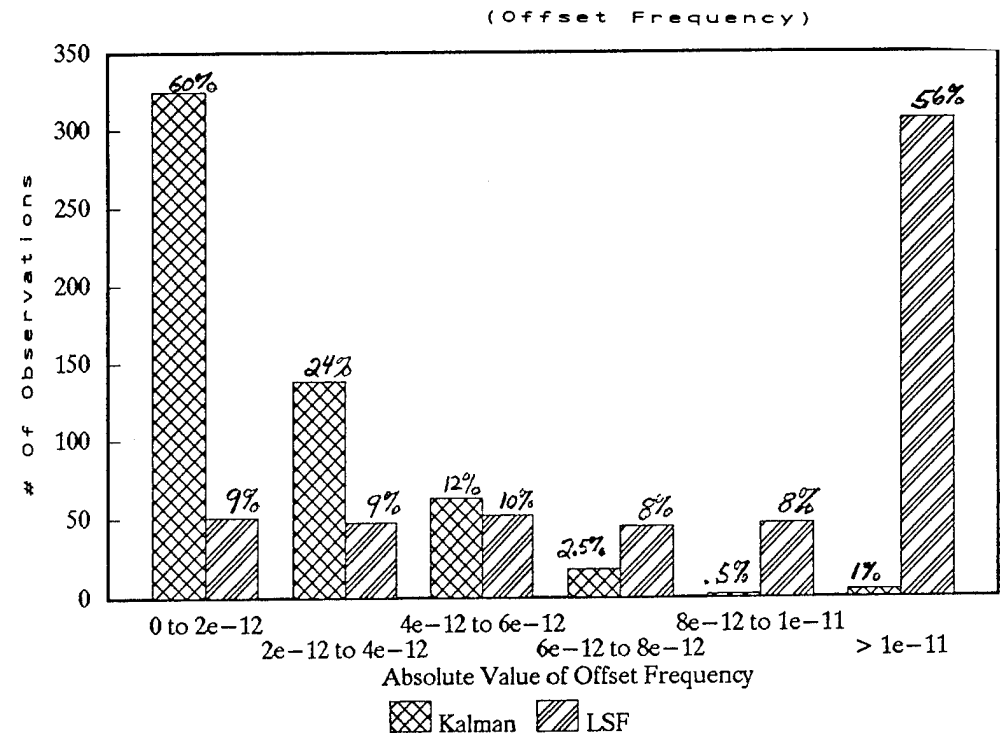


Figure 3. Simulation Results

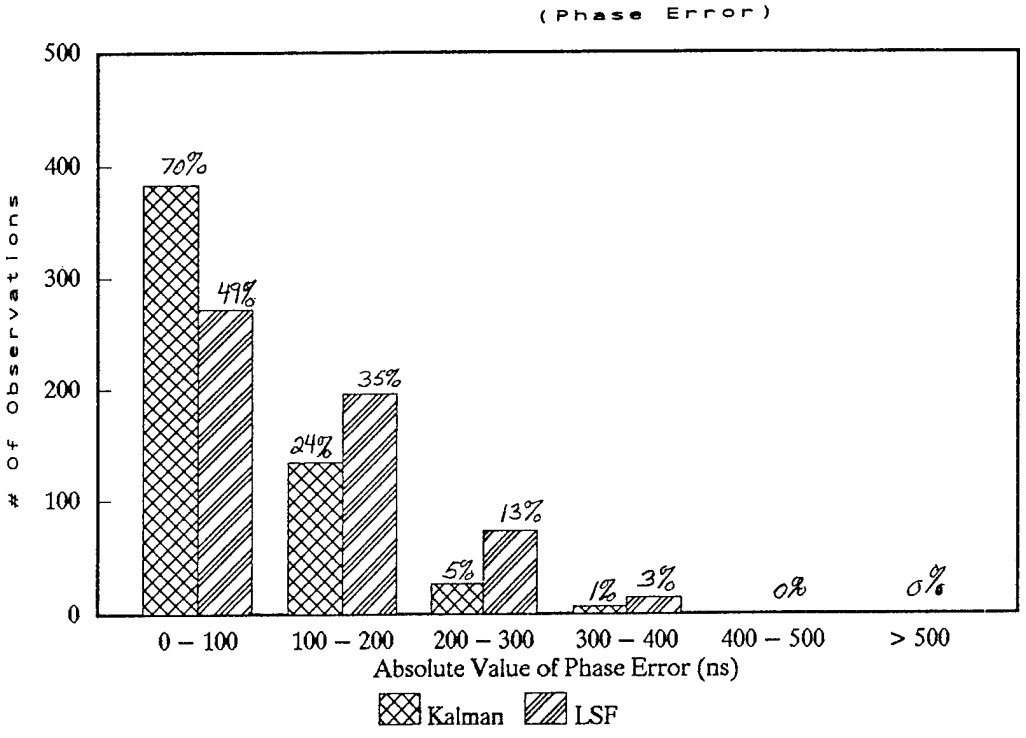


Figure 4. Simulation Results

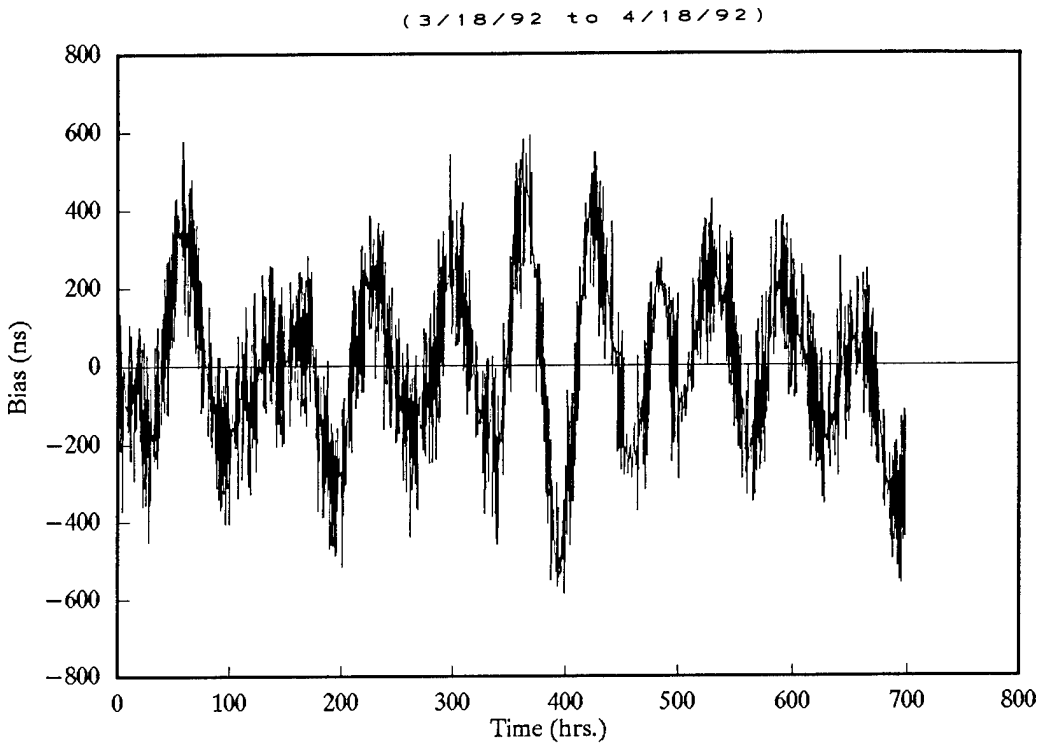


Figure 5. GPS Receiver Bias Data

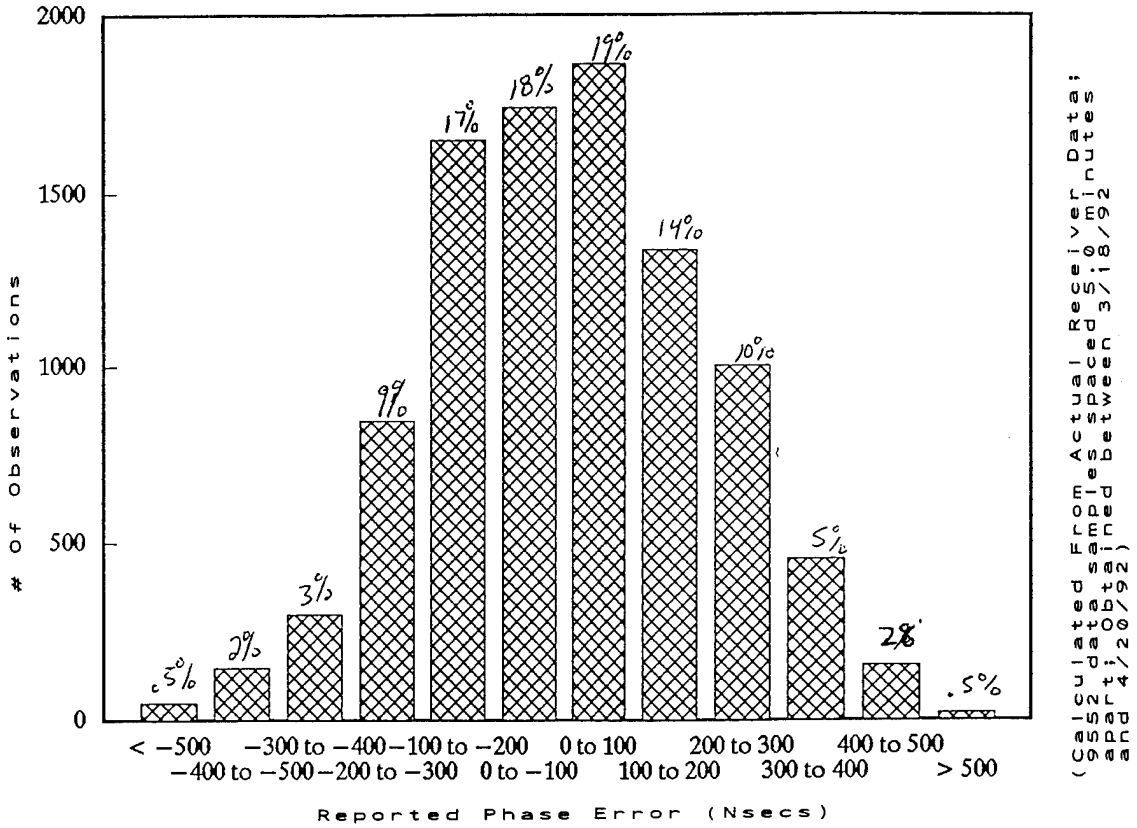


Figure 6. GPS Data Relative Frequency Distribution

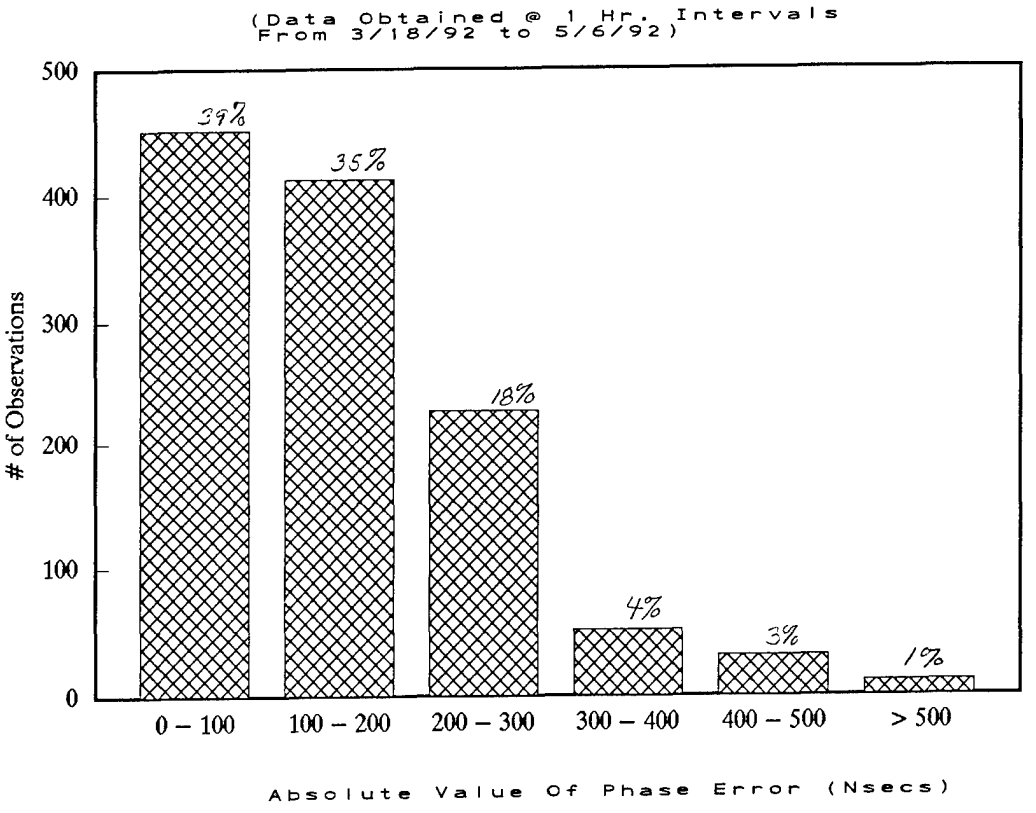


Figure 7. Reference Vs. UUT Phase Error

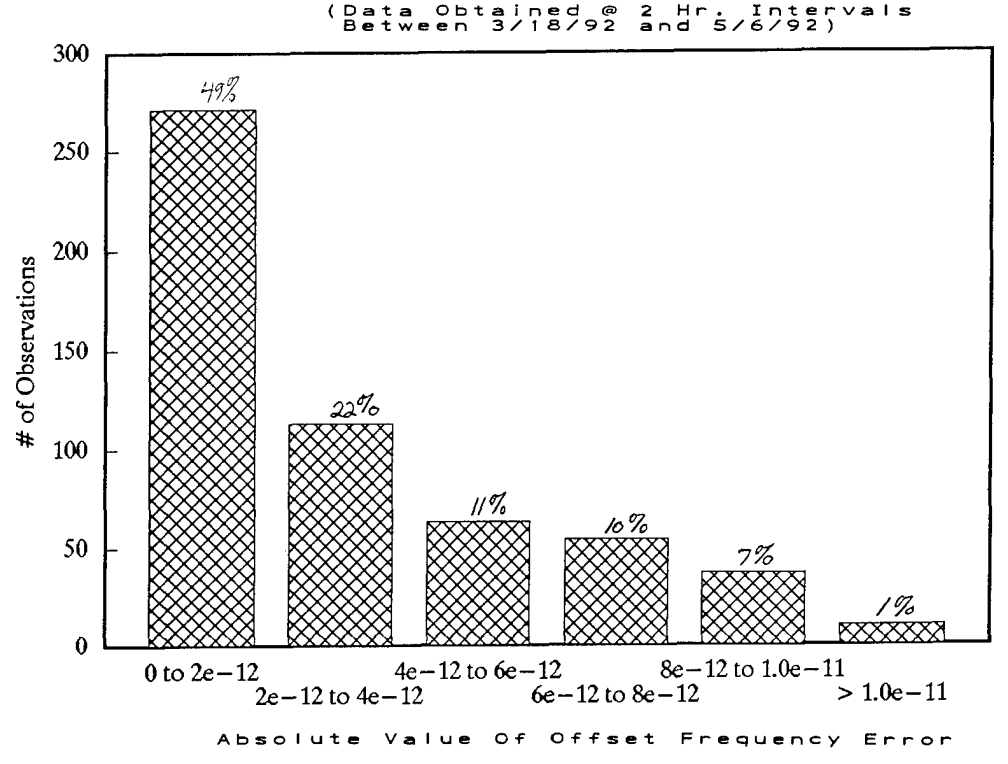


Figure 8. Reference Vs. UUT Offset Frequency

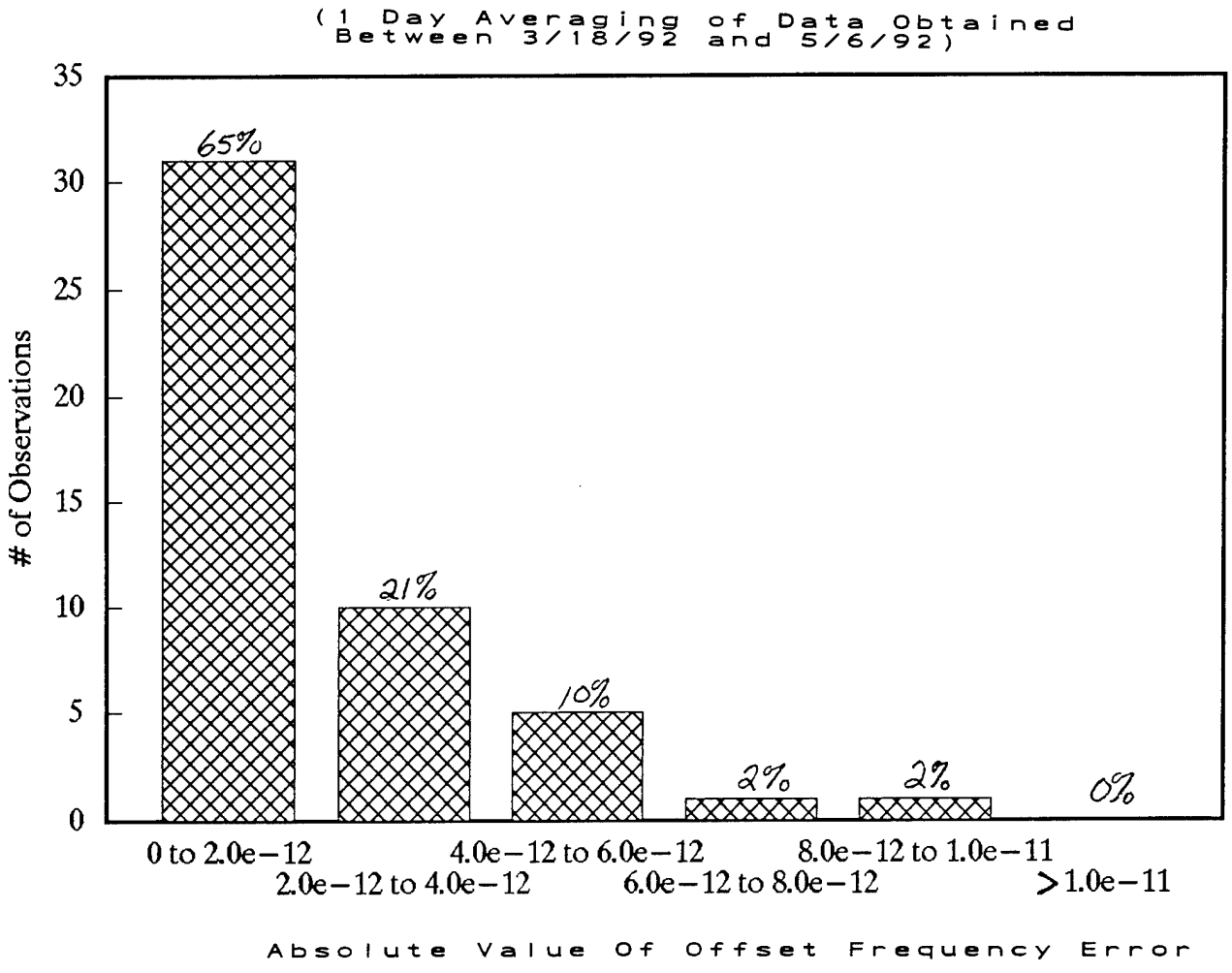


Figure 9. Reference Vs. UUT Offset Frequency

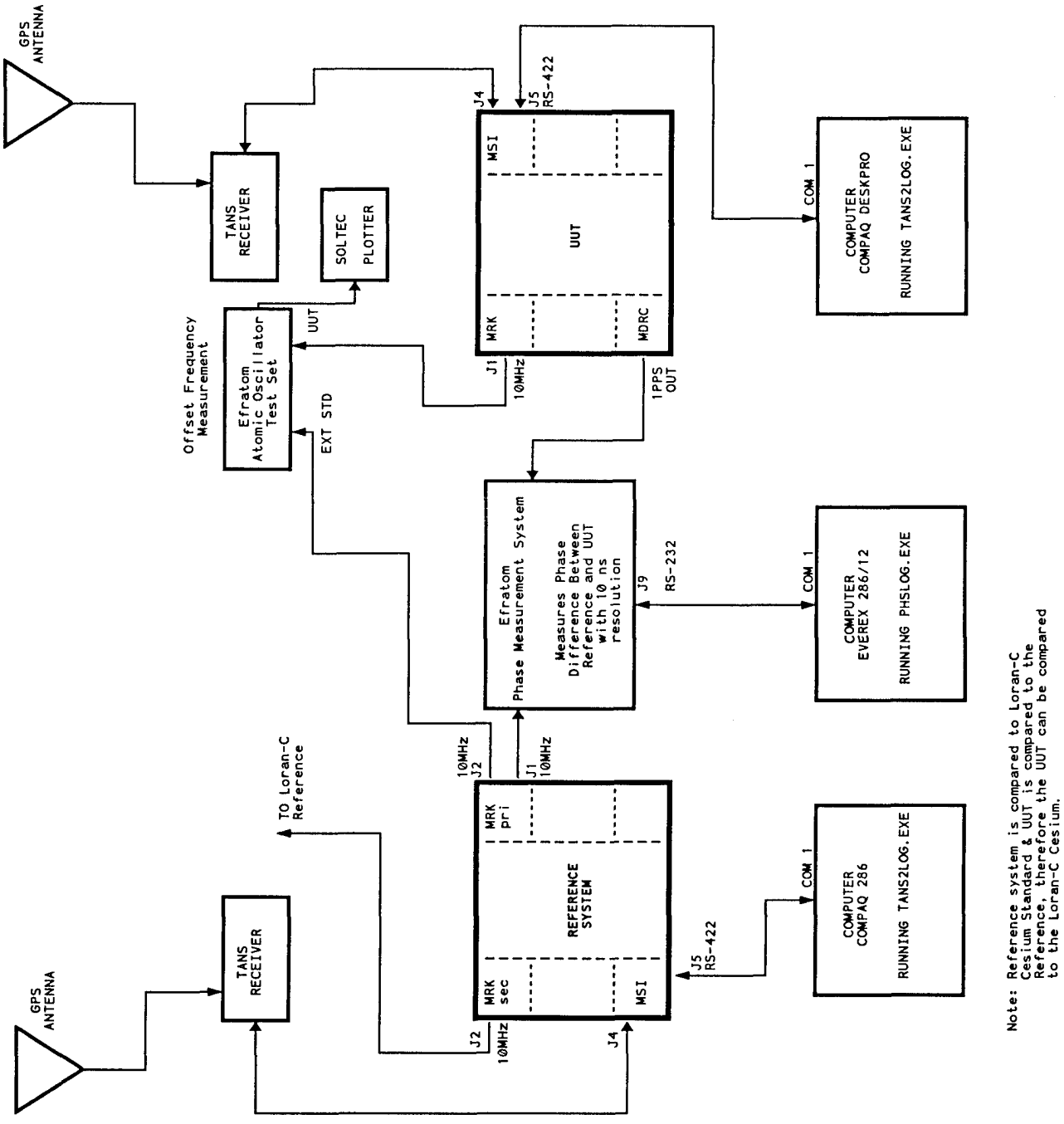


Figure 10. Test Setup

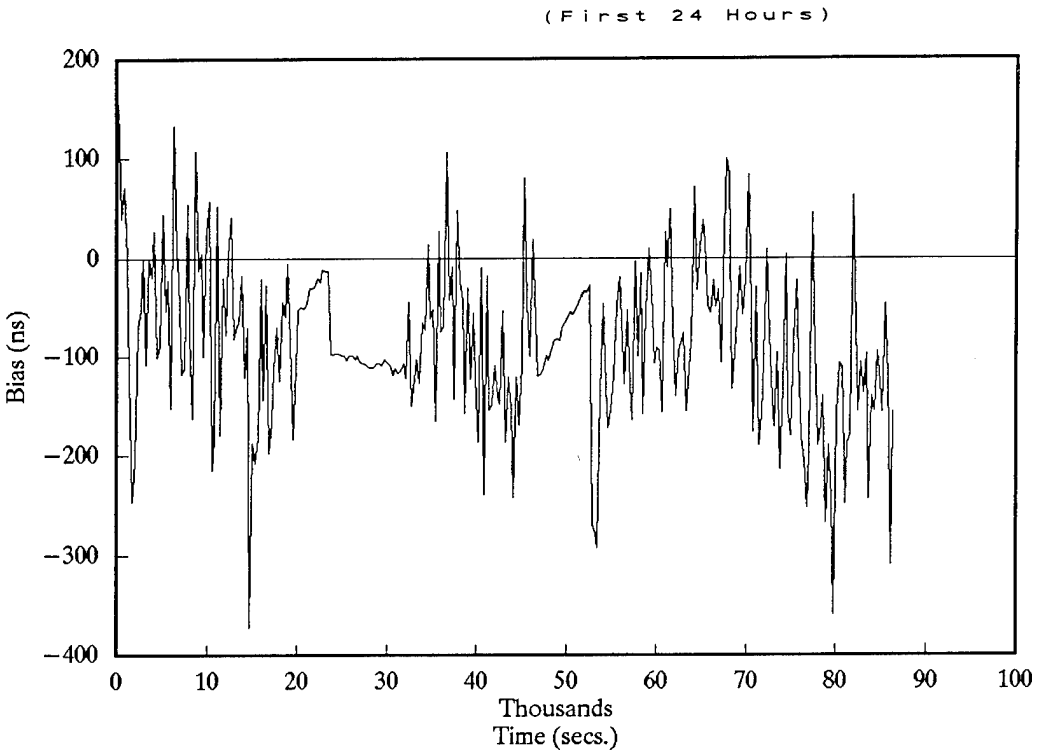


Figure 11. Receiver Bias (GPS Data) Vs. Time

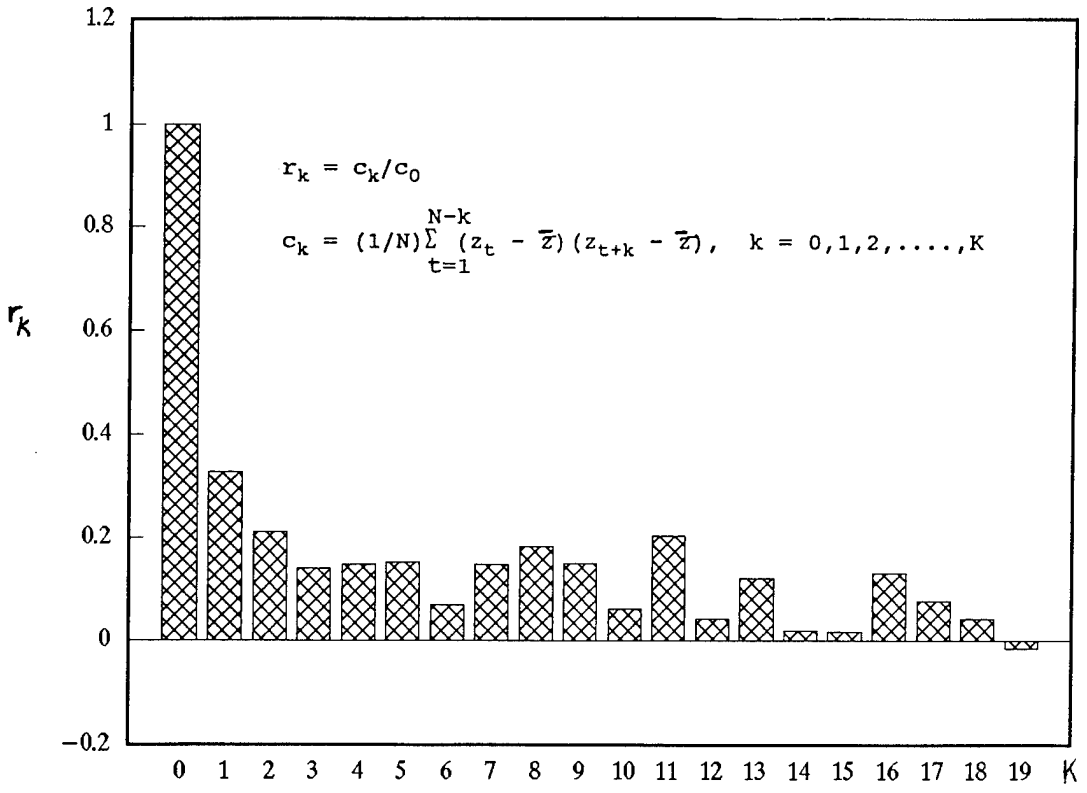


Figure 12. Estimated Autocorrelation Function of GPS Data (r_k is the estimate of p_k)

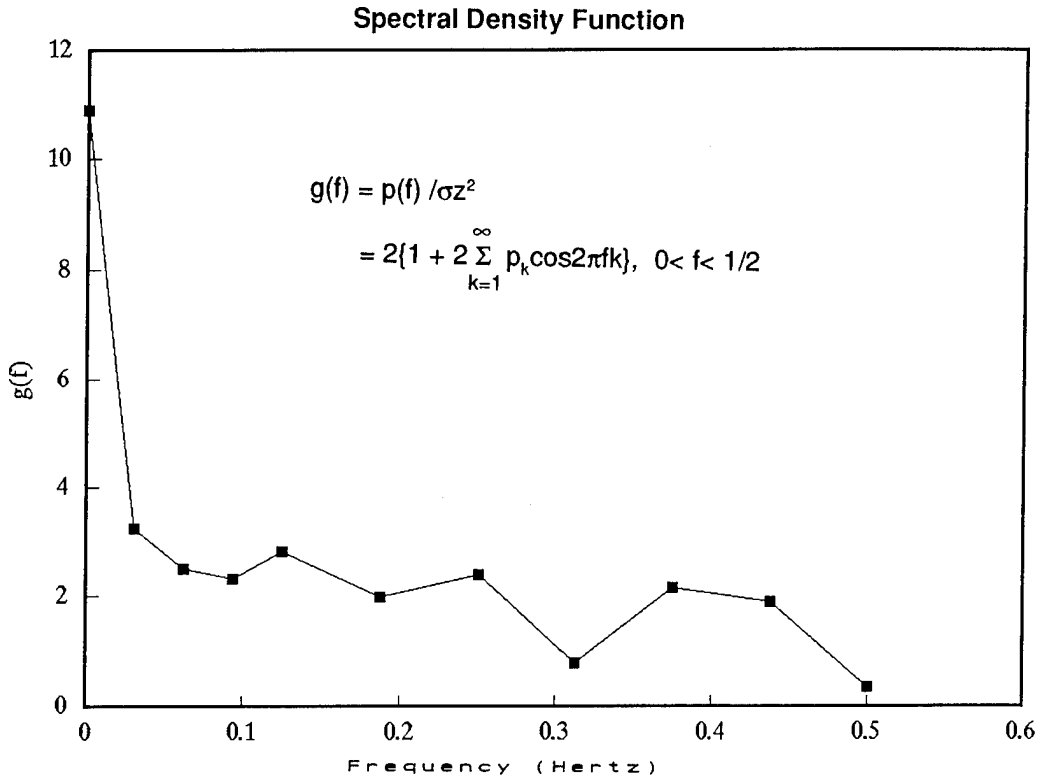


Figure 13. Spectral Density Function Of GPS Data

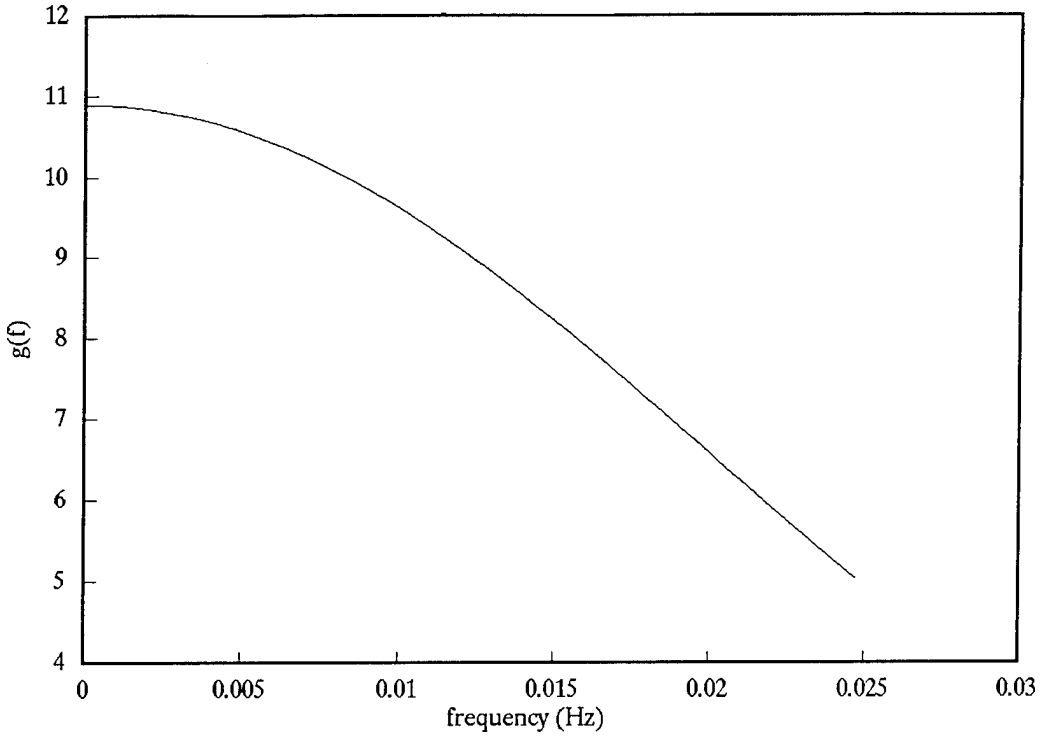


Figure 14. Spectral Density Function Of GPS Data (Looking Close In)

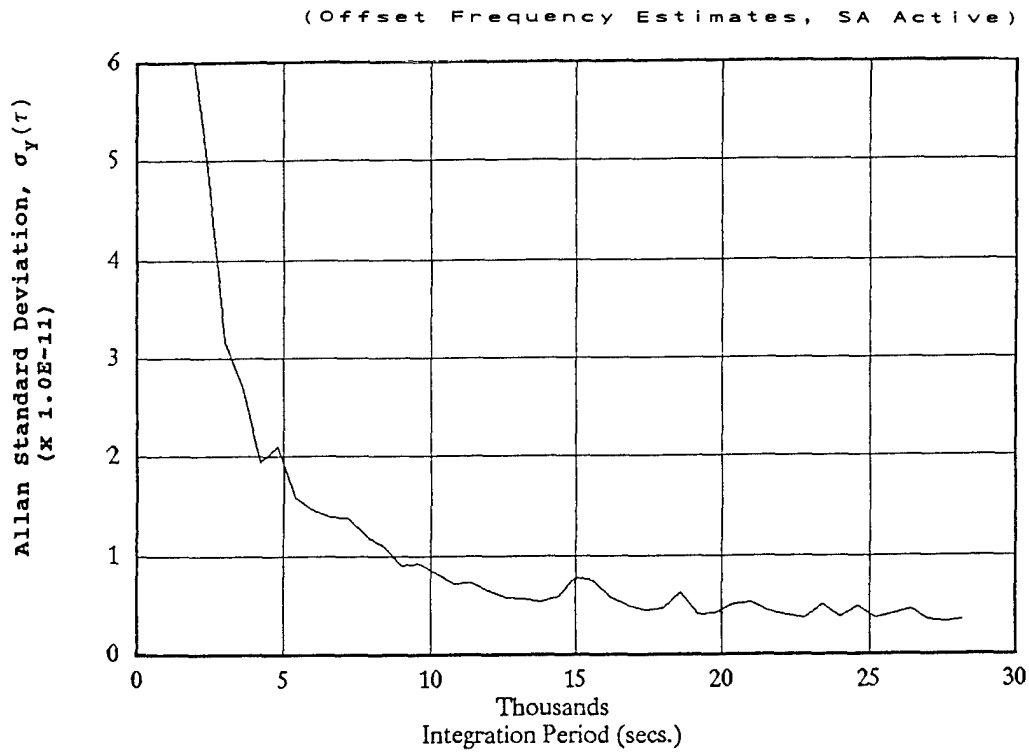


Figure 15. Fractional Frequency Stability Of GPS Data

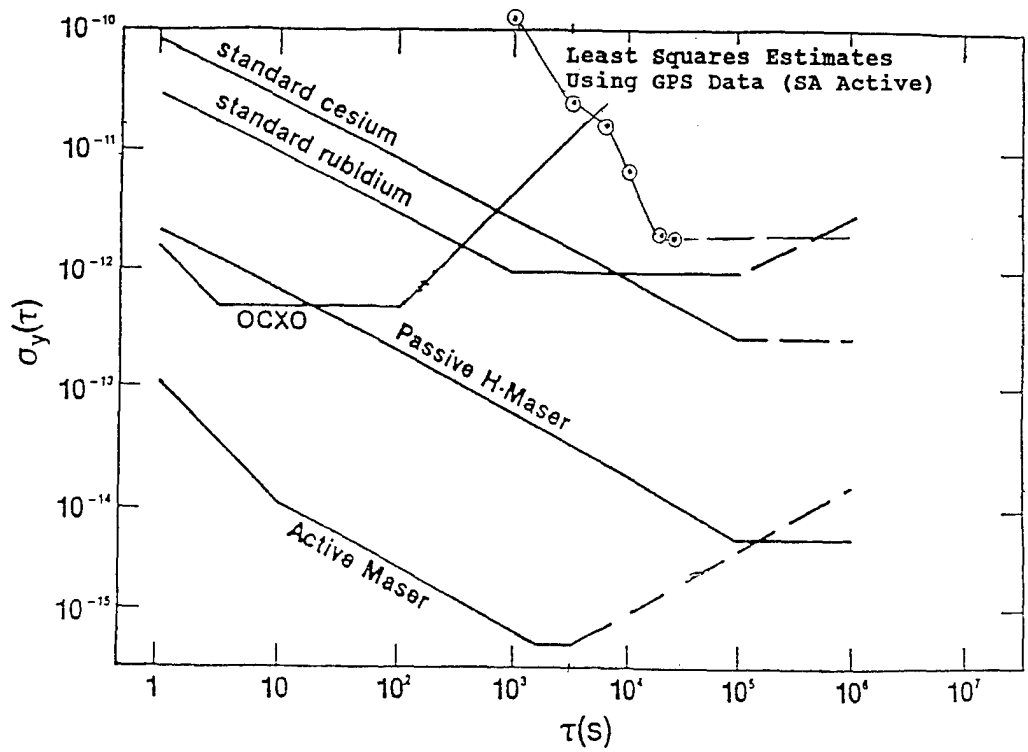


Figure 16. Fractional Frequency Stability Of Several Precision Commercial and Laboratory Frequency Standards

VII. Conclusion

In the above sections the problem of disciplining a rubidium oscillator to GPS in the presence of SA was discussed. Simulation and experimental results of an adaptive Kalman filter solution were presented and discussed. The results show that typical phase and offset frequency performance of 300 nanoseconds and 2.0E-12 (typically) respectively is achievable in the presence of SA.

Furthermore, a brief analysis of recently acquired GPS observation data was given. The analysis showed that the GPS noise originates from a stationary process. The variance of the process contains considerable spectral content distributed at low frequencies (below 1.25 millihertz). Therefore, one must use an integration period much greater than 400 seconds in order to begin filtering the effects of SA.

Further improvement in the test setup will be done by implementing a reference system utilizing a passive hydrogen maser frequency standard and phase data referenced to the National Institute of Standards and Technology (NIST). We shall also further analyze and utilize recently acquired data to tune the phase lock loop to see if performance improvements can be obtained.

Appendix A: Mathematical Formulations

A.1 Kalman Filter Assumptions and Equations

The system model with no control input is given by

(1) $x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k)$

and the measurement model is given by

(2) $z(k+1) = H(k+1)x(k+1) + v(k+1)$.

The following definitions apply:

x(k) is the state vector of the system (nx1)
$\Phi(k+1,k)$ is the state transition matrix (nxn)
w(k) is the excitation vector (px1)
$\Gamma(k+1,k)$ is the excitation transition matrix (nxp)
z(k+1) is the measurement vector (mx1)
H(k+1) is the measurement matrix (mxn)
v(k+1) is the measurement error vector (mx1)

The basic problem of estimation is to determine an estimate of the state vector x(k) at time k when noise corrupted observations z(1), z(2), z(3),.....z(j) are given. If k > j prediction is involved, if k = j filtering is involved, and if k < j interpolation is involved. Only filtering is considered here.

The optimal filtered estimate for a linear discrete system of the type represented by equations (1) and (2) is given by the relation

(3) $\hat{x}(k+1|k+1) = \Phi(k+1,k)\hat{x}(k|k) +$

$K(k+1)[z(k+1) - H(k+1)\Phi(k+1,k)\hat{x}(k|k]$

for k = 0,1,2,.... and

$\hat{x}(0|0) = 0$ and $P(0|0) = P(0) = 0$.

The following additional assumptions are made.

(4) $E\{w(k)\} = 0$ for all k = 0,1,2,.....
(5) $E\{w(j)w'(k)\} = Q(k)\delta_{jk}$ for all k,j = 0,1,2,....

where Q(k) is a positive semi-definite pxp matrix and δ_{jk} is the Kronecker delta matrix ($\delta_{jk} = 0$ for $j \neq k$ and $= 1$ for $j = k$). Note that w'(k) is the transpose of w(k) and w(k) is a gaussian white sequence. Q(k) is the process noise covariance matrix.

(6) $E\{v(k+1)\} = 0$ for all k = 0,1,2,...
(7) $E\{v(k+1)v'(k+1)\} = R(k+1)\delta_{jk}$ for all j,k = 0,1,2,....

where R(k+1) is a positive semi-definite mxm matrix.
R(k+1) is the measurement noise covariance matrix.

(8) $E\{v(j)w'(k)\} = 0$ for j = 1,2,...; k = 0,1,2,...

It is also assumed that x(0) is a gaussian random n vector with

(9) $E\{x(0)\} = 0$
(10) $E\{x(0)x'(0)\} = P(0)$

where P(0) is an nxn positive semi-definite covariance matrix.

(11) $E\{x(0)w'(k)\} = 0$ for all $k = 0,1,2,.....$
(12) $E\{x(0)v'(k+1)\} = 0$ for all $k = 0,1,2,...$

We also have

(13) $P(k+1|k) = \Phi(k+1,k)P(k|k)\Phi'(k+1,k) + \Gamma(k+1,k)Q(k)\Gamma'(k+1,k)$

$P(0|0) = 0.$

(14) $K(k+1) = P(k+1|k)H'(k+1)[H(k+1)P(k+1|k)H'(k+1) + R(k+1)]^{-1}$

(15) $P(k+1|k+1) = [I - K(k+1)H(k+1)]P(k+1|k)$

If a deterministic control input is added it must be accounted for by the optimal estimator. In order for the estimator to remain unbiased, the system model and state estimate must be changed as follows **[4]**:

(16) $x(k+1) = \Phi(k+1,k)x(k) + \Gamma(k+1,k)w(k) + \Lambda(k+1,k)u(k)$

(17) $\hat{x}(k+1|k+1) = \Phi(k+1,k)\hat{x}(k|k) + \Lambda(k+1,k)u(k)$

$+ K(k+1)[z(k+1) - H(k+1)\Phi(k+1,k)\hat{x}(k|k) - H(k+1)\Lambda(k+1,k)u(k)]$
for $k = 0,1,2,....$ and

$\hat{x}(0|0) = 0$ and $P(0|0) = P(0) = 0.$

where u(k) is the control input and $\Lambda(k+1,k)$ is a nxp vector that relates the control inputs to the state vector **[2]**.

The procedures for computing $P(k+1|k+1)$, $P(k+1|k)$ and $K(k+1)$ remain unchanged **[4]**.

A.2 Least Squares Prefilter

By definition the observations

(1) $z_i = x_i + v_i$

and the equation for the least squares line is

(2) $y_i(t_i) = a0 + a1*t_i.$

The least squares line yields a best fit to the observations in the mean squared sense. Therefore, we wish to minimize

(3) $S = (a0+a1*t_1 - x_1 - v_1)^2 + (a0+a1*t_2 - x_2 - v_2)^2..... + (a0+a1*t_N - x_N - v_N)^2.$

Taking the partial derivatives with respect to a0 and a1 we obtain

(4) $\delta S/\delta a0 = a0*N + a1*\Sigma t_i - \Sigma x_i - \Sigma v_i = 0$

and

(5) $\delta S/\delta a1 = a0*\Sigma t_i + a1*\Sigma t_i^2 - \Sigma(t_i*x_i) - \Sigma(t_i*v_i) = 0.$

Solving equations (4) and (5) for a0 and a1 yields:

(6) $a0 = \dfrac{\Sigma x_i \Sigma t_i^2 + \Sigma v_i \Sigma t_i^2 - \Sigma t_i \Sigma(t_i x_i) + \Sigma t_i \Sigma(t_i v_i)}{N\Sigma t_i^2 - (\Sigma t_i)^2}$

(7) $a1 = \dfrac{N[\Sigma(t_i x_i) + \Sigma(t_i v_i)] - \Sigma t_i \Sigma x_i - \Sigma t_i \Sigma v_i}{N\Sigma t_i^2 - (\Sigma t_i)^2}$

The phase at the end of the least squares line is

(8) $z_m(k) = a0 + a1*\Delta T$

where ΔT is the integration period used in the least squares filtering.

By definition we must have

(9) $v_m(k) = z_m(k) - x_m(k)$

Using equations (6), (7), and (8) in equation (9) we obtain

(10) $v_m(k) = \dfrac{\Sigma v_i \Sigma t_i^2 + \Sigma t_i \Sigma(t_i v_i) + N\Sigma(t_i v_i)\Delta T - \Sigma t_i \Sigma v_i \Delta T}{N\Sigma t_i^2 - (\Sigma t_i)^2}$

$+ \dfrac{\Sigma x_i \Sigma t_i^2 - \Sigma t_i \Sigma(t_i x_i) + N\Sigma(t_i x_i)\Delta T - \Sigma t_i \Sigma x_i \Delta T}{N\Sigma t_i^2 - (\Sigma t_i)^2}$

$- x_m(k)$

The first term on the RHS of equation (10) is smoothed observation noise. The second and third terms on the RHS of equation (10) represents the smoothing that the prefilter performs on the process state. In order to use a conventional Kalman filter the variance of the least squares filter end-of-line phase (z_m) must be determined.

A.3 One State Adaptive Kalman Filter With Prefilter

The rubidium oscillator phase as a function of time may be represented as

(1) $x(t) = x(0) + y(0)t + \frac{1}{2}Dt^2 + w(t)$ [8]

where $x(t)$ is the time deviation of the clock at time t
$x(0)$ is the synchronization error at $t = 0$
$y(0)$ is the synchronization error at $t = 0$, which produces a linear ramp in the time deviations
D is the frequency drift term
$w(t)$ contains all of the random fluctuations as a function of t.

Rewriting this equation for a discrete system we obtain

(2) $x(k) = x(0) + y(0)k\Delta T + \frac{1}{2}D(k\Delta T)^2 + w(k)$; $k = 1,2,...$

where k is the time index and ΔT is time between observations.

If we add a control input to the oscillator our equation becomes

(3) $x(k+1) = x(k) + [y(0) - u(k)]\Delta T + \frac{1}{2}D(k)\Delta T^2 + w(k)$.

where u(k) is the control update applied at time index k.

From Appendix A.2 equations (8) and (9) we obtain

(4) $z_m(k+1) = x_m(k+1) + v_m(k+1) = a0(k+1) + a1(k+1)^*\Delta T + v_m(k+1)$

where the time index has been added to a0 and a1 to show that they are functions of the time index and are both updated with each least squares measurement.

The drift of the MRK rubidium oscillator is 4.0E-11 per month which is a negligible contribution to the oscillator phase change in a two hour period. Therefore we conclude for the least squares line measurement that a1(k+1) is the measurement of [y(0) - u(k+1)] which is the oscillator offset frequency at the beginning of the integration period and a0(k+1) is the measurement of x(k) which is the oscillator phase at the beginning of the integration period.

Therefore, $z_m(k+1)$ is the measurement of the oscillator phase with the control input included. Thus we can use $z_m(k+1)$ in place of $z(k+1)$ in Kalman filter equation A.1 (3). However, we need to use $v_m(k)$ (the variance of the least squares line end-of-phase) instead of the variance of the observation data v(i).

Therefore our system and measurement equations become

(5) $x_m(k+1) = x_m(k) + [y(0) - u(k)]\Delta T + w(k)$

$= x_m(k) + a0(k+1) + a1(k+1)^*\Delta T + w(k)$

$k = 0,1,2,...$

(6) $z_m(k+1) = x_m(k+1) + v_m(k+1)$

$k = 0,1,2,...$

where drift is considered to be zero over the integration period.

The quantities in equations (5) and (6) are scalar quantities since we only have one state variable to estimate, namely the oscillator phase.

From (5) we obtain

$\Phi(k+1,k) = 1.0$
$\Gamma(k+1,k) = 1.0$

$H(k+1) = 1.0$

$R(k+1) = \sigma^2_{vm}(k+1)$

$Q(k) = \sigma^2_w$

Therefore our optimal filtered estimate is given by

(7) $\hat{x}(k+1|k+1) = \hat{x}(k|k) + K(k+1)[z_m(k+1) - \hat{x}(k|k]$

for $k = 0,1,2,....$ and

$\hat{x}(0|0) = 0.0$ and $P(0|0) = P(0) = 0.0$.

We also have

(8) $P(k+1|k) = P(k|k) + Q(k)$

(9) $K(k+1) = P(k+1|k)\,[P(k+1|k) + R(k+1)]^{-1}$

(10) $P(k+1|k+1) = [1 - K(k+1)]\,P(k+1|k)$

Since the control problem had to be solved prior to having sufficient data from which to obtain the variance of the least squares line end-of-phase and the Department of Defense (DOD) may vary SPS accuracies should a security threat exist to US/ NATO interests [1] we used a different approach for approximating the variance of $v_m(k)$.

For healthy Block 1 satellites the observation data viewed over a two hour period lie on a straight line whose width is about 10 nanoseconds (the time jitter on the observations). For healthy Block 2 satellites the observation data are similar to noisier portions of figure 11.0. The quieter portions of figure 11.0 are data from Block 1 satellites. We also notice random jumps (biases) when the receiver switches satellites.

Divergence is said to occur when the actual error in the estimate of the state becomes inconsistent with the error covariance predicted by the filter equations and essentially represents a breakdown in the data processing method [9]. The discounting of data in the filtering problem context is achieved by the appropriate selection of the noise covariances [9].

Therefore we decided to implement an adaptive filter based on the variance of the observations from the least squares line. This method causes the Kalman gain to decrease for noisy observations and to increase for relatively noiseless observations by respectively increasing or decreasing the observation noise covariance.

The method requires storing all observation data obtained during each integration period. At the end of each integration period the variance of the observation data from the least squares line is calculated and used as the variance of the least squares line end-of-line phase. Therefore

(11) $R(k+1) = [z(i) - (a0(k) + a1(k)*i)]^2/n$;
$i = 1,2,...n$

where n is the number of observations processed during each integration period.

Since we are using a PID phase lock loop (PLL) we shall make u(k) a function of the estimated oscillator phase at time index k. Therefore we define u(k+1) as

(12) $u(k+1) = K_p \hat{x}(k+1|k+1) + K_d[\hat{x}(k+1|k+1) - \hat{x}(k|k)] + K_i \sum_{k=0}^{M} \hat{x}(k+1|k+1)$

$k = 0,1,2,....$

where K_p is the proportional gain term

K_d is the differential gain term

K_i is the integral gain term

$\hat{x}(k|k)$ is the adaptive Kalman filtered estimate at time index k

M is the total number estimates made including the one at k+1

Appendix B: References

[1] NATO Navstar GPS Technical Support Group under the direction of the NATO Navstar GPS Project Steering Committee, "Technical Characteristics of the Navstar GPS," pp. A-13, June 1991.

[2] H.W. Sorenson, "Kalman Filtering Techniques," Advances in Control Systems Theory and Applications. Academic Press, 1966.

[3] R.E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering, March 1960.

[4] Arthur Gelb, Joseph Kasper, Raymond Nash, Charles Price, and Arthur Sutherland, "Applied Optimal Estimation," (Technical Staff, The Analytic sciences Corporation), pp. 130.

[5] George Box and Gwilym Jenkins, Time Series Analysis, Holden-Day 1976, pp. 24-30, pp. 32, 40-41, pp. 174-176.

[6] B.P. Lathi, An Introduction to Random Signals and Communication Theory, International Textbook Company, 1968, pp. 174.

[7] J. S. Meditch, Stochastic Optimal Linear Estimation Control, McGraw-Hill, 1969.

[8] David W. Allan, "Time and Frequency Characterization, Estimation and Prediction of Precision clocks and Oscillators," National Bureau of Standards, Boulder, Colorado.

[9] H. W. Sorenson and J. E. Sacks, "Recursive Fading Memory Filtering," Information Sciences, Vol. 3, pp. 101-119, Jan. 1971.

[10] Arthur Gelb, Joseph Kasper, Raymond Nash, Charles Price, Arthur Sutherland, "Applied Optimal Estimation," (Technical Staff, The Analytic sciences Corporation), pp. 291-292.

[11] Nasser E. Nahi, Estimation Theory and Applications, John Wiley & Sons, 1969.

[12] Edited by Harold W. Sorenson, Kalman Filtering: Theory and Application, IEEE Press, 1985.

[13] GPS World, pp. 53, September/October 1990.

[14] Lindon L. Lewis, "Performance of Commercial Frequency Standards", (Ball Aerospace Systems Div., Ball Corporation).

Cross-Correlation Phase Noise Measurements

Warren F. Walls

Femtosecond Systems
4120 Evans Drive
Boulder, CO 80303
(303) 431 - 7147

ABSTRACT

An accurate characterization of phase noise in amplifiers, synthesizers, quartz oscillators, and other electronic devices is becoming increasingly important. In particular, extremely low noise quartz oscillators, which are used in many applications including communications and navigational equipment, exceed the present test capabilities of many labs. The noise floor of a typical two channel phase noise measurement system is set by the phase detector (usually a double balanced mixer) and amplifiers that follow the detector. If the phase noise of two oscillators is measured using two of these test systems, the output noise of the equipment is uncorrelated except for the component that is due to the phase noise between the oscillators. Examination of a system which used matched power splitters, two phase detectors, and post amplifiers revealed that an exceptionally low noise floor is achievable. Using a 5 MHz carrier, we achieved better than $\mathcal{L}$(1 Hz) = -160 dBc/Hz and $\mathcal{L}$(10 kHz) = -190 dBc/Hz. The time required to reach a given percentage of precision in the measurement, at these low noise floors, is lengthened due to the need to average away the uncorrelated noise. Details of setup including calibration procedures and confidence intervals for the measurement are discussed.

INTRODUCTION

The typical method used to measure phase noise involves a single mixer that is driven by sources on its RF and LO ports. The IF port is then low pass filtered and amplified before it is connected to a spectrum analyzer or FFT (figure 1). The noise floor of this technique is highly dependent upon the performance of the mixer, filter, and amplifiers[1],[2]. The cross-correlation measurement technique uses two duplicate systems to drive a cross-correlation analyzer. The oscillator signals are split with inductive splitters to provide a total of 2 pairs of input signals. These signals are connected to the LO and RF ports of both double balanced mixers. The output of each mixer has its own low pass filter followed by a set of amplifiers. The output of both channels are then fed into the cross-correlation FFT analyzer (figure 2). The random noise contributed by the individual mixer-filter-amplifier strings is uncorrelated. The correlated noise is the actual noise of the input oscillators plus some leakage due to the imperfect isolation characteristics of the power splitters. The cross-correlation measurement technique increases the amount of time that is required to perform a particular measurement, but the noise floor is lowered due to the cancellation of the incoherent noise of the individual detector systems.

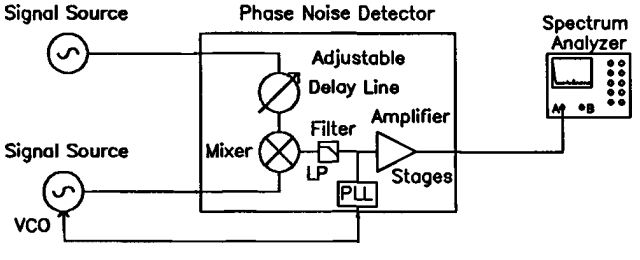


Figure 1 Single Channel Phase Noise Test Set

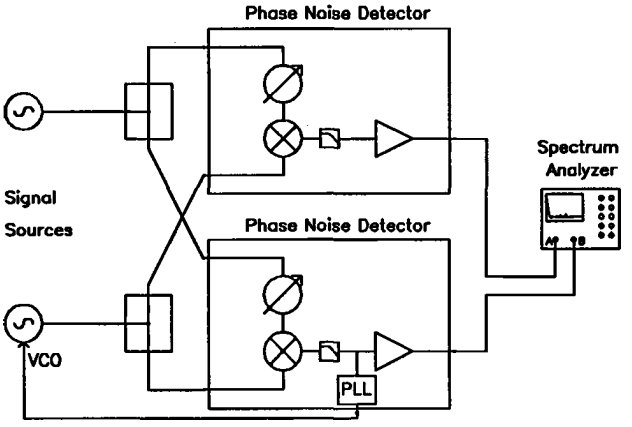


Figure 2 Cross-Correlation Phase Noise Test Set

0-7803-0476-4/92 $3.00 © 1992 IEEE

EQUIPMENT USED

A variety of off-the-shelf equipment was collected to perform the tests discussed in this paper. Two low noise oscillators, one of which had a voltage tune port, were used to drive the mixers for all of the tests. Three power splitters from Mini-Circuits (ZFSC2-4) were used to split the oscillator signals. The system was calibrated (K_d measured) using a typical analog Tektronics oscilloscope. The actual noise measurements were made on an HP cross-correlation FFT analyzer (35665A) and hard copies were made using a HP plotter. Finally, some care was used in selecting the particular cables and connectors appropriate for low noise measurements.

CALIBRATION MODE

In order to calibrate the system, one needs to measure the phase sensitivity of the mixer. This is accomplished by connecting two individual signal sources, which have been split into two pairs, to each of the mixers (figure 3).

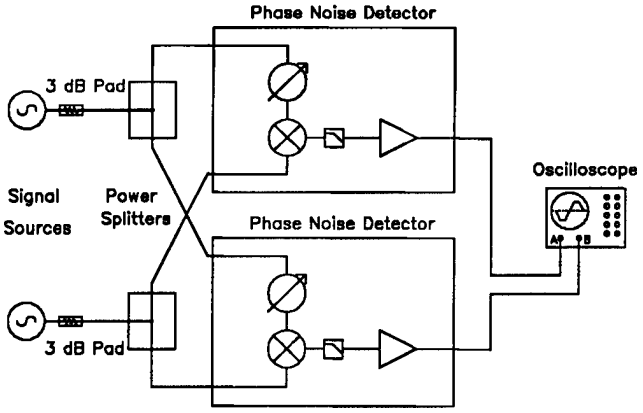


Figure 3 Calibrating the System

The difference in frequency between the two signals produces a "beat note" or "beat frequency". The beat frequency is viewed on the oscilloscope and the horizontal change in radians is measured for a specific voltage change. The ratio in volts/radians is the value of K_d[3].

$$K_d = \frac{\Delta V}{\Delta T} \frac{Period}{2\Pi}$$

MEASURING THE NOISE FLOOR

The signals at the mixers need to be in quadrature when the actual noise measurements are made. This is accomplished by using either a phase locked loop or an adjustable delay line. In order to measure active devices like oscillators and synthesizers, the phase locked loop needs to be used. Measuring the noise floor of the equipment as well as the noise in passive components like amplifiers, requires the use of the adjustable delay line. The sensitivity and thus the calibration of the system is effected by the drive level and length of cable. Therefore the actual measurement needs to be set up and changes minimized when moving from calibration mode to measurement mode. When the delay line is to be used, the system needs to first be set up with the proper amount of delay selected in order to perform the test with one source. After this is achieved, a second source, with an identical drive level and impedance, is inserted in order to produce the desired beat frequency required for calibration (figure 3). Once K_d is determined, the second signal is removed and the original oscillator is reconnected using the same cables. The output of each mixer-filter-amplifier string is connected to the cross correlation FFT analyzer (figure 4). By looking at the power spectral density of each individual channel, one can view the phase noise of the individual detector's noise

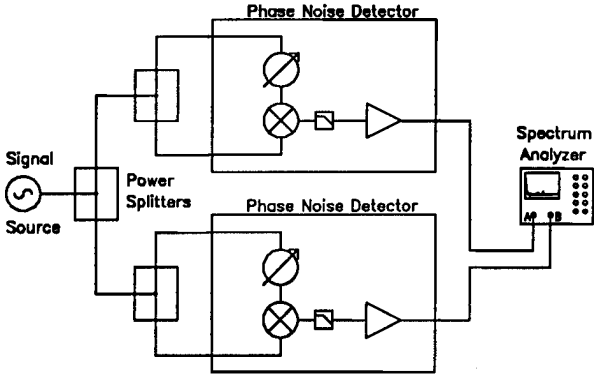
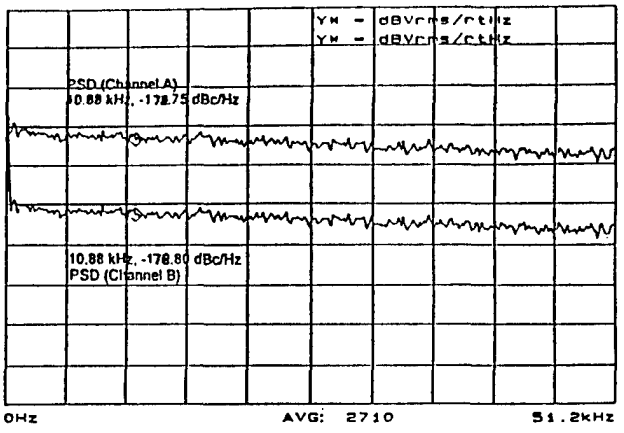


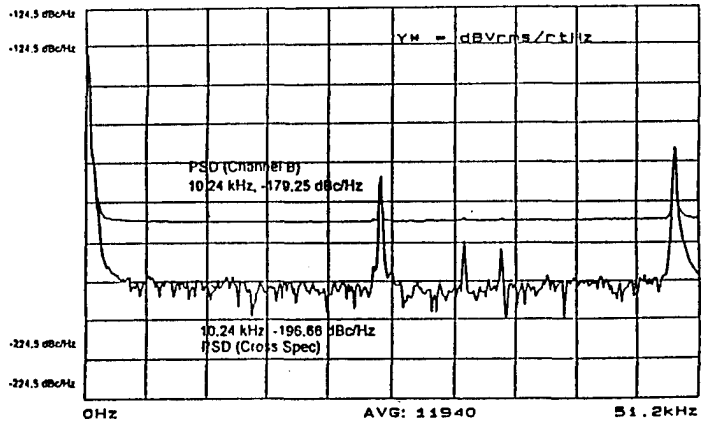
Figure 4 Phase Noise Floor Measurement

floor. This provides a way to double check the system's functionality. After the individual noise floors are verified, the cross correlation power spectral density can be viewed to see the improvement in noise floor (graph 1 versus graph 2 & 3)[4],[5].

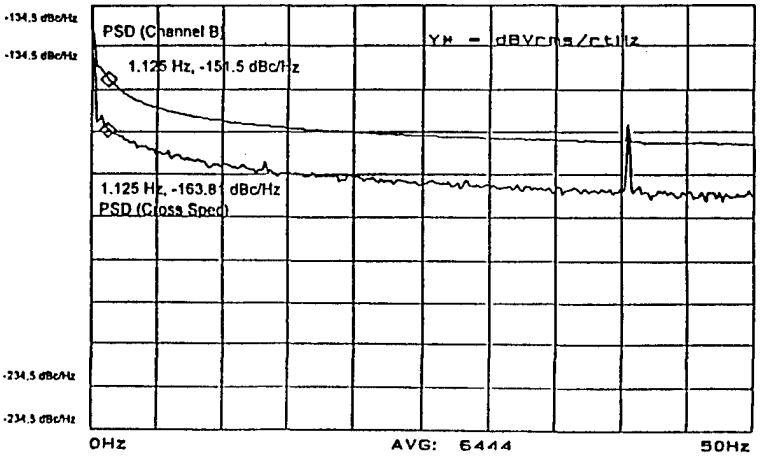
Graph #1 Single Mixer Noise Floor



Graph #2 Cross-Correlation Floor (50kHz)



Graph #3 Cross-Correlation Floor (50Hz)



MEASURING OTHER DEVICES

In order to measure other active devices, the phase locked loop is required. The adjustable delay lines are turned off and the two signals are split and connected to each of the two mixer input ports (figure 3). The difference frequency is then used, as described above, to calibrate the system, resulting in the determination of K_d. Once this is accomplished, the phase locked loop is activated and the signals sources are brought into quadrature at the mixer (figure 2). The output of the two mixer-filter-amplifier strings can then be measured by the cross-correlation analyzer. The phase noise value reported at a particular frequency from the carrier is expressed as $S_\phi(f)$ as shown below in equation 2 [4],[5].

$$S_\phi = \left(\frac{V_n^2(f)}{BW}\right)\left(\frac{1}{K_d^2(f)}\right)\left(\frac{1}{Gain^2(f)}\right)$$

Equation 2 Parameters Required to Report $S_\phi(f)$

MEASUREMENT UNCERTAINTY

The improved noise floor of the cross-correlation phase noise measurement technique does not come without a price. Many more samples are required in order to average out the uncorrelated noise. The confidence interval of a single channel phase noise detector is[6]:

$$S_\phi^{\,s}(f) = S_\phi^{\,m}(f)\left(1 \pm \frac{1}{\sqrt{n}}\right)$$

Equation 3 Confidence Interval for Single Channel

The confidence interval of a dual channel (cross-correlation measurement system) is:

$$S_\phi^{\,x(f)} = S_\phi^{\,m}(f)\left(1 \pm \frac{2S_\phi^{\,s}}{\sqrt{n}}\right)$$

$x = \text{cross}-\text{correlation}$

$m = \text{measured(noise)}$

$s = \text{single channel}$

$n = \text{number of samples}$

Equation 4 Confidence Interval for Dual Channel

Equation 3 shows that for a single channel the confidence interval is ±10% for 100 samples. Equation 4 shows that to obtain the same confidence interval for a phase noise measurement 10 dB below the single channel noise floor 20,000 samples are required.

CONCLUSION

The dual channel or cross-correlation method of phase noise results in a lower floor than the standard single channel method. This does, however, come a cost of measurement speed. A great deal more averages are required to achieve the same level of confidence in a measurement. The lower noise floor that is now achievable using the cross-correlation method provides a level of characterization of extremely good oscillators that was not available using the single channel method.

REFERENCES

[1] D.A. Howe, D.W. Allan, and J.A. Barnes, "Properties of Signal Sources and Measurement Methods," *Proc. of 35th Annual Symposium on Frequency Control*, Philadelphia, PA, May 27-29, 1981, pp. A-1(669)-A47.

[2] F.L. Walls, C.M. Felton, A.J.D. Clements, and T.D. Martin, "Accuracy Model for Phase Noise Measurements," *Proc. of 21st Annual Precise Time and Time Interval Planning Meeting*, Redondo Beach, CA, Nov. 30-Dec. 1, 1990, pp. 295-310; F.L. Walls, A.J.D. Clements, C.M. Felton, and T.D. Martin, "Precision Phase Noise Metrology," *Proc. of National Conference of Standards Laboratories*, Aug. 1991, pp. 257-275.

[3] "Characterization of Clocks and Oscillators," Eds. D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls, NIST Tech Note 1337, 1990.

[4] 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.F.C. Vessot, and G.M.R. Winkler, "Characterization of Frequency Stability," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, pp. 105-120, 1971.

[5] D.W. Allan, H. Hellwig, P. Kartaschoff, J. Vanier, J. Vig, G.M.R. Winkler, and N. Yannoni, "Standard Terminology for

Fundamental Frequency and Time Metrology," *Proc. of 42nd Annual Symposium on Frequency Control*, Baltimore, MD, June 1-4, 1988, pp. 419-425.

[6] F.L. Walls, D.B. Percival, and W.R. Irelan, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," *Proc. of 43rd Annual Symposium on Frequency Control*, Denver, CO, May 31-June 2, 1989, pp. 336-341.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A SHORTCUT FOR COMPUTING THE MODIFIED ALLAN VARIANCE

Charles A. Greenhall

Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr., 298-100
Pasadena, California 91109

Abstract

The computation of modified Allan variance directly from its definition is much more expensive than the computation of conventional Allan variance. A simple observation shows that a program for computing the conventional Allan variance from second differences of phase residuals can easily be embellished to compute the modified Allan variance with only a small increase in code, memory, and time.

Background

The conventional Allan variance, $\sigma_y^2(\tau)$, plays two roles in the study of frequency sources and distribution equipment. First, for a given averaging time τ, it gives a performance measure, a measure of frequency stability for this averaging time. Second, the log-log "sigma-tau" curve, $\log \sigma_y(\tau)$ vs. $\log \tau$, serves to characterize the process of residual phase $\varphi(t)$ or residual time $x(t)$. The shape of the curve often allows the phase process to be identified as a linear combination of independent processes with power-law spectra $S_\varphi(f) = \text{const} \cdot f^{\alpha-2}$. For most processes used as models for residual phase, in particular for power-law processes with $\alpha > -3$, the Allan variance exists, and stable estimates of $\sigma_y^2(\tau)$ can be obtained if the duration of the measurement is much larger than τ.

Two defects of the conventional Allan variance have been recognized. First, in the presence of white phase noise, $\sigma_y^2(\tau)$ depends strongly on the "hardware bandwidth" f_h, the frequency above which the phase noise is attenuated by the measurement system. Second, the sigma-tau curve does a poor job of distinguishing white phase noise, $S_\varphi(f) = \text{const}$, from flicker phase noise, $S_\varphi(f) = \text{const} \cdot f^{-1}$. The Allan variances of these two processes differ by a ratio approximately equal to $\text{const} + \log(2\pi f_h \tau)$, which varies too slowly for easy detection.

The modified Allan variance, $\text{mod}\ \sigma_y^2(\tau)$, was introduced by Allan and Barnes [1]. Within certain limitations, this stability measure lacks the above defects while retaining the range of applicability of the conventional Allan variance. Both variances are obtained by averaging the squares of the outputs of certain discrete-time filters acting on the residual time $x(t)$, as given in the next section. To make the discussion more precise, let us assume that $x(t)$ is sampled with period τ_0. Let $\tau = n\tau_0$, where n is a positive integer. Because $\text{mod}\ \sigma_y^2$ depends in an essential way both on n and on τ_0, we shall write $\text{mod}\ \sigma_y^2(n,\tau_0)$ instead of the more usual $\text{mod}\ \sigma_y^2(n\tau_0)$.

The frequency responses of the filters for $\sigma_y^2(n\tau_0)$ and $\text{mod}\ \sigma_y^2(n,\tau_0)$ (acting on phase or time, not frequency) are given by

$$\frac{8}{(n\tau_0)^2} \sin^4(\pi f n \tau_0) \quad \text{(conventional)},$$

and

$$\frac{8}{(n\tau_0)^2} \frac{\sin^6(\pi f n \tau_0)}{n^2 \sin^2(\pi f \tau_0)} \quad \text{(modified)}.$$

In Fig. 1, these responses are plotted on a decibel scale for $n = 10$. The factor $8/(n\tau_0)^2$ has been removed. The Nyquist frequency $1/(2\tau_0)$ is in the middle of the frequency axis. This picture should be regarded as repeating indefinitely along the frequency axis with period $1/\tau_0$. With this understanding, the theoretical Allan variances (in the absence of linear frequency drift) are obtained by integrating these responses against the spectral density $S_x(f)$ of the residual time $x(t)$ as a continuous-time process.

0-7803-0476-4/92 $3.00 © 1992 IEEE

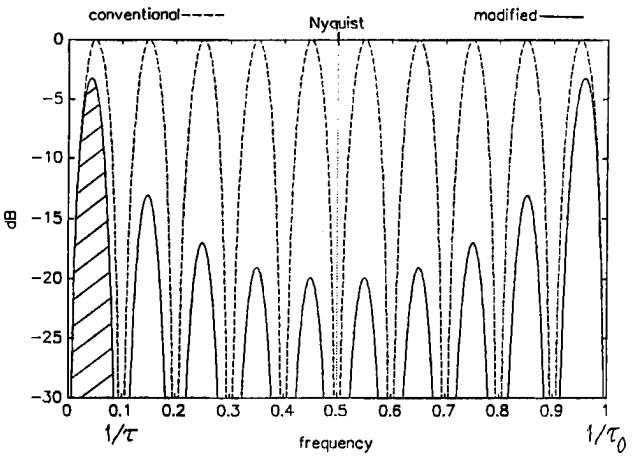


Figure 1. Frequency response of the filters that define conventional and modified Allan variance for $\tau = 10\tau_0$.

Two observations from Fig. 1 will be made. First, because the modified response lies below the conventional response, the theoretical variances always satisfy

$$\text{mod } \sigma_y^2(n,\tau_0) \le \sigma_y^2(n\tau_0).$$

They are equal if $n = 1$. Because of statistical fluctuations, the inequality is sometimes violated by estimates of the variances. Second, the main lobe of the modified response (hatched in the figure) has about 85 percent of the area under the response function between 0 and the Nyquist frequency. Therefore, if the time residual samples x_i satisfy the Nyquist criterion for their sampling rate, so that most of the power in $S_x(f)$ is below frequency $1/(2\tau_0)$, then the modified filter acts roughly as a bandpass filter for a passband centered near $1/(2\tau)$. This explains why mod $\sigma_y^2(n,\tau_0)$ is roughly independent of the hardware bandwidth f_h in the range $1/(n\tau_0) < f_h < 1/(2\tau_0)$, and also explains the ability of the mod sigma-tau curve to distinguish the white and flicker phase processes. On the other hand, if $f_h\tau_0$ is large, as is often the case, then the situation is more complex because several images of the main lobe come into play. See references [1] - [3] for details.

Allan Variance Filters

Let the residual time samples be $x_i = x(i\tau_0)$, $i = 0, 1, \ldots$ For $\tau = n\tau_0$, n a positive integer, the unscaled outputs of the conventional and modified Allan variance filters can be written as

$$z_i = x_i - 2x_{i+n} + x_{i+2n} = \Delta_n^2 x_{i+2n}$$

(the second backward difference of x_{i+2n} with stride n), and

$$Z_i = \frac{1}{n}[x_i + \ldots + x_{i+n-1} - 2(x_{i+n} + \ldots + x_{i+2n-1}) + x_{i+2n} + \ldots + x_{i+3n-1}],$$

respectively. Estimates of $\sigma_y^2(n\tau_0)$ and mod $\sigma_y^2(n,\tau_0)$ are computed by averaging the squares of these quantities over i and dividing by $2\tau^2$. The most common increments for averaging over i are 1 (fully overlapped estimator) and n.

The Shortcut

Because of the summations in the Z_i expression, it appears at first that an estimate for mod $\sigma_y^2(n,\tau_0)$ is about n times as expensive to compute as the corresponding estimate of $\sigma_y^2(n\tau_0)$. We now show that the computation for mod $\sigma_y^2(n,\tau_0)$ is less than twice as expensive as that for $\sigma_y^2(n\tau_0)$, and requires little modification of software. Define the cumulative sum w_i by

$$w_0 = 0, \quad w_i = \sum_{j=0}^{i-1} x_j, \quad i = 1, 2, \ldots$$

In terms of w_i, the sums in the expression for Z_i telescope as follows:

$$nZ_i = w_{i+n} - w_i - 2(w_{i+2n} - w_{i+n}) + w_{i+3n} - w_{i+2n}$$

$$= w_{i+3n} - 3w_{i+2n} + 3w_{i+n} - w_i$$

$$= \Delta_n^3 w_{i+3n},$$

the third difference of w_{i+3n} with stride n. This means that the only extra operations needed for a mod σ_y^2 computation beyond a σ_y^2 computation are the cumulative sum and one more difference. An existing computer program that computes estimates of σ_y^2 by manipulating second differences of x_i needs only to be modified so that it performs the same manipulations on third differences of w_i.

For example, a program for estimating σ_y^2 from x_i in a file may read the data sequentially, accumulating τ-overlap moments of second differences for various n as it goes along, while building a subsampled array of x_i in memory. After all the data have been read, estimates of $\sigma_y^2(\tau)$ for large τ can be computed from second differences of the x values in memory. To obtain a description of a mod σ_y^2 program, insert the operation $w_{i+1} = w_i + x_i$ after the read of x_i, then replace x by w and second difference by third difference in the above description. Note that the memory array of w must be obtained by subsampling the w_i, not by taking cumulative sums of a subsampled x array.

In another example, a simple program for estimating conventional Allan variance from an x array in memory was converted to the equivalent program for modified Allan variance. The computational cost, measured in number of floating-point operations, increased by 38 percent.

Connections with Previous Work

The idea of using the cumulative sum of phase residuals to make an alternate frequency stability measure was proposed by Groslambert and Gagnepain [4]; their "filtered Allan variance" is essentially proportional to the second moment of the second difference of w_i. It is remarkable that one more difference operation gives the modified Allan variance, which can be defined in terms of the third structure function of the discrete-time process w_i with sample period τ_0 (Lindsey and Chie [5]), namely,

$$\text{mod } \sigma_y^2(n,\tau_0) = \frac{1}{2n^4\tau_0^2}\,\mathcal{E}[(\Delta_n^3 w_i)^2],$$

where $\mathcal{E}$ is the operation of mathematical expectation.

D. Allan [6] has suggested another efficient method for computing the fully overlapped estimator of mod $\sigma_y^2(n,\tau_0)$. If one has computed Z_i as given above, then it is easy to compute Z_{i+1}. A direct calculation gives

$$nZ_{i+1} = nZ_i + x_{i+3n} - 3x_{i+2n} + 3x_{i+n} - x_i$$

$$= nZ_i + \Delta_n^3 x_{i+3n}.$$

After computing nZ_0 directly from the first $3n$ values of x, one can compute Z_1, Z_2, ... successively using only four values of x for each. This method has about the same efficiency as the proposed method that uses third differences of w. The proposed method has the disadvantage of requiring extra storage for the w array. On the other hand, the transformation from x to w is reversible. The w array can be generated on top of the x array, which, if needed later, can be recovered in place as the first difference of w. Moreover, once the w array is obtained, it can be used to compute Z_i for any n and with any i increment. The Allan method applies only to the fully overlapped estimator, whose i increment is one.

Conclusion

A program for estimating the conventional Allan variance of a set of phase residuals can be converted into a program for estimating the modified Allan variance of the same data. No extensive reorganization of code is required, and the increase in computational cost is modest.

Acknowledgments

The author thanks W. J. Riley for performance studies of these algorithms.

This work was performed at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] D. W. Allan, J. A. Barnes, "A modified 'Allan variance' with increased oscillator characterization ability", Proc. 35th Ann. Freq. Control Symp., pp. 470-475, Philadelphia, 1981.

[2] P. Lesage, T. Ayi, "Characterization of frequency stability: analysis of the modified Allan variance and properties of its estimate", IEEE Trans. Instrum. Meas., vol. IM-33, pp. 332-336, 1984.

[3] L. G. Bernier, "Theoretical analysis of the modified Allan variance", Proc. 41st Ann. Freq. Control Symp., pp. 116-121, Philadelphia, 1987.

[4] 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.

[5] W. C. Lindsey, C. M. Chie, "Theory of oscillator instability based upon structure functions", Proc. IEEE, vol. 64, pp. 1652-1666, 1976.

[6] D. W. Allan, "Time and frequency metrology: current status and future considerations", Proc. 5th European Frequency and Time Forum, pp. 1-9, Besançon, 1991.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

NOISE INDUCED TIME INTERVAL MEASUREMENT BIASES

E. Rubiola, A. Del Casale, A. De Marchi
Politecnico di Torino, Dipartimento di Elettronica
c.so Duca degli Abruzzi n. 24, I-10129 Torino, Italy

Abstract

Time jitter in the output fronts of hysteretic triggers is not the only effect of additive noise. An average leading time bias error can in fact be observed when large bandwidth noise is added to a poor slope signal. In this paper experimental results are presented, which confirm this statement.

1 Introduction

The resolution of precision time interval measurements is known to be often limited by noise induced jitter of switching fronts at the output of a hysteretic trigger. Such devices are necessary, when turning an analog signal into a usable digital one, in order to avoid multiple switching at the same "zero" crossing, in the presence of noise. The resulting jitter is expected, and is usually estimated from signal slope and S/N measurements.

In this paper such time jitter is shown to be also a possible limitation to the accuracy of high precision time interval measurements, as the average switching time turns out to be always *before* the zero crossing of the ideal noiseless analog signal.

The resulting bias in start-stop type of measurements, in which a noisy signal is compared to a reference, is shown to depend on S/N ratio and on noise bandwidth. An experimental evaluation of this effect is reported, based on time interval measurements in which noise is added to a sinewave signal.

2 Trigger in the presence of noise

Let'consider a signal with a positive slope s which crosses a threshold of 0 V at the time $t = 0$. Being interested in what happens around $t = 0$ when noise is added to the signal, we can focus our attention on a small time interval around the crossing (say a few times the noise induced time jitter), within which the signal usually appears as a straight line.

In this region the signal has the same slope as the sinewave

$$x(t) = V_0 \sin(2\pi\nu_0 t)$$

where $V_0\nu_0 = s/2\pi$. By adding white gaussian noise $n(t)$, represented as *in-phase* and *quadrature* components, the resulting signal can be written as

$$\begin{aligned} y(t) &= x(t) + n(t) \\ &= V_0 \sin(2\pi\nu_0 t) + \\ &\quad n_s(t)\sin(2\pi\nu_0 t) + \\ &\quad n_c(t)\cos(2\pi\nu_0 t) \end{aligned} \qquad (1)$$

where $n_s(t)$ and $n_c(t)$ are time realizations of stochastic processes. If the total rms noise amplitude in the considered bandwidth is σ_n, then the two noise components have rms amplitudes $\sigma_s = \sigma_c = \sigma_n/\sqrt{2}$.

Following a common approach, we get the total rms phase noise ϕ through a phasor representation of (1), in which only the quadrature component gives its contribution

$$\phi = \frac{\sigma_n}{V_0} \qquad (2)$$

0-7803-0476-4/92 $3.00 © 1992 IEEE

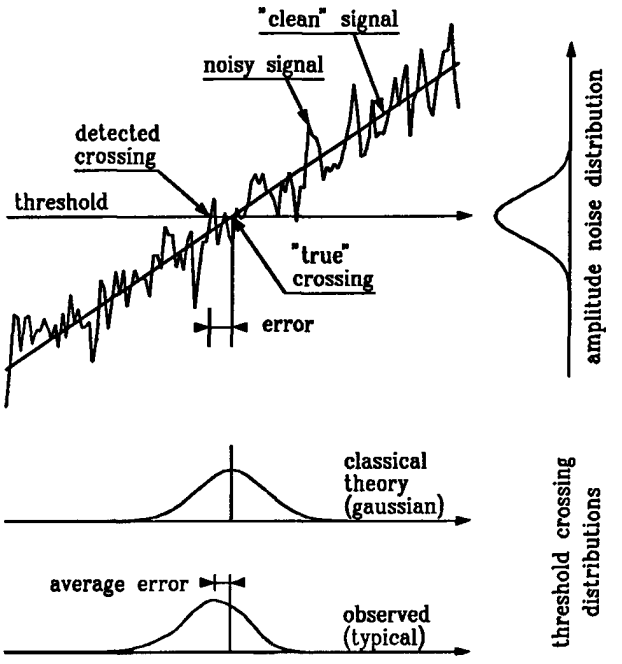


Figure 1: **Top.** A signal combined with wideband gaussian noise crosses a threshold at the time $t = 0$. **Bottom.** Theoretical and observed jitter distribution shapes.

The same concept is usually expressed in a slightly different way, by considering the rms time jitter σ_x (where σ_x is given by $\sigma_x = \phi/2\pi\nu_0$), and remembering that $V_0\nu_0 = s/2\pi$. Thus

$$\sigma_x = \frac{\sigma_n}{s} \tag{3}$$

The latter is the well known formula that describes the behaviour of a counter in the presence of noise. The consequence of noise appears to be time jitter, without any bias. Reality is more complicated.

Equation (3) relies on the noise representation given by (1), which is a (relatively) narrow bandwidth model. Consequently $n_s(t)$ and $n_c(t)$ are about constant around the threshold crossing, thus making the phasor representation meaningful. The noise appears as a phase plus amplitude modulation. For a large sample size, the average triggering time error tends to zero if the noise correlation time is small compared to the measurement time. This last hypothesis is true in most practical situations.

Conversely, wideband noise looks like a series of spikes of random amplitude and timing. Since the trigger switches on the first threshold crossing, it will do so on the first noise peak value of opposite sign and larger than the signal instantaneous amplitude (leading time error); a typical picture of what happens is given in Fig. 1 top. It is also possible that no spike be sufficient to overcome the signal before $t = 0$ and that the noise remain negative for some time after $t = 0$ (lagging time error, not shown in the figure). The first kind of error is more probable than the latter, thus affecting the measured average time.

When analyzing the statistical distribution of the threshold crossing time in the presence of wideband noise (see fig. 1 bottom), we expect to observe three phenomena:

1. for a gaussian distributed noise, the jitter distribution shape is no longer gaussian,
2. the rms time jitter is smaller than the value expected from equation (3),
3. the average measurement is biased.

3 Measurement method

Our experiments are based on start-stop measurements with a sinusoidal signal, averaging on large data sets.

The counter starts on the positive slope zero crossing of a low noise signal. The stop event is the positive slope zero crossing of the same signal, phase inverted and combined with noise. We expect a measured time interval θ given by

$$\theta = \frac{1}{2\,\nu_0} + \lambda + \epsilon \tag{4}$$

where ν_0 is the signal frequency, λ is the result of all the electrical lengths in the measurement scheme (cables, filters etc.), and ϵ is the error due to noise. Through the measurement of θ, we will investigate the statistical characterization of ϵ, depending on the noise bandwidth and on the S/N ratio.

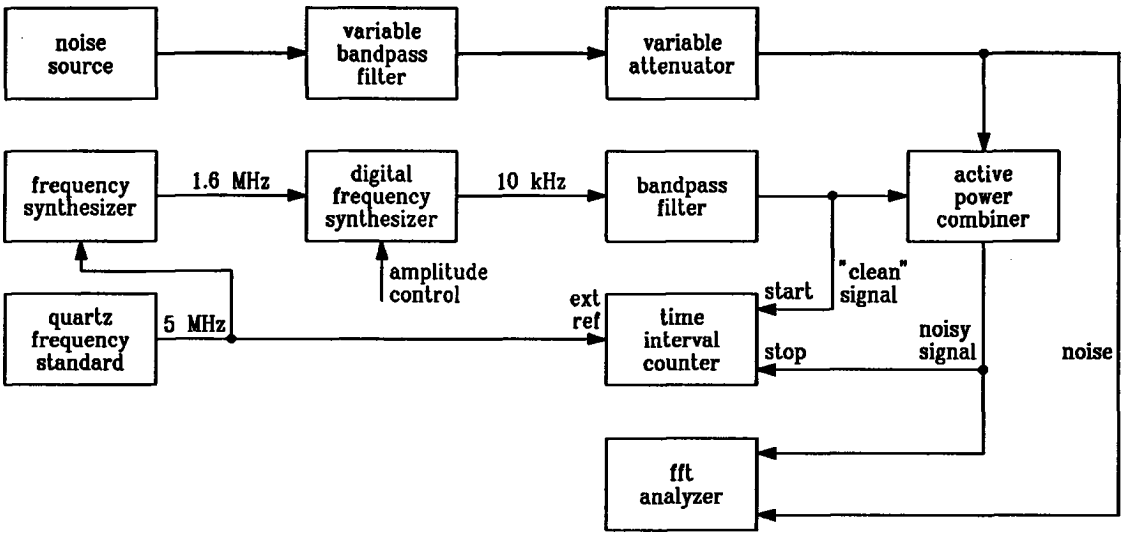


Figure 2: experimental setup.

In the design of the experimental setup (fig. 2), care has been taken to ensure the time stability of the system, thus avoiding confusion between noise induced biases and equipement drift. The counter is locked to the same frequency reference that drives the signal synthesizer. For the best stability, the signal is obtained by digital to analog conversion from memory mapped shape [1]. This technique shows a remarkable phase stability, about $10^{-5}/\nu_0$ in laboratory conditions. Moreover, when measuring the average value of ϵ (i.e. the noise induced bias) the delay λ has been calibrated before and after each set of measurements for each S/N ratio.

High performances counters (like Hewlett Packard HP-5370A [2] and Stanford Research SR-620 [3]) show a resolution of the order of 10 ps, with jitter and accuracy in the region between 100 ps and 1 ns. These values, compared to the jitter evaluated from (3), would limit the signal frequency to 100–1000 kHz (see fig. 3) if measurements free from any problem due to the time jitter or accuracy of the counter must be taken. Since our goal is to gain an experimental evidence on the statistics of ϵ, rather than to achieve the lowest jitter, we have adopted the lower frequency of 10 kHz. This choice brings additional advantages like the availability of FFT analyzers for precision noise measurenents and active filters in

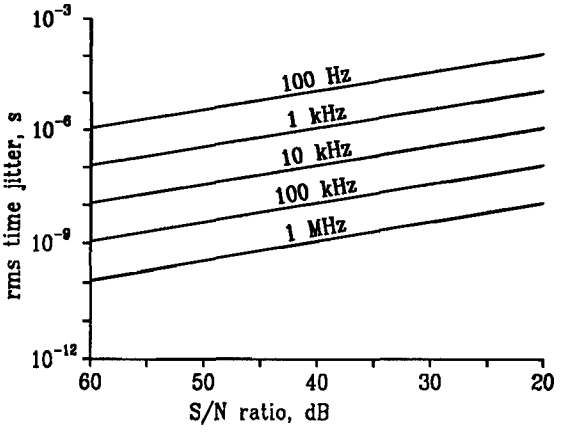


Figure 3: triggering time jitter expected from equation (3).

a wide band as compared to the signal frequency ν_0. In all experiments we have used the SR-620 counter, mainly because of its statistical and interfacing capabilities.

The signal amplitude spans from −13 to +5 dBm, well centered around the recommended input range of the counter. The noise spectral density is flat between two cutoff frequencies variable from a few tens of Hertz to 550 kHz. The noise level spans from −120 to −60 dBm/Hz.

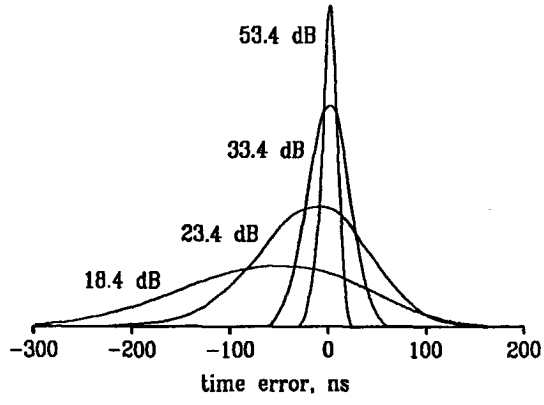


Figure 4: triggering time error probability distributions.

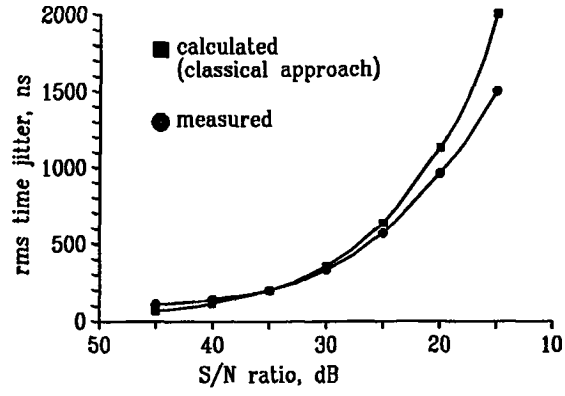


Figure 5: triggering time jitter measured and expected from equation (3).

4 Results

The first experimental evidence of the foreseen phenomena has been obtained collecting sets of 10^5 time interval data for some S/N ratios from about 15 to 65 dB, using a noise spectral density flat from 100 Hz to 550 kHz. The probability distributions for the most meaningful data sets are shown in Fig. 4. In order to make the comparisons between shapes easier, each plot has its own vertical scale; the area under each plot represents the same sample size.

The distribution is nearly gaussian and well centered around zero for the highest S/N ratio, 53.4 dB. As the noise increases (S/N = 33.4 dB), the distribution becomes slightly asimmetrical, and leading errors are privileged. As the noise furthe increases (S/N = 23.4 dB), the presence of a bias error becomes evident; at a closer sight, the left side tail of the distribution is slightly wider and higher than the right one. For the worst S/N ratio, the distrubution is strongly biased.

Figure 5 shows a comparison between the rms time jitter evaluated from (3) and measured with a noise density flat from 100 Hz to 200 kHz. Each experimental point represents the average of 20 blocks of 10^4 data; the standard deviation is well below 1/10 of the average value. For the highest S/N ratios the measured jitter is slightly higher than the evaluated one. This is due to noise of the start signal and to the residual noise of the system, outside of the bandwidth considered but within the trigger bandwidth (up to a few hundredth of megahertz), which have not been taken into account. For the lowest S/N ratios, the measured jitter turns out to be significantly smaller than the calculated one.

In figure 6 the measured average leading time error is reported as a function of the S/N ratio for various bandwidths. Experimental points are shown together with the statistical uncertainty deriving from 20 blocks of 10^4 samples.

It can be seen that at the high end of the S/N scale no average time errors are observed, whereas for low S/N ratios increasing leading time errors show up. The errors start increase at different S/N thresholds depending on noise bandwidth. For much lower S/N ratios, multiple counts start to appear.

References

[1] E. Rubiola, "Omega—VLF Signal Synthesizer," *Proc. 4-th European Frequency and Time Forum*, pp. 243–244, Neuchâtel (Swiss), March 13–15, 1990.

[2] Hewlett Packard, HP-5370A universal time interval counter, operating and service manual.

[3] Stanford Research Systems, SR-620 universal time interval counter manual.

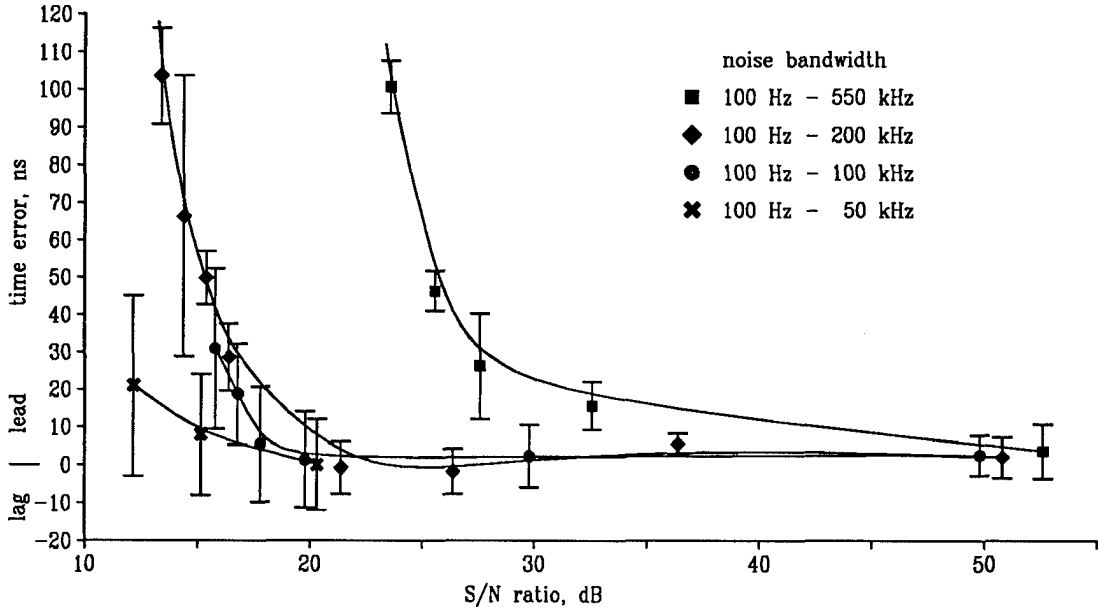


Figure 6: leading time error as function of the S/N ratio.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

THE GREATEST COMMON FACTOR FREQUENCY AND ITS APPLICATION IN THE ACCURATE MEASUREMENT OF PERIODIC SIGNALS

Zhou, Wei

Dept. of Measurement and Instrumentation, Xidian University
Xi'an, 710071, P.R.China

Abstract: The concept of the greatest common factor frequency and its characteristics have special functions in the high accurate measuremnet of periodic signals. The greatest common factor frequency between two frequency signals is similar to the mathematical greatest common factor between two numbers. With the greatest common factor frequency many important characteristics between periodic signals can be essentially recognized. According to these characteristics, the frequency and time measurement instruments designed by a new technique show more than 1000 times accuracy over that of ordinary instruments. If the instruments are designed for some special purpose, their accuracy is much higher than that. They are effective enough to overcome ± 1 word count error of digital measuring instruments. The instruments can suit a very wide frequency range. This paper analyzes the basic characteristics of the greatest common factor frquency, gives several very useful equations and application examples. Some new techniques based on the greatest common factor frequency can be used instead of many frequency and time measuring techniques and instruments. We believe that these new techniques and instruments will play a major role in future measurement and comparison of periodic signals, and they will also have very wide use in other measurements.

1. Introduction

In the research of periodic signals (frequency, period, some time intervals, phase difference, etc.), when the relationship of two same frequency signals is discussed, we often use the method of phase or phase difference analysis. For two signals that have different frequencies, the simple phase relation cannot give us any information. To express this question more clearly, we must study the further periodic rule between two signals. At this time we must pay attention to a very useful concept. It is the greatest common factor frequency f_{maxc} that can express the relatively periodic rule of phase difference between the periodic signals. The period of f_{maxc} is the least common multiple period T_{minc}.

We define the greatest common factor frequency in the following expression. To two frequency signals f_1 and f_2, if $f_1 = Af_c$, $f_2 = Bf_c$, the two positive integers A and B are prime with each other, then f_c is the greatest common factor frequency f_{maxc} between f_1 and f_2.

The greatest common factor frequency can be any frequency. It can be used to express the relationship between two periodic signals. Its period T_{minc} is equal to the multiples of the period values of f_1 and f_2 (A times $1/f_1$ or B times $1/f_2$).

2. The basic characteristics of the greatest common factor frequency

The greatest common factor frequency is the relatively periodic characteristic value between two frequency signals. It can be used to analyse the periodic regularity of the relative phase relationship between two frequency signals. With it we can use the method of phase analysis and measuring techniques to complete measurement between any periodic signals.

In a T_{minc} period, there are A times T_1 (the period of f_1) or B times T_2 (the period of f_2). On the one hand the phase relationship between f_1 and f_2 is influenced by the relative original phase difference between the two signals, on the other hand it is influenced by the relative phase difference change. The relative phase difference change is continuous , but in practical instruments especially in digital instruments, the continuous change is quantized. It has become a noncontinuously quantized change.

If $f_1 = Af_{maxc}$, $f_2 = Bf_{maxc}$, and $A > B$ (if $A < B$, there is the same regularity), in a T_{minc} period there can only be B kinds of quantized relative phase situations. The maximum change of the relative phase difference between f_1 and f_2 is from 0 to T_1. Because of the periodicity of the two frequency signals, in a T_{minc} period, the relative phase situations between the two signals only can be B kinds of different situations. The two kinds of same phase difference situations indicate the end of a T_{minc} period and the beginning of another T_{minc} period. The phase relationship between two periodic signals is descri bed by Figure 1.

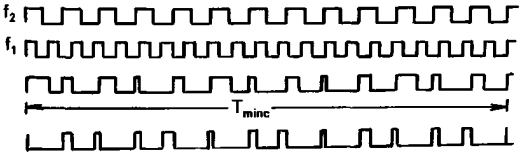


Fig.1

In Fig.1 the waveform 3 indicates the phase difference situations between f_2 signal and following f_1 signal. This waveform is described by "one direction" characteristic. The waveform 4 is the waveform of the quantized closest phase situation between f_1 and f_2, when f_2 is a reference signal. In this situation the range of phase difference change between two signals is from 0 to $T_1/2$.

To find the regularity of the phase change between two frequency signals, we suppose that the relative original phase difference between the two signals is 0. It cannot influence the change regularity of relative phase difference in a T_{minc} period.

From "one direction", in a T_{minc} period, the relative phase difference (time interval) between two frequency signals can be expressed as follows:

$$\begin{aligned} T_1' &= n_1 T_1 - T_2 \\ T_2' &= n_2 T_1 - 2T_2 \\ &\cdots\cdots\cdots\cdots \\ T_B' &= AT_1 - BT_2 = 0 \end{aligned} \qquad (1)$$

where T_1', T_2', ···, and T_B' express the phase difference between f_1 and reference signal f_2 to every specific phase point of f_1. They $\geqslant 0$, and $(n_x-1)T_1 < XT_2$. The n_1, n_2, ···, A and X are positive integers.

In the equations (1), to any equation $T_x' = n_x T_1 - XT_2$, because of $f_1 = Af_{maxc}$, $f_2 = Bf_{maxc}$, that is $T_1 = T_{minc}/A$,

0-7803-0476-4/92 $3.00 © 1992 IEEE

$T_2 = T_{minc} / B$.

$$n_x T_1 - X T_2 = \frac{n_x B - AX}{AB} T_{minc}$$
$$= \frac{Bn_x - AX}{B} T_1$$

where n_x, B, A and X are positive integers, and $T'_x \geqslant 0$. Therefore, $Bn_x - AX$ must be a positive integer or zero. The equations (1) can also be expressed as follows:

$$T'_1 = \frac{Y_1}{B} T_1$$
$$T'_2 = \frac{Y_2}{B} T_1$$
$$\cdots\cdots\cdots\cdots$$
$$T'_B = \frac{Y_B}{B} T_1$$

where Y_1, Y_2, ···, and Y_B only can be positive integers or zero. Y = 0and Y = B have same effect.

Because $0 \leqslant T'_x \leqslant T_1$, in a T_{minc} period T'_1, T'_2, ···, and T'_B cannot be equal, Y_1, Y_2, ···, and Y_B must be B positive integers that are not equal with each other and less than B. Therefore, the values of Y must be 0, 1, ···, B−1. The above description cannot express that what order the B numbers are. It only expresses that in a T_{minc} period, the values of T'_x must be 0, T_1/B, $2T_1/B$, ···, $(B-1)T_1/B$. Maybe the change of T'_x is from large to small or from small to large. Maybe the change is irregular. It depends on the relative relationship between two frequency signals. Undoubtedly, when the two signals have close frequency and a small frequency difference, the arrangement of T'_x is from small to large or from large to small.

Therefore, in a T_{minc} period maybe the phase difference change between any two frequency signals is irregular. If we rearrange the order by size sequence, the change of phase differences for the two rearranged signals between T_2 time interval must be:

$$\Delta T = \frac{1}{B} T_1 \qquad (2)$$

If the relative original phase difference between two frequency signals is not 0, the rearranged T'_x by size sequence are the results of a certain value decided by the original phase difference plus 0, T_1/B, $2T_1/B$, ···, $(B-1)T_1/B$. In this situation equation (2) is also right.

In practical application, the quantized closest phase differences between two frequency signals are often considered. In this situation the values of T'_x are from $-T_1/2$ to $T_1/2$. If the original phase difference between two signals is not 0, in a T_{minc} period the relative phase difference is not symmetric to 0 phase difference point. Therefore, the difference ΔT of two phase differences which have the closest absolute values is less than T_1/B.

From equation (2), we can obtain

$$\Delta T = \frac{T_1}{B}$$
$$= \frac{f_{maxc}}{f_1 f_2} \qquad (3)$$

We can utilize this equation to solve many measuring problems. When we use the methods of equation (3) to complete high accurate measurement, the larger product of two signal frequencies and the lower greatest common factor frequency are favourable.

In a T_{minc} period, there are B phase difference situations which are composed of a time difference originated by the original phase difference added respectively to 0, T_1/B, $2T_1/B$, ···, $(B-1)T_1/B$. The B phase differences have many possible arrangements. If there is general phase change between f_1 and f_2, the φ is changeable $\varphi(t)$. Generally, in every T_{minc} period the T'_x is different from the corresponding T'_x in another T_{minc}. Because of the change of $\varphi(t)$, when the relative phase difference changes T_1/B, in B phase differences of a T_{minc} period a certain phase difference situation must change from $(B-1)T_1/B$ to $BT_1/B = T_1$, i.e. corresponding $T'_x = 0$. The contrary situation is that the phase difference changes from T_1/B to 0. The two situations occur according to the different phase change directions. At the same time other B−1 phase difference situations also increase or decrease a T_1/B value. When $\varphi(t)$ changes in the same direction and same quantity, the above process occurs once more. In this change process of the relative phase difference, after the relative phase difference between two signals changes T_1/B, they complete a periodic full period comparison. The equivalent full period of the phase comparison is:

$$T = \frac{T_1}{B}$$
$$= \frac{f_{maxc}}{f_1 f_2}$$
$$= \frac{1}{ABf_{maxc}}$$

Therefore, the equivalent phase comparison frequency to $f_1 = Af_{maxc}$ and $f_2 = Bf_{maxc}$ is

$$f = ABf_{maxc} \qquad (4)$$

This method of phase comparison was called "Irregular Phase Discrimination" by us.

Because the establishment of equation (4) depends on the characteristic of waveform 3 in Fig.1, the circuit that can complete "Irregular Phase Discrimination" must have "one direction" characteristic. The J–K flip–flop is a better device for this discrimination.

Any ordinary relative phase comparison is only a special situation of equation (4). When two frequency signals are the same in frequency, $f_1 = f_2 = f_{maxc}$ (A = B = 1), the equivalent phase comparison frequency is $f = f_{maxc} = f_1 = f_2$.

3. The application of the greatest common factor frequency and its characteristics in the measurement of the periodic signals

With the simple experiment in the Fig.2 the equation (4) can be proved.

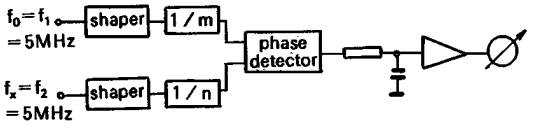


Fig.2

The method in Fig.2 can be used to design a linear phase comparator. Here f_1 and f_2 have same nominal frequency value, and they are high stable frequency standards. The dividers m and n of two frequency dividers are prime with each other. The output is recorded by a recorder after proper zero compensation. Provided m and n are prime with each other, their changes do not influence the full period of phase comparison. The equivalent phase comparison frequency is $f = mn(f_x / mn) = f_x$. The larger m and n are , the less the output voltage is.

Therefore, with the concept and characteristics of the greatest common factor frequency we can have more flexible and better application in method and technique of linear phase comparison. This method and instrument do not require that the two comared signals have the same nominal frequency. According to the relationship between two frequency signals and their f_{maxc}, the phase comparison effect of equation (4) can be obtained. In B relative phase differences of a T_{minc} period, only

a kind of $0\sim T_1/B$ phase difference can influence the linearity of phase comparison. A better linearity than that of phase comparison with same frequency can be obtained.

According to equation (4), in automatic control, frequency synthetic and phase–lock equipment the different frequency signals can be handled by a more direct and simple method. With equation (4) we can explain many regular non–normal phenomena in some circuits and instruments.

According to equation (3) and the above mentioned analysis, in a T_{minc} period there is always a quantized relative phase difference value that is in $0\sim f_{maxc}/f_1f_2$ range. If we use a phase testing circuit that has a larger testing range than f_{maxc}/f_1f_2, the relative phase difference situations that are equal to or less than this range can be tested. If the corresponding phase differences are little enough, we can consider this situation to be the "phase coincidence point" between two signals. The "phase coincidence point" does not mean exact phase coincidence. It means the degree and situation of the near relative phase. In the time interval between two or more "phase coincidence points", there are the period numbers that are very close to many integral periods of the two frequency signals respectively. Using this method to compose measuring gate time, a high accurate frequency and period measurement can be achieved. Fig.3 is the waveform figure of measuring gate time in frequency measuring instrument of wide frequency range designed by the above mentioned new method.

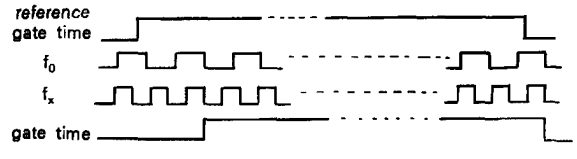


Fig.3

The measuring gate time is controlled by the reference gate time signal (for example: 0.1 second, 1 second, 10 seconds, etc) together with the "phase coincidence points" between the two frequency signals. Therefore, the measuring gate time is close to the reference gate time in length, but its beginning and end moments correspond to two "phase coincidence points" between the two signals. In the measuring gate time the period numbers of two signals are counted at the same time, the measured frequency can be computed accurately by the frequency value of the standard frequency in the two signals. With this method ± 1 word count error can be overcome satisfactorily. This method can be used to measure and handle any periodic signals widely. Because the frequency measurement of a wide range bases on the method of phase testing to the "phase coincidence points", we call it "The Frequency Measuring Technique by Broad–band Phase Detection" and have applied for a Chinese patent (the application number is 90103664.1).

When we measure frequency signal f_x with standard frequency f_0, in the measuring gate time f_0 and f_x are counted respectively and N_0 and N_x are obtained. The measured frequency is:

$$f_x = \frac{N_x}{N_0} f_0 \qquad (5)$$

The ordinary synchronous measuring method of multiple periods has the same equation as equation (5), but its gate time is only synchronized by the multiple periods of measured signal. In its measurement result there is ± 1 period measuring error of standard frequency signal. With "the frequency measuring technique by broad–band phase detection", the gate time is synchronized by the multiple periods of both measured signal and standard frequency signal. Therefore, ± 1 word measuring error is eliminated. The main measuring error is equal to the error of phase detection. The accuracy of phase detection can be made several decades picoseconds to less than several tenths nanoseconds very easily. Therefore with this method and a very simple device, we have obtained $3\times 10^{-9}/0.1s$, $3\times 10^{-10}/1s$ and $3\times 10^{-11}/10s$ measuring accuracy in wide frequency range (from 10kHz to over 100MHz), and $1\times 10^{-11}/1s$ and $1\times 10^{-12}/10s$ measuring accuracy in some frequency ranges. The instrument designed by this method can display a measuring result with 12 figures or more. The measuring accuracy is 1000 times higher than that of ordinary instruments in wide frequency range. If the instruments are designed for special purposes their accuracy is much higher. The instruments can also be used for terminal equipment of a high accuracy measuring system. Therefore, the system accuracy can be greatly enhanced.

According to equateion (3), to obtain the high accuracy, on the one hand less ΔT is required, on the other hand the phase coincidence detection circuit of high distinguishability is also required. In any situation, the high f_0 frequency is always favorable.

The practical gate time is influenced by the reference gate time, the least common multiple period T_{minc} between two signals and the distinguishability of phase coincidence detection circuit. If the distinguishability is about equal to ΔT, in a T_{minc} period only one or two "phase coincidence points" can be detected. If the phase detection range is much larger that ΔT, in a T_{minc} period many "phase coincidence points" can be detected. If T_{minc} is much less than the reference gate time, the practical gate time is influenced very little by T_{minc}. If T_{minc} is close to or larger than the reference gate time, and the distinguishability of the phase coincidence detection circuit is close to ΔT, the practical gate time is different from the reference gate time, even the difference can be very large. If we hope to have strict measuring gate time, we must pay attention to this problem.

Obviously, in the practical measuring instruments, the phase detection range must be larger than ΔT. On the one hand, to suit to the measuring accuracy the phase detection range cannot be too large, and on the the other hand, in the certain T_{minc} period more "phase coincidence points" must be caught to suit to the measuring time response. Except the signals f_0 and f_x having the close frequencies (in this situation, the "phase coincidence points" are concentrated in the some range in T_{minc}), the time response of the measurement can be improved proportionally, when the proportion of the phase detection range to ΔT increases. It is very important. Because the measured frequency signal can be random, the T_{minc} is long in most situations.

When a high accurate integral frequency standard is used to measure any frequency signal, under most situations, very little f_{maxc}/f_1f_2 can be obtained. Therefore, with phase detection circuit which has necessary detective accuracy, the high accurate frequency measurement can be completed. To the very stable integral frequency signals (their frequencies are equal to or close to the frequency of the integral frequency standard or divisors or multiples of its frequency), we can measure them with a frequency synthesizer. Obviously, for the measurement of very wide frequency range, one or two frequency synthesizers that have a single frequency value are enough. The most simple frequency synthesizer is that a little division frequency is added to or subtracted from the frequency of integral frequency standard. If it and the standard frequency signal are used substitutionally according to the situations, in the wide frequency range the demand of f_{maxc}/f_1f_2 can be satisfied. In practical applications, f_0 in equation (5) is equal to the practical frequency of the frequency standard or frequency synthesizer according to the used situation.

The distribution of "phase coincidence points" can make an important contribution to the gate length of short measuring

gate time. The distribution of "phase coincidence points" expresses the distribution situation of the near relative phase in a T_{minc} period. It depends on the frequency values of two compared frequency signals and is relative to the phase detection accuracy of the phase detection circuit. The use of frequency synthesizer enables the equipment to have two kinds of distributions. In measurement, through their compensation with each other, the measurement of wide frequency range can be completed and demanded measuring gate time can be obtained. We shall describe this question in detail in another paper.

The upper frequency measuring limit of the new frequency measuring method is not limited. It is different from the general method of frequency measruement in that the lower–frequency measuring limit of the new method is limited. The low lower–frequency measuring limit can be obtained by three methods. The first is higher standard frequency, the second is longer measuring gate time, and the third is lower measuring accuracy (with the phase detection circuit of lower phase detection accuracy). Because under low frequency, the triggering error of signal becomes the main error, the higher measuring resolution is not necessary. Therefore, we can lower the phase detection accuracy or combine with the method of ordinary multiple periods synchronous measurement.

With the above–mentioned method other periodic signals can also be measured at high accuracy. The time interval and phase difference are the examples. We can detect the "phase coincidence points" between standard frequency signal and two compared signals which are the same in frequency respectively. The measured time interval or phase difference can be calculated by the time differences of the "phase coincidence points" of the two compared signals and standard frequency signal. The waveform figure is shown in Fig.4.

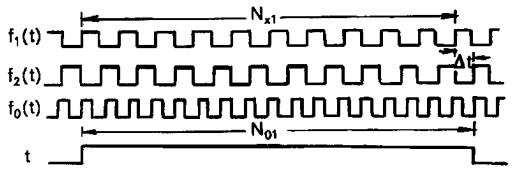


Fig.4

The measuring gate time begins with the "phase coincidence point" between $f_1(t)$ and $f_0(t)$, and ends up with the "phase coincidence point" between $f_2(t)$ and $f_0(t)$. Besides, with the above–mentioned frequency measuring method, one of the equal frequency signals $f_1(t)$ and $f_2(t)$ is measured and its period T_x is obtained. The measured whole time is

$$t = N_{x1} T_x + \Delta t$$
$$t = N_{01} T_0$$

The measured time interval is

$$\Delta t = N_{01} T_0 - N_{x1} T_x \tag{6}$$

When measurand is the per cent in a full period, the phase difference between the two measured signals is

$$T = \frac{\Delta t}{T_x} \tag{7}$$

When the measurand is the angle, the phase difference is

$$\varphi = 360^\circ \frac{\Delta t}{T_x} \tag{8}$$

In the three measurements expressed by equations (6), (7) and (8), the measuring error is much less than ± 1 period of the count pulse, but is the value relating to the quantized phase change between frequency standard and measurand and the phase detection discriminability. Therefore, the measuring accuracy is increased greatly. To the periodic signals which are in action with each other, according to their greatest common factor frequency and its period the further rule of their relationship can be found. The concept of the greatest common factor frequency not only can be used in the high accurate measurement of any periodic signal, but also can be used to explain and deal with many periodic phenomena in nature. As the research proceeds, more very useful measuring methods and instruments can be produced.

References:

1. Zhou, Wei "A Method of High Linear Phase Discrimination for High Frequency–The Phase Discrimination Controlled by Dividing Frequency into a Single Channel" Acta Metrologica Sinica Vol.2, No.1, pp.75–79, 1981

2. Zhou, Wei "A New Principle of Linear Phase Comparison–Irregular Phase Discrimination" 18th Annual PTTI Applications and Planning Meeting, Nov.1986

3. Zhou, Wei "Frequency and Time Metrology" Science and Technology Publishing House, 1986, pp.220–235

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DISCRETE SIMULATION OF POWER LAW NOISE

N. Jeremy Kasdin and Todd Walter

W.W. Hansen Experimental Physics Laboratory
Stanford University, Stanford, CA 94305-4085

Abstract *Power law noise plays an important role in the description of high performance oscillators. Commonly, five types of noise are considered to affect clocks and clock measurements: white phase, flicker phase, white frequency, flicker frequency and random walk frequency. These noise types are distinguished by the slopes of their spectral densities, $S_y(f) \propto f^{\alpha}$ (on a log-log scale). The noise is inherent both to the oscillators and to the measurement systems and defines the limits of stability of the clocks. Accurate simulation of the noise can be important for testing the measurement system and the characterization software. This paper presents a new algorithm and computer code for simulating power law noises with arbitrary α (it is not restricted to the integer values mentioned above). The general theory of noise simulation is investigated to determine the criteria for evaluating and deriving simulation methods. Past techniques are evaluted and the new method is shown to provide improvements, particularly because it results in non-stationary noise sequences that are also scale-invariant and causal and have the proper autospectral densities and Allan variances.*

I. INTRODUCTION

The expression "Colored Noise" is used to refer to any non-white noise stochastic process. That is, a stochastic process whose autospectral density is not a constant but some function of frequency. In many applications, these density functions are described by rational functions of frequency, f, and thus correspond to causal, realizable, linear systems. Such a system representation can then be easily discretized by a number of techniques of varying accuracy and used to generate sample sequences of the noise on a computer [1].

In this paper, we are concerned with a specific category of colored noise – f^{α} or Power law noise. This noise is characterized by an autospectral density function that is not a rational function of frequency but instead is proportional to the frequency raised to an arbitrary power, α. This noise is of interest because of its ubiquity in natural and engineering systems and because of its resilience to theoretical modeling. In particular, because the autospectral density is non-rational, it is impossible to develop a rational transfer function that can exactly describe the noise as the output of a linear system [3]-[10].

Power law noises of this type are particularly prevalent in the output of oscillators. Therefore, models and generating mechanisms are, understandably, of keen interest to the clock community [2],[11],[12]. This paper briefly reviews some of the past methods for modeling and generating power law noise, with special emphasis on those used to describe oscillator stability. A new model, first suggested by Hosking [13], is then proposed, using z-transform theory, for generating these noises with some distinct improvements over past techniques. In particular, it should be pointed out that the power law description considered here, and in the generating mechanisms, is not restricted to integer α, as in most of the clock literature, but is equally valid for all continuous values of the slope.

This noise generator has a number of distinct advantages over the previous methods, including the ARMA model most often used to simulate oscillator noise. In reference [1], which is a broader review of stochastic process simulation, a number of criteria are described that must be met by any simulation algorithm, with additional categories for power law noises. As is described in detail below, the power law description of the autospectral density implies that the noise is both non-stationary (for $\alpha \leq 1$) and scale-invariant, neither of which is satisfied by the ARMA model.

To understand how noise manifests itself in the output of an oscillator, we can represent the output as [2]

$$V(t) = [V_0 + \varepsilon(t)] \sin[2\pi\nu_0 t + \varphi(t)] + V_1(t) \quad (1)$$

0-7803-0476-4/92 $3.00 © 1992 IEEE

V_0 and ν_0 are the respective nominal amplitude and frequency of the output, $\varepsilon(t)$ and $\varphi(t)$ are amplitude and phase fluctuations respectively and $V_1(t)$ is additive noise. Provided ε and V_1 are much smaller than V_0, the instantaneous frequency of the oscillator output can be written

$$\nu(t) = \nu_0 + \frac{1}{2\pi}\frac{d\varphi(t)}{dt} \qquad (2)$$

The instantaneous fractional frequency deviation from nominal may also be defined

$$y(t) \equiv \frac{\nu(t) - \nu_0}{\nu_0} = \frac{1}{2\pi\nu_0}\frac{d\varphi(t)}{dt} \qquad (3)$$

It is the stability of the output frequency that is of primary concern. Stochastic processes that affect this stability will appear in y(t).

Another useful quantity is the phase deviation, in units of time,

$$x(t) \equiv \frac{\varphi(t)}{2\pi\nu_0} \qquad (4)$$

This quantity, the time integral of y(t), is a measure of the time deviations of the oscillator as a clock. Because y(t) must be measured over a finite time interval, we prefer to use the measure x(t), from which we can reconstruct averaged values of y(t).

In high performance oscillators, x(t) is usually dominated by its stochastic components. It is therefore important to understand what the properties of the noise sources imply about the stability of these devices. One excellent and commonly used method for understanding this impact is through computer simulation.

The remainder of this paper describes the new algorithm for modeling and generating f^α noise for use in evaluating oscillator stability. Section II describes some common methods for describing noises in oscillators and presents the criteria against which we measure the quality of a noise simulation algorithm. Section III then presents the required background on linear system theory and general colored noise simulation techniques. Section IV is a discussion of past continuous models for power law noise that form the basis for digital generation techniques. Section V presents the discrete generation method being proposed, followed, in Section VI, by results using this model.

The paper presents a number of computer techniques for implementing this filter rapidly and efficiently. In addition, the model for generating the noise is chosen *a priori* and requires no ARMA fitting of coefficients to data.

II. NOISE CHARACTERIZATION

There are numerous measures used to characterize noises. The most fundamental descriptor of a stochastic process is the joint probability density evaluated at all times. Clearly, this is not a convenient tool as it is often difficult, if not impossible, to evaluate and involves an infinite number of parameters. Instead, the first and second moments of the process, as functions of time, are normally used. These are the mean and autocorrelation function (or autocovariance function). For Gaussian processes (the only ones considered here), these two moments entirely describe the joint density [14].

Although the autocorrelation function completely describes the process, it is not the most useful measure as it is difficult and time consuming to measure in practice. Instead, the autospectral density is used as it is easily computed using FFT algorithms on a sampled sequence. For wide-sense stationary processes, the autospectral density is equal to the Fourier Transform of the autocorrelation function and is thus an equivalent and, therefore, complete description of the Gaussian stochastic process. As we will discuss below, however, this frequency domain definition encounters difficulties when used with power law noises.

II.1 Frequency Domain

Power law, or f^α, noise, the type that is most commonly considered to affect oscillators, is distinguished by the slope of the power spectral densities (measured on a log-log scale). The five traditional noise types used to describe oscillator stability are: white phase, flicker phase, white frequency, flicker frequency and random walk frequency. These correspond to the integer values of α from 2 to -2 respectively. The autospectral density of y(t) is conventionally expressed as a weighted sum of these noise types in the power law model [2]:

$$S_y(f) = \sum h_\alpha f^\alpha \qquad (5)$$

where h_α is the intensity coefficient. Because x(t) is the time integral of y(t), the spectral density of x(t) is related to the spectral density of y(t) by

$$S_x(f) = \frac{S_y(f)}{(2\pi f)^2} \qquad (6)$$

Although it is more common to examine the properties of y(t), we prefer to consider the quantity x(t) and its spectral density. To simplify our notation we will define new quantities β and g_β such that

$$S_x(f) = \frac{h_\alpha}{(2\pi)^2} f^{\alpha-2} \equiv g_\beta f^\beta \qquad (7)$$

The five noise types in this notation correspond to the integer values of β from 0 to -4.

This conventional power law model is an extremely useful tool for grouping oscillator noise into categories and for predicting the limits of precision for a device. It is, however, too restrictive. One problem is that it only considers integer values of α (β). A more significant

difficulty when pursuing noise generation, is inherent in the frequency domain description. To see this, it is helpful to recall the process by which this autospectral density is determined. A typical noise process is sampled, each point separated in time by τ_0, generating an N point sequence that is windowed, Fourier transformed, and then absolutely squared to determine an estimate of its autospectral density. This noisy, non-parametric estimate is used to fit the parametric, power law model in Eq. (5).

An understanding of this process has important implications when generating sample noise sequences. Because of the ubiquity of the autospectral density as a measure, most generating techniques attempt to create a noise sequence with a discrete autospectral density that samples the continuous model above. Such an approach, however, ignores the fact that this equation is only a fit to a noisy, distorted estimate. Any generated sequences must also have the equivalent distortions due to finite length effects (windowing), sampling and aliasing, and causality [1]. Also ignored are the implications of non-stationarity in the power law model. Though a number of past authors have attempted direct frequency domain generation that accounts for some of these effects, none have satisfactorily accounted for all distortions [15]-[17].

This discussion shows the need for an adequate time domain measure against which to measure the quality of a noise simulation. We thus return to the autocorrelation function.

II.2 The Autocorrelation Function

As mentioned, the time domain analog of the autospectral density is the autocorrelation function. In the time domain, the symmetric time lag autocorrelation function of a continuous real process x(t) is given by

$$R_x(t,\tau) \equiv \langle x(t-\tau/2)x(t+\tau/2)\rangle \qquad (8)$$

where the angle brackets $\langle\ \rangle$ denote an expected value or infinite ensemble average. If x(t) is also a zero-mean process then $R_x(t,\tau)$ is equivalent to the autocovariance function. Note that for zero time lag (τ=0) the quantity reduces to the variance at time t. Although it is difficult to obtain a reasonable estimate of the autocorrelation in practice, it is the fundamental benchmark for judging generated noise. Thus we will state our first criteria for an adequate noise simulation:

Criterion 1: A discrete process x_k is said to simulate the continuous process x(t) if the discrete autocorrelation function $R^d_x(k,m)$ samples from the continuous autocorrelation function $R_x(t,\tau)$. That is,

$$R^d_x(k,m) = R_x(k\tau_0 + t_0, m\tau_0) \qquad (9)$$

If x(t) is Gaussian, then a simulation that satisfies Criterion 1 also has the property that the probability density of x_k is equivalent to samples of the joint probability density of x(t). Such a simulation will be called strict-sense. If x(t) is non-Gaussian distributed such that only Criterion 1 is satisfied but the probability distributions do not correspond, then the simulation will be called wide-sense. All noises in this paper will be assumed Gaussian.

Note the important implication of this definition. Because we are sampling directly the autocorrelation function, we are retaining information at all frequencies. That is, the simulation process implicitly aliases the high frequency information in the autospectral density function when producing the discrete autocorrelation function in Eq. (9). This is very reasonable if, for example, one is simulating the output of a dynamic system where the noise action between samples is desired. If, instead, we wish to simulate the output of a sensor after prefiltering at the Nyquist frequency, we must first develop a model that includes this prefilter in the original, continuous autocorrelation function. The simulation then results in a new, bandlimited $R^d_x(t,\tau)$. This new simulation thus still satisfies Criterion 1, but the simulated random variable does not include the effects of aliasing the continuous process. In practice, however, because most processes are bandlimited, the effect of aliasing is small and normally confined to a narrow region near the Nyquist frequency.

A routine satisfying our first criterion is very different from the traditional simulation routines described above which instead try to match the discrete autospectral density to a continuous model. As argued, since we are trying to simulate a time domain sampled quantity this criterion is immune to the sampling distortions that appear in frequency domain measures.

For power law noises, the autospectral density of x(t) has a breakpoint at $\beta = -1$ ($\alpha = 1$). Above this value the noise is stationary, but contains an infinite amount of energy as one includes the higher frequencies, resulting in an undefined variance ($\tau = 0$). Thus, either the model must be modified or a system bandwidth must be imposed to limit the contribution from energy at this end of the spectrum.

A wide-sense stationary noise process is defined by the convergence of the autocorrelation function for large time:

$$\lim_{t\to\infty} R_x(t,\tau) \to R_x(\tau) \qquad (10)$$

Below $\beta = -1$, the noise becomes non-stationary. This property can be seen in the frequency domain by the non-integrable singularity in the autospectral density of x(t) at f = 0. Many authors argue that some second process

must come into play, preventing the autospectral density from becoming infinite at zero frequency. However, a low frequency roll off has never been observed experimentally [3]-[5],[8]. This therefore leads to our second criterion for a power law noise generator:

Criterion 2: The generated process x_k must be non-stationary for $\beta \leq -1$.

A third property of power law noise implied by the spectral density in Eq. (5) is its scale invariance. That is, the model is independent of the scale of observation. If the frequency scale is changed, the original amplitude scaling can be obtained by multiplying by an appropriate constant. This was first noted by Mandelbrot and led directly to his development of fractal geometry [5]-[7],[9]-[10]. Mandelbrot observed the universality of scale invariance and elevated it to a fundamental property of nature. Many authors propose that it is this property that is universal and accounts for the proliferation of power law noises throughout nature.

This property is formally stated as Criterion 3:

Criterion 3: The generated process x_k must be scale invariant.

To be more precise, the generated process must be self-affine. A self-affine process, as defined by Feder [17], is defined by the scale-invariance of its probability density function:

$$p(\bar{x} = b^H x, \bar{t} = bt) = b^{-H} p(x,t) \qquad (11)$$

where the new time scale, $\bar{t}$, has been scaled from the old one by the factor b. Note that a self-similar process is one where the power, H, is unity.

Here, we refer to this as strictly self-affine. As in the stationarity definition, the self-affine property can be made more general by defining a wide-sense self-affine process as one where only the first and second order moments are scale-invariant. That is, the autocorrelation function has the property:

$$R(bt, b\tau) = b^{-H} R(t,\tau) \qquad (12)$$

In this work, we only require that the noise models satisfy Eq. (12). However, for Gaussian noise processes (such as those generated below), wide sense scale-invariance is equivalent to strict scale-invariance.

Our final criterion for the noise generator follows fairly clearly from the spectral density model:

Criterion 4: The generated process x_k must result in an autospectral density estimate with an f^β form over the desired frequency range.

III. LINEAR SYSTEMS

The most common model for noise generation is via linear systems with white noise inputs.† As mentioned earlier, when the autospectral density is a rational function of f, the system can be described by a realizable transfer function and implemented as a simple linear filter. Unfortunately, power law noises result in non-rational transfer functions that raise some difficult issues. In this section we present some fundamental definitions and properties of linear systems necessary for an understanding of the generation algorithm.

III.1 Definitions

We consider white noise convolved with a causal impulse response function h(t)

$$x(t) = \int_0^t h(t-\eta) w(\eta) d\eta \qquad (13)$$

where w(t) is a white noise random process.

The symmetric autocorrelation function can be obtained from Eq. (8) via:

$$R_x(t,\tau) = Q \int_0^{t-\tau/2} h(\eta) h(\eta+t)\, d\eta \qquad (14)$$

For a wide sense stationary process with large t, the one-sided power spectral density function is given by

$$S_x(f) = 2QH(2\pi f) H^*(2\pi f) \qquad (15)$$

where H(s) is the Laplace transform of the impulse response function h(t).

It is possible to replace these continuous relationships with discrete versions. To define a discrete process x_k the convolution integral is replaced by a convolution sum

$$x_k = \sum_{\ell=0}^{k-1} h_{k-\ell} w_\ell \qquad (16)$$

where w_k is an independent and identically distributed (i.i.d.) random variable with discrete autocorrelation function $Q^d \delta_{k0}$.

The discrete autocorrelation function is then given by:

† Mathematically, the concept of white noise is a problematic one. Its standard definition as a process with constant autospectral density implies an infinite variance and is thus physically unrealizable. The more mathematical definition as the derivative of Brownian motion suffers from the question of differentiability. Most of the issues are resolved by either assuming a finite bandwidth or recognizing that we never consider white noise as an independent quantity but only as a distribution appearing in integrals such as Eq. (13) (known as the Ito integral). More detail on these questions can be found in references [1],[14],[18].

$$R_x^d(k,m) = Q^d \sum_{n=0}^{k-m/2-1} h_n h_{n+m} \quad (17)$$

The discrete autospectral density is given by

$$S_x^d(f) = 2Q^d H(z) H(z^{-1}) \quad (18)$$

where H(z) is the z transform of the discrete process.

III.2 Selecting h_k

Eq. (16) is the fundamental discrete equation used to generate noise. However, it is necessary to find the proper h_k such that the discrete autocorrelation in Eq. (17) samples the continuous one in Eq. (14), as required by Criterion 1, and such that the discrete autospectral density has the proper form, as required by Criterion 4. In general there is no definitive technique for choosing h_k. However if H(s) is a rational function of s, we can derive one method for choosing h_k. We can discretize Eq. (14) and find

$$R_x(k\tau_0+t_0, m\tau_0) = Q \sum_{n=0}^{k-m/2-1} \int_{n\tau_0}^{(n+1)\tau_0} h(\eta) h(\eta+m\tau_0) d\eta \quad (19)$$

A simple change of variables yields

$$R_x(k\tau_0+t_0, m\tau_0) = Q \sum_{n=0}^{k-m/2-1} \int_0^{\tau_0} h(\zeta+n\tau_0) h(\zeta+n\tau_0+m\tau_0) d\zeta \quad (20)$$

We can take advantage of the fact that for a rational transfer function $h(t+\tau) = h(t)\, h(\tau)$. For this special case

$$R_x(k\tau_0+t_0, m\tau_0) = Q^d \sum_{n=0}^{k-m/2-1} h_n h_{n+m} \quad (21)$$

where

$$Q^d = Q \int_0^{\tau_0} h^2(\eta) d\eta \quad (22)$$

and

$$h_n = h(n\tau_0) \quad (23)$$

Thus for a rational transfer function the discrete impulse response function directly samples the continuous function.

Unfortunately, for the non-rational power law noises, Eq. (21) does not hold. Even if a continuous impulse response function was available, simply sampling it to generate a discrete power law noise by Eq. (16) would be inadequate. Sections IV and V explore a new technique for finding the proper h_k.

III.3 Batch, Frequency Domain Simulation

Once an h_k is available, the noise sequence, x_k, can be generated. The most straightforward method is to directly perform the discrete convolution in Eq. (16). For each new point x_k to be generated we must compute another coefficient h_k and another white noise element w_k (though the pulse response coefficients can be precomputed and stored). The two resulting time sequences are then convolved to produce the desired noise. The only subtlety is in selecting the proper variance for the white input noise. (Note that this approach is equivalent to an N point FIR filter or purely MA process [19].)

Unfortunately, direct convolution is a very time consuming process, particularly for large N. A preferred approach is to replace the convolution with multiplication in the frequency domain, that is, x_k is the inverse Fourier transform of the product of transforms of h_k and w_k. Thus we compute the N coefficients, h_k and the N i.i.d. points, w_k and discretely Fourier transform them into the frequency domain. Once in the frequency domain we multiply the DFTs and inverse Fourier transform them back to the time domain, generating all N points at once in a batch method. The only remaining detail regards causality. Multiplication of vectors in the discrete Fourier domain is equivalent to a circular convolution rather than a linear one. It is therefore necessary to pad the end of the time domain sequences h_k and w_k with an equal number of zeroes before transforming. This is equivalent to requiring that that the sequences be causal and ensures a linear convolution [19].

The main drawback of this technique is its memory burden. Because it is a batch method, it can require large amounts of computer memory for long time sequences. In addition, normally only sequences whose lengths are powers of two can be generated in reasonable time (a limitation of most FFT algorithms).

IV. CONTINUOUS NOISE MODELING

The final step in producing a simulated f^β sequence is to find an appropriate set of pulse response coefficients, h_k, for use in the algorithm described above. This requires some discussion of previous continuous power law noise models. Continuing with the assumption that the noise generating process is linear, it is natural to assume that the corresponding transfer function has the form $s^{\beta/2}$. Unfortunately, as previously mentioned, this transfer function is irrational and thus does not have a corresponding finite state linear filter that can be used to generate the noise. As a result, a number of approximate techniques have been developed for use as generating mechanisms. The problems associated with direct frequency domain generation (that is, randomizing an f^β spectrum directly and inverting) have already been mentioned. This section discusses the other prominent methods.

IV.1 Finite Pole/Zero Approximation

The most common technique for avoiding the problems of irrationality is to fit the $s^{\beta/2}$ transfer function with a sum of realizable, rational transfer functions, normally first order Gauss-Markov processes. It can be shown that in the limit of an infinite number of poles and zeros, any slope can be achieved exactly with proper selection of the coefficients and root locations [3],[4],[8],[20],[21],[22]:

$$H(s) \propto \sum_{n=0}^{N} \frac{\tau_n}{s+\tau_n} \underset{N\to\infty}{\to} s^{\beta/2} \qquad (24)$$

This continuous model can then be mapped to the discrete domain by any standard method. Alternatively, Barnes finds the coefficients of the discrete linear filter directly by ARMA identification techniques on sample noise sequences [22].

There are three significant problems with this method. The first is its inherent coarseness in describing the spectrum. An infinity of poles is necessary to exactly reproduce the f^{β} form (in his algorithm, Barnes uses only five poles to generate the noise [22]).

The second problem arises when trying to model the low frequency divergent behavior. As more points are generated, more poles must be added. If the number of poles is truncated the generated noise will not satisfy our second criterion.

The final problem is that this model is not scale invariant. The pole locations are inherently functions of time scale. If a new sampling interval (τ_0) is chosen the pole locations are no longer correct and new locations must be found. This violates Criterion 3.

IV.2 Fractional Brownian Motion

Using methods of the fractional calculus, Mandelbrot, and others, as an alternative to the above approximate technique, examined more carefully the irrational transfer function in order to find pulse response coefficients that could be used in the techniques of section III. [5-7],[9-10],[16] Because of its similarity to pure integration ($\beta = -2$) or Brownian motion, Mandelbrot generalized this concept and called f^{β} noise fractional Brownian motion. Thus, just as white noise is often defined as the derivative of a Brownian motion, we define f^{β} noise via the fractional derivative:

$$\frac{d^{-\beta}}{dt^{-\beta}} x(t) = w(t) \qquad (25)$$

While it is beyond the scope of this paper to discuss the fractional calculus in any detail, it can be shown [23] that the anti-derivative corresponding to Eq. (25) is given by the Riemann-Liouville fractional integral:

$$x(t) = \frac{1}{\Gamma(-\beta/2)} \int_0^t (t-\eta)^{-(\beta/2+1)} w(\eta)\, d\eta \qquad (26)$$

This is equivalent to a linear system driven by white noise with impulse response function

$$h(t) = \frac{t^{-(\beta/2+1)}}{\Gamma(-\beta/2)} \qquad (27)$$

The Laplace transform of this equation results in the appropriate transfer function:

$$H(s) = s^{\beta/2} \qquad (28)$$

A number of authors have used the impulse response function in (27) to find an h_k for use in the convolution of section III. However, as pointed out in that section, simply sampling Eq. (27) is incorrect for irrational transfer functions. In addition, for $\beta > -1$, the impulse response is undefined for t=0, raising severe practical and theoretical difficulties. Instead, we take the approach of using this model to understand the correlation properties of power law noise and then find a discrete model directly in Section V that fits these properties.

The symmetric autocorrelation function for the fractional Brownian motion model is:

$$R_x(t,\tau) = \frac{Q}{\Gamma^2(-\beta/2)\,\tau^{\beta/2+1}} \int_0^{t-\tau/2} \frac{\eta^{-(\beta/2+1)}}{(1+\eta/\tau)^{\beta/2+1}}\, d\eta \qquad (29)$$

This integral must be solved separately for $\beta = -1$ and $\beta \neq -1$. For $\beta = -1$ this integral becomes

$$R_x(t,\tau) = \frac{2Q}{\pi} \cosh^{-1}\left[\sqrt{\frac{t+\tau/2}{\tau}}\right] \qquad (30)$$

and for $\beta \neq -1$ it is

$$R_x(t,\tau) = -\frac{2Q/\beta}{\Gamma^2(-\beta/2)\,\tau^{\beta/2+1}} (t-\tau/2)^{-\beta/2} \times {}_2F_1(1+\beta/2, \beta/2; 1/2-t/\tau) \qquad (31)$$

where ${}_2F_1(\alpha,\beta,\gamma;z)$ is the hypergeometric function.

Examinations of these two autocorrelation functions make clear two problems with this model. The non-stationary spectral density does not in general result in an f^{β} form (violating our fourth criterion) and the autocorrelations are not scale invariant (violating our third criterion). Both difficulties arise from the high frequency problem mentioned previously. An exact f^{β} process with $\beta > -1$ out to infinite frequency is impossible – when τ is set to zero the time varying variance becomes undefined (infinite). Some upper bandwidth must be imposed (resulting in some unknown modification to the impulse response function near zero). This has long been recognized as a problem with the continuous model and has raised difficulties with determining h_k from the continuous h(t) (since h(t) is undefined at zero). See, for example, reference [9].

These problems can be partially avoided by examining the autocorrelations for $t >> \tau$ (that is, times

large compared to the start of the process). In this case (30) and (31) are expanded for small t/τ to yield:

$$R_x(t,\tau) \cong \frac{Q}{\pi}(\log 4t - \log \tau) \qquad (32)$$

for $\beta = -1$, and

$$R_x(t,\tau) = \frac{Q\Gamma(1+\beta)}{\Gamma(-\beta/2)\,\Gamma(1+\beta/2)\,\tau^{\beta+1}} + \frac{Q\,\Gamma(-1-\beta)}{\Gamma^2(-\beta/2)\,\Gamma(-\beta)\,t^{\beta+1}} \qquad (33)$$

for $\beta \neq -1$.

For $\beta > -1$ the process is stationary, however, again, the variance is undefined as τ approaches zero.

V. DISCRETE GENERATION METHOD

Finally, we return to discrete noise generation methods. We will take as our starting point the simple discrete differential equation that is used to generate what could be termed random walk phase modulated noise (it is more commonly called white frequency modulated noise)

$$x_k = x_{k-1} + w_k \qquad (34)$$

This equation can be derived via the methods of section III or reference [1] as the discrete simulation of the continuous Brownian motion model. The variance of w_k is found by matching the autocorrelation functions or Allan variances .

The discrete transfer function of this random walk is:

$$H(z) = \frac{1}{1 - z^{-1}} \qquad (35)$$

This corresponds to an impulse response function equal to a unit step.

Just as the continuous noise model for Brownian motion was generalized to fractional order, we can extend the discrete version as well. We propose that the discrete transfer function to be used is

$$H(z) = (1 - z^{-1})^{\beta/2} \qquad (36)$$

This model was first independently proposed by Hosking [13] and called fractional differencing.

This model resolves all of the difficulties of the other generation methods. As we will see, the high frequency problems are avoided because the discrete transfer function results in a well-defined impulse response function for all times. The process defined is both non-stationary and scale invariant satisfying the second and third criteria. A change in the time interval for the transfer function results only in an amplitude change. The discrete spectral density is found from Eq. (18):

$$S_x^d(f) = 2Q^d\tau_0[2\sin(\pi f \tau_0)]^{\beta} \qquad (37)$$

where Q^d is the variance of the i.i.d. input white noise sequence. For frequencies well below the Nyquist frequency ($1/(2\tau_0)$), this can be approximated as

$$S_x^d(f) \cong 2Q^d\tau_0^{\beta+1}(2\pi f)^{\beta} \qquad (38)$$

Thus this process produces the desired shape in the spectral density as per the fourth criterion. This equation also demonstrates the scale invariance and provides a method for choosing Q^d to properly fit the observed continuous autospectral density:

$$Q^d = \frac{g_\beta}{2(2\pi)^{\beta}\tau_0^{\beta+1}} = \frac{h_\alpha}{2(2\pi)^{\alpha}\tau_0^{\alpha-1}} \qquad (39)$$

The discrete pulse response function, h_k, associated with this model can now be found by the power series expansion of the transfer function

$$H(z) = 1 - \frac{\beta}{2}z^{-1} - \frac{\beta/2\,(1-\beta/2)}{2!}z^{-2} + \ldots \qquad (40)$$

The pulse response corresponds to the coefficients of the series. Thus $h_0 = 1$ and

$$h_k = \frac{(-\beta/2)_k}{k!} = \prod_{n=1}^{k}\frac{n-1-\beta/2}{n} \qquad (41)$$

where $(-\beta/2)_k$ is Pochhammer's symbol and is the kth derivative of $(1-x)^{\beta/2}$ evaluated at $x = 0$.

These impulse response values have a recursive relation:

$$h_0 = 1$$
$$h_k = (k-1-\beta/2)\frac{h_{k-1}}{k} \qquad (42)$$

For even-integer β this model can be shown to be exactly the traditional equations for white and random walk phase and frequency noise.

Finally we can use Eq. (17) to find the autocorrelation function for stationary noise ($\beta > -1$) as k goes to infinity. We find it to be

$$R_x^d(m) = \frac{g_\beta\,\Gamma(m-\beta/2)\,\Gamma(1+\beta)}{2(2\pi)^{\beta}\,\tau_0^{\beta+1}\,\Gamma(m+1+\beta/2)\,\Gamma(1+\beta/2)\,\Gamma(-\beta/2)} \qquad (43)$$

Again note that this direct discrete model has solved the bandwidth problem mentioned earlier. For m = 0, where the continuous model is undefined, the discrete model has a finite and well-defined variance. For large m, (or, equivalently, large τ), where the continuous model is valid, Eq. (43) can be shown to converge to exactly samples of Eq. (33), as required by Criterion 1 [1].

VI. RESULTS

The generation routine described in this paper was programmed onto a computer (see the Appendix) and used to generate discrete noise sequences. The noise sequences were examined in both the frequency domain and in the time domain. The time domain measure used to characterize the sequences is the Allan (or two-sample) variance [12]. It is defined by

$$\sigma_y^2(\tau) = \frac{1}{2\tau^2}\langle[x(t+2\tau) - 2x(t+\tau) + x(t)]^2\rangle \qquad (44)$$

This expression bears some resemblance to the autocorrelation function and can be expressed as a linear combination of autocorrelation functions. We therefore use the discrete autocorrelation function (43) to derive the expected value of the Allan variance for noise generated with this routine [24]. This is expressed:

$$\left[\sigma_y^d(m\tau_0)\right]^2 = \frac{h_\alpha \pi \Gamma(\alpha-1)}{m^2(2\pi\tau_0)^{\alpha+1}\Gamma^2(\alpha/2)} \times \left[3 - 4\frac{\Gamma(m+1-\alpha/2)\Gamma(\alpha/2)}{\Gamma(m+\alpha/2)\Gamma(1-\alpha/2)} + \frac{\Gamma(2m+1-\alpha/2)\Gamma(\alpha/2)}{\Gamma(2m+\alpha/2)\Gamma(1-\alpha/2)}\right] \quad (45)$$

Because the discrete autocorrelation function directly samples the continuous function, and the Allan variance can be formed from the autocorrelation function, the discrete Allan variance should also directly sample the continuous model. This holds true except that the Allan variance is also a function of bandwidth and window shape. Neither of these enter into Eq. (45). This issue is discussed in more detail elsewhere [24], but in practice one can assume the routine generates noise that has been passed through a rectangular window with the cutoff frequency set at the Nyquist frequency.

As with the power spectral density, we cannot in practice find the true Allan variance, but rather only estimates for it. The estimates are formed by assuming the noise process is ergodic and replacing the ensemble average in (44) with a time average. The fully overlapping Allan variance estimate is obtained from

$$\hat{\sigma}_y^2(m\tau_0,N) = \sum_{k=1}^{N-2m} \frac{(x_{k+2m} - 2x_{k+m} + x_k)^2}{2(N-2m)(m\tau_0)^2} \quad (46)$$

This estimate is a random variable that is distributed with a mean that corresponds to Eq. (45) and a calculable variance [24].

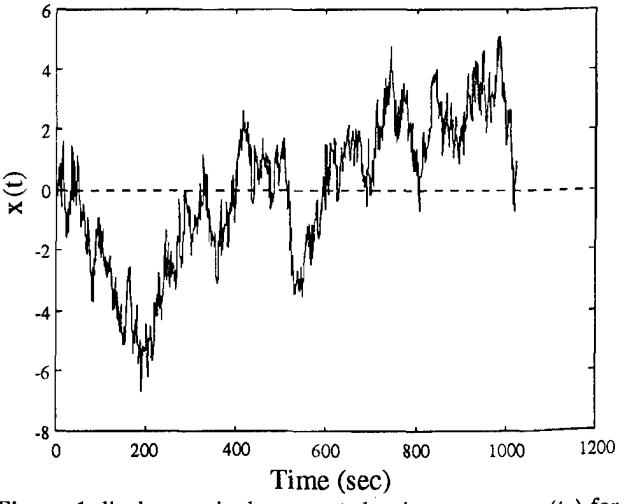


Figure 1 displays a single generated noise sequence $x(t_k)$ for the case of α=0.4, h_α=1 and τ_0=1. The total number of points generated is 1024.

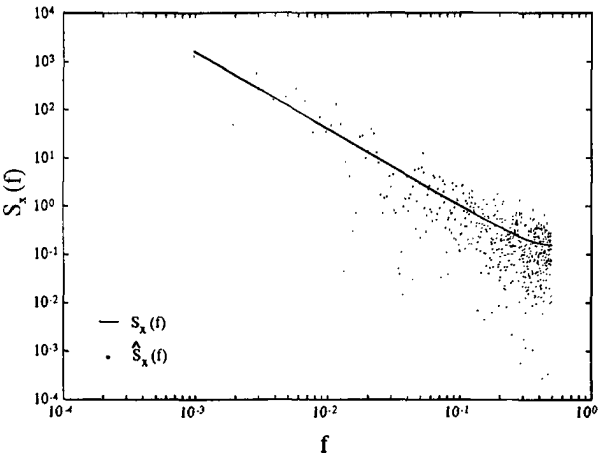


Figure 2 is a plot of the autospectral density estimate obtained via a discrete Fourier transform. The process was first multiplied by a Hanning window. The solid line is the expected autospectral density (37) for noise of the type shown in Figure 1.

An example of a discrete noise sequence can be seen in Figure 1. The time series was generated for the case of α = 0.4 and h_α = 1. A total of N = 1024 points were created, separated in time by τ_0 = 1.

This time series was multiplied by a Hanning window and discretely Fourier transformed to form the power spectral density estimate, which is plotted in Figure 2. It is evident that the spectral density estimate follows the general trend predicted by Eq. (37).

We also calculated the Allan variance estimates for this noise sequence. They are displayed in Figure 3. The error bars correspond to 90% confidence limits [24] and include the expected result.

To further examine the properties of the generated noise sequences, 100 sequences were generated for each of many values of α. For each set of sequences, the mean Allan variance estimates and the sample variance of these estimates were computed and compared to expected results. In each case not only did the means converge to values predicted by Eq. (45) (demonstrating that they are unbiased), but the sample variance of the estimates were in good agreement with predicted values [24]. Figure 4 is an example of these results. It shows the case for α = -1. There is a discrepency between the continuous model (solid line) and the discrete model (dashed line) for small values of m. This can be attributed to the different filter shape assumed in the continuous model. It is evident that the estimate means exactly follow the expected *discrete* values.

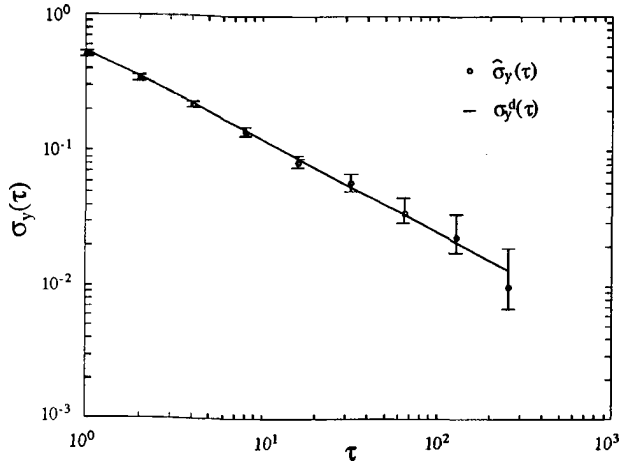


Figure 3 shows the Allan variance estimate formed for the sequence in Figure 1. The solid line is the expected discrete Allan variance (45). The error bars correspond to the 90% confidence intervals using a chi-square distribution.

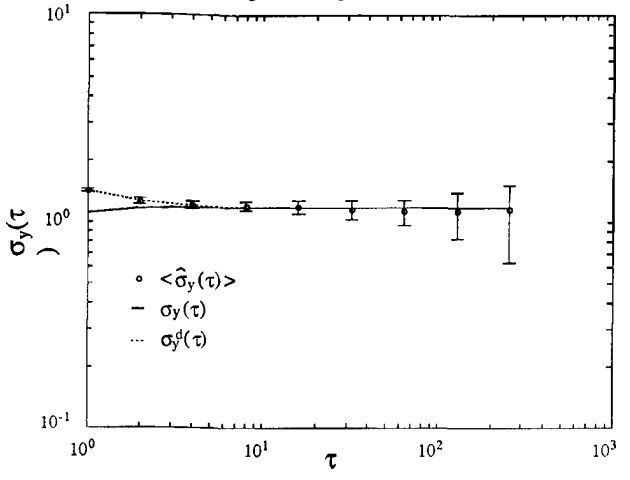


Figure 4 is a plot of the Allan variance and mean Allan variance estimates for α=-1. The dashed line depicts the discrete Allan variance (45) for the case of h_α=1 and τ_0=1. The solid line depicts the continuous Allan variance for the same case with the cutoff set at the Nyquist frequency. The open circles represent the mean Allan variance estimates for 100 generated noise sequences. The sample variances of the 100 generated Allan variance estimates are represented by the error bars.

VII. CONCLUSIONS

In this paper we have argued that it is the time domain measure (autocorrelation function) of noise that is the proper measure used to judge simulated stochastic processes. This is contrary to the more common practice of using the frequency domain measure (power spectral density). We then clearly stated the criteria evaluating power law noise simulators and presented a new method of noise generation that satisfies all of these criteria. Thus the method presented generates processes which are non-stationary ($\beta \leq -1$), scale invariant and that have the desired noise shape. We have also shown that previous methods violate these criteria in several ways.

The routine presented in this paper was tested for a wide variety of conditions, and in each case the output conformed to theoretical expectations. Thus it is possible to use this routine to generate noise with a predetermined Allan variance. The routine is both simple and quick and it can be used to generate noise with both integer *and non-integer* values of α.

This generation routine is beneficial to all clock users who are interested in studying the effects of power law noise. Computer simulation is a powerful technique for testing software used to characterize frequency stability.

VIII. REFERENCES

[1] N. J. Kasdin, "Discrete Simulation of Colored Noise and Stochastic Processes and $1/f^\alpha$ Power Law Generation," in preparation.

[2] J. A. Barnes, et. al., "Characterization of Frequency Stability," *IEEE Trans. Instr. Meas.*, Vol. IM-20, pp. 105-120, May 1971.

[3] N.J. Kasdin, "Precision Pointing Control of the Spinning Gravity Probe B Spacecraft," Ph.d. Dissertation, Stanford University, March 1991, SUDAAR 606.

[4] M.S. Keshner, "1/f Noise" *Proceedings of the IEEE,* Vol. 70, No.3, p. 211, March, 1982.

[5] B.B. Mandelbrot and R.F. Voss, "Why is Nature Fractal and When Should Noises be Scaling?" *Noise in Physical Systems and 1/f Noise*, M. Savelli, G. Lecoy and J-P Nougier, eds. Proceedings of the 7th International Conference on 'Noise in Physical Systems' and the 3rd International Conference on '1/f Noise', Montpellier, May 17-20, 1983. Amsterdam: North Holland Publishing.

[6] B.B. Mandelbrot, B.B., "Some Noises with 1/f Spectrum and a Bridge Between Direct Current and White Noise," *IEEE Transactions on Information Theory*, Vol. IT-13, No. 2, p. 289, April, 1967.

[7] B.B. Mandelbrot and J.W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, Vol. 10, No. 4, p. 422., October, 1968.

[8] V. Radeka, "1/lfl Noise in Physical Measurements," *IEEE Transactions on Nuclear Science*, Vol. 16, p. 17, 1969.

[9] R.F. Voss, "Linearity of 1/f Noise Mechanisms," *Physical Review Letters*, Vol. 40, No. 14, p. 913, 3 April 1978.

[10] R.F. Voss, "Random Fractals: Characterization and Measurement," *Scaling Phenomena in Disordered Systems*, edited by Roger Pynn and Arne Skjeltorp. New York: Plenum Press, 1985.

[11] J. A. Barnes and D. W. Allan, "A Statistical Model of Flicker Noise," *Proc. IEEE*, Vol 54, pp. 176-178, February 1966.

[12] D. B. Sullivan, D. W. Allan, D. A. Howe and F. L Walls, *Characterization of Clocks and Oscillators*, NIST Technical Note 1337, 1990

[13] J.R.M. Hosking, "Fractional Differencing", *Biometrika*, Vol. 68, No.1, pp. 165-176, 1981.

[14] Papoulis, *Probability, Random Variables, and Stochastic Processes.* New York: McGraw-Hill, Inc., 1984.

[15] K.Y.R. Billah and M. Shinozuka, "Numerical Method for Colored-Noise Generation and Its Application to a Bistable System," *Physical Review A*, Vol. 42, Number 12, pp. 7492-7495, 15 December 1990.

[16] H.O. Peitgen and D. Saupe, eds., *The Science of Fractal Images.* New York: Springer-Verlag, 1988.

[17] J. Feder, *Fractals.* New York: Plenum Press, 1988.

[18] T.T. Soong, *Random Differential Equations in Science and Engineering*, New York: Academic Press, 1973.

[19] A.V. Oppenheim, and R.W. Schafer, *Discrete-Time Signal Processing.* Englewood Cliffs, N.J.: Prentice Hall, 1989.

[20] P. Gruber, "1/f-Noise Generator," *Noise in Physical Systems and 1/f Noise - 1985*, A. Diamico and P. Mazzetti, eds. Proceedings of the 8th International Conference on 'Noise in Physical Systems' and the 4th International Conference on '1/f Noise', Rome, September 9-13, 1985. Amsterdam: North Holland Physics Publishing, 1986, p. 357.

[21] R. Saletti, "A Comparison between Two Methods to Generate $1/f^{\gamma}$ Noise," *Proceedings of the IEEE*, Vol. 74, No. 11, p. 1595, November, 1986.

[22] J.A. Barnes, "Simulation of Oscillator Noise," Austron, Inc., Austin, Tx.

[23] K.B. Oldham and J. Spanier, *The fractional calculus; theory and applications of differentiation and integration to arbitrary order.* New York: Academic Press, 1974.

[24] T. Walter, "Characterization of Frequency Stability: A Continuous Power Law Model with Discrete Sampling," in preparation.

IX. APPENDIX

We present a coded implementation of the method outlined in this paper. The routine requires as input: the total number of points to generate (n_pts, *which must be a power of two*), a vector, X[ ] (subscripted from 1 to n_pts), in which to add the generated noise, the appropriate discrete variance Q_d, the desired slope of the noise, b, (β) and an integer seed to initialize the psuedo-random noise generator. The routine is coded in C and makes use of algorithms from Numerical Recipes in C. The important external routines are gas_dev() which returns a zero-mean unit-variance Gaussian random number and realft() which performs the discrete Fourier transform.

```
void f_beta(int n_pts,float X[],float Q_d,
     float b,int *idum)
{
 int i,nn;
 float *hfb,*wfb;
 float mhb,wr,wi;
 nn = n_pts+n_pts;
 mhb= -b/2.0;
 Q_d=sqrt(Q_d);   /* find the deviation of the
                                  noise */
 hfb=vector(1,nn);
 wfb=vector(1,nn);
 hfb[1]=1.0;
 wfb[1]=Q_d*gas_dev(idum);
 for(i=2;i<=n_pts;i++){
/* generate the coefficients h_k */
     hfb[i]=hfb[i-1]*(mhb+(float)
         (i-2))/((float)(i-1));
/* fill the sequence w_k with white noise */
     wfb[i]=Q_d*gas_dev(idum);
 }
/* pad the ends of the sequences with zeroes */
 for(i=n_pts+1;i<=nn;i++){
     hfb[i]=0.0;
     wfb[i]=0.0;
 }
/* perform the discrete Fourier transform */
 realft(hfb,n_pts,1);
 realft(wfb,n_pts,1);
/* multiply the two complex vectors */
 wfb[1]=wfb[1]*hfb[1];
 wfb[2]=wfb[2]*hfb[2];
 for(i=3;i<=nn;i+=2){
     wr=wfb[i];
     wi=wfb[i+1];
     wfb[i]=wr*hfb[i]-wi*hfb[i+1];
     wfb[i+1]=wr*hfb[i+1]+wi*hfb[i];
 }
/* inverse Fourier transform the result */
 realft(wfb,n_pts,-1);
 for(i=1;i<=n_pts;i++)
/* add the final result to X[] */
     X[i]+=wfb[i]/((float)n_pts);
 free_vector(hfb,1,nn);
 free_vector(wfb,1,nn);
}
```

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A NEW MULTI-VARIANCE METHOD FOR THE OSCILLATOR NOISE ANALYSIS

F. Vernotte*, E. Lantz**, J. Groslambert***, J. J. Gagnepain***

*Observatoire de Besançon
**Laboratoire d'Optique P. M. Duffieux, U. A. CNRS 214
***Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.

ABSTRACT

Since the noise altering the output signal of oscillators may be modeled as power laws in the spectral density of frequency deviation, the oscillator noise analysis is the measurement of the level of each power law noise. The principle of this new multi-variance method consists in obtaining the noise type contributions with different variances and different integration time values. The whole data obtained from the different variances with the different integration times are then operated simultaneously. The most probable measurement, in the sense of the least squares, is obtained for each type of noise. Moreover, this method lends itself to an estimation of the uncertainty of the noise type contribution measurement taking into account the dispersion of the variance results.

I. INTRODUCTION

The analysis of random oscillator instabilities is a main issue in various fields. Different types of noises are generally distinguished, each of them corresponding to particular properties. The origins of these noises may be linked to the oscillator environment (temperature changes, electromagnetic fields, shocks, vibrations) but also they may be internal (thermal noise...) [1, 2]. However, the links between these causes and the different types of noises are not still well-known. The reason is the lack of accuracy of the measurement techniques of the contributions characterizing each type of noise.

The aim of the job presented in this paper consists in attempting to improve the measurement accuracy and estimating the measurement uncertainties in order to better distinguish the different types of noises. Following the presentation of the different notion, tools and mathematical models used in the frequency and time field, the classical measurement method using one single variance is recalled. The multi-variance method, based on the use of several variances, is described. Then, a method giving an assessment of the precision is proposed. The performances of this method are discussed.

II. OSCILLATOR NOISE ANALYSIS

Different behaviors of the spectral density of instantaneous normalized frequency deviation $S_y(f)$ have been experimentally observed. It is commonly assumed that 5 different spectral density patterns may be encountered in oscillator output signals according to 5 different power laws [1, 2] :

$$S_y(f) = \sum_{\alpha=-2}^{+2} h_\alpha . f^\alpha \qquad (1)$$

Each of these types of noises corresponds to a physical cause (vibrations, changes in the temperature, electronic agitation). However, the links between these causes and these effects are not well-known. The reason for this lack of knowledge is the mediocre accuracy of the measurement of the noise coefficient h_α.

The IEEE Subcommitee on Frequency Stability has proposed to choose the Allan Variance (two sample variance) as a stantard oscillator measurement method [3]. This variance is defined by the following measurement sequence :

$$\sigma_y^2(\tau) = \frac{1}{2}\left\langle (\bar{y}_2 - \bar{y}_1)^2 \right\rangle \qquad (2)$$

In the frequency domain, the Allan variance may be calculated with its transfer function $H_y(f)$ which is the Fourier Transform of its measurement sequence [1, 2] :

$$\sigma_y^2(\tau) = \int_{-\infty}^{+\infty} S_y(f) . \left|H_y(f)\right|^2 df \qquad (3)$$

where

$$\left|H_y(f)\right|^2 = \frac{2\sin^4(\pi\tau f)}{(\pi\tau f)^2} \qquad (4)$$

Several other variances have been defined by choosing new transfer functions : modified Allan variance, filtered Allan variance, high-pass variance, band-pass variance, etc... [1, 2, 4]. Since the variances exhibit different

0-7803-0476-4/92 $3.00 © 1992 IEEE

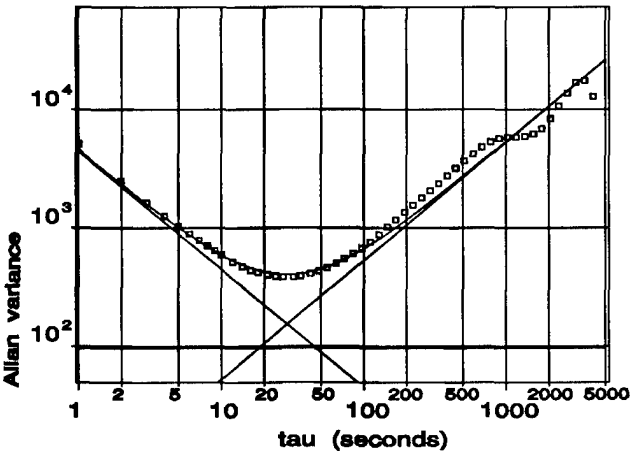


Figure 1 : Allan variance applied to a signal composed of f^0, f^{-1} and f^{-2} noises (h_0=10000, h_{-1}=10, h_{-2}=1).

slopes for the different types of noises versus the integration time τ, the principle of a variance measurement consists in calculating the variance with different integration time τ and then, in plotting those measurements on a log-log graph versus τ (see figure 1).

The identification of each slope on this graph yields the identification of each type of noises in the spectral density $S_y(f)$ of the oscillator output signal. For a f^α noise, the variance measurement graph exhibits a τ^μ slope and generally $\mu = -1-\alpha$ [1, 2] :

$$S_y(f) = h_\alpha . f^\alpha \Leftrightarrow \sigma_y^2(\tau) = C_\mu . \tau^\mu \qquad (5)$$

III. MULTI-VARIANCE MEASUREMENT

The multi-variance method consists in using N different variances for M different integration time values to measure L noise type coefficients h_α [2]. If the number N.M is greater or equal than L, it is possible to calculate the h_α coefficient set which fits at best all the variance measurements. The calculation is performed by minimizing the relative difference between the variance measurements and the theoretical variance curve. The result is then the most probable h_α coefficients in the sense of the least squares. Let us consider N different variances and M different τ values. According to the usual power law model, we will assume in the following that the number of noise types L is equal to 5. Denoting $\sigma^2_i(\tau_k)$ the result of the variance i for the integration time τ_k $\phi_{i\alpha}(\tau_k)$ the response of the variance i for a f^α noise and $\varepsilon_i(\tau_k)$ the dispersion of the variance measurements, it comes :

$$\left\{\begin{array}{l}
\left\{\begin{array}{l}
\sigma_1^2(\tau_1) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{1\alpha}(\tau_1) + \varepsilon_1(\tau_1) \\
\quad \cdots \\
\sigma_i^2(\tau_1) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{i\alpha}(\tau_1) + \varepsilon_i(\tau_1) \\
\quad \cdots \\
\sigma_N^2(\tau_1) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{N\alpha}(\tau_1) + \varepsilon_N(\tau_1)
\end{array}\right. \\
\quad \cdots \\
\left\{\begin{array}{l}
\sigma_1^2(\tau_k) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{1\alpha}(\tau_k) + \varepsilon_1(\tau_k) \\
\quad \cdots \\
\sigma_i^2(\tau_k) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{i\alpha}(\tau_k) + \varepsilon_i(\tau_k) \\
\quad \cdots \\
\sigma_N^2(\tau_k) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{N\alpha}(\tau_k) + \varepsilon_N(\tau_k)
\end{array}\right. \\
\quad \cdots \\
\left\{\begin{array}{l}
\sigma_1^2(\tau_M) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{1\alpha}(\tau_M) + \varepsilon_1(\tau_M) \\
\quad \cdots \\
\sigma_i^2(\tau_M) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{i\alpha}(\tau_M) + \varepsilon_i(\tau_M) \\
\quad \cdots \\
\sigma_N^2(\tau_M) = \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{N\alpha}(\tau_M) + \varepsilon_N(\tau_M)
\end{array}\right. \\
\quad \cdots
\end{array}\right. \qquad (6)$$

The condition for the minimization of the relative difference between the variance measurements and the theoretical variance response is then :

$$\sum_{k=1}^{M}\sum_{i=1}^{N}\left[\frac{\sigma_i^2(\tau_k) - \sum_{\alpha=-2}^{+2} h_\alpha \Phi_{i\alpha}(\tau_k)}{\sigma_i^2(\tau_k)}\right]^2 \text{ is minimum} \qquad (7)$$

Minimization of expression (7) is obtained in the usual way of the least squares method.

IV. ASSESSMENT OF THE PRECISION

As the number of samples is finite, variances are estimated and these estimations can fluctuate. Lesage and Audoin [5] calculated the variance of the Allan variance. Howe, Allan and Barnes [6] showed that Allan variance is chi-square distributed and they calculated the number of degrees of freedom for each type of noise. However, uncertainties on coefficients of noise have not been calculated yet. This problem is addressed here by using a singular value decomposition [7]. The method

was at first developed for the assessment of the precision of optical measurements [8] and has been recently adapted to the oscillator noise analysis [9].
The result is a quantitative assessment of the relative uncertainty dh_α/h_α on each noise coefficient. It is expressed in terms of the variance fluctuations and of the matrix derived from the variance responses. The principal theoretical results can be summarized as follows : great uncertainties on a noise coefficient can be due to :

- great fluctuations of the variances (see references [5] and [6] for the statistical study of these fluctuations) ;
- a small effect of a noise on the variance responses (in the case of a noise type swamped by the other ones [10]) ;
- similar effects of two noises on variance responses.

This third reason was mentioned in the case of Allan variance [4] : flicker phase noise and white noise can not easily be separated with this variance. We can calculate the angle between the vectors corresponding to the variance responses of these noises. We will see in the next paragraph that this angle is small. Such a small angle points out an instability of the inverse transformation from the variance responses to the noise coefficients.

V. RESULTS AND DISCUSSION

In order to check the efficiency of this method, signals of which noise coefficients are perfectly known must be used. Random sequences were generated with a computer and then filtered to obtain different types of noises according to the entered h_α coefficients.
Before to consider the results, let us redefine several concepts in the multi-variance framework. For the accuracy, we will distinguish :

- the *measurement errors*, $\Delta h_\alpha/h_\alpha$, the relative difference between the entered coefficients and the measured coefficients ;
- the *measurement uncertainties*, dh_α/h_α, obtained by the method described in chapter IV.

We must also define :

- the *dynamics*, as the range in which a h_α coefficient can vary without swamping the other ones nor being swamped by the other ones ;
- the *separability*, as the aptitude of a variance to distinguish two noises.

Table I compares the measurement errors for the same signal analysed with the Allan variance, with the modified Allan variance, with the high-pass variance, with the band-pass variance and simultaneously with these 4 variances (multi-variance measurement). We see

Noise coefficient error / Variance name	$\frac{\Delta h_{-2}}{h_{-2}}$	$\frac{\Delta h_{-1}}{h_{-1}}$	$\frac{\Delta h_0}{h_0}$
Allan variance	3.8 %	35 %	0.6 %
Modified Allan variance	5.4 %	86 %	2.7 %
High-pass variance	3.3 %	27 %	0.07 %
Band-pass variance	4.3 %	128 %	2.7 %
Multi-variance (4 variances)	2.4 %	6.9 %	0.05 %

Table I : Measurement errors obtained with different single variances and with a multi-variance system (4 variances) applied to a simulated signal (h_0=10000, h_{-1}=10, h_{-2}=1).

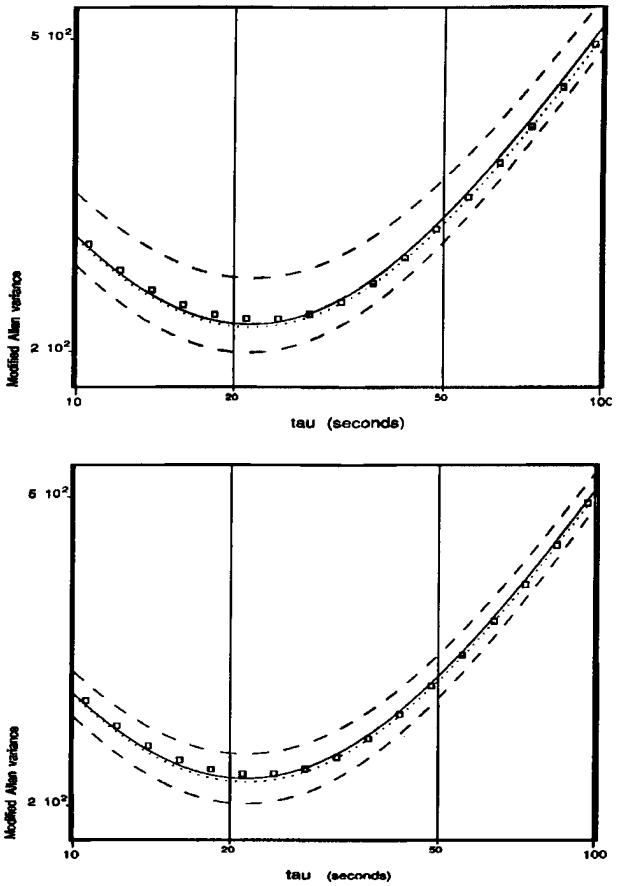


Figure 2 : Modified Allan variance result curves for a signal composed of f^0, f^{-1} and f^{-2} noises (h_0=10000, h_{-1}=10, h_{-2}=1).
Measurement obtained with :
A. (above) only the modified Allan variance
B. (below) a set of 4 variances.
The solid line is the variance curve obtained with the measured h_α, the dotted line is obtained with the entered h_α, the discontinuous lines represent the uncertainty bounds.

Noise type	Entered h_α value	Measured h_α value	Error	Uncertainty	Angle with			
					f^{-2}	f^{-1}	f^0	f^{+1}
f^{-2}	1	0.56	44 %	29 %	0 °	54 °	87 °	89 °
f^{-1}	1000	1130	13 %	12 %	54 °	0 °	58 °	72 °
f^0	10000	7530	25 %	58 %	87 °	58 °	0 °	24 °
f^{+1}	0	6910	100 %	230 %	89 °	72 °	24 °	0 °

f^{-2}	1	0.59	41 %	25 %	0 °	54 °	87 °	90 °
f^{-1}	1000	1090	9.3 %	9.4 %	54 °	0 °	58 °	90 °
f^0	10000	9350	6.5 %	16 %	87 °	58 °	0 °	90 °
f^{+1}	0	0	0 %	0 %	90 °	90 °	90 °	0 °

***Table II** : Results obtained with 4 variances applied to a simulated signal (h_0=10000, h_{-1}=1000,h_{-2}=1) :*
***A.** (above) an erroneous f^1 noise alters the measurement*
***B.** (below) after elimination of h_1 coefficient.*

that the accuracy of the multi-variance measurement is greater than the one of a single variance measurement. The uncertainty measurement is then lower (see figure 2). Moreover, if the sensitivity of the variances choosed for the multi-variance measurement are very different (for instance variances sensitive to low frequency noises mixed with variances sensitive to high frequency noises), the dynamics of the multi-variance measurement increases. The dynamics of the previously described 4 variance system has been estimated at 40 dB for the determination of h_{-1} in a signal composed of f^2, f^1 and f^0 noise (8192 samples) instead of 15 dB for the Allan variance [10].

The second advantage of the multi-variance method concerns the possibility of eliminating a noise coefficient measurement which has a very small influence in the variance curve. Table II.A presents the measurements performed on a signal composed of f^2, f^1 and f^0 noise types. The multi-variance method finds a f^1 component but the uncertainty of its coefficient is very large (>200%). Moreover, the angle between the eigen vectors corresponding to h_0 and h_1 is about 20°. These two noises are then not well separated and a part of the effect of the f^0 noise is assigned to an erroneous f^1 noise. The elimination of the contribution of this f^1 noise improve the measurement accuracy of the other coefficients, particulary the h_0 coefficient (see table II.B). The reconstructed variance curve is not very affected by

Noise type	Entered h_α value	Measured h_α value	Error	Uncertainty	Angle with				
					f^{-2}	f^{-1}	f^0	f^{+1}	f^{+2}
f^{-2}	1	1.45	45 %	22 %	0 °	28 °	63 °	84 °	87 °
f^{-1}	1486	852	43 %	63 %	28 °	0 °	41 °	74 °	82 °
f^0	355800	475 10^3	34 %	27 %	63 °	41 °	0 °	46 °	61 °
f^{+1}	16650000	13.3 10^6	20 %	37 %	84 °	74 °	46 °	0 °	22 °
f^{-2}	149800000	160 10^6	7.2 %	13 %	87 °	82 °	61 °	22 °	0 °

f^{-2}	1	1.52	52 %	17 %	0 °	33 °	69 °	87 °	90 °
f^{-1}	1486	795	47 %	52 %	33 °	0 °	44 °	77 °	88 °
f^0	355800	476 10^3	34 %	19 %	69 °	44 °	0 °	48 °	76 °
f^{+1}	16650000	13.8 10^6	17 %	24 %	87 °	77 °	48 °	0 °	39 °
f^{+2}	149800000	157 10^6	4.9 %	11 %	90 °	88 °	76 °	39 °	0 °

***Table III** : Results obtained with the Allan variances (above) and with the modified Allan variance (below) applied to a simulated signal (h_{-2}=1, h_{-1}=1486, h_0=355800 h_1=16650000, h_2=149800000).*

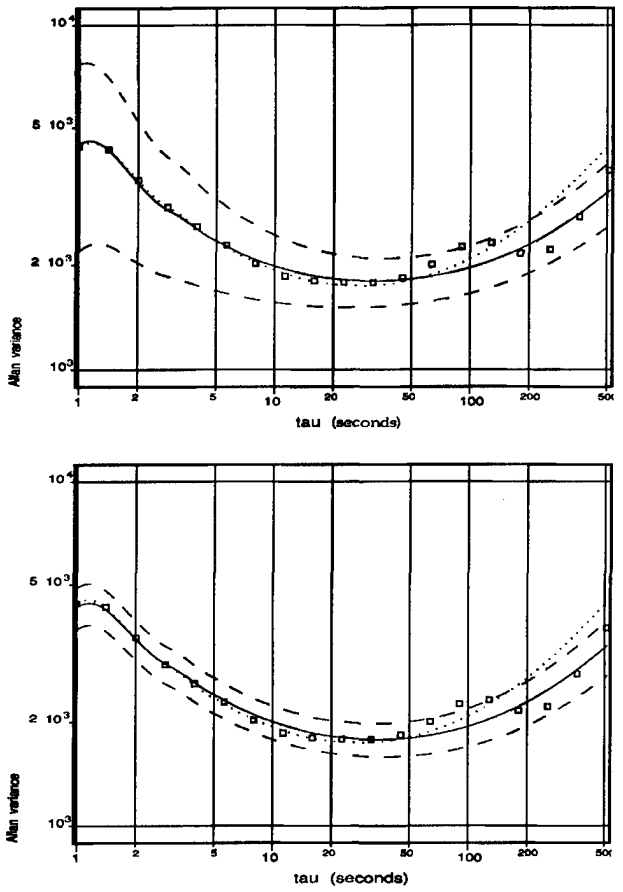


Figure 3 : Allan variance result curves for a signal composed of f^0, f^{-1} and f^{-2} noises (h_0=10000, h_{-1}=10, h_{-2}=1).
Measurement obtained with a set of 4 variances :
***A.** (above) a erroneous f^1 noise alters the measurement*
***B.** (below) after elimination of h_1 coefficient.*
In A the uncertainties are overestimated because the discontinuous curves do not take into account the compensation of the errors between h_0 and h_1.

this modification, but the uncertainties decrease widely (see figure 3). Therefore, the uncertainty estimations, as well as the determination of the angles between the eigen vectors of the different noise types, yield a criterion to eliminate erroneous noise coefficients.

At last, the angle measurement can quantify the separability of the variances. It is well known that the Allan variance does not distinguish easily f^1 and f^2 noises (τ^{-2} slope for both). Allan and Barnes solved this problem with the modified Allan variance (τ^{-2} slope for f^1 and τ^{-3} slope for f^2) [11]. Table III shows the measurements performed over the same signal by the Allan variance and by the modified Allan variance : the uncertainty are smaller with the modified Allan variance, and furthermore the angle between the eigen vectors of the f^1 and f^2 noises is two times greater (40°) with the modified Allan variance than with the Allan variance (20°). Then, the angle determination seems to be an useful tool to estimate the separability of a variance for a couple of noise types.

We can note on table II and III that uncertainties are compatible with actual measurement errors. However, the uncertainties seem underestimated for the h_{-2} coefficient. The estimation of this coefficient is very sensitive to the very first frequency samples. Strong differences appear between variances obtained from several random sequences with the same entered values of noise coefficients. A further study of the specific properties of this coefficient seems necessary.

VI. CONCLUSION

The multi-variance method presents several advantages :

- firstly, its accuracy and its dynamics depends on the number and on the type of the variances choosed and they are at least as well as the ones of each variances. It is possible to use only one variance for a quick estimation, or a set of 5 or more variances for a comprehensive analysis.
- secondly, the determination of the angles between the eigen vectors of the noise types yields a quantitative estimation of the separability of a variance or of a set of variances.
- finally, the multi-variance method gives a criterion to estimate the measurement reliability and to obtain only types of noise that do exist in the signal.

REFERENCES

[1] J. Rutman, "Characterization of phase and frequency instabilities in precision frequency sources : fifteen years of progress", *Proceedings of the IEEE*, vol. 66, no. 9, pp. 1048-1075, September 1978.

[2] F. Vernotte, "Stabilité temporelle des oscillateurs : nouvelles variances, leurs propriétés, leurs applications", *Ph. D. Thesis*, Université de Franche-Comté, N# 199, Besançon, France, February 1991.

[3] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McCunigal, J. A.

Mullen, W. L. Smith, R. L. Sydnor, R. Vessot and G. M. R. Winkler, "Characterization of frequency stability", *IEEE Trans. Instrum. Meas.*,vol. IM-20, pp. 105-120, 1971.

[4] J. Groslambert, J. J. Gagnepain, F. Vernotte and F. Walls, "A new filtered Allan variance and its application to the identification of phase and frequency noise sources", *43rd annual Frequency Control Symposium*, Denver, USA, 1989.

[5] 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.

[6] D. A. Howe, D. W. Allan and J. A. Barnes, "Properties of signal sources and measurement methods", *Proceedings of the 35th annual Frequency Control Symposium*, Ft. Monmouth, USA, May 1981.

[7] A. Lannes, S. Roques and M. J. Casanove, "Stabilized reconstruction in signal and image processing 1 : partial deconvolution and spectral extrapolation with limited field", *Journal of modern Optics*, vol. 34, no 2, pp. 161-226, 1987.

[8] E. Lantz and J. Duvernoy, "Stability of model and selection of parameters : application to metrology in optical microscopy", *Journal of modern Optics*, vol. 36, no 9, pp. 1213-1226, 1989.

[9] F. Vernotte, E. Lantz, J. Groslambert and J. J. Gagnepain, "Oscillator noise analysis : multi-variance measurement", to be published in the *Proceedings of the CPEM 92*, Paris, June 1992.

[10] F. Vernotte and E. Lantz, "Time stability : an improvement of the multi-variance method for the oscillator noise analysis, to be published in the *Proceeding of the EFTF 92*, Noordwijk, Netherlands, March 1992.

[11] D. W. Allan and J. A. Barnes, "A modified "Allan variance" with increased oscillator characterization ability, *Proceedings of the 35th annual Frequency Control Symposium*, Ft. Monmouth, USA, May 1981.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

SECONDARY STANDARD FOR PM AND AM NOISE AT 5, 10, and 100 MHz

F. L. Walls
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

This paper describes a practical implementation of a portable secondary standard for phase modulation (PM) and amplitude modulation (AM) noise at 5, 10, and 100 MHz. The accuracy of the standard for both PM and AM noise is ± 0.14 dB, and the temperature coefficient is less than 0.02 dB/K. The noise floor S_ϕ(10 kHz) of the standard for PM noise measurements, is less than -190 dBC rel to 1 rad^2/Hz at 5, 10, and 100 MHz. The noise floor for AM measurements depends on the configuration. A calibrated level of PM and AM noise of approximately -130 ± 0.2 dB rel to 1 rad^2/Hz (for Fourier frequencies from approximately 1 Hz to 10 % of the carrier frequency) is used to evaluate the accuracy versus Fourier frequency. Similar PM/AM noise standards are under test at 10 GHz. This new standard can also be used as an alternative to the normal method of calibrating the conversion sensitivity of the PM/AM detector for PM/AM measurements. Some types of time-domain measurement equipment can also be calibrated.

Introduction

In the past it has been difficult to evaluate the accuracy and noise floor of phase modulation (PM) and amplitude modulation (AM) noise measurement equipment because no artifact standards were available. Comparisons of measurement systems in different laboratories using commercially available oscillators as transfer standards were typically limited to a repeatability of roughly ± 3 dB due to the temporal variability of the oscillator noise. This paper describes a practical implementation of a portable secondary standard and associated measurement techniques for evaluating the noise floor and accuracy versus Fourier frequency of phase modulation (PM) and amplitude modulation (AM) noise measurement systems [1]. Evaluations of these new PM/AM noise standards at 5, 10, and 100 MHz yield an accuracy of less than 0.2 dB, a temperature coefficient of less than 0.02 dB/K, and a stability of less than 0.4 dB over 1 y. This new PM/AM noise standard can also be used as an alternative to the normal method of calibrating the conversion sensitivity of the PM/AM detector for PM/AM measurements. Some types of time-domain measurement systems can also be calibrated [2-5].

This new PM/AM noise standard produces two high level signals that have exceptionally low residual differential PM and AM noise between them. At 10 MHz, for example, the differential phase noise between the two channels, S_ϕ(10 kHz) = -194 dB rel to 1 rad^2/Hz. This configuration is used for measuring the noise floor of both PM and AM measurement systems.

A calibrated level of passband-limited Gaussian noise can be added to one of the signals to create equal levels of PM and AM noise. The level of added noise is roughly 40 dB above the noise floor of most measurement systems and 60 dB above the residual noise between the two signals. This calibrated noise level is typically constant in magnitude to ± 0.2 dB for Fourier frequencies from dc to 10% of the carrier frequency. This configuration is used to evaluate the accuracy of PM/AM noise measuring equipment as a function of Fourier frequency.

When the AM noise is measured on a single channel, the differential technique applied to S_ϕ(f) does not help to extend the range. The AM noise of the signal contains both the original AM noise of the source after the leveling circuit and the calibrated AM noise. Thus, it might be expected that the spectrum of S_a(f) at the output would rise much faster than the PM noise as f approaches 0. Fortunately, most sources have much lower AM noise than PM noise close to the carrier, so this potential problem is usually eliminated.

When the AM noise is measured using two detectors in a differential mode, most of the AM noise of the

Contribution of the U. S. Government, not subject to copyright.

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

source cancels. This technique, as compared to the single detector methods significantly lowers the contribution of AM noise of the source to the noise floor for AM measurements.

Design of PM/AM Noise Standard

Figure 1 shows a simplified block diagram of the PM/AM noise standard model PMAM 115. A frequency source with very low PM and AM noise is regulated in amplitude and divided into the reference and signal outputs using a passive power splitter or directional coupler. (Active power splitters generally add more PM noise than passive splitters.) A switch turns the carrier signal "on" and "off". The amplitude and phase of these outputs track one another with great fidelity. The residual differential PM noise between the two outputs is typically much smaller than the resolution of available measurement systems [6,7].

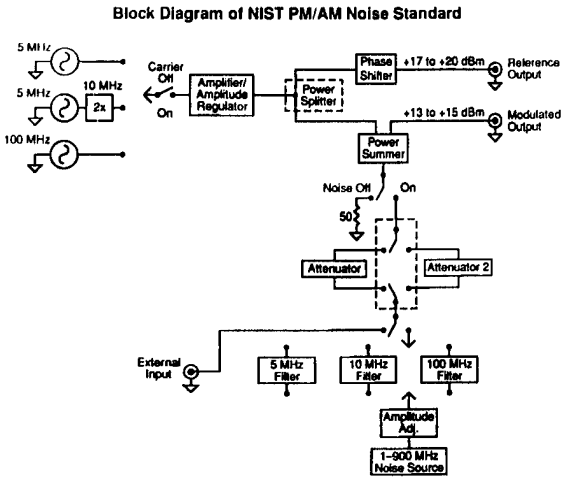


Fig. 1 Block diagram of Calibration Standard PMAM 115 for PM and AM noise.

A broadband power combiner is used to add passband-limited Gaussian noise to a carrier (see Fig. 1). Ideally the noise source is very flat over the Fourier frequency range of interest about the carrier (ν_0-f to ν_0+f) and is independent of temperature. Another switch changes the noise level by inserting known attenuators. This results in a spectral density of PM noise, $S_\phi(f)$ given by

$$S_\phi(f) = \frac{PSDV_N(\nu_0 - f) + PSDV_N(\nu_0 + f)}{2V_0^2}, \qquad (1)$$

where ν_0 is the carrier frequency, V_0 is the amplitude of the carrier, and $PSDV(\nu_0 \pm f) \equiv V_N^2(\nu_0 \pm f)$ per Hz is the power spectral density of voltage noise at frequency $\nu_0 \pm f$. The level of $S_\phi(f)$ is constant from dc to approximately the half-bandwidth of the filter, assuming that the added noise is constant in amplitude from ν_0 - f to ν_0 + f. A switch connects either a 50 Ω termination or the output of the filtered noise source to the summing device. The summing device linearly adds the noise from either the 50 Ω termination or the filtered noise source with the signal to create a modulated output. Since there is no phase coherence between the signal and the noise, the resulting modulated output has precisely equal AM and PM noise with the condition given by Eq. (2) (assuming negligible amplitude compression in the power summer which is easily attained) for the same ranges in analysis frequencies about the carrier [5-7]. These noise spectra can be made very nearly constant over a very wide temperature range by stabilizing the carrier and the noise separately using traditional approaches. The level of the added noise is such that

$$\int_0^\infty S_\phi(f) < 0.1. \qquad (2)$$

This ensures that the compression of the measured $S_\phi(f)$ is smaller than 0.04 dB [5-7]. The residual differential PM noise between the two signals when the noise source is off is generally much smaller than the PM noise of the source. The ratio of the differential PM noise to the source PM noise can approach -100 dB at low Fourier frequencies, degrading to approximately -16 dB at $f = \nu_0/10$ [7]. This insensitivity to source noise improves the accuracy of the PM standard, especially at low Fourier frequencies where the PM noise of the source becomes important.

Although not necessary in some measurement configurations, a phase shifter can be used to provide a 90° phase shift between the reference and modulated output. The phase shifter can be on either the signal or reference side of the power summer. The most common form of phase shifter is a length of transmission line although many other types of phase shifters are possible.

Measuring $PSDV_N$ and V_0^2 separately calibrates the level of $S_\phi(f)$ at the modulation output relative to the reference signal, as specified by Eq. (1). Turning the filtered noise off allows good resolution and accuracy in the measurement of V_0^2 by eliminating or greatly reducing V_N. Turning the noise on and the carrier off optimizes the resolution and dynamic range for measuring $PSDV_N$. Usually the attenuation of the

noise signal is set to a minimum for this measurement. The components are chosen to minimize the voltage-standing-wave-ratio (VSWR) so that changing the state of the various switches does not significantly alter the impedance or phase of the output to the measurement system. The attenuators are calibrated using a source routed through the external port.

Implicit in these discussions is the assumption that the phase detector in the PM noise measurement system under test has sufficient discrimination against AM noise and that the amplitude detector in the AM noise measurement system has sufficient discrimination against PM noise. These assumptions are not excessively restrictive since many devices to be calibrated also have similar PM and AM noise [7]. Discrimination of 15 dB reduces the unwanted effects below 0.14 dB. This discrimination level is easily met by virtually all AM and PM noise measurement techniques in use today [6,7] (typical levels of discrimination are 25 dB for PM and even higher for AM measurement systems.)

Calibration Considerations

A. Calibration of PM Noise in PM/AM Noise Standard

If a scanning narrow-band receiver is used to measure the relative value of V_0^2 and $PSDV_N$ on the modulated output, only the equivalent noise bandwidth BW of the receiver, and its accuracy for relative measurements from the level of V_0^2 to $(PSDV_N)BW$ must be calibrated to determine $S_\phi(f)$. The form of Eq. (1) is such that small errors in the alignment of the noise passband filter resulting in odd-order variations of $V_N^2(\nu)$ about ν_0 are averaged away. The only limit in setting $S_\phi(f)$ is that $S_\phi(f)$ should be high enough that it is far above the noise floor of the measurement system and yet small enough to satisfy Eq. (2) (15 dB margin reduces the errors to approximately 0.14 dB). The smaller the bandwidth for $S_\phi(f)$, the higher the level can be set without violating Eq. (2). Typically $S_\phi(f)$ is constant for $0 < f < \nu_0/10 \leq 1$ GHz and the integrated phase noise is in the range from 10^{-2} to 10^{-4} rad^2. We must also check that all of the amplifiers used in the evaluation are operated in the linear gain region. The error in the calibration for the circuit in Fig. 1 is typically 0.05 dB due to random errors and $\pm$ 0.05 dB due to uncertainties in the linearity of the receiver, 0.04 dB due to uncertainties in determining the bandwidth, yielding an overall root-sum-square (RSS) uncertainty of $\pm$ 0.14 dB.

B. Measurement of Errors in PM Noise Measurement Systems

Figure 2 shows the preferred method for determining the accuracy of a PM noise measurement system using the new PM/AM noise standard [5-7]. As explained above, the PM noise of the driving source cancels out to a very high degree. This approach, requires that the phase noise measurement system be able to adjust the phase between the two signals at the mixer to 90° so that the output of the mixer is close to O V. Not all commercial units have this provision, and some may require the use of phase shifter in the phase noise standard. The noise floor of the measurement system is determined with noise modulation off and the calibrated PM noise of the standard is measured with the noise modulation on. Differences between the measurement of $S_\phi(f)$ and the measurement determined using the method in Section IV A. represent errors in the PM noise measurement system under test. The contribution of the noise floor of the PM noise measurement generally needs to be considered only near the carrier, (that is, as f approaches 0 Hz) [7].

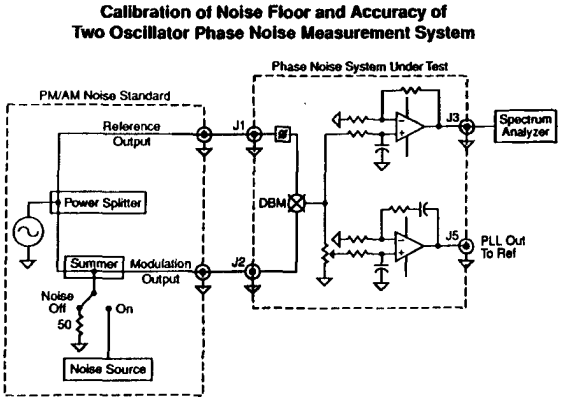


Fig. 2 Preferred calibration method for two oscillator type phase noise measurement systems.

A configuration similar to Fig. 2 is useful for measuring the PM noise added by elements such as frequency multipliers and amplifiers. The element to be tested is inserted in one or both of the signal paths from the PM/AM noise standard to the PM measuring system. PM measurements with the calibrated noise modulation on are used to calibrate the system sensitivity for converting phase fluctuations to voltage fluctuations versus Fourier frequency. PM measurements with the noise

modulation off yield the PM noise (noise floor of the system plus the added noise of the element under test) relative to the calibrated PM noise.

Another method of calibration is shown in Fig. 3, where a second oscillator is used. The PM noise of both oscillators and the calibrated PM noise are then detected in the measurement. This approach can also lead to accurate evaluations of the measurement system errors versus Fourier frequency if the oscillator PM noise is small enough [7]. The AM/PM noise standard accommodates this approach by providing oscillators with very low residual AM and PM noise and the ability to turn the calibrated PM noise off to check the noise floor of the measurement. As mentioned above, the measurement sequence is to measure the residual PM noise in the measurement with the calibrated PM noise off and then to measure the PM noise with the calibrated PM noise on. From these two measurements we can determine the contributions of the uncancelled PM noise of the oscillators and the measurement system and their contribution to the measurement of the PM noise standard. Unlike the measurement technique outlined in Fig. 2, it is not generally possible to verify the noise floor of the measurement system because the phase noise of the reference oscillators always contributes to the measured noise. Also the measurement cannot be carried out at small values of f since the PM noise of the reference oscillators dominates over the added noise. This problem can be alleviated somewhat by increasing the added noise and decreasing the bandwidth to satisfy Eq. (2).

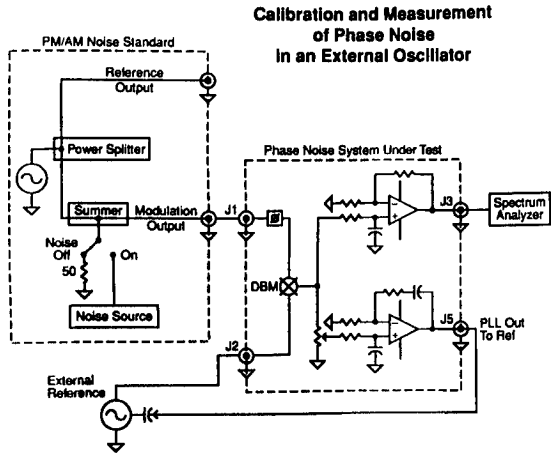


Fig. 3 Calibration method for two oscillator phase noise measurement systems.

The configuration of Fig. 3 can be used to measure the phase noise of an external oscillator. PM measurements with the calibrated noise modulation on are used to calibrate the system sensitivity for converting phase fluctuations to voltage fluctuations versus Fourier frequency. PM measurements with the noise modulation off yield the PM noise (noise floor of the system plus the added noise of the oscillator under test) relative to the calibrated PM noise.

The PM/AM noise standard can also calibrate the accuracy of single oscillator (frequency discriminator) phase noise measurement systems by the configuration shown in Fig. 4. The PM noise of both the oscillator and the single-oscillator measurement system is detected in the measurement. This approach can lead to accurate measurements if the oscillator PM noise is small enough [7]. The AM/PM noise standard accommodates this approach by providing oscillators that have very low residual AM and PM noise and the ability to turn the calibrated PM noise on and off. The procedure is to measure the residual PM noise in the measurement with the calibrated PM noise off and then to measure the PM noise with the calibrated PM noise on. Unlike the technique of Fig. 2, it is not generally possible to verify the noise floor of the measurement system for these calibration methods because the phase noise of the reference oscillator always contributes to the measured noise level. Another limitation is that the measurement cannot be carried out at low values of f since the PM noise of the reference oscillator generally dominates the added noise. This problem can be alleviated somewhat by increasing the added noise and decreasing the bandwidth to satisfy Eq. (2).

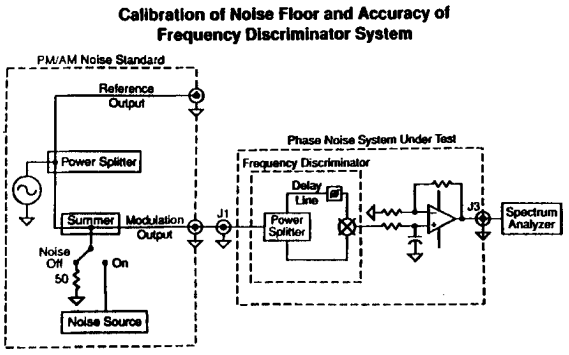


Fig. 4 Single oscillator phase noise calibration method using a frequency discriminator (see [5]).

C. Calibration of AM Noise Level in PM/AM Noise Standard

The AM noise in the standard is determined using the procedure described in section III.A and Eq. (1), since in the PM/AM noise standard,

$$S_a(f) = S_\phi(f) = \frac{PSDV_N(\nu_0 - f) + PSDV_N(\nu_0 + f)}{2V_0^2}. \quad (3)$$

D. Measurement of Errors in AM Noise Measurement Equipment

Amplitude noise can be measured by several different methods. Each method is characterized by a detector that converts the amplitude of the rf carrier signal to a dc voltage. A working definition of $S_a(f)$ is

$$S_a(f) = \frac{\delta V^2(f)}{V_o^2} \frac{1}{BW}, \quad (4)$$

where $\delta V^2(f)$ is the mean squared fluctuation of the carrier rf amplitude measured at Fourier frequency f from the average frequency of the carrier in a noise bandwidth BW, and V_0 is the average amplitude of the carrier [5]. The most uncertain part of the process is determining the sensitivity of the detector k_a for converting the fractional changes in the carrier amplitude to a dc voltage fluctuation, since k_a is a function of V_0, f, and the carrier frequency ν_0.

1. *Calibration of Simple AM Noise Measurement Systems*

The signal input of the AM noise measurement system to be measured is connected to the modulated output of the PM/AM noise standard as shown in Fig. 5. The spectral density of AM noise is calculated as

$$S_a(f) = \frac{V_n^2(f)}{BW}\left(\frac{1}{(k_a(f))^2}\right), \quad (5)$$

where $V_n^2(f)/BW$ is the PSD voltage noise measured by the spectrum analyzer, and $k_a(f)$ is the sensitivity of the detector for converting fractional changes in amplitude to voltage fluctuations. (k_a also includes any gain associated with the detection.)

When the noise modulation is on, the measured level as computed with the aid of Eq. (6) should agree with the results of the calibrated level as given by Eq. (4). The differences between the measured results and that determined from Eq. (4) yield the error of the AM measurement system under test as a function of Fourier frequency. This measurement can also be used to determine $k_a(f)$ except at very low values of

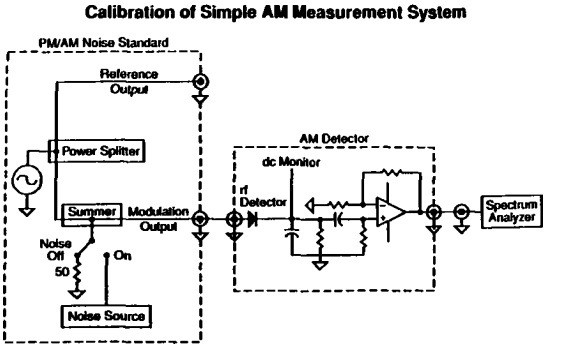


Fig. 5 Simple calibration method for AM noise measurement systems.

f where the AM noise of the source might mask the added noise. This problem can be alleviated somewhat by increasing the level of the added noise and decreasing the bandwidth to satisfy Eq. (2).

The variation of $k_a(f)$ with f is extremely difficult to determine by traditional methods since few sources can be accurately amplitude-modulated over a broad range. This issue becomes even more serious as the carrier frequency and the offset frequency from the carrier increase.

2. *Determining the AM Noise Floor of an AM Noise Measuring System*

The noise floor for AM measurements can be estimated by turning the noise modulation off, measuring $V_n(f)$, and computing $S_a(f)$ using Eq. (5). The noise floor determined in this manner includes the true noise floor of the AM noise measurement system being calibrated plus the AM noise of the internal source in the PM/AM noise standard. In most cases, the noise floor of the standard is lower than the detectors being tested.

In cases where the AM noise of the internal source is so high that it is difficult to evaluate the AM noise floor of a measurement system as described in A, a differential measurement technique can be used as shown in Fig. 6. This technique takes full advantage of the very low differential AM noise between the

reference and modulated outputs of the PM/AM noise standard by using two similar detectors. The rf input of one detector to the reference output, and the rf input of the other detector is connected to the modulated output. The dc outputs of the two detectors are preferably adjusted to be the same within approximately 1% using attenuators on the

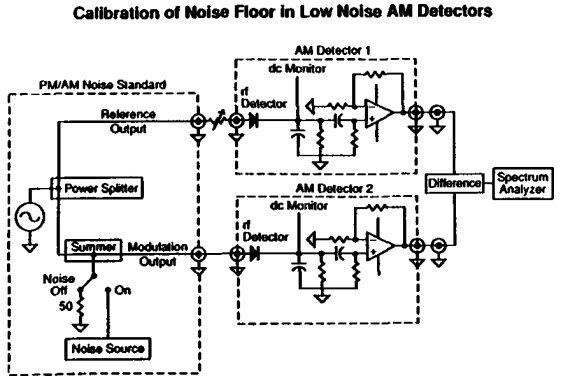


Fig. 6 Differential AM calibration method to determine the noise floor of AM detectors and measurement systems.

input of the detectors. The gain following the detectors may also be adjusted; however, the cancellation of the residual AM noise in the source may not be as complete as when input attenuators are used. The outputs are then subtracted in a conventional device such as a differential amplifier (see Fig. 6). Some spectrum analyzers also can process the difference between the two input signals. With either of these methods, the AM noise of the internal source is substantially reduced, perhaps as much as 20-40 dB. (At very high Fourier frequencies it may be necessary to adjust the phase slightly on the incoming signal to one of the AM detectors to optimize the cancellation.) The noise floor measured in this manner is that of both detectors (and their associated amplifiers). If they were identical, we would subtract 3 dB from the measured result to obtain the noise floor of a single unit. A more accurate value of the AM noise floor of the individual detectors (and amplifiers) requires that measurements be made on 3 detectors with all combinations (detectors 1-2, 1-3, and 2-3) [4,6]. The determination of $S_a(f)$ of unit 1 is

$$S_a^1(f) = \tfrac{1}{2}(S_a^{1\text{-}2}(f) + S_a^{1\text{-}3}(f) - S_a^{2\text{-}3}(f)), \quad (6)$$

since the random noise in the detectors is uncorrelated [4,6].

The variation of $k_a(f)$ with f of a differential AM noise measurement system can be determined by turning the added differential noise on, measuring $V_n(f)$, and comparing the results obtained from Eq. (3) with these from Eq. (6).

Using the PM/AM noise standard in the manner prescribed above calibrates all of the measurement errors even at very low values of f where the AM noise of the internal source might swamp the noise floor of the measurement system under test or the added differential noise, if we used the technique described in A.

A typical method of determining $k_a(f)$ prior to measuring AM noise is shown in Fig. 7. First, the modulated output of the PM/AM noise standard is connected to a detector and the detected dc level is noted. Second, the output of an auxiliary source that can be amplitude modulated at a known level is connected to the detector and the rf level, impedance (50 Ω), and the frequency are adjusted to that chosen for the PM/AM noise standard. The auxiliary source is then AM modulated at a convenient frequency and level. A common choice is 1% modulation at 1 kHz. The level of the modulated signal from the port labeled "OUTPUT POWER" in Fig. 7 is then recorded as "AM REF LEVEL." The detector sensitivity multiplied by the gain of the amplifier is then found from

$$k_a^2(f) = \frac{(AM\ REF\ LEVEL)^2}{\tfrac{1}{2}\left(\frac{\%AM}{100}\right)^2}. \quad (7)$$

In general the AM REF LEVEL, and hence $k_a(f)$, is a function of ν_0, f, and V_0.

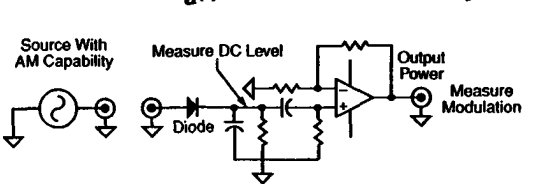


Fig. 7 One method of determining the sensitivity for converting small changes in amplitude to small changes in voltage.

E. Calibrations of Time-Domain Measurement Systems

Measurement of frequency stability in the time-domain is another common method used to characterize the frequency stability of oscillators and other components [2-5]. In the past a convenient method to calibrate the accuracy of such systems has not been available because the measured time-domain frequency stability depended on the measurement bandwidth, the averaging time, and the type of phase noise. The phase noise standard provides a differential level of PM noise that is constant in amplitude (white PM) for measurement bandwidths from 1 Hz to 10% of the carrier. Under these conditions the fractional frequency stability between the reference and modulated outputs as characterized by the square root of the Allan variance [5,8] is given by

$$\sigma_y(\tau) = \frac{3f_h S_\phi(f)}{2\pi\nu_0\tau_o}, \qquad (8)$$

where f_h is the measurement bandwidth and τ is the measurement time. When a time-domain frequency stability measurement system [4] is connected between the two outputs, $\sigma_y(\tau)$ should vary as τ^{-1} for measurement times from approximately 0.1 μs to 0.5 s. Comparing the results with Eq. (8) yields the measurement bandwidth, since Sϕ(f) is known. The noise floor of time-domain measurement system can be evaluated by turning the added PM noise off. Time-domain calibrations can also be done by comparing the modulated output to an external reference. In this case the time-domain stability may deviate from the τ^{-1} behavior for the longer measurement times due to noise in the external and the internal reference [5].

Results

The first PM/AM noise standard model PMAM 115 operates at 5, 10, and 100 MHz. Switches select the proper carrier frequency and matching pass band filters. The noise source is flat within a few tenths of 1 dB from less than 0.1 MHz to about 1 GHz and is stable with temperature to 0.02 dB/K. The differential noise added to the carrier ν_0 is exceptionally flat due to the symmetry of Eq. 1.

Figure 8 shows the performance of the PM/AM standard at 100 MHz. In curve A, the phase noise of the 100 MHz reference oscillator; in curve B, the level of differential PM added noise between the reference and modulated outputs at 100 MHz; in curve C the noise floor of the PM/AM standard at 100 MHz using the method of Fig. 2 with the noise modulation off.

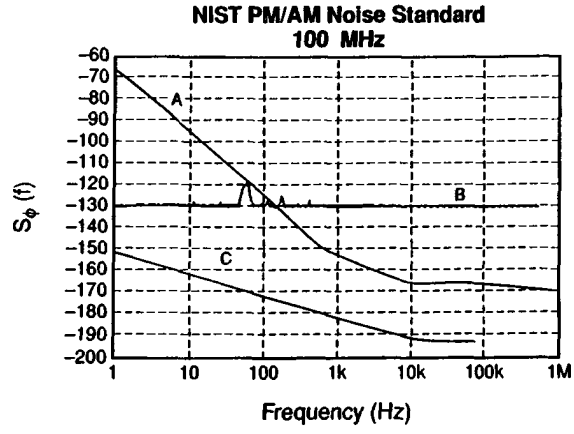


Fig. 8 Measurement results for PM/AM Standard model PMAM 115 at a carrier frequency of 100 MHz. Curve A shows the PM noise of the 100 MHz reference oscillator. Curve B shows the calibrated level of PM and AM noise. Curve C shows the typical noise floor for PM noise in the PM/AM Noise Standard at 100 MHz. Somewhat better results are obtained at 5 and 10 MHz.

The cancellation of the 100 MHz reference oscillator phase noise by using the method of Fig. 2 (the difference between curves A and C) reaches approximately 85 dB at 1 Hz. The large suppression of the reference oscillator PM noise makes it possible to use the PM/AM noise standard over a much larger range of Fourier frequencies about the carrier than any technique that is sensitive to the PM of the reference such as that shown in Figs. 3 and 4. At 100 MHz, the difference is having an accurate PM standard for Fourier frequency offsets from 1 Hz to 10 MHz versus 300 Hz to 10 MHz. The peaks at 60, 120, and 180 Hz in curve B are due to spurious signals originating from the power lines.

Curve A of Fig. 9 shows the PM noise in the PM/AM noise standard at 100 MHz determined using a high-dynamic-range scanning receiver in the calibration mode described in Section III A above. Curve B of Fig. 9 shows the measurement of the PM noise level using a totally separate and independently calibrated phase noise measurement system PNMS [7,11]. The noise level is constant to $\pm$ 0.05 dB for Fourier frequencies from dc to 5 MHz and gradually decreases 0.3 dB from 5 to 10 MHz. Curve C of Fig. 9 shows the AM noise determined using the

simple measurement technique of section III C. The divergence of curve C from curves A and B for Fourier frequencies below 100 Hz is easily shown, by the AM noise floor measurements, to be due to noise in the diode detector used for these measurements. The fluctuations in curve B of Fig. 9 versus frequency are primarily due to the finite number of samples taken to estimate the spectral density [10]. The differences between curves A and B of Fig. 9 are consistent with the estimated errors in the internal phase modulator used in the calibration of the PNMS. These estimated inaccuracies are shown in Fig. 10 as a function of Fourier frequency for carrier frequencies of 5 MHz, 10 MHz, 300 MHz, 10 GHz and 42.7 GHz. Table 1 summarizes the PM/AM calibrated noise and noise floor at 5, 10, and 100 MHz. The totally independent data from the PNMS confirm the calibration of the PM/AM standard to ± 0.5 dB.

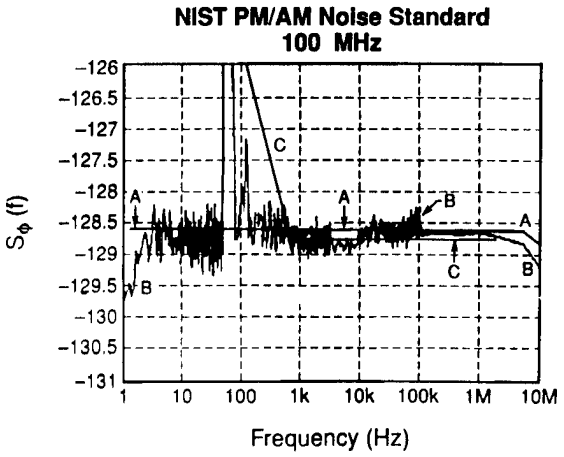


Fig. 9 Measurement results for PM/AM Standard model PMAM 115 at a carrier frequency of 100 MHz. Curve A shows the PM noise level of the standard determined using the technique of Section III. Curve B shows the measurements of PM noise using the totally independent NIST Phase Noise Measurement System (see [7,11]). Curve C shows the level of AM noise of the standard at 100 MHz using the method shown in Fig. 5. The peaks at 60 and 120 Hz are due to the power line. Similar results are obtained at 5 and 10 MHz.

The techniques illustrated here can be extended to high frequencies and to higher accuracies using presently available technology. Preliminary results at 10.6 GHz look promising for Fourier frequencies out to at least 500 MHz. The PM/AM noise of the first PM/AM standards has remained constant within ± 0.4 dB at all three carrier frequencies (5, 10, and 100 MHz) over the past year. Accuracies in excess of ± 0.1 dB are possible by using components with lower VSWR, circuits with better level control and more extensive measurements on receiver bandwidth and linearity.

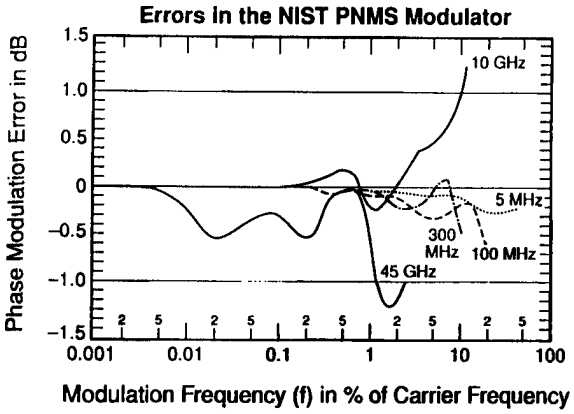


Fig. 10 Typical errors in the internal calibration of the phase modulator in the NIST Phase Noise Measurement System (see [7,11]).

TABLE I. SYSTEM NOISE FLOOR AND PHASE NOISE OF SOURCES

NOMINAL SOURCE PHASE NOISE/CHANNEL ± 3 dB dBC/Hz								
	FOURIER FREQUENCY							
SOURCE FREQUENCY	1 Hz	10 Hz	100 Hz	1 kHz	10 kHz	100 kHz	1 MHz	10 MHz
5 MHz	-121	-151	-163	-171	-174	-174	-174	
10 MHz	-115	-145	-157	-165	-168	-168	-168	
100 MHz	-70	-100	-130	-156	-170	-170	-173	-173
MAXIMUM RESIDUAL NOISE BETWEEN CHANNELS dBC/Hz								
5 MHz	-162	-172	-182	-190	-194	≤-175	≤-175	
10 MHz	-161	-176	-183	-191	-197	≤-175	≤-175	
100 MHz	-152	-162	-172	-182	-193	≤-175	≤-175	≤-175
DIFFERENTIAL PM/AM NOISE LEVEL ± 0.14 dBC/HZ								
5 MHz	-128.3	-128.3	-128.3	-128.3	-128.3	-128.3		
10 MHz	-129.1	-129.1	-129.1	-129.1	-129.1	-129.1	-129.2	
100 MHz	-128.9	-128.9	-128.9	-128.9	-128.9	-128.9	-128.9	-129.2

Conclusion

A new portable PM/AM noise standard (model PMAM 115) has been developed, tested, and used to evaluate the noise floor and accuracy of PM and AM noise ($S_\phi(f)$ and $S_a(f)$) measurement equipment. The differential PM noise in the PM noise standard used to test the noise floor is typically much less than the noise floor of present measurement systems. The PM noise in the standard was evaluated for Fourier

frequencies from 1 Hz up to 10% of the carrier frequency to an accuracy of ± 0.14 dB. The temperature coefficient of the PM and AM noise level is less than 0.02 dB/K. These calibrations were independently verified at carrier frequencies of 5, 10, and 100 MHz to ± 0.5 dB using the Phase Noise Measurement System [7,11]. The design of the standard is such that the fractional amplitude noise $S_a(f)$ is equal to $S_\phi(f)$ over a very wide range of Fourier frequencies. The new secondary standard has maintained the same level of PM and AM noise to within ± 0.4 dB over the past year [1].

Although the PM/AM noise standards described above are at frequencies from 5 to 100 MHz, the principle can be applied to any other frequency from audio to the visible. Prototypes are presently working at 10.6 GHz [1].

Acknowledgements

The author thanks F. Ascarrunz, C. Nelson, and P. Pond for help in construction. This work was supported in part by the calibration coordination group of JCTG/GMT-JLC and the CECOM Center for Space Systems TSSRF, Ft. Monmouth.

References

[1] F. L. Walls, U. S. Patent applied for.

[2] 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. F. C. Vessot, and G. M. R. Winkler, "Characterization of Frequency Stability," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-20, pp. 105-120, 1971.

[3] D. W. Allan, H. Hellwig, P. Kartaschoff, J. Vanier, J. Vig, G. M. R. Winkler, and N. Yannoni, "Standard Terminology for Fundamental Frequency and Time Metrology," *Proc. of 42nd Annual Symposium on Frequency Control*, Baltimore, MD, June 1-4, 1988, pp. 419-425.

[4] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of Signal Sources and Measurement Methods," *Proc. of 35th Annual Symposium on Frequency Control*, Philadelphia, PA, May 27-29, 1981, pp. A-1(669)-A47.

[5] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[6] 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," *Proc. of 42nd Annual Symposium on Frequency Control*, Baltimore, MD, June 1-4, 1988, pp. 279-283.

[7] F. L. Walls, C. M. Felton, A. J. D. Clements, and T. D. Martin, "Accuracy Model for Phase Noise Measurements," *Proc. of 21st Annual Precise Time and Time Interval Planning Meeting*, Redondo Beach, CA, Nov. 30-Dec. 1, 1990, pp. 295-310; F. L. Walls, A. J. D. Clements, C. M. Felton, and T. D. Martin, "Precision Phase Noise Metrology," Proc. of National Conference of Standards Laboratories, August, 1991, pp. 257-275.

[8] F. L. Walls and A. DeMarchi, "RF Spectrum of a Signal after Frequency Multiplication; Measurement and Comparison with a Simple, Calculation," *IEEE Transactions on Instrumentation and Measurement*, vol. IM-24, pp. 210-217, 1975.

[9] F. L. Walls, D. B. Percival, and W. R. Irelan, "Biases and Variances of Several FFT Spectral Estimators as a Function of Noise Type and Number of Samples," *Proc. of 43rd Annual Symposium on Frequency Control*, Denver, CO, May 31-June 2, 1989, pp. 336-341.

[10] F. L. Walls, J. Gary, A. O'Gallagher, L. Sweet, and R. Sweet, "Time-Domain Frequency Stability Calculated from the Frequency Domain: An Update," *Proc. of 4th European Frequency and Time Forum*, Neuchatel, Switzerland, March 13-15, 1990, pp. 197-204; J. Gary, A. O'Gallagher, L. Sweet, R. Sweet, and F. L. Walls, "Time Domain Frequency Stability Calculated from

the Frequency Domain Description: Use of the SIGINT Software Package to Calculate Time Domain Frequency Stability from the Frequency Domain," NISTIR 89-3916 Revised, 1989.

[11] F. L. Walls, "Method and Apparatus for Wide Band Phase Modulation." U. S. Patent No. 4,968,908 issued 6 Nov. 1990, F. L. Walls, "Wide-band Phase Modulation." U. S. Patent No. 5,101,506 issued 31 Mar. 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

CONFIDENCE ON THE SECOND DIFFERENCE ESTIMATION OF FREQUENCY DRIFT
A Study based on Simulation

Marc A. Weiss, David W. Allan, and David A. Howe
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, CO 80302

Abstract

We use simulation to compare the confidence in estimating frequency drift in the presence of stochastic noise between two different estimators: (1) a mean second difference using neighboring data, and (2) a single second difference over the entire data set. In each case we simulate 100 s data sampling and 10,000 samples (11 and 1/2 days of data). The two estimators show a similar confidence when the noise is purely random walk frequency modulation, but when there is also white frequency modulation, as is common in oscillators, method (2) is more efficient. An advantage of (1) is that it provides an internal confidence to an estimate on a single data set. As a practical example, when we simulate a typical rubidium gas cell frequency standard, the confidence of the drift estimate was ten times better for method (2) over method (1).

Theory

We studied the confidence on two second difference estimators of frequency drift in the presence of white noise frequency modulation (WHFM) and of random walk frequency modulation (RWFM) using simulation methods in Monte Carlo runs. We generated time series simulating clock data (10,000 points) taken every 100 s against a perfect reference, continuing for 10^6 s, about 11 and 1/2 days. In each case we repeated the simulation 100 times, that is, we generated 100 different data sets with the same stochastic and deterministic parameters but with a different starting seed number. We repeated our two different second difference estimators on each data set, taking the mean and standard deviation across all data for a given set of parameters. In this paper we refer to a "knee" in the Allan variance plot of $\sigma_y(\tau)$ versus τ. The knee is the τ value where we see a transition from a region where one power law dominates, such as $\sigma_y(\tau) \sim \tau^{1/2}$ for RWFM, to the region where another process dominates, such as $\sigma_y(\tau) \cong \tau^1$ for frequency drift.

Contribution of the U. S. Government, not subject to copyright.

We simulated clock data for two types of experiments. The first type may be typical of some quartz crystal oscillators. In this case, there was a frequency drift of 1 part in 10^{10} per day, and the stochastic noise was purely RWFM. This was done using four different levels of RWFM to make the knee in the Allan variance curve appear at 10^3 s, 10^4 s, 10^5 s, and equivalent of 10^6 s, if we had generated more data. This allowed us to study the ability of the second difference estimator to find the true drift when the random walk effects dominated increasingly more of the data set. In a second type of experiment we introduced a combination of WHFM and RWFM with frequency drifts of either 1 part in 10^{14} per day or 1 part in 10^{13} per day, modelling the performance of two types of rubidium gas-cell frequency standards.

The second difference operator, Δ^2, estimates drift by computing the change in average frequency from one interval to the next. The second difference was computed in two ways: (1) a mean second difference using neighboring data, and (2) a single second difference over the entire data set. For (1), the mean second difference of a data set, we take first differences over the minimum time interval, τ_0, the average frequencies over each τ_0, and then difference them again to obtain an estimate of drift, the change in frequency, for the interval τ_0. We then average these second differences over the entire data set. In this case the individual estimates of average frequency which are used to estimate drift are poor in the presence of the random walk frequency process, compared to those for a longer averaging time as in (2). But the second difference operator converts the RWFM process to a white noise process, hence allowing averaging of the individual drift estimates in

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

method (1), since in our simulation, the drift remains constant throughout the data length. This compares with method (2) where we take the average frequency over the first half of the data set and subtract it from the average frequency over the second half. In this case we have a better estimate of the average frequency, thus a better estimate of drift in the single second difference. It turns out from our simulation that methods (1) and (2) provide similar confidence on the estimates in the case of pure RWFM. However in real oscillators we usually find stochastic noise which is a mixture of RWFM as well as WHFM or what is called flicker frequency modulation. In these more realistic situations, method (2) is significantly more efficient as a drift estimator. We will demonstrate the increased efficiency if WHFM is present. An advantage of method (1) is that it provides an internal confidence of the estimate, the standard deviation of the mean of the series of second differences.

If WHFM noise is included with RWFM and drift, then the application of the second-difference operator Δ^2 yields a non-white process. In a WHFM process the frequency is a white noise process of deviations around a mean frequency. Since a WHFM process is equivalent to a random walk phase modulation (RWPM), then the first difference of phase of WHFM data yields a white noise process. The longer we average white noise, the better the estimate of the mean. This means, the larger the value of τ we use to compute first differences, as long as the dominant noise process is WHFM, the better the estimate of mean frequency. As the interval τ increases for the Δ^2 operator, the estimate of average frequency $\overline{y}$ improves because high-frequency energy contained in the WHFM is "filtered" by a greater extent. In the extreme case in which τ is 1/2 the length of the full data set and there is only a single Δ^2 to compute (i.e., three points as shown in figure (1), then the high-frequency energy is maximally filtered. Hence the knowledge of the frequency component of the data set is the best and is better than computing $\overline{\Delta^2}$ over shorter τ intervals. Though as τ increases, the effects of RWFM and then drift become more dominant, yet this can explain why method (2), using a single over-all second difference, is more practically efficient than method (1).

From a variance point of view, the variance of frequency is largest at τ_0, the minimum data spacing, and decreases as τ. This is because with white noise, data are random uncorrelated. From a frequency-domain viewpoint, the spectrum resulting from applying Δ^2 to WHFM phase data yields a non-white process -- having increased power at high Fourier frequencies compared to low Fourier frequencies.

We will consider a few subtleties of Δ^2 in the presence of RWFM. The Δ^2 operator applies a sequential "difference-of-slopes" operation to a data set. If the set is the time difference of two clocks (or phase error data) from, say, a time-interval counter, then the Δ^2 operator represents the "difference-of-frequencies" (since the slope of the phase is proportional to frequency) determined from three sequential time difference readings. Subsequent three-somes can be taken and the Δ^2 operator applied to each; an average over the full data set can then be computed as Barnes has discussed [1]. The Δ^2 operator approximates the familiar $\frac{d^2}{dt^2}$ operator used in a continuous function.

To say that noise on a signal is consistent with a model of RWFM has several implications. We assume that there is a continuous process which is the double integral of a white noise process. Any measurements on the signal, however, are discrete band-limited measurements. Thus, if we doubly difference the discrete measurements, we might not obtain the theoretical underlying white noise process. There are at least three reasons for this: (1) initial data points are lost for each difference operation, i.e., there are N-1 differences for N data samples; (2) the functions involved are sampled at τ_0 thus the data always include power at the sampling frequency of $1/\tau_0$ which has an associated window function, and hence has power introduced within the bandwidth of our operators; and (3) as shown by Greenhall [2], adjacent sets of data can be highly correlated in RWFM, yielding a bias when double differencing. These problems can be mollified. If the initial data points on differencing are maintained, then item (1) is not a concern. Regarding item (2), the sampling window function can be shaped to reduce sidelobe leakage [3]. The correlation effect of item (3) may still cause the second difference of the sampled data to yield a different white-noise process than the theoretical, continuous process. As an aside, the addition of the drift term to the RWFM can perturb the estimate of WHPM using the Δ^2 operator, since it is added and subsequently subtracted at different steps in the process. This problem can also be resolved. Thus, with proper care we can find that double differencing yields a white process, though we

lose some information about the theoretical underlying continuous process.

Experiment

Measurements are commonly taken by a time-interval counter, producing a sampled, usually equally spaced, time series with interval τ_0. The WHFM and RWFM models are invariably useful for studying the performance of Cs, Rb, Qz, H, and other frequency standards [4]. Optimal estimation involves converting the noise term to a white process so that a simple average and standard deviation can be computed. The $\overline{\Delta^2}$ operator will do this to pure RWFM. This paper shows that the result is essentially the same for pure RWFM plus drift if we apply Δ^2 to each data three-some (at interval τ_0) or if we compute Δ^2 by simply taking the first, middle, and last points as shown in figure 1.

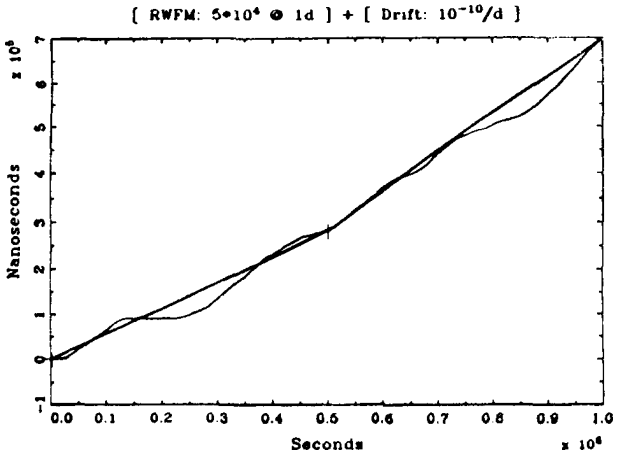


Figure 1. This figure illustrates the three points used in a simple second difference for estimating the frequency drift. The random line is simulated RWFM super-imposed on a linear frequency drift. The two straight lines indicate the slopes corresponding to the two frequencies from which the frequency drift is estimated.

A Monte Carlo method was used to test the concept on a set of simulated data. The simulated data consist of drift equal to 10^{-10} per day with RWFM added at four levels of amplitude. The RWFM data were generated by twice integrating band-limited white noise. The first integration converts white noise into random walk noise. This output can be used to model RWPM, which is the same as WHFM. Using a different "seed" and two integrations converts white noise into a model of the phase for RWFM. These noises are then added to a quadratic term whose slope increase corresponds to a frequency drift ranging from 10^{-14} per day to 10^{-10} per day to generate the final data set. The analysis is then applied to these data to compute the drift and confidence. Since we know the drift a priori, we can test the accuracy of the drift and confidence estimation.

A family of plots of $\sigma_y(\tau)$ showing (theoretically) four levels of RWFM and drift of 10^{-10} per day are shown in figure 2. RWFM and drift go as $\tau^{+½}$ and τ^{+1} respectively. The "knee" (or change of slope in the $\sigma_y(\tau)$ plot) occurs at 10^3, 10^4, and 10^5 s. We note that in the highest level of RWFM used in our simulation, the knee occurs at a sample time equal to the length of the data set, 10^6 s; so in this case, the RWFM essentially masks the drift.

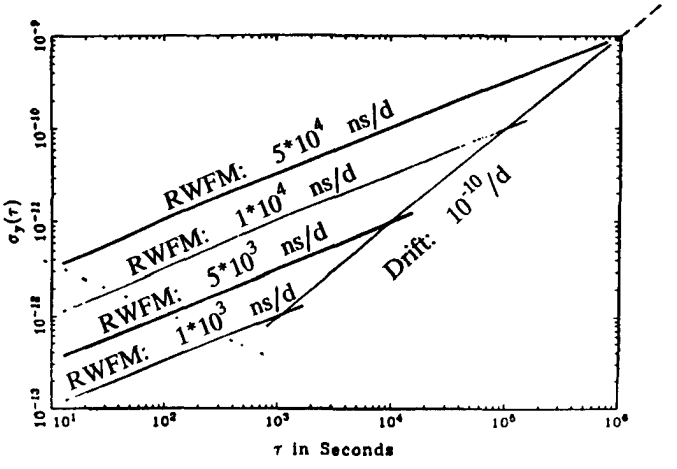


Figure 2. An illustration of a frequency stability plot, $\sigma_y(\tau)$, for different simulated levels of RWFM and a fixed frequency drift of 1 part in 10^{10} per day as is typical for some precision quartz-crystal oscillators. Notice the intersections "the knee" between the RWFM curves and the frequency drift curve occur at 1000 s, 10,000 s, 100,000 s, and at 1,000,000 s.

Also shown in figure 2 is white frequency modulation (WHFM) (the single dotted line with $\tau^{-½}$ behavior) at a level common to high-performance commercial cesium-beam frequency standards and to rubidium gas-cell frequency standards. Note, however, that this much frequency drift would be unusually poor performance for a cesium standard. Depending on the level of RWFM, we may or may not see its $\tau^{+½}$ behavior if masked by the WHFM plus drift. If there is no RWFM and we have pure WHFM and drift, then Δ^2 applied to the phase data will not convert the

WHFM to a white process. In this case, the residuals around a linear regression to the frequency are white, and this regression line is the appropriate estimator of drift. Note, WHFM is the same as RWPM, the first integral of a WHPM process.

Figure 3 is a model with a WHFM level of 10 ns and a RWFM of 30 ns/d, both at 1 d, and a drift of 10^{-13} per day. The last two numbers are typical for commercial rubidium gas-cell frequency standard and the first number for traditional cesium-beam frequency standard. Though not plotted, we also did the simulation for the anticipated performance of the rubidium frequency standards that are planned for the GPS Block 2-R satellites to be used in cross-link ranging for that system. The model elements in this second case were 3 ns/d for the WHFM, 4 ns/d at 1 d for the RWFM and 10^{-14} per day for the frequency drift.

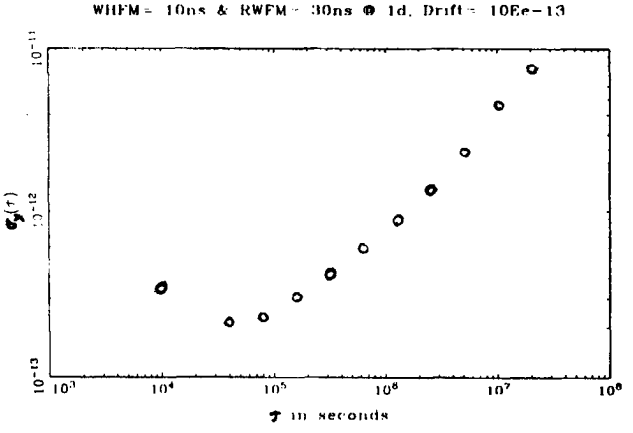


Figure 3. An illustration of a frequency stability plot, $\sigma_y(\tau)$, for simulated data as may be typical in some atomic frequency standards. The noise model includes both WHFM and RWFM. The frequency drift model is typical of commercial rubidium frequency standards.

Results

Results of using the simulated data and computing a single second difference (first, mid, and last points only) are shown in Table I. One can imagine the data representing four different experiments each involving 100 independent clocks and each clock having the same level of RWFM (and value of drift, of course). The third column is the average of each of the 100 cases; the results closely match the expected value of 10^{-10} per day with a +18% error being the worst in the last (highest level) case of RWFM. The fourth column represents the standard deviation over each 100 cases.

TABLE I
100 Single 2nd Difference Estimates of
10^{-10}/d Drift + RWFM

RWFM @ 1d	"Knee" [s]	Mean of 2d's	S.D. of 2D's
1000 ns/d	10^3	$1.00*10^{-10}$	$.004*10^{-10}$
5000 ns/d	10^4	$1.00*10^{-10}$	$.190*10^{-10}$
10000 ns/d	10^5	$1.02*10^{-10}$	$.344*10^{-10}$
50000 ns/d	10^6	$1.18*10^{-10}$	$2.07*10^{-10}$

Table II shows the results of computing the mean second differences using the same original data. The average of the mean second differences over each of the 100 cases is shown in column three. Again the worst match is in the last (highest level) case of RWFM being in error by +12%. The fourth column shows the standard deviations of the mean second differences used in column three and is a measure of confidence on the result. The fifth column is the RMS of the standard deviations of the means and these are essentially identical to the standard deviations because the Δ^2 operator yields a pure white process from the pure RWFM data set.

TABLE II
100 Mean 2nd Difference Estimates of
10^{-10}/d Drift + RWFM

RWFM (1d)	"Knee" [s]	Mean of M2d	S.D. of M2D	RMS(SD Mean)
1000 ns/d	10^3	$1.00*10^{-10}$	$0.035*10^{-10}$	$0.034*10^{-10}$
5000 ns/d	10^4	$0.99*10^{-10}$	$0.173*10^{-10}$	$0.170*10^{-10}$
10000 ns/d	10^5	$1.00*10^{-10}$	$0.333*10^{-10}$	$0.340*10^{-10}$
50000 ns/d	10^6	$1.12*10^{-10}$	$1.670*10^{-10}$	$1.670*10^{-10}$

A typical plot of the simulated phase data with the "knee" at 10^4 s is shown in figure 4. We see the obvious quadratic trend and the added slight RWFM noise accompanying it. The corresponding $\sigma_y(\tau)$ plot is shown in figure 5. Here we see the $\tau^{+½}$ RWFM behavior over most of the plot and then the curvature in long-term toward τ^{+1} frequency drift. It is interesting to note that even with the "knee" being comparable to the data length, the frequency drift is still reasonably measurable.

For the typical rubidium frequency standard data (see figure 3) where the modeled drift was 10^{-13} per day, we obtained (0.61 $\pm$ 0.42) x 10^{-13} per day for the external estimate of the confidence using method (1) and (0.91 $\pm$ 0.12) x 10^{-13} per day for the estimate using method (2). For the improved performance rubidium planned for the GPS Block 2-R with a modeled drift of 10^{-14} per day we estimated (0.3 $\pm$ 1.4) x 10^{-14} per day from the internal estimate of the confidence using method (1). The external estimate of the confidence gave a very large uncertainty. For method (2) the frequency drift estimate was (0.88 $\pm$ 0.17) x 10^{-14} per day.

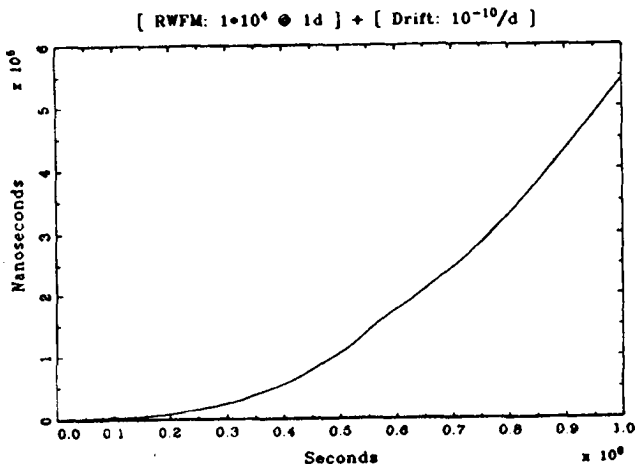


Figure 4. A typical plot of the simulated phase data with the "knee" at 10^4 s. We see the obvious quadratic trend and the added slight RWFM noise accompanying it.

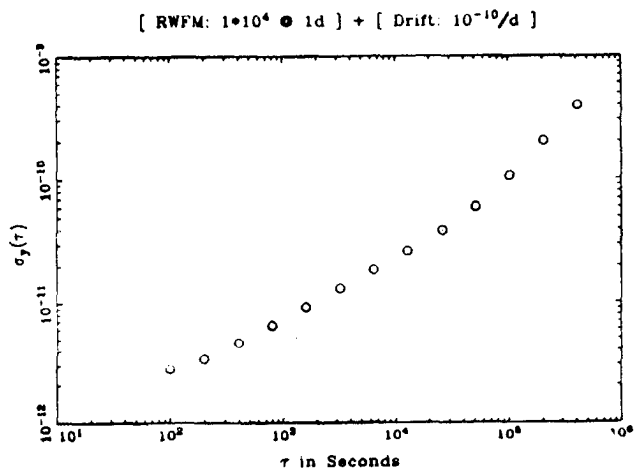


Figure 5. This frequency stability plot, $\sigma_y(\tau)$, corresponds to the data shown in figure 4. One sees the $\tau^{1/2}$ (RWFM) behavior transitioning into the τ^{+1} (frequency drift) behavior beyond 10^4 s.

Conclusions

We have compared two different estimators of drift in the presence of pure RWFM, and the more realistic combination of WHFM and RWFM. We find with pure RWFM that even though one is computationally simpler, they both produce estimates of drift with comparable confidence. Thus, the drift can be estimated with confidence using only three points from this data set: the first, the middle, and the last points. The real advantage of this simple second-difference frequency drift estimator appears when a significant level of WHFM is present. Then the confidence on this estimate becomes significantly better than computing second differences over τ_0, then computing the mean of these. We note that similar simulation would show this second difference estimator giving significantly better estimates of drift with WHFM and RWFM stochastic noise, than linear regression on the frequency or quadratic regression on the phase.

In general, taking the results of this paper, which are based on pure simulation and modeling, along with the results of Barnes and Allan (see ref. [1] and [4]) which are based on real data, we have the following conclusions concerning the estimation of frequency drift. If the predominate noise process is white noise phase modulation (WHPM), then the frequency drift can be best determined by calculating a quadratic-least-squares to the phase (or the time) data values. If the predominate noise process in white noise frequency modulation (WHFM), then the frequency drift can be best determined by a linear-least-squares to the frequency data values. Otherwise, the simple second difference, determined from the first, middle and last phase (or time) data values will be a more efficient estimator of frequency drift for typical frequency standards.

References

[1] J. A. Barnes, "The Measurement of Linear Frequency Drift in Oscillators," in NIST Technical Note 1337, pp. 264-295, 1990.

[2] C. Greenhall, "A Structure Function Representation Theorem with Application to Frequency Stability Estimation," IEEE Trans. Instrum. Meas., vol. IM-32, no. 2, pp.364-370, 1983.

[3] D.A. Howe, D.W. Allan, J.A. Barnes, "Properties of Signal Sources and Measurement Methods," in NIST Technical Note 1337, pp. 14-60, 1990.

[4] D.W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," in NIST Technical Note 1337, pp. 121-128, 1990.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ACCURACY ESTIMATION OF GLONASS SATELLITE OSCILLATORS

Prof. Yuri G.Gouzhva, Prof. Arvid G.Gevorkyan
Dr. Pyotr P.Bogdanov

Russian Institute of Radionavigation and Time (RIRT)
2, Rastrelli square, St.-Petersburg, 193124, Russia

Abstract

Being used GLONASS satellites onboard are GEM atomic frequency standards (AFS) with daily frequency instability not worse than 5.10^{-13}. Results are given for measurement of onboard oscillators' accuracy parameters under works' conditions as well as under operation conditions. It is shown the parameters do not change within the preset limits over several years. The estimate of the accuracy parameters of oscillators installed the GLONASS satellites onboard was performed on the basis of processing the measured offsets of satellites' times generated by oscillators' signals with respect to the system common time generated by the central synchronizer during operation.

Introduction

Time/frequency reference installed onboard the satellite entering into the GLONASS constellation is one of the main elements determining the accuracy of synchronizing the phases of navigation signals emitted by satellites. For adequate system functioning, it is necessary to estimate accuracy parameters of oscillator onboard under operation conditions in order to confirm its compliance with preset requirements. The every-day measurements (3-5 results) are processed on the least-square approximation basis using a linear model for time offsets, and then individual values of oscillator's fractional frequency uncertainty are determined. Thereafter, in terms of those values, for every 10-day observation period, computed are such parameters as a fractional frequency uncertainty of oscillator, $\Delta\hat{f}/f$, and a fractional root-mean-square frequency fluctuation for τ_{obs} = 1 day (root square of N - sample variance), $\hat{\sigma}$.

Of great importance is a problem of accuracy and reliability of results obtained depending on time offsets' measurement error, central synchronizer's accuracy parameters and adopted measurement technique.

Performances of the space-borne time/frequency reference

From 1982 to 1985, at the first stage of GLONASS deployment, the satellite-borne time/frequency reference comprised two rubidium - 87 - beam frequency standard with daily frequency instability not worse than 5.10^{-12} and life time of 0.5 year. From the beginning of the second stage (1985, May), the onboard reference consisted of three cesium - 133 - beam frequency standards with daily frequency instability not worse than 5.10^{-13} and life time of 1 year. An external view of GEM frequency standard as well as its main technical specifications are shown in Fig. 1.

The results of AFS ground tests after its works' production have shown that 80 per cent of all the sets have production margins as follows:

- in terms of fractional frequency uncertainty, $\Delta f/f$ - two-and-half time (see corresponding histogram in Fig.2);
- in terms of frequency instability, $\hat{\sigma}$, - three times for τ_{meas} = 1 sec and two times for τ_{meas} = 100 sec and τ_{meas} = 1 hour (Fig.3, 4, 5).

The results estimating AFS accuracy performances under operating conditions have shown that 90 per cent of AFS had the daily frequency instability of order $(2 - 3).10^{-13}$. The data generalized over 10 - AFS functioned onboard are presented in table 1. One can see that AFS have a good frequency reproducibility over sufficiently long time periods between production and operation.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The estimate of the accuracy parameters of the oscillators installed the satellites onboard is performed on the basis of processing the measured offsets of satellites' times generated by oscillators' signals with respect to the system common time generated by the central synchronizer. The frequency instability of the central synchronizer is not worse than 5.10^{-14} over 1 day. Used is a linear model for time offset:

$$\Delta T(t) = \Delta To + \frac{\Delta \hat{f}}{f} t + \xi(t), \qquad (1)$$

where To is initial time offset,

$\frac{\Delta \hat{f}}{f}$ is fractional frequency uncertainty,

$\xi(t)$ - random time variations due to the corresponding frequency variations, $\nu(\tau)$

$$\xi(t) = \int_0^t \nu(\tau)\, d\tau .$$

While taking into account time comparison (drift measurement), we have

$$\Delta \tilde{T}(t_i) = \Delta T_o + \frac{\Delta \hat{f}_i}{f} t_i + \xi(t_i) + \eta(t_i), \qquad (2)$$

Where $\eta(t_i)$ is comparison error,

t_i is comparison instant.

The comparison error is described by the normal distribution law with zero-mean value and constant variance, σ^2_{meas}.

Over 1 day performed are from 3 to 5 measurement with session duration of 15 minutes.

Since the time drift over a session interval is negligible, the results of time comparison are averaged over a session.

$$\Delta T(t_k) = \frac{1}{N} \sum_{i=1}^{N} \Delta T(t_i), \qquad (3)$$

where $\Delta T(t_i)$ are the values of individual time comparisons over the session with the rate of 1 see,

t_k is session time .

This values are processed on the least-square fit basis using linear model for time drift, thus providing with sample estimate for oscillator's frequency shift, $\frac{\Delta \hat{f}}{f}$ j, over j - day. Over the 10-days interval, there will be 10 estimates for fractional frequency shift, $\frac{\Delta \hat{f}}{f}$ j, for τ_{meas} = 1 day, τ_{sample} = 1 day, τ_{obs} = 10 days, thus giving a possibility to compute a mean value:

$$\frac{\Delta \hat{f}}{f} = \frac{1}{10} \sum_{j=1}^{10} \frac{\Delta \hat{f}}{f} j \qquad (4)$$

and instability:

$$\hat{\delta}_{daily} = \sqrt{\frac{1}{10-1} \sum_{j=1}^{N} \left(\frac{\Delta \hat{f}}{f} - \frac{\Delta \hat{f}j}{f}\right)^2} . \qquad (5)$$

All the measurements are performed with a specified accuracy, a number of measurements is finite, and models are used which are non-coincident with a real process. All the results are sample estimates for which one can have confidence intervals.

The instrument error (rms) for fractional frequency shift estimate over 1 day in the case considered will be:

$$\sigma\left[\frac{\Delta \hat{f}}{f_j}\right] = \frac{2\sqrt{3(K-1)}\,\sigma_{meas}}{\sqrt{K(K+1)}\,\tau_{meas}} , \qquad (6)$$

Where σ_{meas} = 5 nsec (comparison error),

K = (3-5) (number of estimates over 1 day),

τ_{meas} = 86400 sec (measurement time, 1 day)

Consequently, $\sigma\left[\frac{\Delta \hat{f}}{f_j}\right] \approx 7{,}3 \cdot 10^{-14}$.

a) The confidence interval for an estimate of the fractional frequency shift for the instrument error of individual estimates being less than frequency instability, i.e.

$$\left|\frac{\Delta f}{f} - \frac{\Delta \hat{f}}{f}\right| \leqslant t(\rho, K_1) \cdot \hat{\delta} \text{ day}, \qquad (7)$$

is determined using the tables through the probability interval for the specified confidence probability for ρ = 0,95 $t(\rho)$ = 1,96 $K_1 = N - 1 = 9$.

b) The confidence interval for frequency instability estimate

$$(Z_1 - Z_2)\,\hat{\delta} \text{ day}, \qquad (8)$$

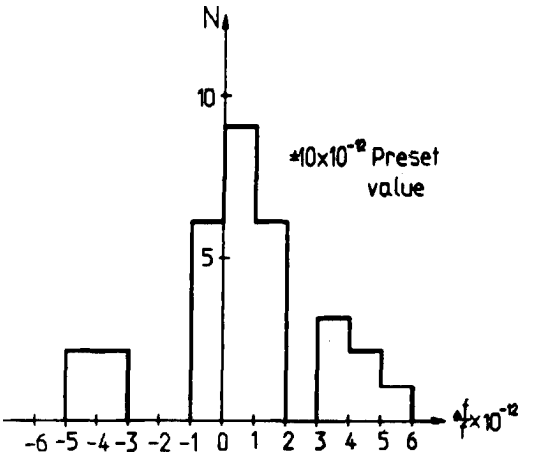
where Z_1 and Z_2 are determined using the tables for ρ = 0,95 and $K_1 = N - 1$, which results (0,688 - 1,826) $\hat{\delta}$ (1 day).

Fig.6 gives the plot of the decade-basis frequency values for one of GLONASS platforms over the time interval from January 1, 1991 till January 31, 1991.

Table 1

GEM AFS reproducibility parameters

Number of the set	$\frac{\Delta f}{f_g} \cdot 10^{-12}$ (ground tests)	$\frac{\Delta f}{f_o} \cdot 10^{-12}$ (operation conditions)	Variation $\left\lvert \frac{\Delta f}{f_g} - \frac{\Delta f}{f_o} \right\rvert \cdot 10^{-12}$	Time interval between estimates, years
1	1,5	- 0,3	1,8	4,5
2	- 0,86	- 1,2	- 0,34	3,1
3	1,3	1,7	- 0,4	3,9
4	0,05	- 1,0	- 1,0	1,4
5	2,1	0,1	2,0	1,4
6	- 0,62	- 0,8	0,18	1,7
7	- 0,75	1,0	- 1,75	1,2
8	- 2,1	6,0	- 8,1	1,3
9	- 2,5	- 3,1	2,2	2,0
10	- 1,4	0,7	- 2,1	1,7

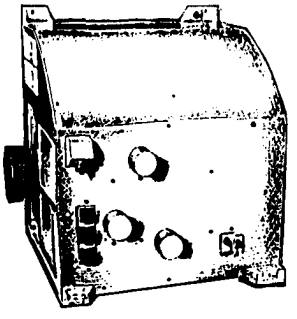


Histogram of distribution of relative frequency accuracy values for GEM atomik beam frequency standard

Fig. 2

GEM

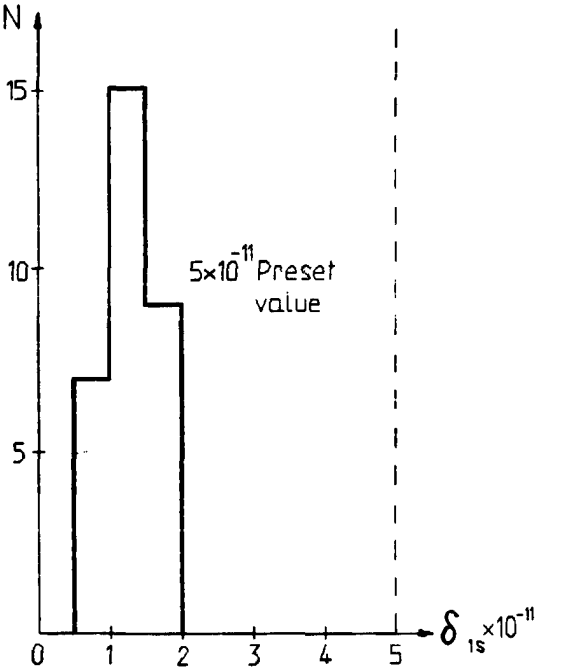
Cesium Beam Frequency Standard for on-Board Space Equipment



Specifications

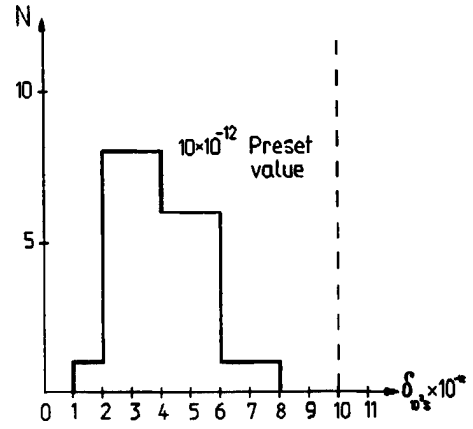
Output frequency	5 MHz
Relative frequency accuracy over life time	$\pm 1 \times 10^{-11}$
Stability, averaging time:	
1 s	5×10^{-11}
100 s	1×10^{-11}
1 hour	2.5×10^{-12}
1 day	5×10^{-13}
Temperature change of frequency, 1/° C	5×10^{-13}
Temperature range, °C	0 to + 40
Power Supply, DC	27 V
Power Consumption	60 W
Dimensions, mm	370x450x500
Weight, kg	39.6
Operational life, hours	17,500

Fig.1



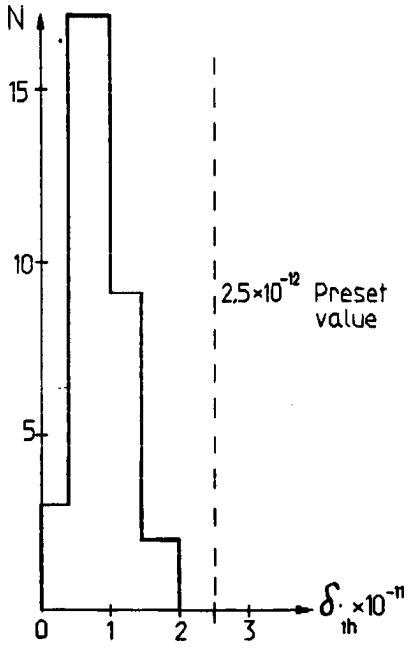
Histogram of distribution of δ_{1sek} stability values for GEM atomik beam frequency standard

Fig. 3



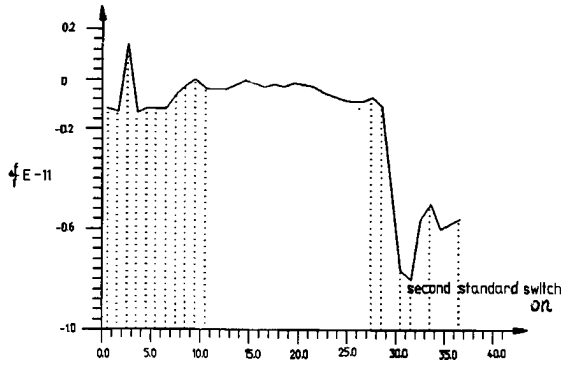
Histogram of distribution of $\delta_{10^2 sek}$ - stability values for GEM atomik beam frequency standard.

Fig. 4



Histogram of distribution of $\delta_{1\ hour}$ - stability values for GEM atomik beam frequency standard.

Fig. 5



Plot for dekade - basis frequency values for one sattelite within the time interval from january 1, 1991 to December 31, 1991.

Fig. 6

Conclusion

At present, installed the GLONASS satellites onboard are AFS with the fractional frequency uncertainty not worse than 2.10^{-11} and with fractional frequency root-mean-square fluctuations not worse than 5.10^{-13} for τ_{obs} = 1 day. Therefore the adopted technique of their accuracy parameters estimating can be considered as valid.

References

[1.] Rumshinsky L.Z. Matematicheskaya obrabotka resultatov eksperimenta (Mathematical processing of experimental results). Moscow, "Nauka", 1971.

[2.] Granino A., Korn P. Mathematical Handbook. McGraw-Hill Book Company, N.-Y.-San Francisco-Toronto-London-Sydney, 1968.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

HIGH STABILITY MICROWAVE FIBER OPTIC SYSTEMS: DEMONSTRATIONS AND APPLICATIONS[1]

Ronald T. Logan Jr. and George F. Lutes

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, California 91109
(818) 354-2471

Abstract

Measurements are presented for microwave fiber optic transmission systems that have much lower phase noise than high quality oscillator signals which have been multiplied to microwave frequencies. Progress in the development of wide-band microwave fiber optic links for direct transmission of the S- and X-band low-noise amplifier outputs in JPL/NASA Deep Space Station antennas will also be discussed. A test bed microwave fiber optic system using a semiconductor-diode-laser pumped neodymium-yttrium-aluminum-garnet (Nd:YAG) laser and a lithium niobate electro-optic modulator is described. This system was used to demonstrate transmission of signals from 2.5 GHz to 12 GHz over 29 km between two Deep Space Stations at the NASA/JPL Deep Space Communications Complex at Goldstone, California. Phase noise performance measured at X-band on this system is presented, and factors affecting phase noise performance are discussed.

Introduction

Previously, ultra-stable analog fiber-optic transmission systems have been developed to transmit the outputs of hydrogen maser frequency references at 100 MHz over distances up to 29 kilometers in the JPL/NASA Deep Space Network (DSN) [1], in support of tests of special relativity [2], radio science, and antenna arraying [3]. A 58 kilometer round-trip link between antenna sites had measured differential fractional frequency stability $\delta F/F = 1 \cdot 10^{-15}$ at 1000 second averaging times. For many applications, it would be desirable to increase the frequency and bandwidth of these ultra-stable fiber optic systems to enable, for instance, transmission of the output of the low-noise amplifier (LNA) in the front-end area of a deep space receiving station to a remote location, or to eliminate the need for frequency multipliers at multiple locations where local oscillator signals must be generated.

Historically in microwave space communications systems, down-converting of the received microwave signals in the front-end area of large antennas was employed to circumvent the deficiencies of metallic waveguides. Excessive losses and frequency dispersion at frequencies in the microwave range (>1 GHz) require down-conversion of these signals to a lower, intermediate frequency (IF) in the 50-550 MHz band for transmission on coaxial cable to the control room. A transmission medium that could provide for direct transmission of the LNA output to the control room would be preferable, but simply did not exist previously. Also, ultra-stable microwave fiber optic systems would permit distribution of high-quality microwave local-oscillator (LO) signals to several antennas from a central location, thereby eliminating the need for frequency multipliers in the front-end areas of antennas.

Previous reported ultra-stable fiber-optic systems [1] employed injection-modulated laser diodes as the optical source, in which an RF input current is converted to an intensity modulated optical signal. The higher intensity noise and interaction of the laser linewidth with fiber dispersion frustrates attempts to use injection-current-modulated laser diodes as fiber-optic sources for long-distance, high-quality transmission of microwave signals. In this paper, a new type of fiber-optic transmitter developed at JPL is described, in which the intensity of a CW laser is modulated using an external device. The noise, modulation bandwidth, and spectral characteristics of this improved transmitter system allow high-stability transmission of microwave frequency signals over longer distances than

[1] This work was performed by the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

previously possible using semiconductor lasers. Results of laboratory measurements and field demonstrations for both state-of-the-art commercial injection modulated laser diodes, and a novel externally modulated solid-state Nd:YAG laser transmitter are presented.

Microwave Distribution Requirements

In the JPL/NASA DSN, antenna sites are separated by distances as great as 29 kilometers. An ultra-stable frequency reference is needed at each antenna site to support precision navigation and radio science for deep space missions. It is advantageous to generate the reference frequencies for the entire complex at a single location and distribute them to users, since atomic frequency standards are costly facilities to install and maintain.

Next generation microwave frequency standards, such as the superconducting cavity maser oscillator coupled to a mercury ion trap [4], promise to provide stabilities approaching 10^{-16} for averaging times from 1 to several hundred seconds, with $\tau^{-1/2}$ performance beyond 1000 seconds.

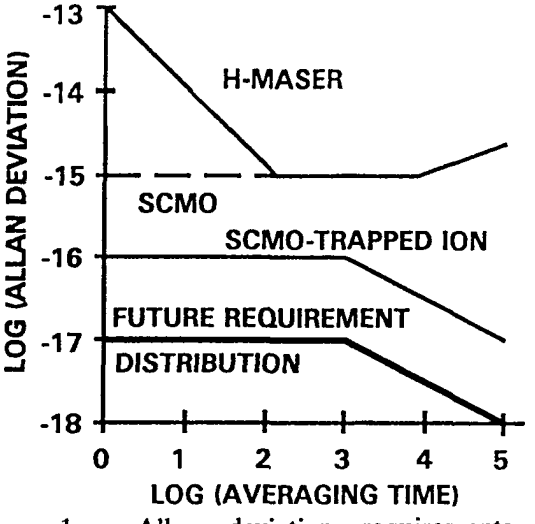


Figure 1. Allan deviation requirements for microwave frequency distribution.

As the DSN moves to spacecraft uplink and downlink channels at Ka-band (32 GHz), the need for stable distribution of microwave signals increases. Ideally, a distribution system should provide differential stability ten times greater than the signal source, in order to preserve the original signal stability. The requirements on future microwave signal distribution systems are illustrated in Figure 1, for averaging times longer than 1 second, and in Figure 2, for phase noise.

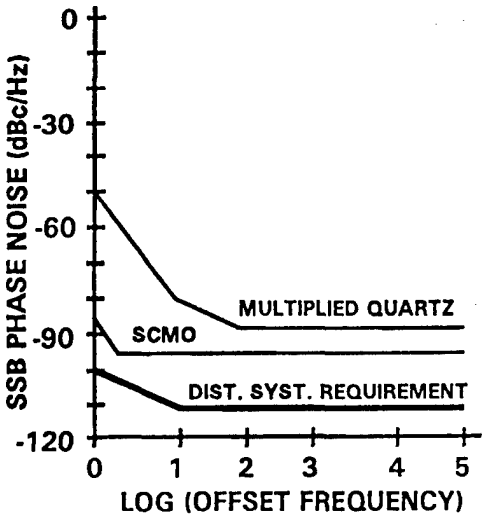


Figure 2. Phase noise requirements for microwave distribution systems in the DSN.

Injection-current Modulated Laser Diodes as Fiber Optic Transmitters

At present, injection-modulated laser diodes are employed for transmission of 100 MHz frequency reference signals between antenna sites separated by up to 29 kilometers in the DSN [1]. These lasers have a multi-mode optical emission spectrum several nanometers wide. Although the laser wavelength is close to the minimum dispersion point of the fiber, the interaction of fiber dispersion and laser modes limits the frequency response of a fiber optic link to [5]

$$f_{-3dB} \sim \frac{10^{12}}{2LD\Delta\lambda} \quad [Hz], \qquad (1)$$

where L is the fiber length in kilometers, D is the fiber group delay dispersion in psec/(nm·km), and $\Delta\lambda$ is the width of the laser emission spectrum in nm. Thus, for typical parameters of D = 3.5 psec/(nm·km), and $\Delta\lambda$ = 5 nm (for a Fabry-Perot laser), the theoretical bandwidth-length product is 28.6 GHz·km. For a 29 km link, the fiber dispersion limits the bandwidth to less than 1 GHz. The exact bandwidth will depend sensitively on the characteristics of the optical fiber and the center wavelength and optical mode structure of the laser. Therefore, it is difficult to precisely predict the signal-to-noise ratio of a long microwave fiber-optic link for modulation frequencies above 1 GHz. To circumvent these problems, a narrow-linewidth optical source would be required.

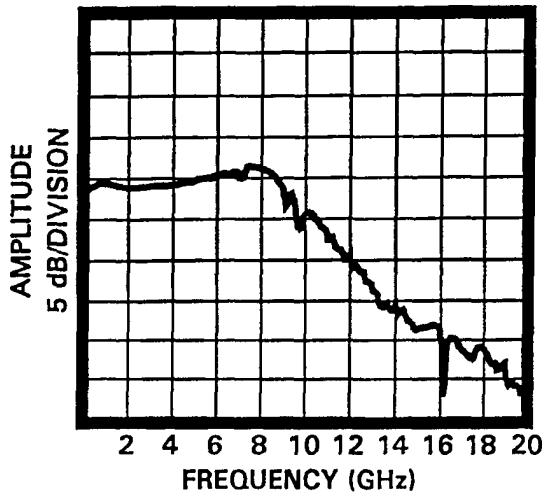


Figure 3. Frequency response of state-of-the-art semiconductor laser diode, corrected for photodiode response.

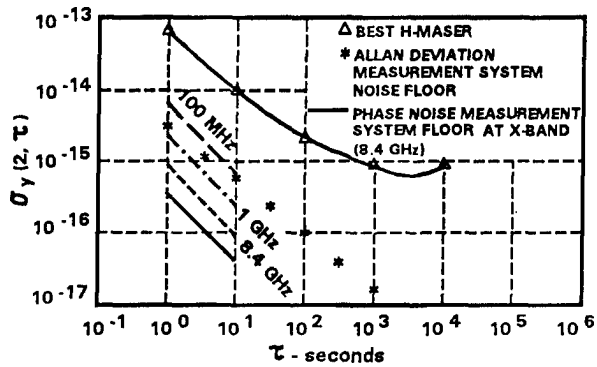


Figure 4. Allan deviation at 1 second for semiconductor laser measured at 100 MHz, and calculated from phase noise at 1 GHz and 8.4 GHz.

For distances within the bandwidth described by Equation 1 above, the performance of injection-modulated laser diodes is adequate for many applications up to X-band (10 GHz). The frequency response of a state-of-the-art laser diode is plotted in Figure 3. This laser has an intrinsic 3-dB modulation bandwidth of approximately 9 GHz, and spectral width of approximately 5 nm. From Equation 1, typical fiber dispersion of 3.5 psec/(nm·km) limits the transmission distance of a 9 GHz signal to less than 3.2 km.

It has been observed that the additive CTC phase noise of laser diodes does not increase as the ratio of the modulation frequencies, as it does for frequency multiplied signals. Therefore, the Allan deviation of a fiber optic link employing a laser diode actually improves as the modulation frequency is increased. Figure 4 depicts the Allan deviation of this semiconductor laser at 100 MHz, 1 GHz, and 8.1 GHz, showing the improvement in Allan deviation with increasing modulation frequency.

A test at 1 GHz was conducted over a 14 kilometer buried link at the Goldstone Deep Space Communications Complex of the DSN, using a Fabry-Perot type laser diode commercially available from Ortel Corporation [6]. The test setup is depicted in Figure 5. From Equation 1, this link has a calculated -3 dB bandwidth of approximately 2 GHz. The differential fractional frequency stability of this link was not observable above the $2 \cdot 10^{-14}/\tau$ noise floor of the measurement system. The Allan deviation for this link, calculated from phase noise measurements, is $8 \cdot 10^{-15}$ at 1 second.

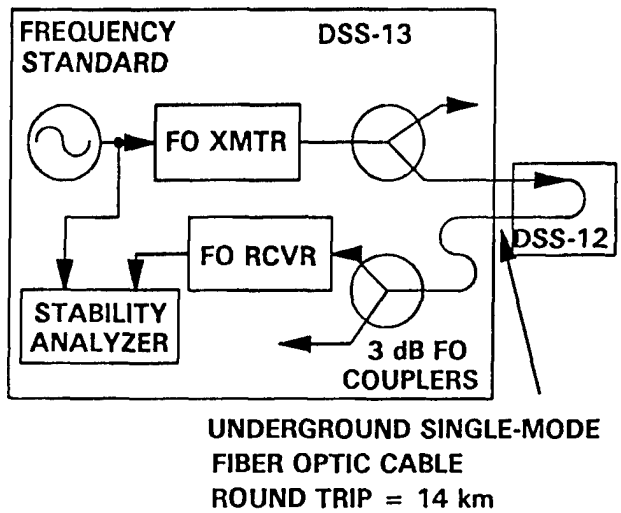


Figure 5. 1 GHz, 14 km distribution experimental setup.

From the earlier discussion, it is apparent that the length-bandwidth product of an injection-modulated laser diode microwave fiber optic link is limited by two factors: the intrinsic modulation bandwidth of the laser diode (X-band, at present), and the interaction of fiber dispersion with the laser modes (28 GHz·km, for Fabry-Perot lasers). The effect of fiber dispersion can be minimized by utilizing an optical source with a narrower linewidth, such as a distributed feedback (DFB) laser diode [7]. However, the intrinsic modulation bandwidth of DFB lasers is typically less than that of Fabry-Perot lasers, and they are much more expensive. Thus, at present, for modulation frequencies above X-band or for long-distance microwave distribution, laser diodes are not well-suited for use in wide-band fiber optic transmitters.

Improved optical transmitter design

To increase the length-bandwidth product of fiber optic systems, first a narrow-linewidth optical source is necessary to minimize the deleterious effects of fiber dispersion. Although the effect of fiber dispersion can also be minimized by tuning the center of the laser emission spectrum to the minimum dispersion point of the fiber, long-term drift of the laser wavelength can still degrade performance. A second problem is the limited intrinsic modulation bandwidth of laser diodes. This problem can be circumvented by performing the intensity modulation external to the laser, using an electro-optic modulator.

An improved fiber optic transmitter testbed system developed at JPL that combines these features is depicted in Figure 6. In this system, a solid-state diode-pumped neodymium:yttrium-aluminum-garnet (Nd:YAG) laser supplies the optical carrier at a wavelength of 1.3 microns, while a $LiNbO_3$ Mach-Zehnder travelling-wave electro-optic modulator impresses the RF input signal onto the carrier as an intensity modulation [8]. The linewidth of the Nd:YAG laser is approximately 10 kHz.

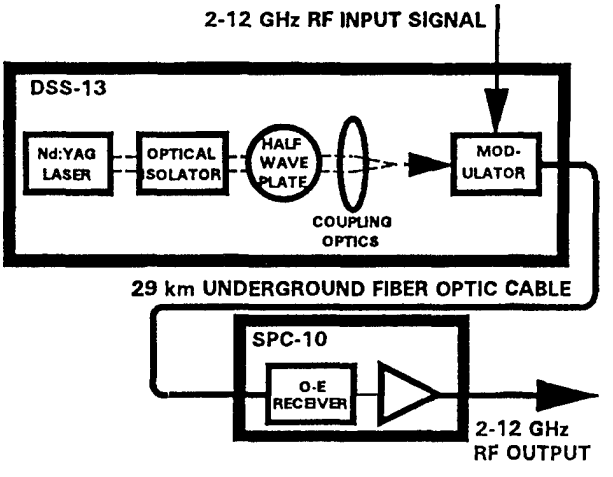


Figure 6. Setup for 29 km microwave transmission experiment using externally-modulated Nd:YAG microwave fiber optic transmitter testbed.

The width of the optical spectrum at the modulator output is then essentially $2F_m$, where F_m is the RF modulation frequency. Thus, for a sinusoidal 8.4 GHz modulation signal, the spectral width at the output of the modulator is simply given by $\Delta\lambda = (2F_m\lambda^2)/c = 0.01$ nm, where c = speed of light in vacuum, yielding a much-improved length-bandwidth product of 14,286 GHz·km. Rearranging Equation 1, the fiber length that produces a dispersion-induced low-pass response of $f_{-3dB} = 8.4$ GHz for this transmitter is found to be $L_{YAG} = 1700$ km. Therefore, the transmission distance for this system will be limited by the attenuation of the optical fiber, and not by fiber dispersion. By contrast, for the Fabry-Perot semiconductor laser, the transmission distance at 8.4 GHz is only $L_{semi} = 3.4$ km.

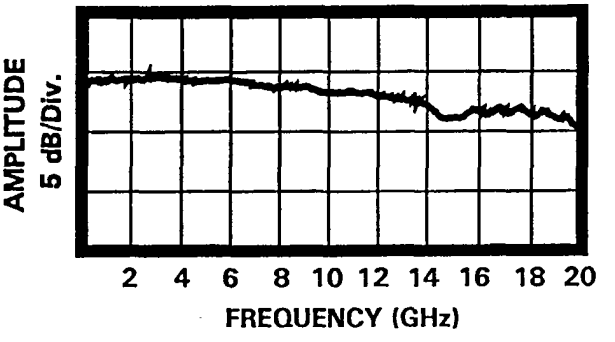


Figure 7. Electro-optic modulator frequency response, corrected for photodiode response.

In contrast to the semiconductor laser link, the bandwidth of the externally-modulated testbed fiber optic link is limited by the frequency response of the electro-optic modulator. This modulator has a measured -3dB bandwidth of 20 GHz, as shown in Figure 7. The testbed system was used to transmit test signals from 2 GHz to 12 GHz over a 29 kilometer distance in the DSN; the result for an 8.4 GHz tone is shown in Figure 8. The measured signal-to-noise ratio (SNR) at 8.4 GHz was 100 dB·Hz at the output of the 29 km fiber. The SNR was limited by the optical receiver noise, due to poor optical coupling at the fiber input. Similar SNR was obtained for test signals at 2 GHz, 10 GHz and 12 GHz. Improved optical coupling is expected to result in a 20 dB improvement in SNR.

The CTC phase noise of this link, measured at 8.1 GHz for a 10 meter link in a laboratory environment, is shown in Figure 9. Note that the phase noise sidebands are only visible above the measurement system noise for offset frequencies less than 1 Hz.

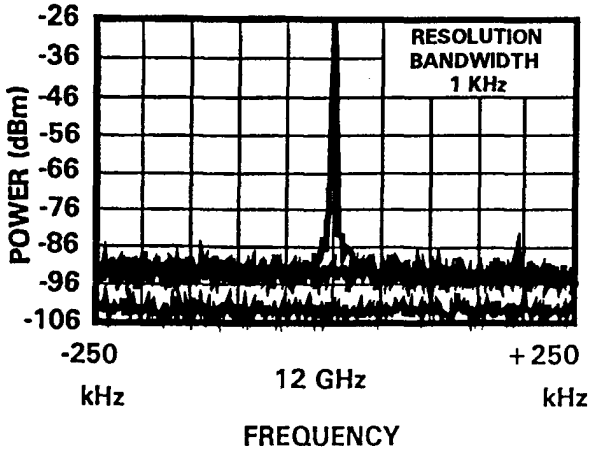


Figure 8. Upper trace: 8.4 GHz received signal at output of 29 km fiber link between DSS-13 and SPC-10 at Goldstone DSCC. Lower trace: Spectrum analyzer noise floor.

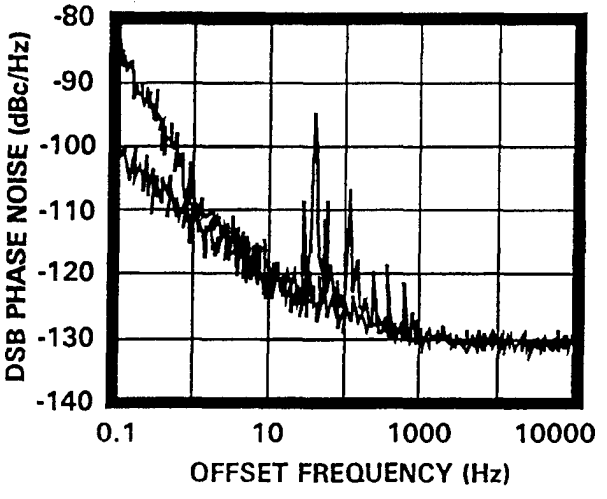


Figure 9. Upper trace: Phase noise of externally-modulated Nd:YAG transmitter at 8.1 GHz. Lower trace: Test system noise floor.

CTC Phase Noise Considerations

It was found experimentally that optical reflections in the path between the laser and modulator can seriously degrade the CTC phase noise performance of externally-modulated Nd:YAG fiber optic transmitters [9]. The mechanism for producing the degradation is different than in semiconductor lasers [10]. The narrow linewidth of the Nd:YAG laser implies that the phase of the optical carrier is coherent (deviates by less than 1 radian) over a distance $L_c \sim = c/(n\delta\nu)$. This provides a mechanism for producing low frequency fluctuations of the unmodulated optical carrier intensity, via interference between multiply-reflected signals. A simple scenario in which such an "unintended" interferometer is formed is sketched in Figure 10.

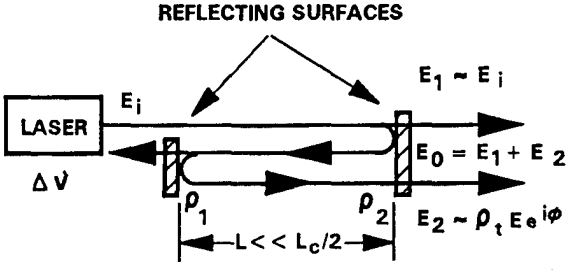


Figure 10. Sketch of an unintended interferometer formed from two partially-reflecting surfaces in the fiber optic path.

Consider two reflecting surfaces in the optical path of an unmodulated laser, such as lenses, or fiber optic connectors, separated by a distance L less than $L_c/2 = c/(2n\delta\nu)$, where n is the group index of the fiber, and $\delta\nu$ is the linewidth of the laser. To first order, the total electric field at the output will be

$$E_o(t) \sim E_i\left(1 + \rho_t e^{i\phi(t)}\right) \qquad (2)$$

Equation 2 is the sum of the input field amplitude E_i, and the doubly-reflected field that has experienced a phase shift of $\phi(t) = 4\pi nL(t)/\lambda$ radians, where λ is the wavelength of the optical carrier (1.3 microns in the testbed system), and ρ_t is the product of the amplitude reflection coefficients for each surface. $L(t) = L + \delta L(t)$ incorporates thermally-induced variations in the optical path length. The intensity at the output is then given by

$$\begin{aligned} I_o(t) &\sim E_o^* E_o \\ &\sim I_i(1 + 2\rho_t \cos\phi(t)) \end{aligned} \qquad (3)$$

where I_i is the intensity of the input field. From the expression for $\phi(t)$ above, it is apparent that a small change in L(t) on the order of 1 micron is sufficient to cause $\phi(t)$ to change by several radians. Standard optical fiber has a thermal coefficient of delay of $7 \cdot 10^{-6}$/degree C, so a temperature change of 1

degree on a 1 meter fiber will cause a length change δL of 7 microns. If careful attention is not paid to minimizing ρ_t in the optical path between laser and modulator, the optical carrier will experience a low-frequency intensity modulation that will be up-converted to interferometrically-induced phase noise sidebands on the RF modulation. Also, this interferometer will act as an optical frequency discriminator, transforming excursions of the optical carrier frequency to intensity modulation. This effect may become dominant for large L, and produce higher frequency noise components than the thermal fluctuations.

The phase noise results of Figure 9 were obtained when the components in the path between the Nd:YAG laser and modulator were oriented at slight angles, or were angle-polished, so that the reflected optical power from any surface was suppressed by at least 60 dB. Further testing is required to determine whether the residual phase noise shown in Figure 9 is due to interferometrically-induced low-frequency noise, the intrinsic noise of the laser source, the bias voltage on the modulator, the optical receiver, or some other source. In any case, it is noteworthy that this effect occurs completely external to the laser, and is not alleviated by simply increasing the amount of optical isolation to the laser. By contrast, reflection-induced degradation of CTC phase noise in injection-modulated laser diodes is caused by instability of the laser output in the presence of external reflections [10]. Interferometrically-induced RF phase noise has not been observed in a Fabry-Perot laser diode transmitter, due to the broad spectral width and hence extremely short coherence length of the optical carrier.

Conclusion

A new type of microwave fiber optic transmitter was described that circumvents limitations of previous laser diode fiber optic transmitters for high-quality, long distance microwave transmission. The new transmitter employs a solid-state Nd:YAG laser and a $LiNbO_3$ external intensity modulator. Phase noise data for a state-of-the-art laser diode transmitter at 100 MHz and 8.4 GHz were presented, as well as field measurements of differential Allan deviation for a 1 GHz 14 km distribution system. 29 km distribution field test results were also presented for the improved Nd:YAG laser transmitter. The single-sideband phase noise power spectral density of the Nd:YAG laser transmitter was measured to be L(f) = -115 dBc/Hz at 1 Hz from an 8.1 GHz carrier. The length-bandwidth product of the Nd:YAG laser transmitter is calculated to be 14,286 GHz·km, compared to 28 GHz·km for the semiconductor laser system. A new interferometrically-induced phase noise mechanism was discussed. It was shown that with adequate suppression of optical reflections, the Nd:YAG laser fiber optic transmitter is superior to previous semiconductor laser diode transmitters for long-distance distribution of high-quality microwave signals over single-mode optical fiber.

References

[1] L.E. Primas, R.T. Logan, G.F. Lutes, L. Maleki, "Distribution of ultra-stable reference frequency signals over fiber optic cable," Department of Defense Fiber Optics Conference '90, McLean VA, March 20-23, 1990, pp. 353-356.

[2] T.P. Krisher, L. Maleki, G. Lutes, L.E. Primas, R.T. Logan, J.D Anderson, C.M. Will, "Test of the isotropy of the one-way speed of light using hydrogen maser frequency standards," Physical Review D, Vol. 42, No. 2, pp. 731-734, 15 July 1990.

[3] T.T. Pham, M.K. Simon, T.K. Peng, M.H. Brockman, S.S. Kent, and R. Weller, "A carrier-arraying demonstration at Goldstone for Receiving Pioneer-11 Signals," Jet Propulsion Laboratory TDA Progress Report 42-106, pp. 307-334, 15 August 1991.

[4] J.D. Prestage, R.L. Tjoelker, G.J. Dick, L. Maleki, "Hg^+ trapped ion standard performance with the superconducting cavity maser oscillator as L.O.," Proc. of the 46th Annual IEEE Frequency Control Symposium, Hershey PA, 27-29 May 1992.

[5] C.J. Palais, *Fiber Optic Communications*, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1984.

[6] The laser transmitter used was a model 3530A from Ortel Corporation, Alhambra, CA, with an additional 35 dB optical isolator fusion spliced to the fiber pigtail output.

[7] K. Peterman, *Laser Diode Modulation and Noise*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1991.

[8] The laser used was a model ALC-50S from Amoco Laser Co., Naperville, IL; the modulator was

a GEC-Marconi model Y-35-8808-01 (angle-polished input fiber).

[9] R.T. Logan Jr., L.E. Primas, G.F. Lutes and L. Maleki, "Modulation signal stability considerations in analog fiber optic systems," DARPA/RADC Symposium on Photonics Systems for Antenna Applications, Monterey CA, December 13-15, 1990.

[10] R.T. Logan Jr., G.F. Lutes, L.E. Primas, and L. Maleki, "Design of a fiber optic transmitter for microwave analog signal transmission with high phase stability," Department of Defense Fiber Optics Conference '90, McLean VA, March 20-23, 1990, pp. 359-365.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

NRL/USNO TWO WAY TIME TRANSFER MODEM DESIGN AND TEST RESULTS

G. PAUL LANDIS
IVAN GALYSH

Naval Research Laboratory
Code 8322
4555 Overlook Ave. S.W.
Washington, D.C. 20375-5000
202-404-7061

Abstract

The Naval Research Laboratory (NRL) has developed a two way time transfer modem for the United States Naval Observatory (USNO). This modem in conjunction with a Very Small Aperture Terminal (VSAT) and a communication satellite can achieve sub nanosecond time transfer performance. The purpose of this paper is to present an overview of the hardware and software design of the digital modem showing its unique features and results of satellite testing. Time transfer performance is achieved in the hardware by a combination of stability, matching and calibration. Hardware stability is achieved by using synchronous digital methods in both the transmitter and receiver sections. Analog components are kept to a minimum and are wide band where possible to promote delay stability. Identical narrow band filters are used in the transmitter and the receiver section to help match temperature dependent delays. Time transfer between sites is a master slave technique, a master site transferring time to a network of slave sites. The master transmits first, specifically addressing a slave site and the slave site responds. This allows the master to operate on a predetermined schedule based on satellite availability, and the target to respond based only on signals it receives. Once communications has been established, data is exchanged in both directions through the satellite, thus eliminating the need for an additional communication path. A personal computer (PC) is used for operator interface, modem control, data collection and processing. With this method, both sites can be automated. Initial tests have been conducted to evaluate and verify modem performance and reduce the effects of other systematic errors. Modem tests at a common location were performed by direct connection and through two VSAT's. Tests used a common oscillator to the two modems to examine optimum comparison conditions. The test results will be presented.

Design Philosophy

Figure 1 shows a block diagram of the principle of the two way measurement. If the hardware sections are matched and corrections are made for the Sagnac effect on the propagation path, the clock difference can be found by taking half the difference in the measurements.

The system uses a Spread Spectrum Pseudo-Random Number (PRN), Code Division Multiple Access (CDMA), Binary Phase-Shift Keying (BPSK), modulation to transmit timing pulses and data between modems. The hardware was designed with as much digital hardware as practical for stability and compatibility with other modem designs or other CDMA

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

systems.

Hardware

Time transfer performance is achieved in the hardware by a combination of stability, matching and calibration. Hardware stability is achieved by using synchronous digital methods for most of the system in both the transmitter and receiver sections, along with wide band analog components to promote delay stability and matched narrow band filters in transmitter and receiver sections for canceling temperature dependent delays. See figure 2.

Analog Hardware

The analog components of the receiver are a relay for testing, a filter to prevent aliasing, an attenuator to control signal levels and two amplifiers for gain. The analog parts of the transmitter are a free-running oscillator at seventy MHz, a bi-phase modulator to add the code and data on the signal, an attenuator to control drive to the VSAT, an amplifier for gain, a filter to limit bandwidth and a relay for testing. Except for the filters, all of the analog signal parts are wide band to achieve delay stability. The narrow band filters are identical and located close together to help canceling temperature dependent delays.

Analog Digital Interfaces

The interfaces between the analog and digital sections are a D flipflop that drives the bi-phase transmitter modulator and an Analog to Digital Convertor (ADC) in the receiver. All the interface functions are clocked with the same twenty-five MHz.

Digital Hardware

The digital section can be divided into two areas, the real time digital hardware functions and programmable functions.

The real time Digital Signal Processing (DSP) hardware functions consist of a transmit section, a receive section and a timing section.

Digital Transmitter Section

The transmitter section consists of a transmit code rate Numerically Controlled Oscillator (NCO) and code generator. The equation below is the standard equation for the output frequency of a 32 bit NCO.

Output = Clock * Control / 2^32

With a 32 bit NCO and a clock of twenty-five MHz the frequency steps are ~0.0058 Hz. The wanted frequency might be off by as much as half this step. To overcome this problem the control word to the NCO is switched for four clocks on each output cycle of the NCO to a value that is X higher than the control word. This reduces the effective size of the accumulator by 4X to

2^32 - 4X.

The equation for the output frequency becomes

Output = Clock * Control /(2^32-4X)

This allows both the numerator and denominator to be programmed giving almost an unlimited selection of output frequencies. In this system the transmit code rate NCO can be set at any frequency in steps of 1/n where n is any integer up to 42. One Hz steps are used to generate the one Pulse Per Second (PPS) synchronization pulse in the transmitter system. See figure 3.

Digital Receiver Section

The receiver contains a receive code rate NCO, code generator, carrier to baseband down conversion and correlators. The baseband down conversion uses a NCO to generate an In (I) phase and a Quadrature (Q) phase Local Oscillator (LO) signal to drive a pair of digital mixers (multipliers) forming the I and Q baseband signals. The I and Q baseband signals are both despread by an early and a late reference code in four accumulators or correlators. The four correlators output are an Early In phase (EI), a Late In phase (LI), an Early Quadrature phase (EQ), and a Late Quadrature phase (LQ). The correlator outputs go to the programmable DSP processor. See figure 4.

Timing Section

The timing section is synchronized to the external one PPS and five MHz and is used to control the transmitter and to time tag the received data. Part of the modem start-up is to synchronize the internal timing section to the external one PPS. The first step in synchronizing is to measure the peak amplitude of the external one PPS and set an input threshold based on the peak level. The level is measured by using a comparator with one input connected to the one PPS and the other input connected to a digital to analog convertor (DAC) . The output of the DAC is increased until the output of the comparator is stable and then the comparator is set to a fraction of this level. The internal clock is twenty-five MHz and is derived by multiplying the five MHz input by five. The five MHz signal passes through an eight bit digitally controlled variable delay with a total range of about 50 nanoseconds and a resolution of approximately 0.2 nanosecond. The twenty-five MHz is connected to the D input of a D flipflop with the one PPS connected clock of the flipflop. The D flipflop samples the phase of the twenty-five MHz on the second. The output of the flipflop can be monitored by the DSP processor. By varying the five MHz delay and monitoring the phase of the twenty-five MHz at the D flipflop, the internal twenty-five MHz clock can be synchronized to the one PPS independently of the phase of the five MHz, the phase shift in the multiplier and the amplitude of the one PPS. If the phase of the five MHz is known to be stable with respect to the one PPS, this system can be used to monitor the phase stability of the twenty-five MHz multiplier. This method has demonstrated a repeatability of .2 nanoseconds. See figure 6.

Most of the real time digital hardware can be configured by the software. Codes, code rates, code lengths, code types and sample rate are all programmable.

Hardware Test Points

The modem contains two digital test points and two analog test points. The test points were designed for product development and trouble shooting. The digital test points selection is controlled from the keyboard. Epochs, codes, code clocks for both transmitter and receiver, the receiver carrier nco frequency, and the one PPS input and output are some of the signals that are available at the digital test ports. There also are two analog test points. The analog test points are driven by a dual twelve-bit Digital to Analog Convertor (DAC). This DAC is driven in real time from the TMS320C30. Output rates can be in excess of fifty thousand samples per second. With the pair of DAC's it is possible to monitor the digital outputs of the correlators on an oscilloscope in real time.

Dual Processor Configuration

The system uses a dual processor configuration that communicate through a dual port memory. One processor is a Texas Instrument TMS320C30 DSP integrated circuit capable of 33 Million FLOating Point Operations Per Second (MFLOPS) on 32 bit words. It is located in the backplane of the host processor an International Business Machine (IBM) compatible Personal Computer (PC), the other processor. The PC must have sixteen bit or larger backplane. The PC is used for operator interface, for data storage and to control the DSP processor. Both processors were programmed in C.

Codes

The modem uses a code generator chip with three independent code generators. The generators each use a 32 stages shift register with programmable taps to generate PRN codes. Also included in each code generators is a programmable initiation register that defines a start code and a programmable epoch register that generates a pulse when the generator reaches this pattern. Three programmable counters are also available. Any epoch or counter can be programmed to initialize any generator or counter. With this chip it is possible to generate maximum length sequences, non-maximum as Gold codes used by the Global Positioning System (GPS) System or other Ranging systems or truncated codes as used by the Hartl/Mitrex modem. The codes used in the modem are maximum length type with a length of 4095, and are selected to have low a cross correlation characteristic which has a peak value of 127. The feedback tap positions for these codes are 100000101001, 101000011000 and 110010100000 and the decimal equivalents 2089, 2584 and 3232. Figures 6 and 7 show some of the cross correlation characteristics of the codes. The cross correlation of the above codes can induce errors in the tracking loops equal to the code chip period times one half the ratio between the cross correlation and the auto correlation functions. With a chip frequency of 2,489,760 Hz, this error could be as large as 6.2 nanoseconds. This error is based on only one code period and with equal carrier frequencies. Simulations on these codes show that carrier offsets can increase the cross correlation by about 3 db or up to about 8.8 nanoseconds. The cross correlation in the system is not just the cross correlation of the codes for one code period but is also changed by the difference in the carrier of the signals, the data modulation on the code and is an averaged over about five minutes of data. For the system where the code repeats at 608 Hz rate, the cross correlation of 182,400 code periods over 300 seconds should have a net effect significantly below the one nanosecond range.

Sampling Rate and Code Chip Rate

There are several factors driving the System Configuration. The bandwidth of the satellite channel limited the chip rate to approximately 2.5 MHz. A 600-bit-per-second data rate is needed to transfer data and general communication. Words of 32 bits help match the computer architecture. The 2.5 MHz bandwidth and the 600-bit-per-second rate lead to a code length of 4095 and 600 bits per second lead to 19 words of 32 bits per word. The product of 19, 32 and 4095 give a chip rate of 2,489,760 Hz. The selection of the ratio sampling rate and code chip rates is critical. The system code rate is 2,489,760 Hz and the sampling rate is twenty-five MHz. The highest common frequency is 160 Hz which corresponds to 156,250 clock cycles and 15561 code cycles. The sampling rate of

25,000,000 samples per second can lead to a problem with the Hartl/Mitrex modem. This Hartl/Mitrex modem code rate is 2,500,000 which is also the highest common frequency with the sampling rate. This commensurability problem would be a quantization of TOA measurements at 40 nanoseconds. Additional hardware is available in the modem to add phase noise to the output of the receive code rate NCO to correct this problem. This noise will have a uniform distribution over and a peak-to-peak range of forty nanoseconds. This addition takes place before the time quantization of forty nanoseconds.

Signal Search

A model of the four components of the correlator outputs is given below.

CORRELATOR OUTPUTS and
LOOP FILTER INPUTS

M = SIGNAL MAGNITUDE
R = CARRIER PHASE ERROR RADIANS
C = CODE OFFSET ERROR IN CHIPS
D = DATA MODULATION IN RADIANS

EI = IN PHASE EARLY
EQ = QUADRATURE PHASE EARLY
LI = IN PHASE LATE
LQ = QUADRATURE PHASE LATE

For $-.5 < C < .5$
$EI = M (.5 - C) \cos (R + D)$
$EQ = M (.5 - C) \sin (R + D)$
$LI = M (.5 + C) \cos (R + D)$
$LQ = M (.5 + C) \sin (R + D)$

SIGNAL LEVEL SQUARED
$M^2 = (EI + LI)^2 + (EQ + LQ)^2$

CODE LOOP INPUTS
$C = (LI^2-EI^2+LQ^2-EQ^2)/(2M^2)$

CARRIER LOOP INPUTS
$R = \arctan((EQ+LQ)/(EI+LI))$

To acquire the signal, the correct carrier frequency and phase and code phase must be found. The carrier frequency can be offset from nominal by the seventy MHz crystal at the transmitter, the VSAT oscillators, the satellite oscillator or the clock at receiver. The largest of these is the satellite and next the VSAT. The modem can be tuned over several MHz but the software search for the carrier is restricted to plus or minus 19 KHz from nominal. The code phase is unrestricted and a full code period search of 1.6 milliseconds is performed. To decrease acquisition time, a Fast Fourier Transform (FFT) is used to search for the unknown frequency while the reference code is slid to find the received code phase. The samples of the FFT are calculated by adding the early and late components of I and Q to form the complex samples. Normally during tracking the accumulators are sampled at the end of each code period for 608 samples per second rate, but during acquisition the sampling rate is increased 64 times to 38,912 samples per second. Blocks of 64 complex samples are saved, the length of one code period. The first sample of the block is synchronized with the code period to prevent possible cancellation if a data bit transition occurred in the middle of the block. An FFT is calculated on the block of 64 complex points. The output of the FFT is scanned to find the maximum frequency. If the energy at this frequency is significantly greater than average energy in all of the non-adjacent frequencies, a signal is assumed to be present. The search is performed over the entire code period to insure that the maximum peak signal is found and that the modem does not lock on to a lesser peak. After the peak is found an estimate on the frequency is calculated by looking at the magnitude of the largest frequency found by the FFT and the magnitude of the largest adjacent frequency. The carrier NCO is tuned by this calculation to

reduce this frequency from the output of the correlators to much less than 304 Hz. A second carrier NCO adjustment is made by watching for over a second the complex spinning rate on the output of the correlators. At this point the code phase and carrier frequency are close and the phase locked loops for both the carrier and code can be enabled.

Code and Carrier Tracking Loops

The error signal for the carrier loop is R and the code loop is C in the equations for the correlator outputs. A double argument arctangent routine could be used to solve for angle of the carrier and data phase errors, R+D, but the single argument arctangent routine produces an output that is independent of D. The PI ambiguity in the routine is the same as the data modulation. The DSP software calculates the value of C and R for carrier and code and they are used as inputs to two digital filters. The output of the filter is added to the approximate center frequency value and feeds the results are feed to the NCO. This forms the carrier or code phased locked loop. If C and D are within bounds, both loops are independent of M and D and the carrier loop is independent of C and the code loop is independent of R. This operation is done at the 608 Hz rate of the code epochs. Bit Synchronization The purpose of the bit synchronization is to mark the second point in the message. The synchronization word is 32-bits long for convenience. It is sixteen zeros followed by sixteen ones. This pattern was chosen to resolve the 180-degree phase ambiguity problem with bi-phase modulation, to use most of the ASCII character and to match the word size in the processor. Synchronization is recognized by looking for sixteen zeros followed by sixteen ones or by looking for sixteen ones followed by sixteen zeros. The significance of the latter is that the carrier loop tracks 180-degrees out of phase. If this happens data is decoded by simply inverting each bit.

Time Of Arrival Calculations

The Time Of Arrival (TOA) is not measured directly. The code loops generates a reference code that tracks the incoming signal. The timing of the reference loop is measured at a 608 Hz rate. This measurement is made at each code epoch. The epoch time is measured to 40 nanoseconds by counting the twenty-five MHz clock. The phase of the reference code clock is measured by reading the accumulator of the reference code clock NCO. With a code clock of 2,489,760 Hz and a twelve-bit phase measurement the time of the epoch can be measured to less than one hundred picoseconds. By looking for the synchronization pulse the one second epoch can be recognized. The measurements are reduced by using a least squares linear fit calculated on the 608 measurements centered around the one second epoch. The variance is also calculated. The TMS320C30 is only a single precision (32 bit) floating point processor. The mantissa of a 32-bit floating point word has 24 bits of precision which is not enough precision for our results. By removing the known slop characteristics of the data before the linear fit the results can be calculated and interpolated by the TMS320C30.

Operation

Time transfer between sites is a master slave technique, a master site transferring time to a network of slave sites. The master transmits first, specifically addressing a slave site and the slave site responds. This allows the master to operate on a predetermined schedule

based on satellite availability, and the target to respond based only on signals it receives. Once communications has been established, data is exchanged in both directions through the satellite, thus eliminating the need for an additional communication path. The PC is used for operator interface, modem control, data collection and processing. With this method, both sites can be automated.

Data

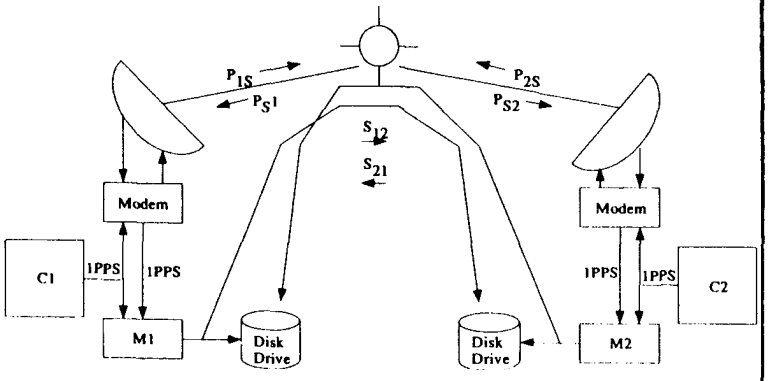
Data is transmitted between the two modems. This data contains a synchronization pulse that repeats every second, time of day, measurement, and station identification information. There is also area in the format for general purpose communication. The source of the general purpose communication data is taken from a disc file on the PC and is passed through the dual port ram to the DSP system. The DSP merges this data with the synchronization pulse, time of day and measurements. Using the Direct Memory Access (DMA) interrupt and serial port features of the TMS320C30, a serial bit stream drives an exclusive or circuit in the transmit code generator. The received data is processed by the DSP system. Working with the data from the correlators it searches for the synchronizing pulse and resolves the 180 degree ambiguity of the data. This information is then passed through the dual port memory to the PC where it is split and the parts are stored in different files.

Test Results

Initial tests have been conducted to evaluate and verify modem performance and reduce the effects of other systematic errors. Modem tests at a common location were performed with single and dual VSAT's. All tests used a common oscillator and location to the two modems to examine modem stability. A pair of power dividers were used at seventy MHz and the single VSAT case to connect the two modems to the VSAT. No attempt was made to match cable lengths or calibrate the modems. The test did not examine the effects of cross correlation because of the common clock and common up-link distance. Future tests to examine cross correlation will use an offset between the oscillators of the two modems to look at the non-linear characteristics in time transfer data. Two distance locations will also show cross correlation problems but additional information will be needed to separate clock and range changes from the cross correlation problems. For most of the test a 200 pico seconds noise level was showed on the data with day-to-day repeatability below this level. Two problems have been seen in the data. The data on a pass-to-pass basis will occasionally jump twenty nanoseconds due to a twenty-five MHz clock slip in one unit, half a clock period. Faster chips are going to be used to fix this problem. A second problem is a discrete change of one to two nanoseconds. This problem has not been pinpointed at the present time. Figures 8 and 9 show the performance of the modem and figures 10 and 11 show the two modem problems.

Figure 1

Delays and Measurements



$M_2 = C_1 + T_1 + P_{1S} + S_{12} + P_{S2} + R_2 - C_2$

$M_1 = C_2 + T_2 + P_{2S} + S_{21} + P_{S1} + R_1 - C_1$

$\frac{M_2 - M_1}{2} = C_1 - C_2 - \frac{T_1 - R_1}{2} - \frac{T_2 - R_2}{2} + \frac{(P_{1S} - P_{S1})}{2} + \frac{(P_{S2} - P_{2S})}{2} + \frac{(S_{12} - S_{21})}{2}$

M = Measurement
C = Clock
P = Path
S = Satellite
T = Transmit
R = Receive

Figure 2

Hardware Stability Design Method

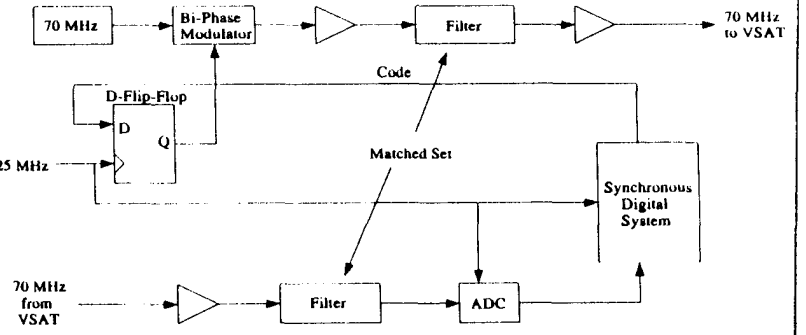
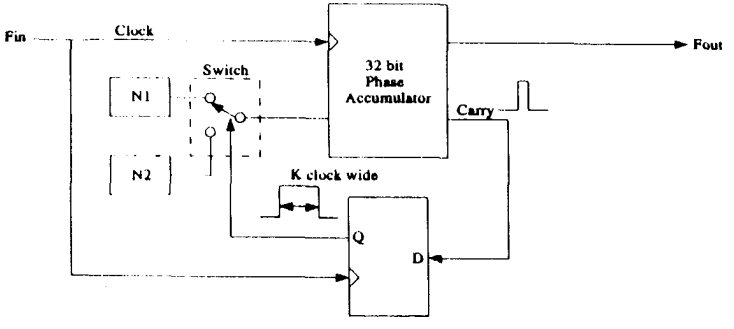


Figure 3

NCO with Programmable Numerator and Denominator



$$Fout = \frac{Fin \cdot N1}{2^{32} - K[N2 - N1]}$$

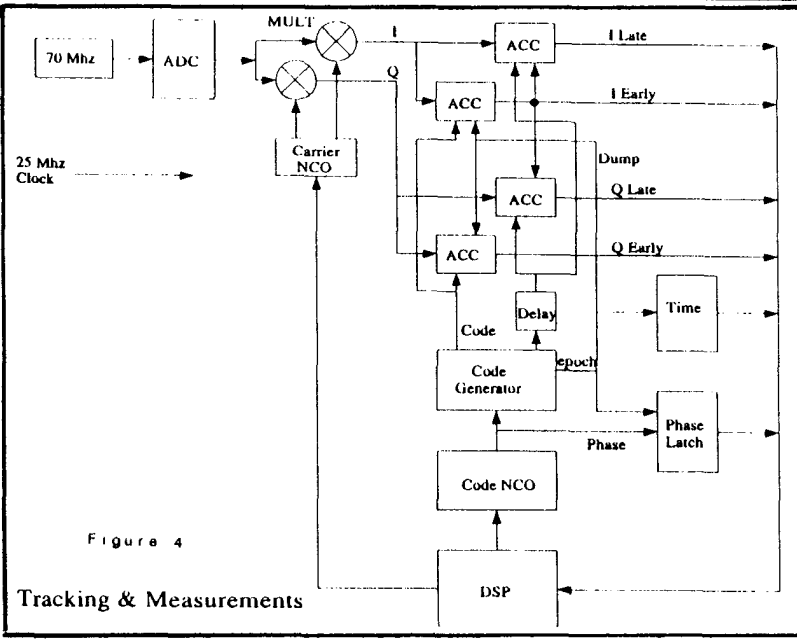


Figure 4

Tracking & Measurements

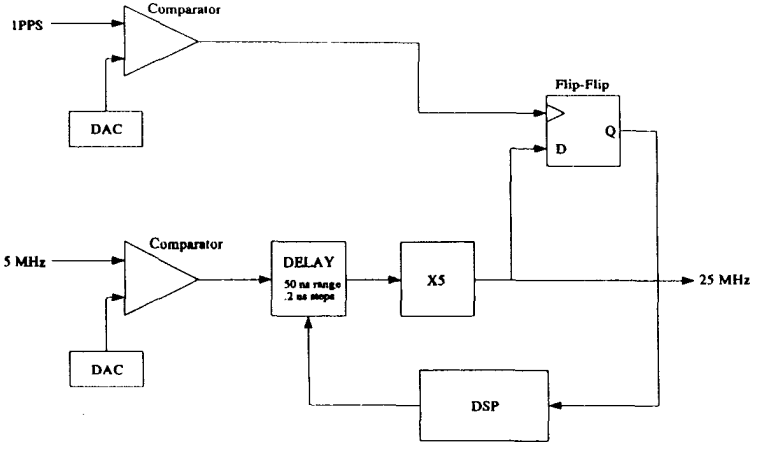


Figure 5

Reference Input Circuit and Internal Clock Synchronization

Figure 6

Cross Correlation, Code Taps: 2089, 3232

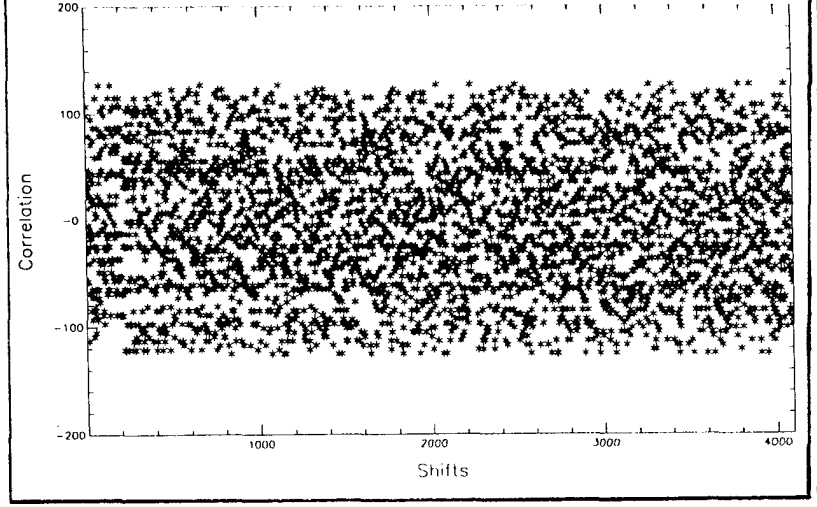


Figure 7

Cross Correlation, Code Taps: 2089, 2584

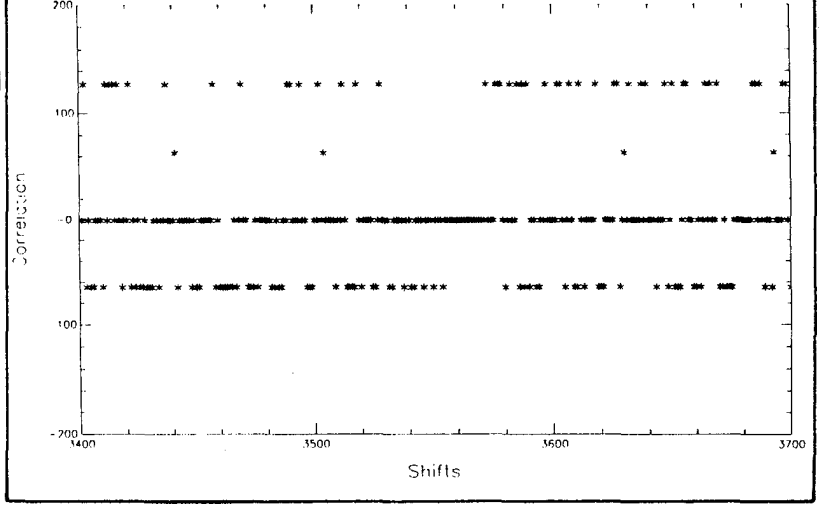


Figure 8

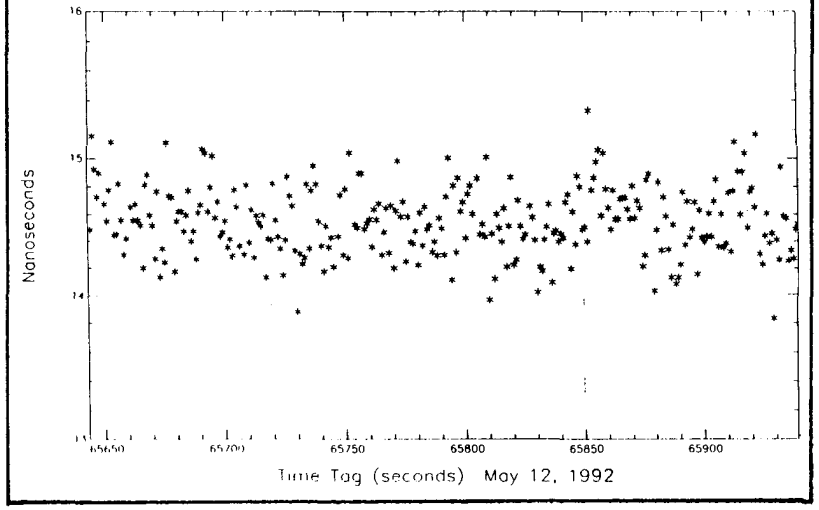


Figure 9

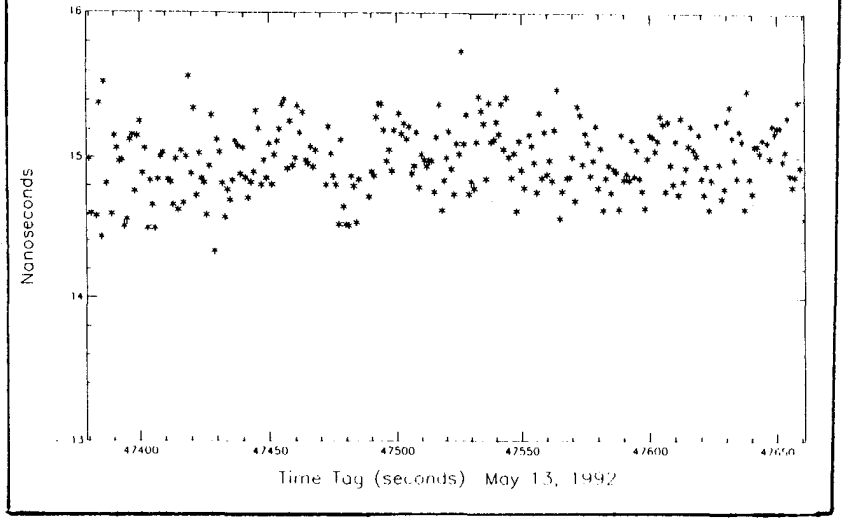


Figure 10

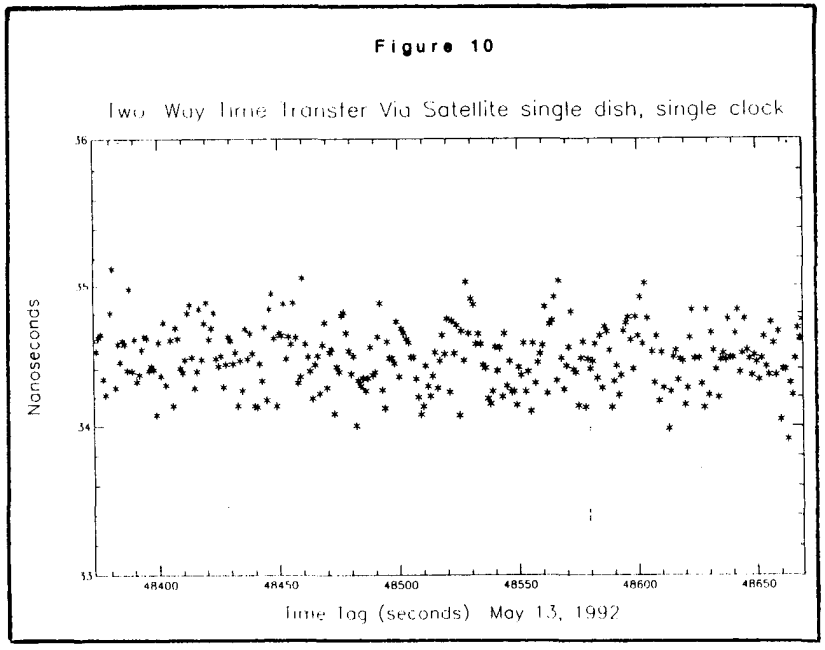
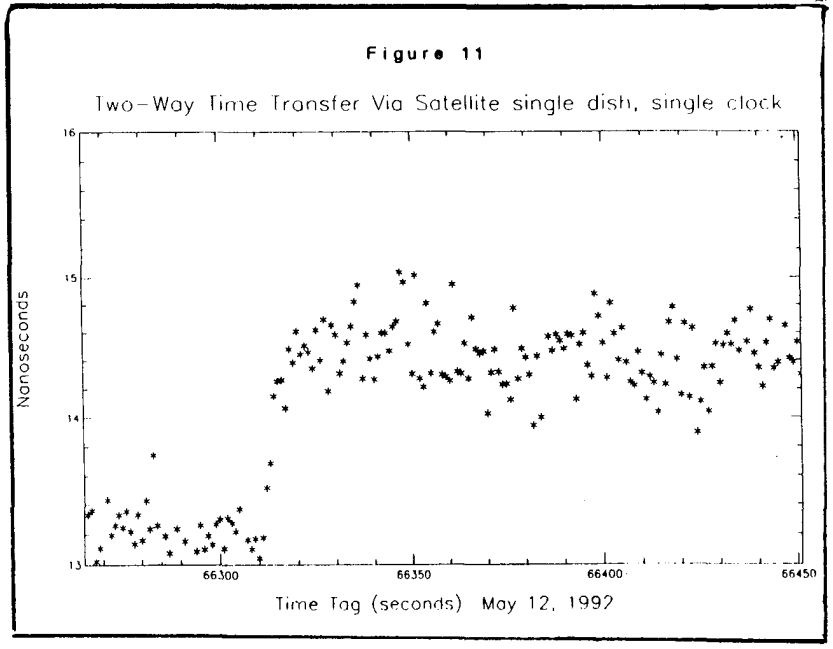


Figure 11



1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A NEW MODEL OF 1/F NOISE IN BAW QUARTZ RESONATORS

F. L. Walls*, P. H. Handel#, R. Besson@, and J.-J. Gagnepain+

* Time and Frequency Division, National Institute of Standards and Technology, Boulder Colorado, USA; #Dept of Physics, University of Missouri, St Louis, Missouri, USA;@ Ecole Nationale Superieure de Mecanque et des Microtechniques, Besancon, France; +Laboratoire de Physique et Metrologie des Oscillateurs du CNRS, Besancon, France

Abstract

This paper presents a new model for predicting the 1/f (flicker) frequency noise in quartz resonators as a function of the unloaded resonator quality factor Q and volume under the electrodes for bulk acoustic wave (BAW) resonators. The functional form of this model originates from a quantum 1/f theory for scattering of phonons from the primary oscillator mode. Using this new model, we are able to match the 1/f frequency noise observed in the best quartz-controlled oscillators and resonators. Quite unexpectedly, this model indicates that the amplitude of 1/f frequency noise might be improved by making resonators with smaller electrodes. BVA resonators show approximately a factor of 3 improvement in 1/f frequency noise ($S_y(f)$ over electroded resonators with the same unloaded Q-factor and electrode volume.

Introduction

The amplitude of 1/f or flicker frequency noise in quartz resonators is a very important parameter of oscillators used in a wide range of applications. Although work has been done in this area for more than 20 years, it is difficult to find data where the operating conditions and all of the resonator parameters are well known. In addition, some workers report S_ϕ(1 Hz), while others report σ_y(1 s or 100 s). For a given resonator Q, many different levels of frequency stability have been reported. Some of the variation may be due to random walk and drift which were not removed from the data and thereby bias the estimate the amplitude of 1/f frequency noise. In some cases the electronics may limit the noise. What is needed for accurate modeling is the amplitude of the 1/f frequency noise in the resonator, independent of electronic noise. The variation in $S_y(f)$ for the same Q may also indicate that some other variable significantly affects the 1/f frequency noise as the acoustic losses become small. Given these problems and uncertainties in the available data, it is difficult for any model to fit all the data. The 1/f contribution to frequency stability is best obtained by observing the stability over a range of measurement times in the time domain or a range of frequencies in the frequency domain. From the extended data we can fit a flicker-of-frequency model to the data that exclude the biases due to random walk FM and drift present in many resonators and oscillators (see Fig. 1).

Contribution of the U. S. Government, not subject to copyright.

Gagnepain [1] was one of the first to systematically study 1/f noise as a function of geometry, temperature, and Q. He found that the 1/f contribution to the spectral density of fractional frequency fluctuations, $S_y(f)$, varies as $1/Q^4$ for resonators between approximately 1 and 25 MHz. As the temperature of a resonator changes, Q changes. This makes it possible to exclude the effect of many other factors. Additional work by Parker, however, showed that the data from both BAW and Surface Acoustic Wave (SAW) devices could be roughly fit to the same model if one assumes a $1/Q^4$ dependence for $S_\phi(f)$ instead of $S_y(f)$ [2]. The fit is not particularly good, for the best resonators (see Fig. 2).

From a theoretical viewpoint, work on the general problem of 1/f fluctuations in systems has long pointed toward a $1/Q^4$ dependence for $S_y(f)$ [1,3,4,5]. Work on many systems other than quartz has yielded very good quantitative agreement between theory and experimental data for 1/f quantum noise [5]. The $1/Q^4$ dependence of the 1/f contribution to $S_y(f)$ is, however, in apparent conflict with the 1/f noise of the best quartz resonators over a wide frequency range where the dependence is between $1/Q^2$ and $1/Q^5$.

This paper refines the previous theoretical work on 1/f noise in quartz to suggest a better framework for predicting the amplitude of 1/f noise in quartz resonators over a wide range of frequencies and Q [1,3,4,5].

0-7803-0476-4/92 $3.00 © 1992 IEEE

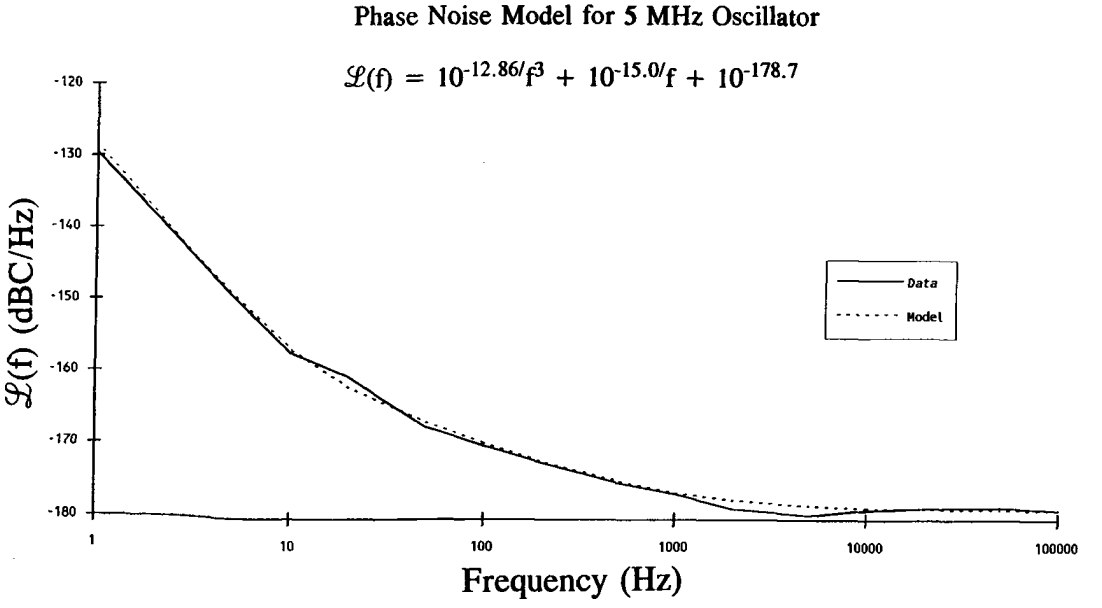


Figure 1. Phase noise of a 5 MHz quart z oscillator as a function of Fourier frequency. The coefficient of the f^{-3} component corresponds to 1/f (flicker) frequency noise.

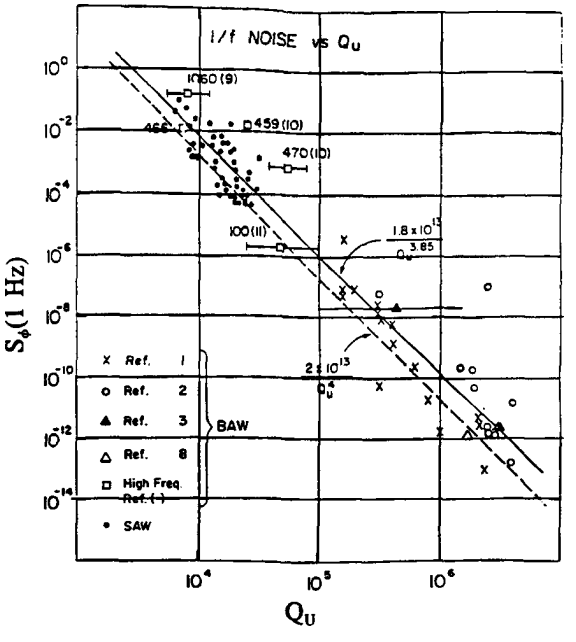


Figure 2. 1/f noise level at 1 Hz (S_ϕ(f) of quartz acoustic resonators as a function of unloaded Q [2]. Reference numbers are from [2].

Condensed Theory of 1/f Noise in Quartz Resonators

According to the general quantum 1/f formulation [5], $\Gamma^{-2}S_\Gamma(f) = 2\alpha A/f$ with $\alpha = \epsilon^2/\hbar c = 1/137$ and $A = 2(\Delta J/ec)^2/3\pi$ 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 the spectral density

$$\Gamma^{-2}S_\Gamma(f) = 4\alpha(\Delta\dot{P})^2/(3\pi e^2c^2), \qquad (1)$$

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. These fluctuations in the rate Γ are obtained by scattering on a phonon from any other mode of average frequency $<\omega>$, or through a two-phonon process at a crystal defect or impurity, involving a phonon of average frequency, $<\omega'>$. To calculate it, we write the energy W of the interacting resonator mode $<\omega>$ in the form

$$W = n\hbar<\omega> = 2(Nm/2)(dx/dt)^2 =$$

$$(Nm/e^2)(edx/dt)^2 = (m/Ne^2)\epsilon^2(\dot{P})^2. \qquad (2)$$

The factor 2 includes the potential energy contribution. Here m is the reduced mass of the elementary oscillating dipoles, θ their charge, ϵ a

polarization constant, and N their number in the quartz crystal between the electrodes. Applying a variation $\Delta n = 1$, we get

$$\Delta n/n = 2|\Delta\dot{P}|/|\dot{P}|, \text{ or } \Delta\dot{P} = \dot{P}/2n. \quad (3)$$

Solving Eq. (2) for $\dot{P}$ and substituting, we obtain

$$|\Delta\dot{P}| = (N\hbar<\omega>/n)^{1/2}(e/2\epsilon). \quad (4)$$

Substituting $\Delta\dot{P}$ into Eq. (1), we get

$$\Gamma^{-2}S_\Gamma(f) = N\alpha\hbar<\omega>/(3n\pi mc^2f\epsilon^2) = \Lambda/f. \quad (5)$$

This result is applicable to the fluctuations in the loss rate Γ of the quartz. In the presence of a damping term Γ, the frequency of a harmonic oscillator of unperturbed angular frequency ω_0 is

$$\omega = \sqrt{\omega_o^2 - 2\Gamma^2} \text{ where } Q = \frac{\omega}{2\Gamma}. \quad (6)$$

The fractional variation of ω_0 due to fractional changes in Γ is

$$\frac{\Delta\omega}{\omega_0} = \frac{2\Gamma^2}{\omega_0\sqrt{\omega_0^2-2\Gamma^2}}\frac{\Delta\Gamma}{\Gamma} \sim \frac{1}{2Q^2}\frac{\Delta\Gamma}{\Gamma}. \quad (7)$$

The spectral density of frequency fluctuations of the quartz resonator is [4]

$$\omega^{-2}S_\omega(f) = (1/4Q^4)(\Lambda/f) = N\alpha\hbar<\omega>/(12n\pi mc^2f\epsilon^2Q^4), \quad (8)$$

where Q is the unloaded quality factor of the single-mode quartz resonator considered, and $<\omega>$ is not the circular frequency ω_0 of the main resonator mode, but rather the nearly 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 also has to be included in principle, but is negligible because of the very large number of phonons present in the main resonator mode.

Eq. (7) can be written in the form

$$S_y(f) = \beta V/(fQ^4), \quad (9)$$

where, with an intermediary value $<\omega> \sim 2 \times 10^{11}$/s, $n = kT/\hbar<\omega>$, T = 300 K and $kT = 4 \times 10^{14}$, $\beta \approx (N/V)\alpha\hbar<\omega>/(12n\pi\epsilon^2mc^2) = 10^{22}(1/137)(10^{-27}10^8)^2/(12kT\pi(2 \times 10^{-24}) 9\times10^{20}) \approx 1$.

The form of Eq. (8) shows that the amplitude of 1/f frequency noise depends not only on Q^4 as previously proposed but also on the volume between the electrodes. This model qualitatively fits the data of Gagnepain et al. [1,3] and the recent data of El Habti and Bastien [6], where Q was varied with temperature in the same resonator (but not frequency or volume).

The model also provides the basis for predicting how to improve the 1/f noise in resonators beyond just improving the Q, which has been known for many years. Since the amplitude of 1/f noise depends on active volume, we use the lowest overtone and smallest electrode diameter consistent with other circuit parameters.

Experimental Measurements and Analysis of 1/f Noise in Quartz Resonators

The 1/f frequency noise in quartz resonators has been measured using phase bridges and in complete oscillators [1,2,6-12]. Unfortunately much of the data in the literature are unusable for modeling because the unloaded Q is unknown. (Our case is even more restrictive because we also need to know the electrode size). The advantages of the phase bridge approach are that the unloaded Q can be easily measured and the noise in the measurement electronics can be evaluated independent of the resonator. If resonator pairs are used, driving source noise can generally be neglected and the pair can operate at virtually any frequency [9]. The oscillator approach makes it possible to compare many different resonators one at a time. The noise of individual oscillators can be derived by measuring the phase noise between 3 oscillators [14].

Figure 3 taken from [2] is one of several studies showing that the amplitude of 1/f frequency noise is virtually independent of the loaded Q. This is in complete agreement with the theoretical model. In practical oscillators dependence on loaded Q occurs only when the phase noise of the sustaining electronics contributes to the overall noise level.

We have analyzed 1/f frequency noise as a function of unloaded Q, volume under the electrodes, and frequency. For a given resonator geometry and manufacturer, we have taken the best values reported of $S_y(f)$ to remove the effects of poor crystals or electronics. In Fig. 4 we have taken all of the precise data available with unloaded Q, electrode volume, and frequency stability and plotted it

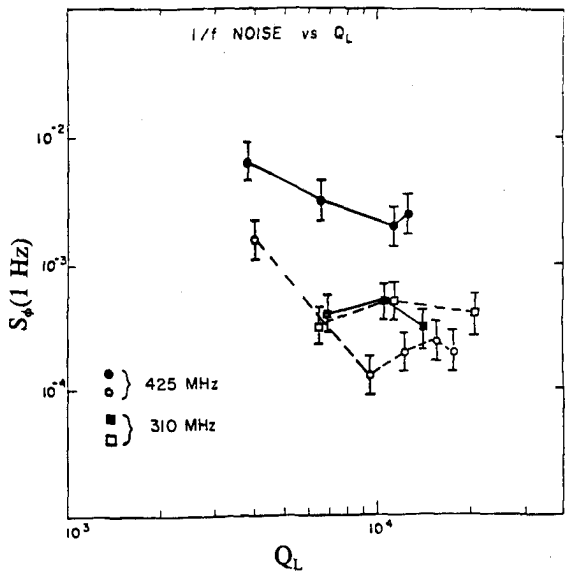


Figure 3. 1/f noise level at 1 Hz (S_ϕ(f) of 4 SA[7] resonators as a function of loaded Q [2].

according to the three models. Except for the 2.5 MHz resonator where $Q\nu_o$ = 0.95 x 10^{13}, the $Q\nu_o$ product for all resonators plotted is near 1.2 x 10^{13} (this is close to the material limit for AT and SC cut resonators). The curve labeled K_y shows the fit of the data to the model [1]

$$S_y(f) = K_y/f\,(3 \times 10^{-5}/Q^4). \qquad (10)$$

K_y varies about a factor of 500 for Q between 10^5 and 3.8 x 10^6 (with resonator frequencies between 2.5 and 100 MHz). The curve labeled K_ϕ shows the fit of the same resonator data to the model [2]

$$S_\phi(f) = K_\phi/f^3\,(3 \times 10^{10}/Q^4). \qquad (11)$$

K_ϕ varies about a factor of 10 for the same range in Q. Curves β_e and β_b show the fit of the same resonator data to the model

$$S_y(f) = \beta/f\,(Vol/Q^4), \qquad (12)$$

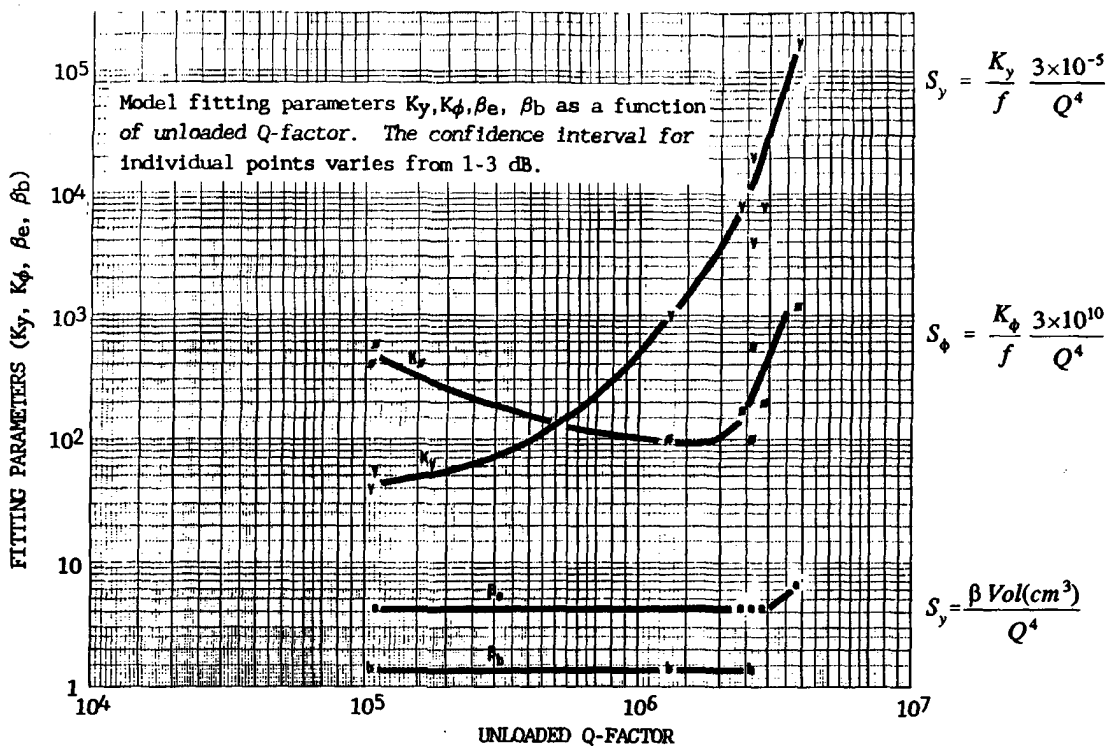


Figure 4. 1/f frequency noise for 9 resonators plotted according to the three different models using the fitting parameters Ky, Kϕ, βe, and βb as a function of the unloaded Q-factor. The confidence interval for the individual points varies from 1 to 3 dB. Only resonators with the lowest level of 1/f noise are reported for each type [8, 11, 13].

where $ß_e$ is for SC and AT resonators with electrodes plated on the resonator and $ß_b$ is for BVA-style AT and SC resonators [7]. Volume between the electrodes (in cm^3) is used to approximate the volume of quartz contributing to the scattering of power from the primary resonator mode. The ß factors are remarkably constant for Q from 10^5 to 3.8×10^6.

Figure 5 shows the dependence of β on Q for 3 different types of resonators as measured by Norton [8]. The wide variation in $ß_e$ and β_b for the same type of resonator and Q indicates that acoustic loss is not the only mechanism contributing to the noise level. The data for this graph were taken from measurements of $\sigma_y(100\ s)$ in similar oscillators. The difference between the various resonators of a given type can only be due to differences in the resonators. The reference oscillator for these measurements was a hydrogen maser. The data of Fig. 4 used only the smallest value of β for each resonator type.

Discussion

The 1/f frequency noise of the most stable resonators is in excellent agreement with the functional form of Eq. (8). The agreement between the theoretical and the experimentally measured values of β to the same order of magnitude is remarkable considering the rough approximation of both the average phonon frequency $<\omega>$ and the number N of oscillating dipoles (through volume) under the electrodes contributing to the scattering processes. The data suggest that the correct volume is that between the electrodes and not the volume of the oscillating mode. Further investigation of this question would help in interpreting the theory. One method would be to measure the 1/f frequency noise of a high performance resonator as a function of the diameter of the electrodes, since many other variables would be held constant. We are not surprised that $ß_e$ and $ß_b$ are different for the two types of resonators, since energy trapping and electrode stress are considerably different.

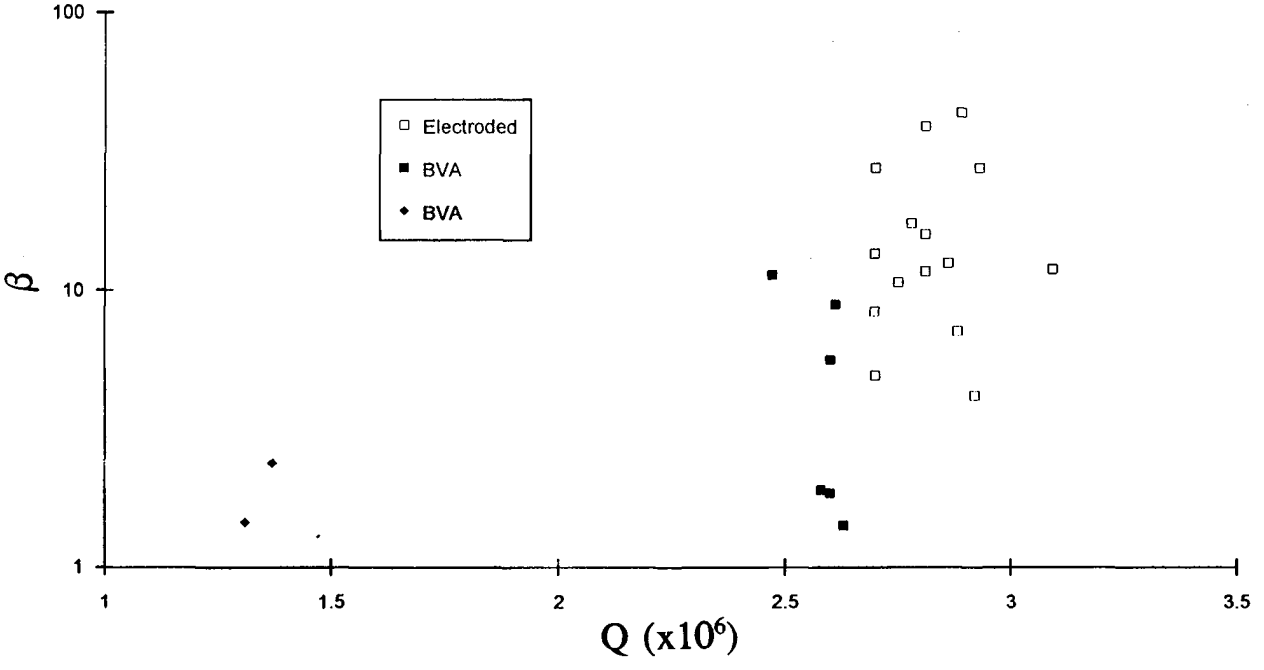


Figure 5. Fitting parameter β as a function of unloaded Q-factor for three types of resonators. The resonators in each group were matched in all known electrical parameters except Q-factor and 1/f noise [15].

Figure 5 shows that there is at least one other noise process besides acoustic loss that affects the 1/f frequency noise in some resonators. The magnitude of β in Fig. 5 is larger for the electroded resonators than for the BVA resonators. This suggests that the extra noise source is associated with the electrode-quartz interface. The fractional variation in β is roughly comparable for the two types of resonators. Much more data on resonators from the same material with the same surface preparation are needed to make any further conclusions.

Although we have analyzed only the data for a few resonators, the consistency of β_e and β_b over a factor of 40 in Q and resonator frequency and the general agreement for the magnitude of ß between theory and experiment give us confidence that this new model can be used to predict the best performance of different resonator geometries and as a basis to analyze other 1/f noise processes in quartz resonators.

This new volume model predicts that a resonator having smaller electrodes would have lower 1/f frequency noise than another with the same frequency and Q but with larger-diameter electrodes. The decrease in electrode area would increase the impedance and degrade the wide-band noise somewhat. For most resonators the wideband noise is dominated by the electronics and not the resonator. The increase in series resistance, obtained by decreasing the electrode area by a factor of 4, would probably be tolerable from the standpoint of wideband noise, but might require a change in loop gain.

BT-cut resonators are potentially useful in that they offer a $Q\nu_o$ product approximately 3 times higher than that of AT- and SC-resonators. BT-cuts are nearly as sensitive to temperature transients as AT cuts. Therefore to achieve parts in 10^{-14} frequency stability with BT cuts would require temperature stabilities of order 10^{-9} K/s or 100 times better than is required for SC-cut resonators [14]. The phase noise requirements of the sustaining electronics would be less than for other resonators due to the increase in Q.

We conclude based on these early observations that the amplitude of 1/f frequency noise in quartz may yet be improved to 10^{-14} by applying one or more of the following techniques: reducing the electrode area, using lower-overtone resonators, using BVA type resonators, going to lower frequencies, and using BT-cut resonators. Acceleration-induced effects, however, become more dominant as the stability improves [16].

Acknowledgements

We thank R. Michael Garvey of Frequency and Time Systems; Stan Kaat of Bliley Electric Co.; David Lind of Colorado Crystal Corp.; Jerry Norton of the Applied Physics Laboratory, Johns Hopkins University; Charles Stone of Frequency Electronics; and John Vig, of SLCET-EQ, U. S. Army for fruitful discussions and for sharing unpublished data with us. The majority of the data in Fig. 4 and all of the data in Fig. 5 were graciously supplied by Jerry Norton.

References

[1] J.-J. Gagnepain, "Fundamental Noise Studies of Quartz Crystal Resonators," Proc. 30th Ann. Frequency Control Symp., 1976, pp. 84-91.

[2] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", Proc. 41st Ann. Frequency Control Symp., 1987, pp. 99-110.

[3] 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," Proc. 35th Ann. Frequency Control Symp., 1981, pp. 476-483.

[4] P. H. Handel, "Nature of 1/f Frequency Fluctuations in Quartz Crystal Resonators Solid State Electronics," vol. 22, pp. 875-876, 1979.

[5] P. H. Handel "Quantum Approach to 1/f Noise," Phy. Rev., vol. 22A, pp. 745-757 1980.

[6] U.R. Peier and G. Jenni, "Space qualified 100 MHz BVA-VHF Resonator," Proc. 2nd European Frequency and Time Forum, 1988.

[7] R. Besson, "A New Piezoelectric Resonator Design," Proc. 30th Ann. Frequency Control Symp., 1976, pp. 78-83.

[8] J. Norton, "BVA-Type Quartz Oscillator for Spacecraft," Proc. 45th Ann. Frequency Contol Symp., 1991, pp. 426-430.

[9] F. L. Walls and A.E. Wainright, "Measurement of the Short-Term Stability of

Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications," IEEE Trans. on Inst. and Meas., vol. IM-24, pp. 15-20, 1975.

[10] J.-J. Gagnepain "A Comparison of Frequency Noise of Quartz Resonators," Proc. 39th Ann. Frequency Control Symp., 1985, pp. 135-137.

[11] A. El Habti and F. Bastien, "Low Temperature Limitation on the Quality Factor of Quartz Resonators," Proc. 46th Ann. Frequency Control Symp., 1992, to be published.

[12] J.-J. Gagnepain, M. Olivier, and F. L. Walls, "Excess Noise in Quartz Crystal Resonators," 37th Ann. Symp. on Freq. Cont., 1983, pp. 218-225.

[13] M. B. Bloch, J. C. Ho, C. S. Stone, A. Syed, and F. L. Walls, "Stability of High Quality Quartz Crystal Oscillators: An Update," 43rd Ann. Symp. on Freq. Cont., 1989, pp. 80-84.

[14] "Characterization of Clocks and Oscillators," Eds. D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls, NIST Tech Note 1337, 1990.

[15] F. L. Walls and J.-J. Gagnepain, "Environmental Sensitivities of Quartz-Crystal-Controlled Oscillators, IEEE UFFC, 1992, to be published.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

REDUCTION OF QUARTZ CRYSTAL OSCILLATOR FLICKER-OF-FREQUENCY AND WHITE PHASE NOISE (FLOOR) LEVELS AND ACCELERATION SENSITIVITY VIA USE OF MULTIPLE RESONATORS

M. M. DRISCOLL
Westinghouse Electronic Systems Group
Baltimore, MD

ABSTRACT

Through the use of N number of series-connected, quartz crystal resonators in an oscillator circuit, a $10\log_N$ reduction in both flicker-of-frequency noise and white phase noise (floor) levels is possible and has been demonstrated. The reduction in flicker noise occurs as a result of the uncorrellated short-term frequency instability in each of the resonators, and the reduction in noise floor level is a simple result of the increase in net, allowable crystal drive level [1].

This technique has been used in 40-MHz, 80-MHz, 100-MHz, AT, BT, and SC-Cut crystal oscillators using low flicker-of-phase noise, modular amplifier sustaining stages, and four series-connected crystals. Total (four-crystal) power dissipations of up to 30mW have been utilized. State-of-the-art, flicker-of-frequency noise levels characterized by $\mathcal{L}(100Hz) = -156dBc/Hz$ at 40MHz and -143 dBc/Hz at 80MHz and 100MHz have been obtained with noise floor levels (80MHz) as low as $-180dBc/Hz$.

Further, the use of four crystals allows crystal mounting in fixed, relative orientations so that, if the direction and amplitude of each individual crystal's acceleration sensitivity vector is identical, exact cancellation occurs for all directions [2]. In practice, unit-to-unit sensitivity vector amplitude and direction are non-identical. However, even under these circumstances, a four- to five-fold reduction in acceleration sensitivities has been demonstrated.

INTRODUCTION

As indicated in figure 1, the output signal phase noise spectrum of a low-noise, acoustic resonator-stabilized oscillator may be divided into two distinct regions: a near-carrier region and a (white phase noise) floor region. If the resonator itself were perfectly stable, the near-carrier noise level would be characterized by flicker-of-frequency noise resulting from the conversion of oscillator sustaining stage signal (open-loop) flicker-of-phase noise to (closed-loop) flicker-of-frequency noise. As shown by curves 1 and 2 in figure 1, the effect of the conversion is a 20dB/decade increase in near-carrier noise for carrier offset frequencies less than the reciprocal of the closed loop signal (i.e., resonator) group delay. In this regard, the resonator loaded Q may thought of as a measure of its ability to suppress the effects of near-carrier phase noise in the oscillator sustaining stage circuitry [3].

Silicon bipolar transistor (sustaining stage) amplifier flicker-of-phase noise performance has improved to a point where oscillator output-signal, near-carrier noise characteristics are not limited by the sustaining stage circuit, but by short-term frequency instability in the resonator itself (figure 1, curve 3) [4-9]. Curve 4 in figure 1 shows additional, near-carrier spectral degradation that can occur as a result of resonator frequency sensitivity to environmental stress such as vibration [10-12]. In addition, attainable oscillator signal noise floor levels are more related to limitations in resonator maximum drive level than to small variations in sustaining stage amplifier noise figure. This is especially true in the case of conventional, bulk-wave, quartz crystal resonators whose maximum drive level (3-dBm to 8-dBm dissipation) capability is smaller, compared to that of surface-acoustic-wave resonators.

Figure 1 indicates that, for low noise oscillators employing high Q acoustic resonators, limitations in attainable signal spectral performance can be traced to corresponding limitations in resonator performance parameters such as maximum drive level, short-term frequency stability, and vibration sensitivity. Oscillator spectral performance improvement is possible via: (1) improvement in individual resona-

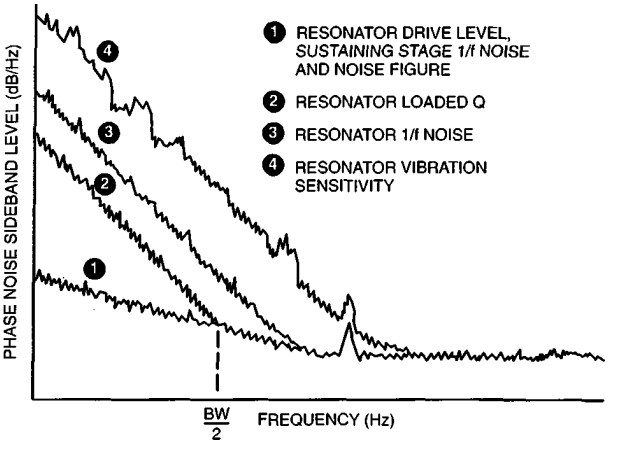


Figure 1. Resonator and Sustaining Stage Parameters Affecting Oscillator Short-Term Frequency Stability

0-7803-0476-4/92 $3.00 © 1992 IEEE

tor characteristics, (2) use of multiple, phase-locked oscillators, or (3) use of multiple resonators in a single oscillator.

USE OF N SERIES-CONNECTED RESONATORS

Figure 2 shows a method for obtaining reduced levels of oscillator noise. As shown in the figure, the in-phase carrier signal outputs of N phase-locked oscillators, all locked to a common reference signal, can be summed using sets of two-way power combiners. At carrier offset frequencies in excess of the narrow loop bandwidths, the phase noise spectra of each oscillator is uncorrelated, and half the noise appears at each combiner internal load. Thus, there is a net reduction of $10\log_N$ in noise level in the final (summed) output. The disadvantages associated with this method are (1) large overall circuit component count and power consumption, and (2) the fact that spectral improvement is not obtained for carrier offset frequencies within the loop bandwidth. A much simpler, more efficient solution consists of use of multiple resonators in a single oscillator circuit.

Figure 3 shows a series connection of N individual quartz crystal resonators, each fabricated to have near-identical impedance characteristics. It is apparent that, in the vicinity of the (desired overtone) series resonant operating frequency, the impedance characteristic of the resonator combination is simply N times that of each resonator. Further, the net series-resonant frequency of the combination is the average of the individual resonator series-resonant frequencies. As shown the figure, anti-resonating individual crystal static capacitance with parallel inductors provides a means of accommodating larger value, unit-to-unit resonant frequency variation. Undesired series resonant frequencies that can occur between non-coincident resonator anti-resonant frequencies can be made desirably lossy via inclusion of relatively large-value, parallel resistors.

Referring to figure 3, if the individual resonator short-term, resonant frequency instability spectra are uncorrelated and denoted by $S_{Y1}(f)$, $S_{Y2}(f)$, etc, then the resonant frequency instability for the four-crystal combination is:

$$S_{YT}(f) = S_{Y1}(f)/N^2 + S_{Y2}(f)/N^2 \ldots\ldots + S_{YN}(f)/N^2 \quad (1)$$

For resonators exhibiting near-equal levels of short-term frequency instability, there is a net $10\log_N$ stability improvement for the multiple N series-connected resonator combination. This improvement is realized in the near-carrier portion of the oscillator output signal spectrum, where resonator flicker-of-frequency noise constitutes the dominant source of instability.

For the case where the instability of one of the resonators is much higher than that of the others, a $20\log_N$ noise level reduction is achieved, compared to that exhibited by the high-noise unit. Therefore, in less severe spectral performance applications, the scheme depicted in figure 3 has an additional advantage of allowing use of otherwise unacceptably high-noise resonators.

With regard to oscillator output signal (white phase) noise floor performance, a $10\log_N$ improvement is also realized. This improvement simply results from being able to drive the multiple-resonator combination at an N-times higher drive level. For a given resonator circuit insertion loss (i.e., resonator loading factor), this means that N-times higher relative sustaining stage amplifier drive level can be utilized. For a fixed, sustaining stage amplifier noise figure, then, the signal-to-noise (floor) ratio also improves by a factor of N.

Certain sources of individual resonator frequency instability, such as those resulting from environmental stress, are somewhat correlated. Individual resonator frequency change due to acceleration can be characterized by a vibration sensitivity vector, Γ [10-14]. If one considers an ideal (albeit unrealistic) case of identical amplitude and direction resonator Γ vectors, four series-connected resonators, may be oriented with respect to each other such that exact Γ vector cancellation occurs. Unlike the case of a single pair of resonators that may be oriented for anti-parallel Γ vector cancellation [13], cancellation can be obtained in the four-resonator case regardless (without measurement) of the vector direction. Figure 4 shows one example of such an orientation, with each resonator mounted on the four faces of a cube-type resonator assembly. Similar, in-plane

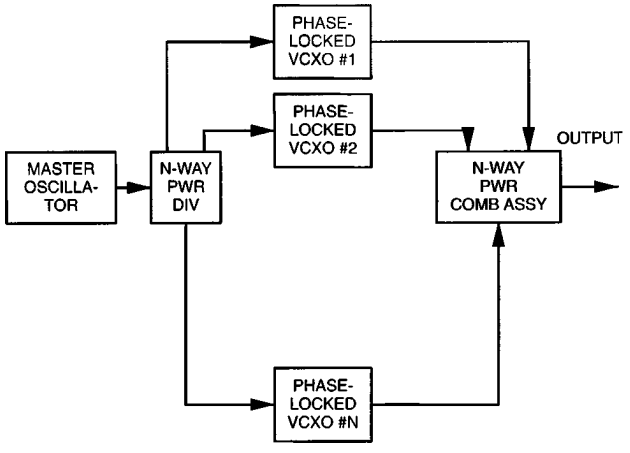


Figure 2. Use of Multiple, Phaselocked Oscillators for Short-Term Frequency Stability Improvement (In Region $f_m \geq$ PLL BW)

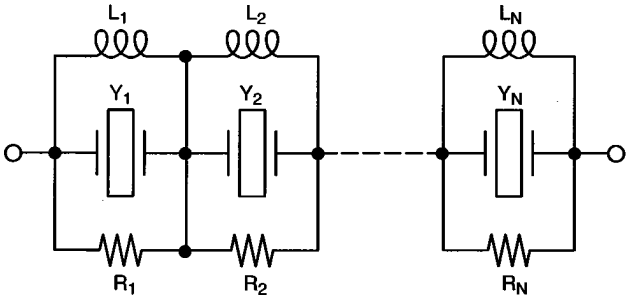


Figure 3. Use of Multiple, Series-Connected, Quartz Crystal Resonators for Reduction of (Net) Resonator Short-Term Frequency Instability

orientations can also be used. Unfortunately, resonator processing and mounting inaccuracies are such that unit-to-unit uniformity in Γ-vector amplitude and direction are not achieved. A more realistic situation is one where Γ-value amplitude variations of 2:1 are encountered, and Γ-vector direction variations are typically such that, at best, unit-to-unit direction variations are confined to a single spherical quadrant. In spite of these variations, typical reductions in (four-resonator combination) vibration sensitivity by factors of four to five have been achieved using the four-resonator configuration.

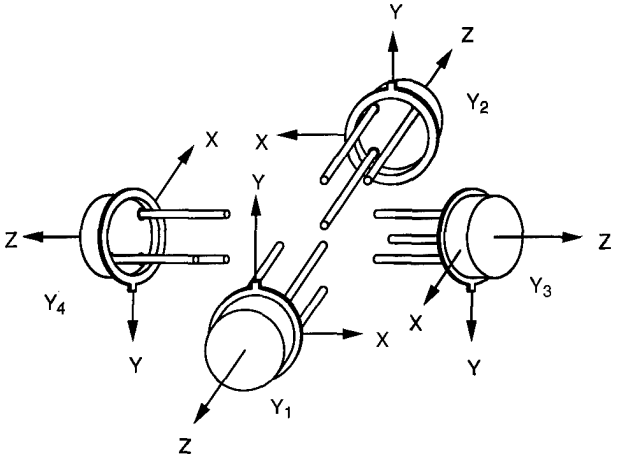


Figure 4. An Example of a Four-Crystal Resonator Mounting Orientation for Vibration Sensitivity Reduction

MULTIPLE CRYSTAL, PROTOTYPE OSCILLATOR DESIGN

The design used for the multiple crystal oscillator is nearly identical to that described in an earlier paper by the author [8]. As shown in figure 5, the sustaining stage incorporates: (1) a low flicker-of-phase noise, 14-dB gain, modular amplifier, (2) a phase shift circuit used to set $2N\pi$ closed-loop phase shift at the operating frequency, (3) a tuned circuit providing requisite frequency selectivity in order to prevent oscillation in undesired crystal resonator overtone modes, and (4) a schottky diode limiter providing symmetrical waveform clipping and assuring oscillator steady-state operation just below amplifier compression.

The crystal resonators are connected to the sustaining stage via a quarter-wavelength transmission line that inverts the four-crystal, series-resonant impedance characteristic to a parallel-resonant characteristic whose resonant parallel resistance is on the order of 20-22 ohms. All or a portion of the transmission line may be synthesized in lumped element form. For vibration measurements, however, a portion of the line is composed of coaxial cable, and this allows individual placement of the resonator and/or sustaining stage assemblies on the shake table for individual measurement of the vibration sensitivity associated with each assembly. As was described in an earlier paper by the author, this configuration also allows separate determination of resonator and coaxial cable vibration effects [14]. A photograph of one of the test oscillators is shown in figure 6.

MEASUREMENT RESULTS: NOISE REDUCTION

Figures 7 through 10 show phase noise measurement results for oscillators utilizing four series-connected resona-

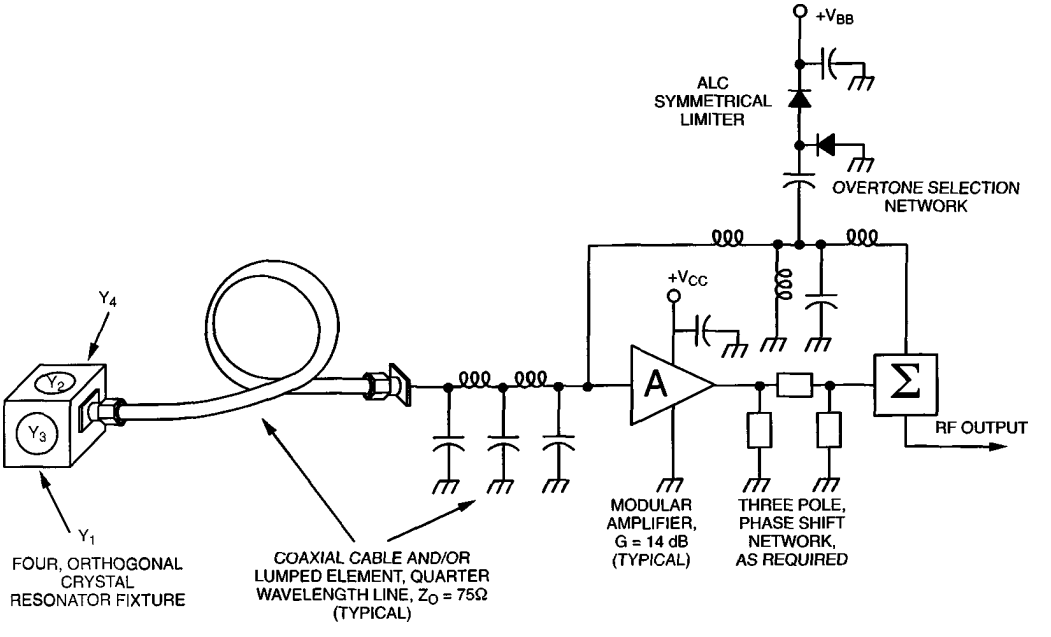


Figure 5. Simplified, Four-Crystal Oscillator Schematic Diagram

tors: (1) 40-MHz, 5th-overtone, AT-cut; (2) 80-MHz, 3rd-overtone, SC-cut; (3) 100-MHz, 3rd overtone, SC-cut; and (4) two series-connected, 100-MHz, 5th-overtone, BT-cut crystal resonators. The data shown in the figures are entirely for cases where all of the resonators (fabricated by Piezo Crystal Co. and Croven Crystal, LTD) were selected for lowest flicker-of-frequency noise level. The reduction in near-carrier noise level was uniformly achieved as predicted in equation (1), compared to single crystal oscillator performance. Because 100-MHz, BT-cut crystal series resistance values were twice that of the SC-cut crystals, the BT-cut crystal oscillator was evaluated using two (rather than four) crystals in order to use a common, 100-MHz sustaining stage design without modification. In the 80-MHz oscillator, a high-level sustaining stage amplifier was used to allow 9dBm-per-resonator (15dBm total) dissipation with 12-dBm amplifier input drive level. The results, shown in figure 8, when referred to the oscillator frequency prior to times-four frequency multiplication, and taking into consideration the noise floor contributions of the cascaded, high-level doublers and doubler interstage amplifiers, correspond to achievement of a per-oscillator noise floor level of −180dBc/Hz. The noise peaking in figure 8 at 6-MHz offset frequency is a result of doubler interstage bandpass filter passband ripple response.

MEASUREMENT RESULTS: VIBRATION SENSITIVITY REDUCTION

Vibration measurements were made for individual sets of 40-MHz, 5th-overtone, AT-cut and 3rd-overtone, SC-cut crystals. The results are shown in tables 1 and 2. As indicated in the table, unit-to-unit variations in AT-cut resonator Γ values were more uniform but larger in amplitude, compared to the SC-cut units.

The results of the measurements show that a typical four-fold reduction in vibration sensitivity is obtained using the four-crystal mounting orientation shown in figure 4. As indicated in table 2, unit-to-unit Γ-direction variations for the SC-cut resonators were confined to a single spherical quadrant, with the exception of resonator #05.

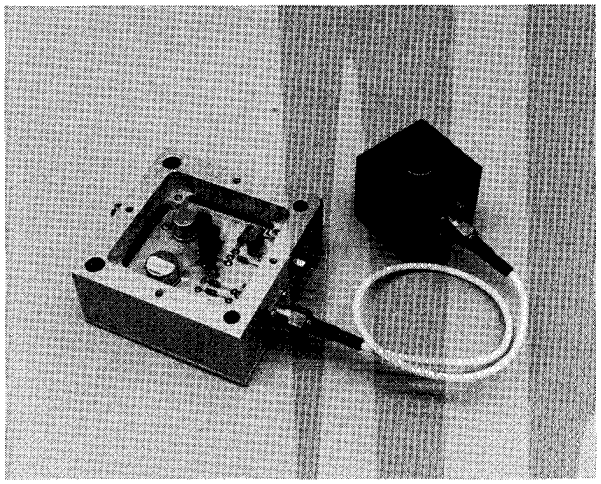


Figure 6. Photograph of the Four-Crystal Oscillator

Table 1. Vibration Test Results: 40-MHz, 5th-Overtone, AT-Cut Crystals (Γ Values in Parts in 10^{-10})

Crystal No.	ΓX	ΓY	ΓZ	ΓT
25	20.4	2.0	6.0	21.4
27	21.9	4.1	6.0	23.1
28	22.3	7.0	0.1	23.4
29	18.8	1.2	2.4	19.0
	ΓX′	ΓY′	ΓZ′	ΓT
Series Combination 25+27+28+29	3.0	3.3	1.7	4.8

Table 2. Vibration Test Results: 40-MHz, 3rd-Overtone, SC-Cut Crystals (Γ Values in Parts in 10^{-10})

Crystal No.	ΓX	ΓY	ΓZ	ΓT
01	0.6	7.6	0.9	7.7
02	0.5	3.8	1.4	4.0
05	−0.8	6.1	−3.2	7.0
15	0.3	5.7	3.3	6.6
	ΓX′	ΓY′	ΓZ′	ΓT
Series Combination 01+02+05+15	1.5	0.9	0.7	1.9

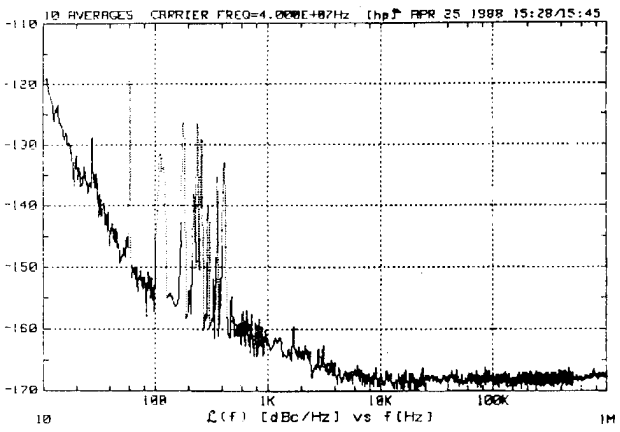


Figure 7. Measured Phase Noise Spectra for Two Phaselocked, 40-MHz (5th-Overtone, AT-Cut), Four-Crystal Oscillators

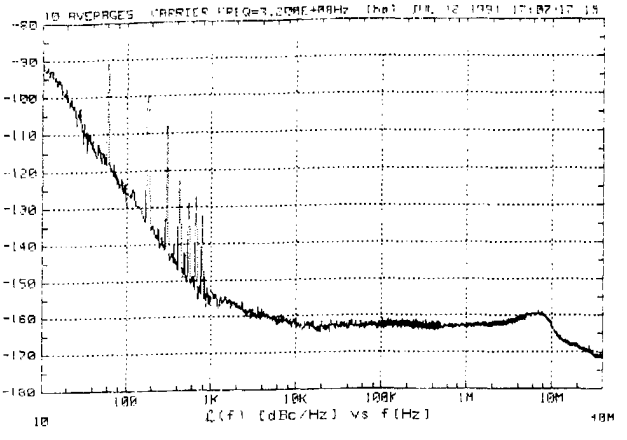


Figure 8. Measured Phase Noise Spectra for Two Phaselocked, 80-MHz (3rd Overtone, SC-Cut), Four-Crystal Oscillators After Frequency Multiplication to 320MHz

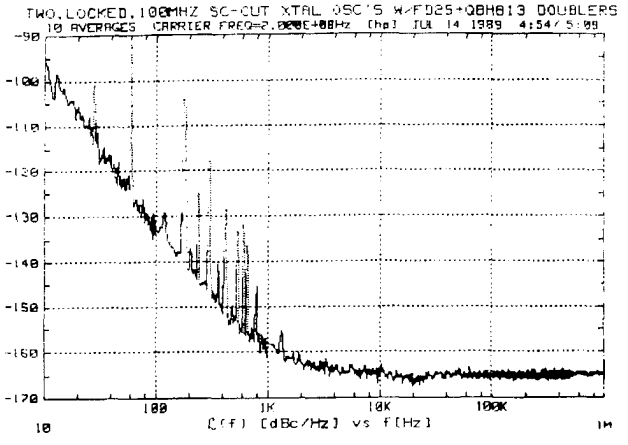


Figure 9. Measured Phase Noise Spectra for Two Phaselocked, 100-MHz (3rd-Overtone, SC-Cut), Four-Crystal Oscillators After Frequency Multiplication to 200MHz

CONCLUSIONS

Utilization of multiple series-connected resonators in a quartz crystal-controlled oscillator can provide significant reduction in near-carrier, flicker-of-frequency noise and (white phase) noise floor levels.

Using this technique, other non-correlated portions of resonator frequency instability, including random-walk and long-term drift, may be reduced as well.

At least four resonators may be oriented with respect to one another in a repeatable manner in order to obtain reduction in (net resonator) vibration sensitivity.

The technique described here applies to oscillators incorporating other types of frequency control elements (such as surface acoustic wave resonators) as well.

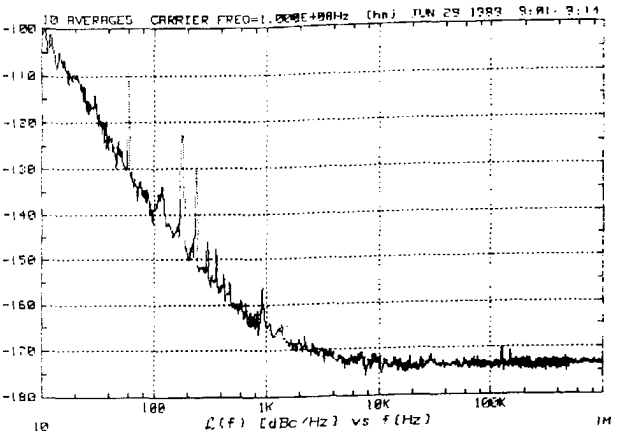


Figure 10. Measured Phase Noise Spectra for Two Phaselocked, 100-MHz (5th-Overtone, BT-Cut), Two-Crystal Oscillators

REFERENCES

[1] U. S. Patent No. 4,851,790.

[2] U. S. Patent Application submitted 6/13/91.

[3] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proc. IEEE*, Vol. 54, No. 2, Feb. 1966, pp. 329–330.

[4] F. W. Walls and A. E. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design," *IEEE Trans. Instrum. & Meas.*, Vol. IM-24, No. 1, Mar. 1975, pp. 15-20.

[5] J. J. Gagnepain, "Fundamental Noise Studies of Quartz Resonators," *Proc. 30th Freq. Contr. Symp.*, June, 1976, pp. 84-87.

[6] T. E. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators," *Proc. 39th Freq. Contr. Symp.*, May 1985, pp. 97-106.

[7] M. M. Driscoll and B. W. Kramer, "Spectral Degradation in VHF Crystal-Controlled Oscillators Due to Short-Term Instability in the Quartz Resonator," *Proc. 1984 Ultras. Symp.*, Oct. 1985, pp. 340-344.

[8] M. M. Driscoll, "Low Noise Crystal Oscillators Using 50-Ohm, Modular Amplifier, Sustaining Stages," *Proc. 40th Freq. Contr. Symp.*, May 1986, pp. 329-335.

[9] G. K. Montress, et. al., "Extremely Low Phase Noise SAW Resonator Oscillator Design and Performance," *Proc. 1987 IEEE Ultras. Symp.*, Oct. 1987, pp. 47-52.

[10] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," *Proc. 41st Freq. Contr. Symp.*, May 1987, pp. 398-408.

[11] J. R. Vig, "Quartz Crystal Resonators and Oscillators: A Tutorial," *U. S. Army Electronics Technology and Devices Laboratory Report SLCET-TR-88-1 (Rev. 4.2)*, March 1991, Section 4.

[12.] J. R. Vig, et. al., "The Effects of Acceleration on Precision Frequency Sources (proposed for IEEE Standards Project P1193)," *U. S. Army Electronics Technology and Devices Laboratory Report SLCET-TR-91-3 (Rev. 1)*.

[13] R. L. Filler, "Acceleration-Resistant Crystal Resonator," U. S. Patent 4,410,822, Oct. 1983.

[14] M. M. Driscoll, "Quartz Crystal Resonator G Sensitivity Measurement Methods and Recent Results," *IEEE Trans. UFFC*, Vol. 37, No. 5, Sept. 1990, pp. 386-392.

SPECTRAL PURITY OF ACOUSTIC RESONATOR OSCILLATORS

Thomas E. Parker and Gary K. Montress

Raytheon Company
Research Division
131 Spring St.
Lexington, MA 02173 USA

Abstract

This paper surveys the spectral purity of state-of-the-art oscillators that use acoustic resonators in the frequency range of 5 to 500 MHz, with the primary emphasis placed on oscillators that use Bulk Acoustic Wave and Surface Acoustic Wave devices on single crystal quartz. The phase noise characteristics of low noise oscillators are analyzed in the context of Leeson's model of feedback oscillator phase noise and particular attention is given to flicker frequency noise levels and white phase noise floors. Procedures and methods for reducing phase noise levels are discussed along with an analysis of reported attempts to obtain improved spectral purity.

Introduction

Oscillators using Bulk Acoustic Wave (BAW) resonators fabricated on single crystal quartz have been a standard for high stability frequency sources in the 1 to 100 MHz range for many years [1]. More recently, Surface Acoustic Wave (SAW) quartz resonators have been used to extend the realm of high performance oscillators employing acoustic resonators to frequencies approaching 1 GHz [2]. The performance of state-of-the-art laboratory oscillators using BAW or SAW resonators is reviewed and their spectral purity is analyzed in the context of Leeson's model of oscillator phase noise [3],[4]. A particular emphasis is placed on the flicker frequency and white phase noise regions of the noise spectrum. From this analysis one is able to identify areas where possible improvements can be made and how this would affect oscillator spectral purity. The performance levels achieved by several attempts to obtain enhanced spectral purity through techniques to increase Q are also evaluated.

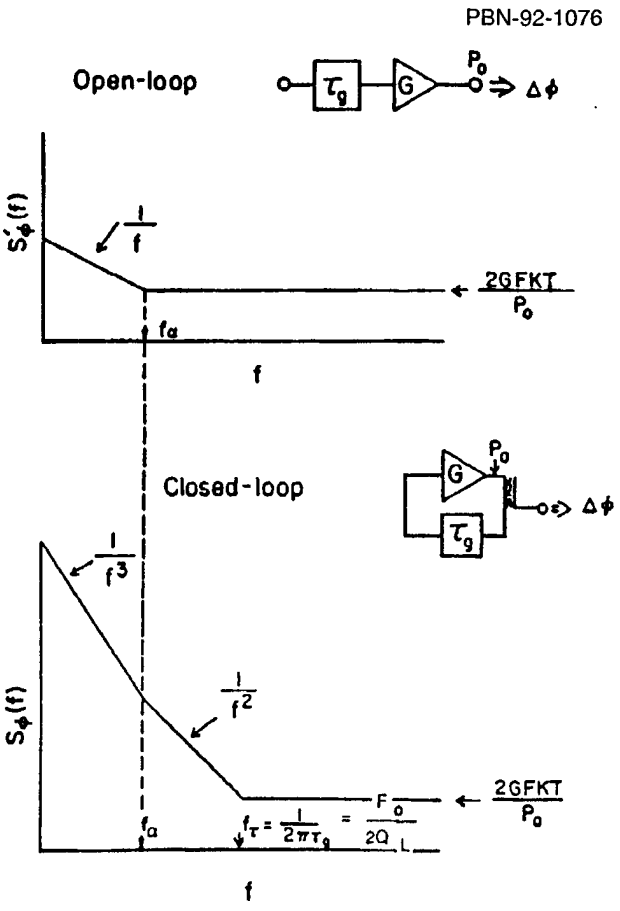


Figure 1. Leeson's model of oscillator phase noise.

Oscillator Phase Noise Basics

Figure 1 illustrates Leeson's model of phase noise for both the open- and closed-loop configurations of a simplified oscillator circuit. $S_\phi(f)$ is the spectral density of phase noise at the output of an oscillator, while $S'_\phi(f)$ represents the open-loop (or residual) phase noise

0-7803-0476-4/92 $3.00 © 1992 IEEE

spectral density. This model has been discussed many times in the literature but it is still worthwhile pointing out a few highlights. The white phase noise floor in Fig. 1 is equal to $2GFKT/P_0$, where G is the compressed gain of the loop amplifier, F is the noise factor of the loop amplifier, K is Boltzmann's constant, T is the temperature in °K, and P_0 is the carrier power level (in watts) at the output of the amplifier. Note that the white phase noise floor is the same in both configurations and is caused by thermal noise that is present whether the carrier is there or not. The flicker noise region occurs at noise frequencies below f_α, where $S'_\phi(f)$ increases as 1/f. Below the noise frequency $f_\tau = F_0/(2Q_L)$ the closed-loop phase noise, $S_\phi(f)$, begins to increase as $1/f^2$, and then transitions to $1/f^3$ below f_α. F_0 is the carrier frequency in Hz and Q_L is the loaded Q of the resonator. Equation 1 shows the relationship between open- and closed-loop phase noise for noise frequencies below f_τ.

$$S_\phi(f) = \left[\frac{F_0}{2Q_L}\right]^2 \frac{S'_\phi(f)}{f^2} \qquad (1)$$

The region where $S'_\phi(f)$ increases as 1/f is fundamentally different from the white noise floor because flicker noise is caused by direct phase or frequency modulation in either the amplifier or the resonator or both. A significant consequence of this fact is that 1/f noise is not directly dependent on P_0 [4]. Furthermore, the 1/f noise power contribution from each component adds to the overall oscillator noise in direct proportion to its level, unlike the amplifier noise figure which is dominated by the first amplifier stage. There is also an important distinction between the resonator and the amplifier (or other electronic components in the oscillator loop) as a source of flicker noise. The $1/f^3$ noise in the oscillator that is caused by the resonator has no Q_L dependence (it depends only on the unloaded Q, Q_u, or F_0). This occurs because the open-loop phase noise observed in a resonator, $S'_{\phi R}(f)$, is proportional to Q_L^2, when Q_L is changed by impedance matching circuit elements [4]. This characteristic, when applied in Eq. 1, results in the cancellation of the Q_L dependence for $S_\phi(f)$. However, this is not the case with the residual noise of the amplifier, $S'_{\phi E}(f)$, which has no direct dependence on the loaded Q of the resonator. Using $S'_{\phi E}(f)$ in Eq. 1 shows that the effect of the residual 1/f noise of the amplifier on the oscillator noise is reduced if the resonator loaded Q is increased.

It is not recommended that f_α be used as a measure of the flicker noise level of an oscillator or component because this corner frequency depends on both the flicker of phase noise level and the white phase noise level. Since they are independent processes, one can be changed without affecting the other. Thus a low corner frequency could be obtained simply by increasing the white phase noise level. A more useful measure of the flicker noise is the value of $S'_\phi(f)$ or $S_\phi(f)$ at a particular noise frequency, such as 1 Hz.

The approximate phase noise of an oscillator, based on the Leeson model, can be expressed in equation form as shown below [4].

$$\begin{aligned} s_\phi(f) = &\left[\alpha_R F_o^4 + \alpha_E (F_o/(2Q_L))^2\right]/f^3 \qquad (2)\\ &+\left[(2GFKT/P_o)(F_o/(2Q_L))^2\right]/f^2\\ &+(2\alpha_R Q_L F_o^3)/f^2\\ &+\alpha_E/f + 2GFKT/P_o \end{aligned}$$

The flicker noise contributions from the resonator and amplifier are expressed as constants α_R and α_E, which are defined by Eqs. 3 and 4.

$$\alpha_R = \left[\frac{1}{(2Q_L F_o)^2}\right] f S'_{\phi R}(f) \qquad (3)$$

$$\alpha_E = f S'_{\phi E}(f) \qquad (4)$$

The f^{-3} section in Eq. 2 is made up of contributions from the resonator and the amplifier and reflects the different way in which each affects the oscillator noise level. The f^{-2} section has two terms, one from the white phase noise, $2GFKT/P_0$, at frequencies less than f_τ, and one from the flicker noise of the resonator outside the 3 dB bandwidth of the resonator [5]. The f^{-1} section comes solely from the flicker noise of the amplifier and the noise floor is simply the white phase noise to carrier ratio as discussed earlier.

The F_0^4 dependence of the resonator flicker noise in the f^{-3} term in Eq. 2 simply reflects an experimental observation that flicker noise in low noise oscillators increases roughly as the fourth power of the resonator frequency and does not imply any fundamental physical basis for this dependence. Since the $Q_u F_0$ product is approximately constant for high performance BAW and SAW resonators on quartz, the resonator flicker noise term in Eq. 2 could alternatively be expressed in terms of the unloaded Q, Q_u. It is very likely that the form of the flicker noise terms in Eq. 2 for both the resonator and amplifier will change as greater understanding of the causes of flicker noise is achieved [7].

The Leeson model in Eq. 2 works very well for feedback loop oscillators using 50 Ω modular amplifiers, but is more difficult to apply to some oscillators using

custom amplifiers, or to negative resistance oscillators. A limitation of Leeson's model is that it does not handle properly the fact that the white phase noise at the amplifier output may be dependent on the resonator impedance, which changes significantly with the frequency offset from resonance. Some oscillators are designed to take advantage of this characteristic to achieve lower noise floors [6], but the modular amplifiers that the authors most commonly use show only a modest change in white noise level at the amplifier output of only about +/-1 dB over a wide range of input impedance values. A second limitation of the Leeson model is that it is not easily applied to negative resistance oscillators since many of the necessary parameters, such as loaded Q, noise factor or amplifier gain, are not readily quantifiable. Never-the-less, it is a model that is conceptually easy to understand and can be very useful in identifying sources and characteristics of oscillator noise. In this regard it can be very useful in identifying areas where improved oscillator performance may be achievable.

Present State-of-the-Art Oscillator Phase Noise Levels

The 5 MHz SC-cut BAW resonator has become a standard for high stability acoustic resonator oscillators. It offers excellent long-term frequency stability, outstanding temperature stability, and very low acceleration sensitivity. Also, significant improvements in flicker noise levels have been achieved over the years. Bloch, et al. [8] reported in 1989 a flicker noise level of $\mathcal{L}(f)$ = -134 dBc/Hz at 1 Hz, and a noise floor of -157 dBc/Hz on an oscillator using a 5th overtone 5 MHz resonator as shown in Fig. 2. (Note that $\mathcal{L}(f) = 10\text{Log}[S_\phi(f)/2]$.) In 1991, Norton [9] reported similar flicker noise levels and a white phase noise floor of -160 dBc/Hz on oscillators using 5 MHz, 3rd overtone, BVA resonators. Norton also measured a time domain stability on the order of $6\text{x}10^{-14}$ for time periods in the 10 to 100 second range. This data is shown in Fig. 3. 5 MHz oscillators with flicker noise levels of -130 dBc/Hz at 1 Hz have also been measured with a noise floor of -179

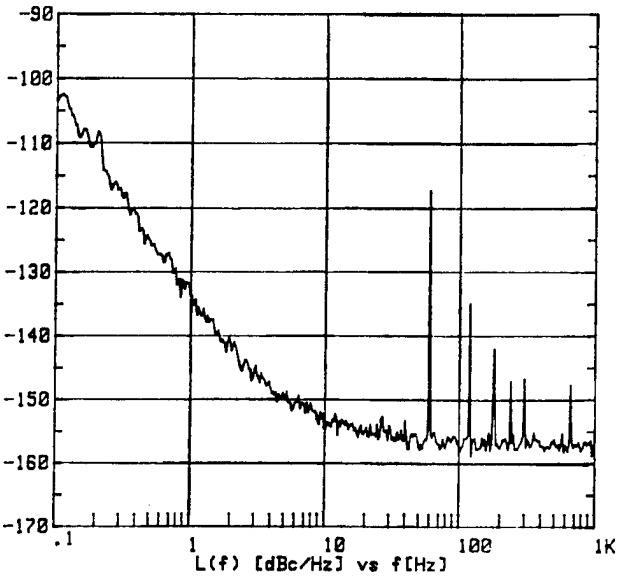


Figure 2. Phase noise of a 5 MHz oscillator, from Ref. 8.

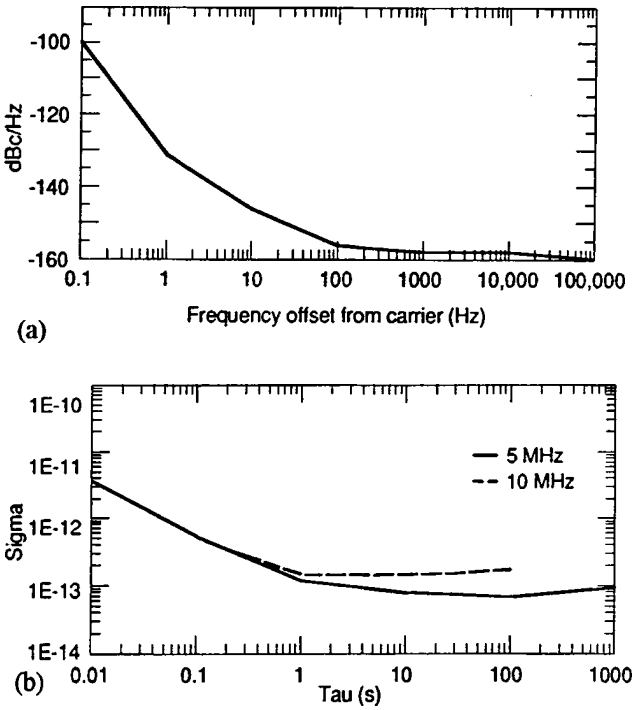


Figure 3. Phase noise (a) and Allan deviations (b) of a pair of 5 MHz BVA resonator oscillators, from Ref. 9.

dBc/Hz [7]. At higher resonator frequencies, state-of-the-art performance has been achieved on 80 and 100 MHz BAW resonator oscillators. Driscoll [10] has reported on 80 MHz oscillators with a flicker noise level of -72 dBc/Hz extrapolated to 1 Hz and a floor of -173 dBc/Hz, as shown in Fig 4. Peier and Jenni [11] have measured 1 Hz flicker noise levels of -76 dBc/Hz on oscillators using 100 MHz BVA resonators but with floors of only -145 dBc/Hz. Recently, Driscoll has reported noise floors of -180 dBc/Hz on 80 MHz oscillators [12]. At even higher frequencies, the best

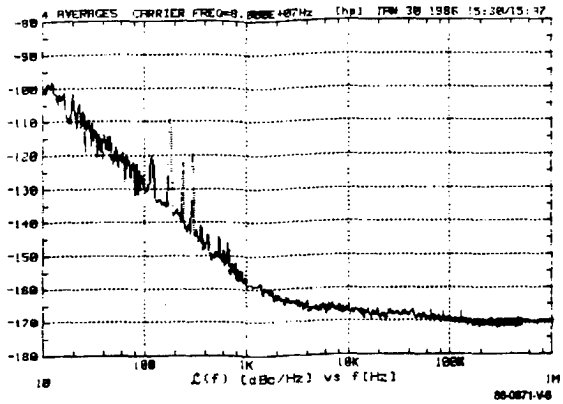


Figure 4. Phase noise of a pair of 80 MHz oscillators, from Ref. 10.

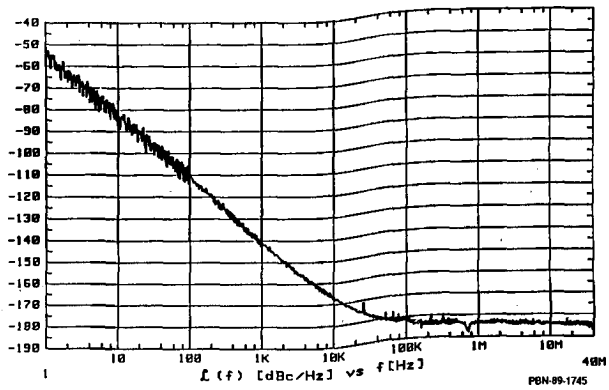


Figure 5. Low noise 500 MHz SAW resonator oscillator.

Table 1.

State-of-the-Art Phase Noise Characteristics

Frequency MHz	$Q_L=Q_m/2$	$\mathcal{L}(1)$ dBc/Hz	$\mathcal{L}$(floor) dBc/Hz	σ_y(~1 sec.)	$\mathcal{L}'(1)$ dBc/Hz
5	1.5×10^{6}	-134	-179	8×10^{-14}	-138
100	7.5×10^{4}	-76	-180	3×10^{-12}	-133
500	1.0×10^{4}	-55	-183	8×10^{-12}	-143

results have been obtained on SAW resonator oscillators. The authors have reported 500 MHz SAW resonator oscillators [13] with a 1 Hz flicker noise level of -55 dBc/Hz and a white phase noise floor of -183 dBc/Hz, as shown in Fig. 5.

Table 1 summarizes the present state-of-the-art in spectral purity for acoustic resonators in the frequency range of 5 MHz to 500 MHz. $\mathcal{L}(1)$ represents the best flicker noise level of the oscillators and may have been directly measured at 1 Hz or extrapolated to 1 Hz from a higher noise frequency. $\mathcal{L}$(floor) is the observed white phase noise floor and it may not have been achieved simultaneously on the same oscillator with the tabulated flicker noise level. All of the oscillators had resonators with unloaded Q's near the material limit, Q_m, and the open-loop noise levels, $\mathcal{L}'(1)$, are normalized to the loaded Q values in column 2. In some cases the open-loop noise levels were actually measured and in others the open-loop values were derived from oscillator measurements. However, in all cases the resonator appeared to be the dominant source of the flicker noise. As indicated in Eq. 1, the oscillator flicker noise level increases very rapidly relative to the open-loop flicker noise level as the resonator frequency is increased because of the F_0^2 dependence in Eq. 1 and because the loaded Q will decrease as the frequency is increased. This is clear in Table 1 where there is almost an 80 dB difference between the 5 MHz and 500 MHz oscillator noise levels. However, it should also be noted that the open-loop flicker noise levels vary by only 10 dB over the 5 to 500 MHz range. In fact, the 5 MHz noise level is only 5 dB higher than the 500 MHz level. The approximate F_0^4 dependence in the resonator flicker noise term in Eq. 2 follows from the fact that the open-loop residual flicker noise term does not vary significantly with resonator frequency.

The data in Table 1 indicates that $\mathcal{L}'_R(1) \approx$ -133 to -143 dBc/Hz (for $Q_L=Q_m/2$) for the best laboratory oscillators. This corresponds to $S'_{\phi R}(1) \approx 1$ to 10×10^{-14} radians^2/Hz, and gives an effective resonator flicker noise constant of $\alpha_R \approx 1$ to 5×10^{-40} (radians/Hz)^2. The best 5 MHz BAW resonators and the best 500 MHz SAW

resonators have a flicker noise constant of approximately $1x10^{-40}$, while the constant for 100 MHz resonators is about a factor of 5 larger. It follows from the approximate F_0^4 dependence of $S_\phi(f)$ in Eq. 2 that $\sigma_y(\tau)$ should have an approximately linear dependence on F_0 [4], and this is consistent with the data in Table 1.

From the available data in the literature it also appears that the lowest flicker noise levels for silicon bipolar amplifiers are on the order of $\mathcal{L}'(1)$ = -145 to -150 dBc/Hz. This gives a value for α_E of approximately 2 to $6x10^{-15}$ radians^2.

Improving Oscillator Spectral Purity

To illustrate the effect on spectral purity of changes in oscillator component characteristics, Eq. 2 is used to calculate oscillator phase noise levels for a 100 MHz oscillator with typical state-of-the-art phase noise parameters. Then a comparison is made to calculated oscillator phase noise levels when improved parameters are used. The dashed line in Fig. 6 is a calculated phase noise curve based on the following parameters: $\mathcal{L}'_R(1)$ = -130 dBc/Hz, $\mathcal{L}'_E(1)$ = -145 dBc/Hz, F_0 = 100 MHz, Q_L = 75,000 and $\mathcal{L}'$(floor) = 10Log(GFKT/P_0) = -180 dBc/Hz. The dashed curve in Fig. 6 matches closely the phase noise characteristics observed on laboratory oscillators [10, 12]. Note that the phase noise is dominated by the white noise floor and the flicker frequency region ($1/f^3$), which is typical of low noise oscillators.

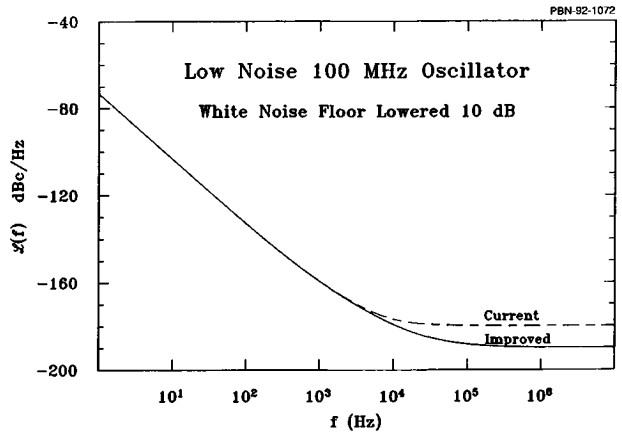


Figure 6. Comparison of the calculated phase noise spectrum of a current 100 MHz oscillator with the spectrum of an oscillator with reduced noise floor.

Lower Noise Floors

The solid curve in Fig. 6 illustrates the phase noise characteristics of an oscillator with the white phase noise level, $\mathcal{L}'$(floor), reduced by 10 dB and with the other parameters unchanged. The oscillator noise floor is, of course, reduced by 10 dB but there is virtually no change below 10 kHz. This occurs because the resonator flicker noise level has no direct dependence on white phase noise [4],[14] and because there is very little $1/f^2$ region.

Of course, the main question is how such a reduction in the white phase noise level can be achieved. Two basic techniques have been used in the past. One is to use resonators that can handle higher power levels, such as SC-cut BAW resonators or the over-moded SAW resonators with Cu doped Al transducers that the authors use [2]. Second, the large change in resonator impedance that occurs off resonance can be used to reduce the white noise at the amplifier output as discussed earlier. This technique has been used on low frequency BAW resonator oscillators, where custom amplifier designs are common and resonator power dissipation levels much above 1 mW are not acceptable. Oscillator noise floors approaching -180 dBc/Hz with 5 MHz BAW resonators have only been achieved by using this technique. Future reductions may be achieved by extending these techniques or by employing new ideas.

One very interesting new technique is the use of multiple resonators [12], and its full potential may not yet have been realized. At frequencies above 100 MHz, Surface Transverse Wave (STW) devices may have some potential for handling higher power levels than SAW devices [15],[16]. Also, the High overtone Bulk Acoustic wave Resonator (HBAR) may offer some advantage at high frequencies due to its highly over-moded structure, which can result in a lower acoustic power density. Through the use of new resonator designs, new resonator materials or cuts, multiple resonators, new transducer materials, and improved circuit designs, it is entirely possible that oscillator noise floors approaching -190 dBc/Hz may be achieved in the next 10 years. However, all techniques for achieving lower noise floors must be carefully evaluated to determine if degradation in other important parameters such as flicker noise or long-term frequency stability may have occurred. A common method that is also used to reduce the white phase noise floor is post filtering, but this technique will not be addressed here since the present discussion is confined to the oscillator loop.

Reduced Flicker Noise

Figure 7 shows the same curve for the current state-of-the-art oscillator phase noise spectral density as shown in Fig. 6, but here the improved curve illustrates a 10 dB reduction in the open-loop flicker noise level. For this curve the flicker noise parameters are, $\mathcal{L}_R(1)$ = -140 dBc/Hz and $\mathcal{L}_E(f)$ = -155 dBc/Hz, and all other parameters are unchanged. Now the situation is reversed from that in Fig. 6 because only the frequencies below 10 kHz are affected. A reduction in both the amplifier and resonator flicker noise levels is required to achieve the full 10 dB improvement, though the improvement in the amplifier need not be 10 dB since the resonator is the major source of flicker noise.

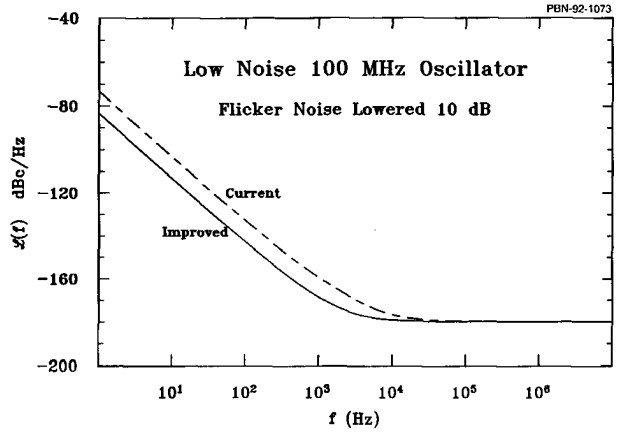


Figure 7. Comparison of the calculated phase noise spectrum of a current 100 MHz oscillator with the spectrum of an oscillator with reduced flicker noise.

Over the past 10 to 20 years a steady improvement in flicker noise levels by 10 to 20 dB has been achieved [17],[18] as oscillator designers have come to understand more about the sources of close-to-carrier noise in oscillators. These improvements have been made in spite of the fact that the basic physics of flicker noise is still poorly understood, and in many cases the details of how these improvements were accomplished have not been published. The authors have found that subjecting a SAW resonator or delay line to a higher than normal incident RF power level of approximately 1 W can bring about a permanent reduction in SAW device flicker noise by as much as 5 to 10 dB [18]. Also careful selection of amplifier designs and vendors, along with screening of the RF transistor wafers has significantly increased the availability of low flicker noise amplifiers [13]. At this time it is still not certain whether there is a fundamental lower limit to 1/f noise in resonators and amplifiers, or if it could be totally eliminated through the eventual removal of all impurities and defects. Obviously, a greater understanding of the basic physical causes of 1/f noise is required before reasonable estimates of ultimate lower limits can be made, but an important step may have been taken with the observation of an apparent acoustic volume dependence in BAW resonators [7]. This volume dependence is consistent with a quantum mechanical origin for 1/f noise. However, it is also clear that there are other, non fundamental, causes of flicker noise since noise levels vary substantially from device to device [4],[7], and can be reduced by improved processing [18].

Even if there may be fundamental lower limits to 1/f noise in any particular device or design, there are techniques that can still be used to further reduce the noise level. One is to use multiple resonators [12] or transistors to take advantage of the uncorrelated nature of the phase fluctuations in individual devices. This technique gives a 10Log(N) reduction in the noise level, where N is the number of independent devices that are operated in series or in parallel, and has been used to produce the lowest observed flicker noise levels on oscillators in the 40 to 100 MHz range [12]. The theory in Ref. 7 indicates that alternative approaches may also be available. In this theory it is proposed that $S_y(f) \propto V/Q_u^4$, which, if proven true, means that 1/f noise can also be reduced by decreasing acoustic volume, V, or by increasing the material Q. If increased Q's, or decreased

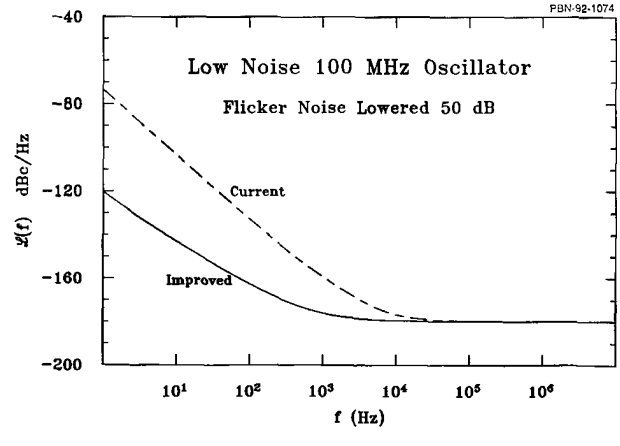


Figure 8. Comparison of the calculated phase noise spectrum of a current 100 MHz oscillator with the spectrum of an oscillator with flicker noise nearly eliminated.

volumes, can be achieved without increasing other sources of flicker noise, then a reduction in α_R, or $S'_{\phi R}(f)$, would be obtained. A more detailed discussion of techniques to increase Q is presented in the next section.

Before beginning the discussion on Q, however, it is of interest to illustrate the effect on spectral purity of a near total elimination of 1/f noise. This is done in Fig. 8 where the affect of a 50 dB reduction in open-loop flicker noise (both resonator and amplifier) is shown and all other parameters are unchanged, including loaded Q. The close-to-carrier noise is now predominantly $1/f^2$, and is as much as 46 dB lower than the noise levels in the reference curve. Clearly, substantial benefit is to be gained if 1/f noise levels can be reduced, even if there is no change in Q.

Increased Material Q

Figure 9 shows a comparison between the current 100 MHz oscillator phase noise spectral density and the calculated performance obtained from a factor of 10 increase in loaded Q, all other parameters, including open-loop flicker noise, being the same as for the current performance curve. As expected, the close-to-carrier noise is reduced by 20 dB.

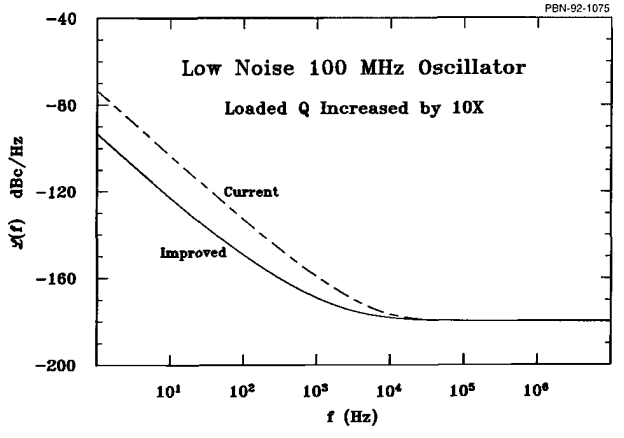


Figure 9. Comparison of the calculated phase noise spectrum of a current 100 MHz oscillator with the spectrum of an oscillator with increased Q.

Over the years there have been a number of attempts to obtain higher material Q's and they have basically fallen into two categories. One is to use cryogenic temperatures to increase the Q of quartz [19]-[22], and the other has been to use high Q acoustic materials other than quartz at room temperature [14],[23],[24]. To obtain significant increases in the Q of quartz resonators, the temperature must be reduced to below 10 K. In recent years the unloaded Q has been observed to be as high as 91×10^6 at 4.2 K and 150×10^6 at 1.5 K on 5 MHz resonators [20]. Though very high Q's have been observed, the frequency stabilities achieved have not yet exceeded the best room temperature performance. A factor of ten reduction in σ_y, (20 dB in $S_\phi(f)$), has been observed when the temperature is reduced from room temperature to 1.5 K, but the values of σ_y obtained at low temperatures were only in the low 10^{-13} range [19],[20]. As discussed earlier, the best room temperature performance is in the upper 10^{-14} range. It still remains to be demonstrated that performance superior to the best room temperature stabilities can be obtained at very low temperatures. A major problem with using low temperatures to increase Q is that environmental sensitivities also play a significant role in frequency stabilities, and adequately controlling these factors at low temperatures is not a simple process.

At higher resonator frequencies the main approach to obtaining increased Q values has been to use materials with higher room temperature Q's than quartz. $LiTaO_3$, a piezoelectric material, has a material Q about 3 time higher than SC or AT quartz, and also has temperature stable BAW cuts. However, the measured flicker noise levels on 400 MHz oscillators using over-moded $LiTaO_3$ resonators [14] were no better than the best quartz SAW oscillator noise levels. The only case that the authors know of where phase noise performance better than the best quartz results have been obtained is with the HBAR structure which uses a non piezoelectric material, such as YAG, and ZnO transducers in a highly over-moded resonator [23],[24]. Figure 10 shows the measured phase noise from Ref. 24 on a pair of approximately 640 MHz HBAR oscillators along with a comparison to the estimated performance of a pair of state-of-the-art 640 MHz SAW oscillators. The unloaded Q's of the HBAR resonators where 160,000, which is about ten times larger than that of 640 MHz SAW resonators, yet the noise improvement was only 10 dB. The reason for this worse than expected performance is that the open-loop flicker noise levels of the HBAR resonators were about 10 dB higher than that of the best SAW resonators. Thus, from Eq. 1, there is only a 10 dB improvement.

The HBAR oscillators are a good example to use for discussing the difficulties in using new materials to obtain higher Qs. One major problem is controlling flicker noise levels. Though there may be fundamental lower limits to flicker noise, it is also very clear that material quality can also strongly influence flicker noise. In the case of the HBAR resonators, it is probably not the

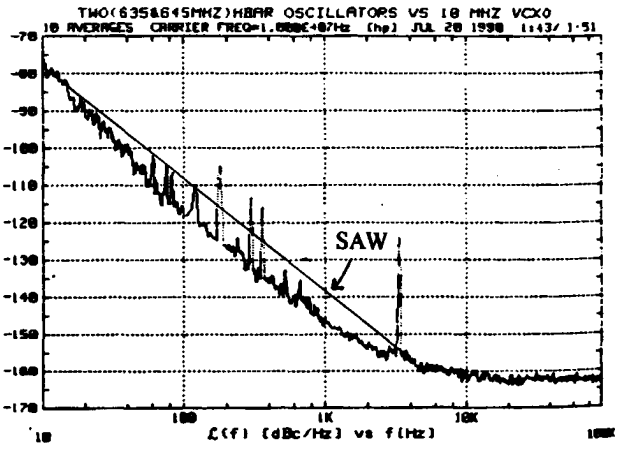


Figure 10. Measured phase noise of a pair of 640 MHz HBAR oscillators (from Ref. 24) compared to state-of-the-art 640 MHz SAW resonator oscillators.

YAG material that is the problem but rather the ZnO transducers. To fully exploit the high Q of the YAG material, the flicker noise level in the transducers will have to be reduced, and this can be a long process. Also, the YAG material has a very large temperature coefficient of 30 PPM/°C. This increases frequency instabilities caused by temperature fluctuations (possibly the cause for the increased phase noise at 10 Hz in Fig. 10) and also increases the difficulty of maintaining an accurate operating frequency.

Other Considerations

Rarely are there applications were spectral purity is the only consideration, and most of the time many other factors, such as long-term stability, frequency settability, temperature stability, power consumption, size, cost, etc., must also be taken into account. In addition, there are the environmental sensitivities such as vibration sensitivity, acoustic sensitivity, and temperature stability, which may directly affect the spectral purity, depending on the environment in which the oscillator is intended to operate. In any program to improve oscillator spectral purity all of these factors must be considered in the overall evaluation of the technology used to make the improvements. In this respect, a reduction in flicker noise levels through process improvements is an attractive approach to obtaining improved spectral purity since it generally has minimal impact on other important oscillator characteristics.

Conclusions

Improvements in oscillator phase noise levels have been steadily occurring and it can be expected that they will continue to occur. Phase noise floors near -180 dBc/Hz have now been demonstrated on quartz resonator oscillators from 5 MHz to 500 MHz, and open-loop resonator flicker noise levels are in the -133 to -143 dBc/Hz range (for $Q_L = Q_m/2$) over the same resonator frequency range. It is reasonable to expect that noise floors will continue to move down slowly and that further reductions in flicker noise will also be achieved through continued improvements in resonator design and processing. However, a better understanding of the basic physical causes of flicker noise is very much needed before a truly systematic attempt to minimize resonator flicker noise can be made. If $\mathcal{L}'_R(1)$ and $\mathcal{L}'_E(1)$ could be reduced to the -180 dBc/Hz level, some very impressive close-to-carrier oscillator phase noise levels could be achieved, even without an increase in Q. Attempts to achieve and exploit higher Q values have been only partially successful, but these efforts will most likely continue. Using new materials or low temperature techniques introduces many new problems and it will take time to solve them. However, these problem must be solved before improved frequency stabilities are achieved, and all other real system requirements can be met.

References

[1] E.A. Gerber and A. Ballato, Editors, Precision Frequency Control. Orlando, FL: Academic Press, 1985.

[2] T.E. Parker and G.K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 3, pp. 342-364, May 1988.

[3] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proc. of the IEEE, vol. 54, no. 2, pp.329-330, February 1966.

[4] T.E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", in *Proc. of the 41st Annual Symp. on Frequency Control*, 1987, pp. 99-110.

[5] V.F. Kroupa, "The State-of-the-Art of the Flicker Frequency Noise in BAW and SAW Quartz Resonators", *IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 35, no. 3, pp. 406-420, May 1988.

[6] M.M. Driscoll, "Low Noise, Microwave Signal Generation Using Bulk and Surface Acoustic Wave Resonators", in *Proc. of the 42nd Annual Symp. on Frequency Control*, 1988, pp. 369-377.

[7] 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",in *Proc. of the IEEE Frequency Control Symp*, 1992, (in this Proc.).

[8] M.B. Bloch, J.C. Ho, C.S. Stone, A. Syed, and F.L. Walls, "Stability of High Quality Quartz Crystal Oscillators: an Update", in *Proc. of the 43rd Annual Symp. on Frequency Control*, 1989, pp. 80-84.

[9] J.R. Norton, "BVA-Type Quartz Oscillator for Spacecraft", in *Proc. of the 45th Annual Symp. on Frequency Control*, 1991, pp. 426-430.

[10] M.M. Driscoll, "Low Noise Crystal Oscillators Using 50-ohm, Modular Amplifier Sustaining Stages", in *Proc. of the 40th Annual Symp. on Frequency Control*, 1986, pp. 329-335.

[11] U.R. Peier and G. Jenni, "Space Qualified 100 MHz BVA-VHF-Resonator", in *Proc. of the 2nd European Frequency and Time Forum*, 1988, pp.701-713.

[12] 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", in *Proc. of the IEEE Frequency Control Symp*, 1992, (in this Proc.).

[13] G.K. Montress and T.E. Parker, "Design Techniques for Achieving State-of-the-Art Oscillator Performance", in *Proc. of the 44th Annual Symp. on Frequency Control*, 1990, pp. 522-535.

[14] M.M. Driscoll, R.A. Jelen, R.W. Weinert, S.V. Krishnaswamy, and B.R. McAvoy, "Low Noise, UHF Oscillators Utilizing High Overtone, Lateral-Field Excitation, Lithium Tantalate Resonators", in *Proc. of the IEEE Ultrasonics Symp.*, 1991, Vol. 1, pp. 453-457

[15] T.L. Bagwell and R.C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q",in *Proc. of the IEEE Ultrasonics Symp.*, 1987, Vol. 1, pp. 319-324.

[16] I.D. Avramov, "Microwave Oscillators Stabilized with Surface Transverse Wave Resonant Devices", in *Proc. of the IEEE Frequency Control Symp.*, 1992, (in this Proc.).

[17] J.J. Gagnepain, "Quartz Crystal Oscillators From Their Design to Their Performances", in *Proc. of the 4th European Frequency and Time Forum*, 1990, pp.121-129.

[18] 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 Trans. on Ultrasonics, Ferroelectrics, & Frequency Control*, vol. 35, no 6, pp. 657-667, November 1988.

[19] J.J. Gagnepain, J. Ubersfeld, G. Goujon, and P. Handel, "Relation Between 1/f Noise and Q-Factor in Quartz Resonators at Room and Low Temperatures", *Proc. of the 35th Annual Symp. on Frequency Control*, 1981, pp. 476-483.

[20] G. Robichon, J. Groslambert, and J.J. Gagnepain, "Frequency Stability of Quartz Crystal at Very Low Temperatures: Preliminary Results", *Proc. of the 38th Annual Symp. on Frequency Control*, 1984, pp. 201-205.

[21] B. Komiyama, "Quartz Crystal Oscillator at Cryogenic Temperature", in *Proc. of the 35th Annual Symp. on Frequency Control*, 1981, pp. 335-339.

[22] A. El Habti and F. Bastien, "Low Temperature Limitation on the Quality Factor of Quartz Resonator", in *Proc. of the IEEE Frequency Control Symp*, 1992, (in this Proc.).

[23] D.S. Bailey, M.M. Driscoll, R.A. Jelen, and B.R. McAvoy, "Frequency Stability of High Overtone Bulk Acoustic Resonators", in *Proc. of the IEEE Ultrasonics Symp.*, 1990, Vol. 1, pp. 509-512.

[24] M.M. Driscoll, R.A. Jelen, and N. Matthews, "Extremely Low Phase Noise UHF Oscillators Utilizing High-Overtone, Bulk Acoustic Resonators", in *Proc. of the IEEE Ultrasonics Symp.*, 1990, Vol. 1, pp. 513-518.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

MICROWAVE OSCILLATORS FOR SUPERIOR SHORT TERM STABILITY AND ULTRA-LOW PHASE NOISE*

G. J. Dick
California Institute of Technology, Jet Propulsion Laboratory
4800 Oak Grove Drive, Bldg 298
Pasadena, California 91109

Abstract

A survey is presented of the wide variety of available microwave oscillator capabilities, with emphasis on their phase noise and short term frequency stability. These oscillators show unique capabilities for frequency generation, making possible the lowest phase noise floors and short term stabilities of any technology. They range in complexity from a free-running transistor oscillator with an ultra-low noise floor, to cryogenic oscillators with higher frequency stability than any other source for measuring times up to $\approx$ 100 seconds. Their advantages derive primarily from high operating frequency and from recent advances in resonator and circuit technology.

Introduction

Microwave oscillators show unique capabilities for frequency generation, making possible the lowest phase noise floors and short term stabilities of any technology. Because phase fluctuations multiply and divide as operating frequencies are similarly transformed, a high operating frequency gives microwave oscillators an inherent advantage with respect to phase noise floor. The noise floor for any given oscillator is determined by its output power together with the noise figure of the follower amplifier, and does not vary strongly with operating frequency. Thus, for any required frequency of use, a microwave oscillator has a relatively lower noise floor, being multiplied less (or divided more) compared to an oscillator operating at a lower frequency.

*This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

While obtaining an inherent advantage with respect to noise floor, microwave oscillator performance is typically surpassed by that of lower-frequency oscillators for close-in noise ($f \leq$ 10kHz), due to lower resonator Q's and poor semiconductor noise performance at the higher frequencies. This state of affairs is now changing, due to recent advances in resonator capability and in microwave circuitry.

A remarkable variety of resonators have been developed which make possible high Q's at both room and cryogenic temperatures, and, at the lowest temperatures, show a "freezing out" of thermal expansion for ultra-high frequency stability. They include copper cavities resonators at room temperature[1], room temperature and cooled high-epsilon DRO's[2], low temperature superconducting (LTS) resonators at liquid helium temperatures[3], high temperature superconducting (HTS) resonators[4,5], sapphire resonators using both conventional[6] and whispering gallery[7,8] modes, and sapphire/LTS resonators using conventional[9] and whispering gallery[10] modes. While most of the newly developed technologies are associated with research laboratories, at least two [7,10] are offered commercially.

Close-in phase noise for microwave oscillators typically shows a $S_\phi(f) \propto 1/f^3$ dependence on the offset frequency, where $S_\phi(f)$ is the power spectral density of phase fluctuations. This dependence is due to flicker (1/f) fluctuations of phase in the active device itself, as transformed by the resonant circuitry, and is reduced by increasing resonator Q. With improvements in both Q and flicker noise, practical microwave oscillators now promise to surpass Quartz crystal oscillators in performance for all offset frequencies greater than 1 Hz.

0-7803-0476-4/92 $3.00 © 1992 IEEE

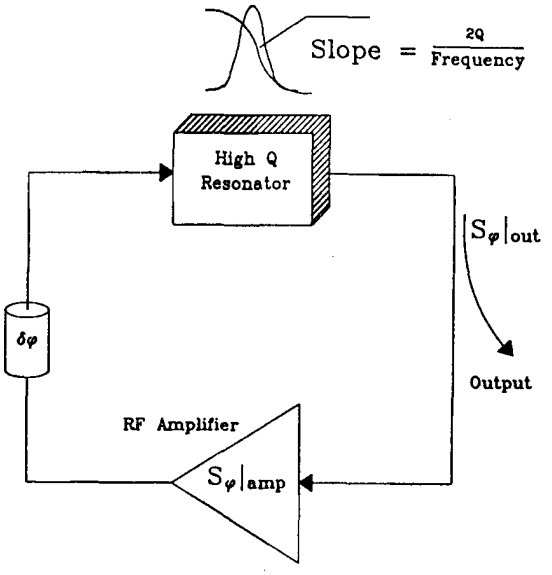


Figure 1: Block diagram of simple oscillator with direct RF feedback. Output phase noise is derived from amplifier noise together with phase slope of resonator. Phase is adjusted to give $2 \cdot n\pi$ radians around the loop at the center of the passband.

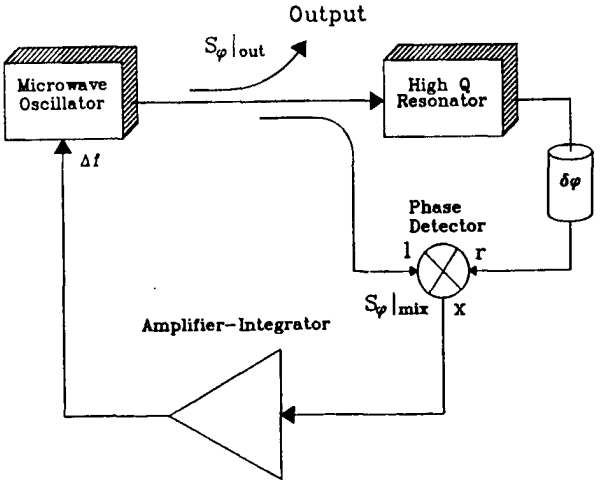


Figure 2: Block diagram of stabilized local oscillator (STALO) with double-balanced mixer phase detector. Mixer noise plays the same role as amplifier noise in Figure 1. Phase is adjusted to give l and r signals in quadrature.

Even the short-term frequency stability of quartz is surpassed by cryogenic microwave oscillators at helium temperature, which show a capability that is expected to become more practical as the technology advances. Their frequency is in some sense undetermined, depending on the actual (and accidental) physical dimensions of the resonator. However a freezing out of the mechanical expansion coefficient, as required by the third law of thermodynamics, makes possible the generation of frequencies with ultra-high stability for resonators at temperatures below about 10 Kelvins. These low temperature sources provide the highest stability available today from any RF or microwave source for measuring times less than about 100 seconds.

Background

While the noise sources in, e.g., quartz oscillators are primarily due to processes within the resonator itself, the source of noise performance in microwave oscillators is generally due to flicker fluctuations in an associated active semiconducting device. This can take the form of a transistor which is used to excite oscillations in the resonator in an active oscillator; or it may be a phase comparator (mixer) which is used together with the resonator form a very sensitive and low noise frequency discriminator. In either case these semiconducting fluctuations give rise to the characteristic flicker frequency noise associated with oscillator performance for small offset frequencies.

In the self-excited oscillator shown in Figure 1, the oscillation condition requires that the phase shift around the complete feedback loop comprising the amplifier, resonator, and interconnections be a multiple of 2π. With this condition satisfied, any phase fluctuation in the microwave amplifier must be accompanied by an opposite shift of equal magnitude in the resonator. For slow phase fluctuations ($f \ll \nu/Q$), the characteristic phase slope of the resonator $\partial\phi/\partial\nu = 2Q/\nu$ implies a corresponding slow fluctuation in the frequency of the oscillator. Here f represents the fluctuation frequency, ν the microwave frequency, Q the quality factor of the resonator and ϕ the phase of the microwave signal. In this way, a power spectral density of phase fluctuations for the amplifier $S_\phi(f)|_{amp}$ results in oscillator output frequency noise

$$S_y(f)|_{osc} = \frac{S_\phi(f)|_{amp}}{(2Q)^2} \qquad (1)$$

or the mathematically equivalent output phase fluc-

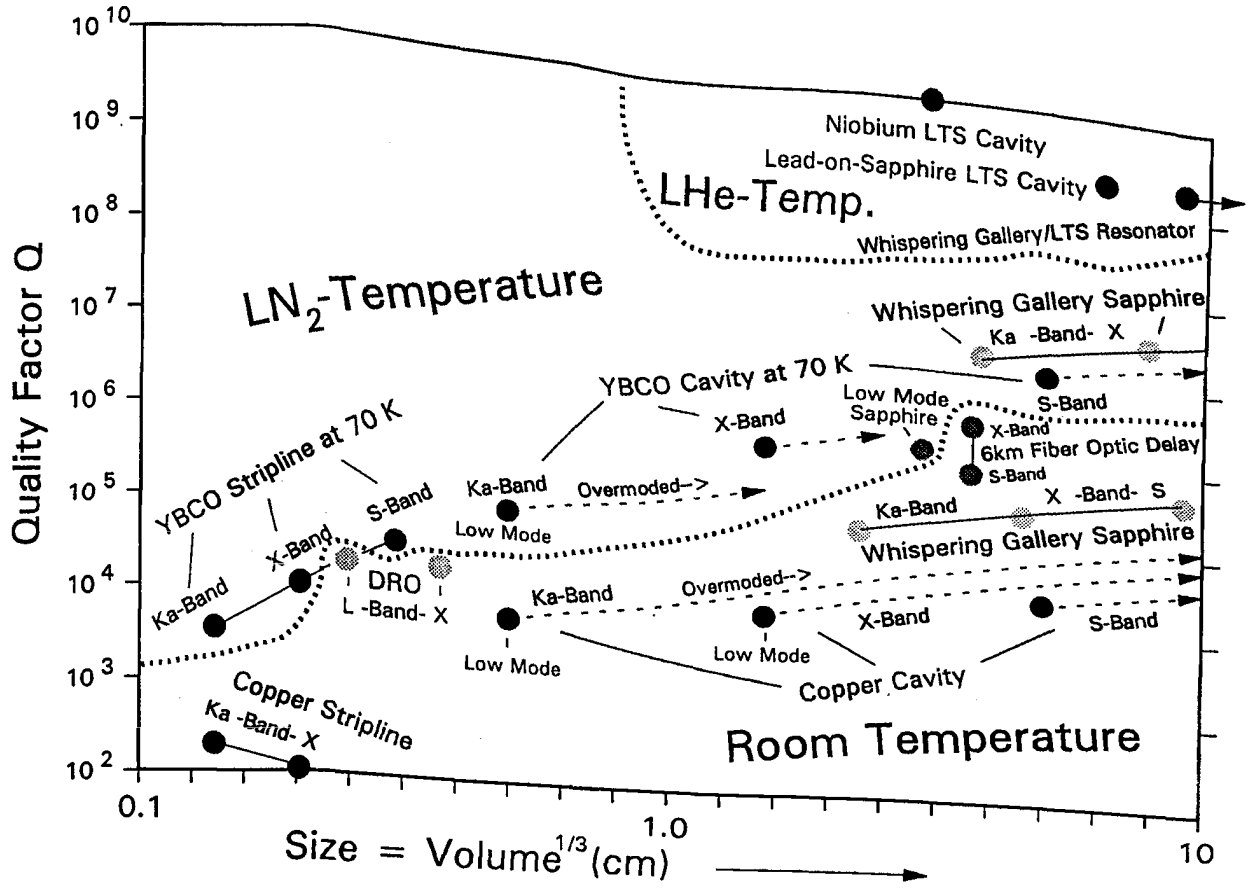


Figure 3: Resonator Q for various technologies plotted as a function of resonator size. Three temperature regimes are delineated by dotted boundaries. Other things being equal, a figure of merit would increase toward the upper left of this graph with small size and high Q. Additionally, a slight performance advantage is obtained with lower operating frequency due to improved semiconductor noise performance.

tuations

$$S_\phi(f)|_{\rm osc} = \left(\frac{\nu}{2Q}\right)^2 \frac{S_\phi(f)|_{\rm amp}}{f^2} \qquad (2)$$

where $y \equiv \delta\nu/\nu$ is the fractional frequency deviation. For flicker phase noise in the active device, $S_\phi(f)|_{\rm amp} \propto 1/f$, and Eq. (2) gives the $1/f^3$ dependence for oscillator phase noise mentioned earlier.

It is useful to include an additional step in the analysis, converting this achieved performance at the oscillator frequency to that which will be available to the user at a required fixed frequency $\nu_{\rm user}$. Using the conventional scaling of phase with frequency multiplication we can write

$$S_\phi(f)|_{\rm user} = S_\phi(f)|_{\rm osc} \times \left(\frac{\nu_{\rm user}}{\nu_{\rm osc}}\right)^2 \qquad (3)$$

$$= \left(\frac{\nu_{\rm user}}{2Q}\right)^2 \frac{S_\phi(f)|_{\rm amp}}{f^2} \qquad (4)$$

from Eq. (2). This equation contains no reference to the operating frequency of the oscillator itself. Only the resonator Q and the phase noise of the active device are left. An appropriate figure of merit might be expressed e.g. as Q divided by the phase noise of the amplifier, expressed as radians in a 1 Hz bandwidth at 1 Hz offset from the carrier.

Figure 2 shows the schematic diagram for a STALO in which the frequency variations of a noisy microwave source are cancelled by a feedback loop that detects the consequent phase shifts across a high–Q resonator to generate a frequency correction voltage. In the limit of large loop gain, stable equilibrium requires that the phases at the two input ports of the mixer be in quadrature (mixer output = zero). An important advantage of the STALO approach is that X–band mixers are available with lower phase noise than can be attained by active devices. Furthermore, several "circuit tricks" have been adapted

to this type of approach to further reduce the effects of semiconductor noise.

In the following section we will deal in more detail with the issues brought up here that together determine oscillator performance; resonator Q, semiconductor performance and circuitry; and finally, will review phase noise performance which has been attained by various laboratories using these devices.

Resonator Q

Figure 3 shows resonator Q performance for various technologies presently available or under development. The choice of *size* for the X–axis is motivated both by the disadvantage of size to any application or cost, and the advantage of size with respect to performance. Assigning a size to some of the resonators was problematical and a little arbitrary, e.g. should the size of required coupling circuitry be assigned to a DRO resonator, etc.

Some of the points are calculated from measured materials performance, when representative comparable data was not available. Thus, for example, both copper and YBCO high temperature superconducting cavities are assumed to have a figure of merit of $QR_s = 300\Omega$, with Q's based on well measured values for surface resistance R_s[5]. Stripline resonators were assumed to have a figure of merit $QR_s = 3\Omega$ at 10 GHz, a typical value[4] for a $\lambda/2$ mode. This geometry was then fixed, varying only the strip length for the other frequency bands.

Because the drop in surface resistance in YBCO is barely under way at a temperature of 77K, a temperature of 70K was used in this case, while available whispering gallery sapphire resonator results are reported at 77K.

Low mode sapphire Q performance is extrapolated to that expected in a HTS container[6]. Measured results with a copper container are roughly 4 times lower.

The general features shown by this display are:

- Increased performance may be purchased for any given technology at the cost of increased size. More generally, this same trade–off holds for all available technologies taken together, at either 300K or 77K.

- Substantially increased performance is available from advanced technologies, and even more prominently, for reduced operating temperatures.

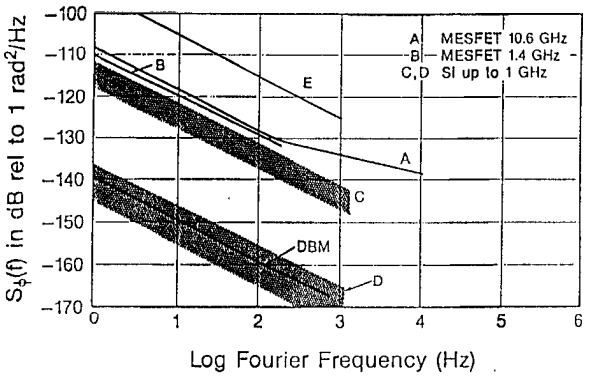


Figure 4: Microwave amplifier phase noise performance from reference [11]. Double balanced mixer (DBM) performance at frequencies up to X–band matches that of available active devices at lower frequencies. Curve E shows fiber optic transmit/receive phase noise performance at 10 GHz[12].

Both fundamental mode (points) and overmoded (arrows) operation are shown for cavity resonators, and here a sharp difference is shown between copper and YBCO HTS resonators. Interestingly, for any given size, the highest Q in a copper resonator may be obtained at the highest available operating frequency; here the advantages of overmoding surpass the disadvantage due to increased surface resistance as the frequency is raised. In contrast, each (point) low mode HTS resonator shows better performance than the (arrow) overmoded higher frequency resonator of the same size. The difference is that the HTS surface resistance shows a much stronger ν^2 dependence on frequency ν compared to the $\sqrt{\nu}$ dependence characteristic of copper at room temperature.

Clear "winners" in the various size ranges are DRO and whispering gallery sapphire resonators at room temperature, and HTS and sapphire resonators at LN_2 temperatures. A surprising performer is the fiber optic delay line. However, the high Q in this case is offset by higher flicker noise associated with required additional transmitter and receiver, compared to a semiconducting phase comparator.

Semiconductor Device Noise

Figure 4 shows representative active device noise for microwave devices. Active device performance is substantially degraded at higher frequencies compared to that available from silicon devices at lower frequencies. However, the performance of the lower

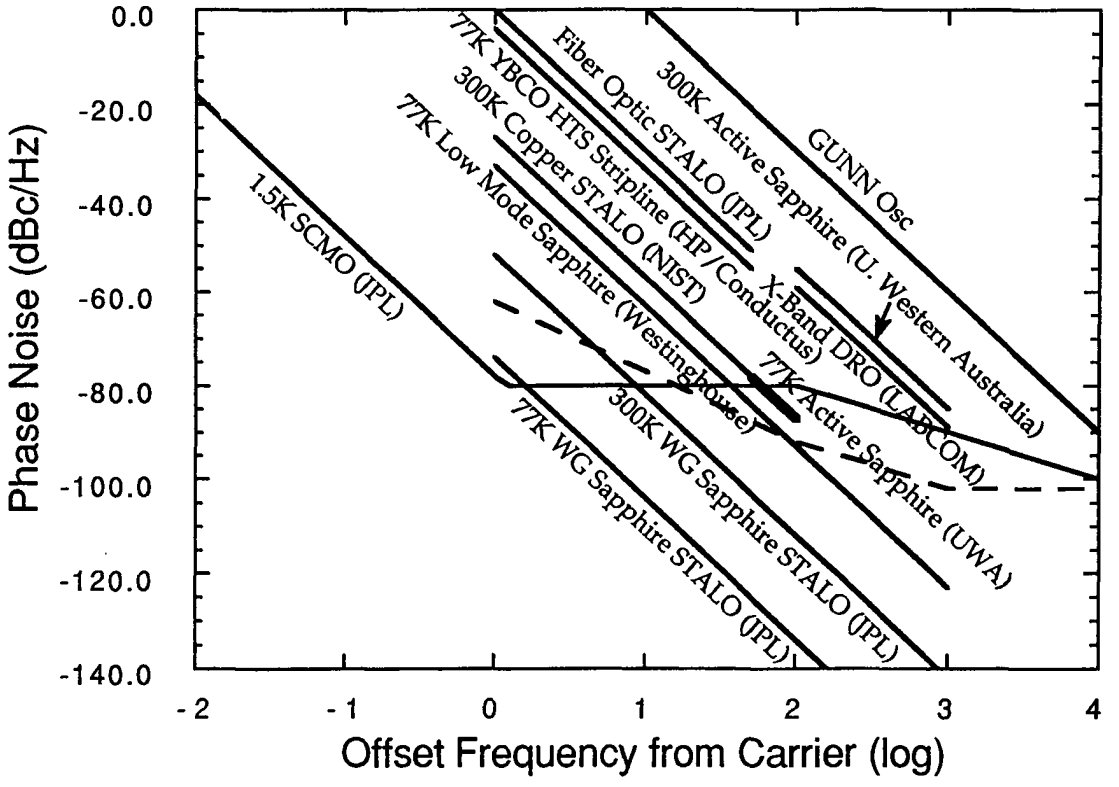


Figure 5: Representative best oscillator phase noise performance for various microwave technologies, referred to 10 GHz. Performance of a high-quality quartz oscillator is shown by the dashed line for reference. Results for the WG (whispering gallery) and low mode sapphire, as well as for the YBCO stripline are based on open loop measurements; all others are direct measurements of oscillator output frequency fluctuations.

frequency devices can be recovered at microwave frequencies by the use of a passive STALO configuration incorporating a double balanced mixer as a phase detector.

Additionally, several circuit approaches have been developed that provide greater phase sensitivity for the DBM output without increasing its noise. One of these uses a phase multiplying circuit, comparing the phases of signals at resonator output and at the reflected signal from an undercoupled drive in the DBM[1,11]. A second approach amplifies a nominally zero reflected signal at critical coupling to achieve a similar effect[8]. The effectiveness of these methods is shown in the resulting oscillator performance in the next section.

Optical fiber component noise, while not as low as that of purely electronic devices, is advancing rapidly, and is offset by the very high effective Q's obtained[12].

Oscillator Noise Performance

Figure 5 shows oscillator phase noise performance for examples of various technologies available. An attempt has been made to choose the best performance in each case. The very large range from a commercial GUNN oscillator to that of, e.g., the SCMO operating at liquid helium temperature[9] is spanned by various resonator and electronic capabilities. It is clear that the electronic approach is as important as resonator performance. For example, the NIST STALO operating at room temperature[1] substantially matches the performance of the 77K active sapphire oscillator by UWA[10]. On this basis it seems that a passive approach to DRO design might also result in important advances.

However, size and operating temperature are important cost drivers for most applications, and these together favor the LABCOM DRO[2], and the 300K and 77K WG sapphire STALO[8]. For size alone, the HP/Conductus 77K HTS stripline shows the ad-

vantage[5], while at both 300K and 77K, the WG sapphire STALO designs show the best numerical performance.

For ultimate performance, operation at liquid helium temperatures makes possible electromagnetic resonator Q's and stabilities which are not possible any other way. The result is the superior performance shown by the JPL SCMO very close to the carrier ($0.01\text{Hz} \leq f \leq 1\text{Hz}$). Higher operating power for the SLOSC from UWA[6], another liquid helium design using a sapphire/LTS resonator, results in a substantially lower noise floor, at the expense of somewhat poorer performance very close to the carrier. Not shown, but also of interest, are results for a Tunable SLOSC operating at 40K which essentially matches Westinghouse sapphire resonator measurements for low-mode operation at 77K.

Surprising performance is shown by the HTS stripline with high performance in spite of its relatively low Q, and by the 77K WG sapphire STALO, which matches and complements the performance of much more complex systems operating at liquid helium temperature.

Conclusions

The range of capabilities between relatively simple GUNN and DRO oscillators on the one hand and the "super" oscillators at liquid helium temperature on the other, is being spanned by new generation of oscillators with a wide variety of resonator and electronic features. They make possible a variety of new capabilities which will complement and eventually replace more conventional oscillators. Several new ideas have recently appeared, including small, low mode sapphire resonators combined with a HTS container for small size with high Q, and an optical delay line which shows low phase noise with wide range fast tuning capability together with temperature compensation due to sophisticated fiber design. Application of advanced circuitry and resonators together make possible performance far superior to that which was available only recently. New approaches are needed, however, to enable frequency compensation for sapphire and HTS resonators at temperatures above 10K in order to match the stability of quartz and DRO oscillators that has made them standards of the industry.

Acknowledgements

The author would like to thank L. Maleki, R.T. Wang, D.G. Santiago, J.D. Prestage, A. Kirk, R.L. Sydnor, and P.F. Kuhnle for their many contributions to the technical and expository perspectives embodied by this work.

References

[1] F.L. Walls, C.M. Felton, and T.D. Martin "High Spectral Purity X-band Source," *Proc. 44th Symposium on Freq. Control*, 542–547 (1990).

[2] M. Mizan and R. McGowan, "Ultra-low-noise 8.3 Ghz Dielectric Resonator Oscillator," *Proc. 45th Symposium on Freq. Control*, 693–699 (1991).

[3] S.R. Stein and J.P. Turneaure, "Superconducting Cavity Stabilized Oscillator with Improved Frequency Stability," *Proc. IEEE*, **63** 1249 (1975).

[4] D.E. Oates, W.G. Lyons, and A.C. Anderson, "Superconducting Thin-Film $YBa_2Cu_3O_{7-x}$ Resonators and Filters," *Proc. 45th Symposium on Freq. Control*, 460–466 (1991).

[5] 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).

[6] 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); also see M.M. Driscoll and R.W. Weinert, "Low Noise, Microwave Signal Generation using Cryogenic Sapphire Dielectric Resonators: an Update," this conference.

[7] 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.

[8] D.G. Santiago and G.J. Dick, "Microwave Frequency Discriminator with a Cooled Sapphire Resonator for Ultra–Low Phase Noise," this conference; also see G.J. Dick and J. Saunders, "Method and Apparatus for Reducing Microwave Oscillator Output Noise," U.S. Patent #5,036,299, July 30, 1991.

[9] R.T. Wang and G.J. Dick, "Improved Performance of the Superconducting Cavity Maser," *Proc. 44th Symposium on Freq. Control*, 89–93 (1990).

[10] A.J. Giles, S.K. Jones, D.G. Blair and M.J. Bukingham, "A High Stability Microwave Oscillator based on a Sapphire Loaded Superconducting Cavity," *Proc. 43rd Symposium on Freq. Control*, 83–89 (1989); also see 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 Frequency Control*, 495–499 (1990); available commercially through Sapphire Technologies, technical marketing for the University of Western Australia.

[11] 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," *Proc. 42nd Annual Symposium on Frequency Control*, 432–441 (1989).

[12] R.T. Logan, Jr. and L. Maleki, "Ultra–Stable Microwave and Millimeter Wave Photonic Oscillators," this conference; also see R.T. Logan, Jr., L. Maleki, and M. Shadaram, "Stabilization of Oscillator Phase Using a Fiber–Optic Delay–Line," *Proc. 45th Symposium on Frequency Control*, 508–512 (1990).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DESIGN AND PERFORMANCE OF A LOW NOISE, WIDE TUNING RANGE AQP SAW DELAY LINE VCO*

G. K. MONTRESS, T. E. PARKER, D. W. KRESS+, and J. A. KOSINSKI++

Raytheon Research Division, Lexington, MA 02173
+Naval Air Warfare Center, Aircraft Division, Indianapolis, IN 46219
++U.S. Army Research Laboratory, Ft. Monmouth, NJ 07703

ABSTRACT

A hybrid circuit surface acoustic wave delay line based oscillator was developed in response to an airborne system application. The 160.7 MHz oscillator's electronic frequency tuning range is extremely wide, approximately ±275 kHz (±1700 ppm). In addition, analog electronic temperature compensation to better than ±30 ppm at the center of the oscillator's voltage tuning range was incorporated into the design. A phase noise level of better than -80 dBc/Hz at 100 Hz carrier offset frequency was demonstrated on several prototypes. Finally, a vibration sensitivity less than $1x10^{-9}/g$ per axis was achieved in order that the oscillator would operate satisfactorily in the comparatively high vibration airborne environment without significantly degraded phase noise performance.

I. INTRODUCTION

Recently, state-of-the-art phase noise performance has been demonstrated for both surface acoustic wave (SAW) resonator and delay line stabilized oscillators [1-5]. The same basic feedback loop oscillator design philosophy [6] was applied in each case in order to achieve these results. This paper will describe in detail the design, component selection, fabrication, and performance of a low noise All Quartz Package (AQP) SAW delay line voltage controlled hybrid circuit oscillator. The 160.7 MHz oscillator's electronic frequency tuning range is extremely wide, approximately ±275 kHz (±1700 ppm), consistent with the All Quartz Packaged (AQPd) SAW device's nominal 0.9 μs delay. In addition, analog electronic temperature compensation [7] of the oscillator's inherent fractional frequency variation over the -54°C to +71°C temperature range, approximately a 3 or 4:1 reduction compared to the free running oscillator's fractional frequency variation, was incorporated into the oscillator's design. Finally, in order to accommodate the oscillator's specified bipolar voltage control signal (0v ± 5v) and properly interface with the feedback loop's electronic phase shifter voltage control range of 0v to +8v, while still operating satisfactorily from a single +15v dc power supply, it was necessary to include a charge-pump integrated circuit (MAXIM 637) in combination with a dual voltage regulator integrated circuit (SG 1502) in the oscillator's dc power conditioning circuitry. This presented several unique problems, including the need to adequately filter the dual voltage regulator's output voltage noise spectra while simultaneously minimizing coupling of the rather significant approximately 50 kHz (and many harmonics thereof) spurious outputs of the charge-pump circuit to the oscillator's rf output as either AM or PM sidebands.

The residual phase noise properties of an oscillator's electronic components, for example RF amplifiers and electronic phase shifters (those used for electronic frequency agility as well as for electronic temperature compensation) are extremely important if reproducibly low white PM ($S_\phi(f) \sim f^0$) and flicker FM ($S_\phi(f) \sim f^{-3}$) oscillator noise levels are to be achieved. For SAW delay line based oscillators in particular, it goes without saying that the residual flicker (f^{-1}) noise of the SAW delay line must also be minimized if truly state-of-the-art performance is to be achieved. Residual phase noise measurements performed on a number of 160.7 MHz AQP SAW delay lines indicate that the typical 1 Hz intercept for $\mathcal{L}(f=1\text{ Hz})$ is approximately -130 dBc/Hz. This is consistent with an oscillator flicker FM ($S_\phi(f) \sim f^{-3}$) phase noise level of -82 dBc/Hz at 100 Hz carrier offset, which was indeed typically observed. This suggests that the SAW delay line device rather than the loop amplifiers (three Avantek UTO-111 type) is the limiting source of flicker FM ($S_\phi(f) \sim f^{-3}$) noise in the oscillator. The residual flicker (f^{-1}) noise levels of the hyperabrupt varactors used in the temperature compensation circuitry have been found to degrade the oscillator's typical flicker FM ($S_\phi(f) \sim f^{-3}$) phase noise spectrum of -82 dBc/Hz at 100 Hz carrier offset, potentially by as much as 40 to 50 dB at 100 Hz carrier offset in several instances. Usually, only 15 to 25 dB degradation at 100 Hz carrier offset due to the hyperabrupt varactors has been measured.

The oscillator's design will be described, and details of its performance will be given, including: phase noise, voltage tuning characteristics, fractional frequency stability and output power variation versus temperature, load pulling, dc power supply

*This work was partially supported by the Naval Air Warfare Center (NAWC), Aircraft Division, Indianapolis, Indiana, under contract number N00163-91-C-0060.

0-7803-0476-4/92 $3.00 © 1992 IEEE

pushing, three-axis vibration sensitivity, and harmonic/spurious signal levels. In addition to the oscillator's phase noise level specification at 100 Hz carrier offset ($\mathcal{L}(f) < -80$ dBc/Hz) and required electronic frequency tuning range (> ±275 kHz), its vibration sensitivity was to be better than $1x10^{-9}/g$ per axis. These three performance criteria in particular, when taken together, presented a truly state-of-the-art SAW oscillator design challenge. As such, quite often trade-offs must be made between the various performance criteria that might otherwise be desired. These trade-offs will be discussed in the context of the oscillator's design, including particular details of the dc bias circuitry.

II. AQP SAW DELAY LINE CHARACTERISTICS

The particular oscillator specifications which directly influenced the design of the SAW delay line device were:

1) Center frequency: 160.7 MHz
2) Tuning range: ±275 kHz (±1700 ppm)
3) Phase Noise: -80 dBc/Hz at 100 Hz offset (maximum)
4) Deviation from linear tuning: ±25 kHz (maximum)

In order to achieve the desired tuning range, while simultaneously assuring that the oscillator wouldn't jump modes, the SAW delay line was designed for a 1 dB bandwidth of approximately 600 kHz and a group delay of 0.9 μsec. The oscillator's mode spacing is therefore approximately 1.11 MHz and the delay line's design provides more than 10 dB mode suppression at either tuning extreme. Figure 1 illustrates the basic features of the SAW delay line design, including the inclusion of "dummy" fingers to minimize amplitude and phase ripple due to mechanical reflections. The IDTs have been thinned into four and five rungs, respectively, in order to eliminate the phasing degeneracy and achieve better mode suppression. A measured insertion gain and phase characteristic for a typical device is shown in Fig. 2(a), while Fig. 2(b) illustrates the measured deviation from a linear phase characteristic for the same device. The measured deviation from linear phase of less than ±2° is such that the SAW delay line itself would only contribute approximately ±6.2 kHz deviation from a linear frequency tuning curve, if the SAW device were the only source of such errors. This was ultimately found to be quite satisfactory in meeting the oscillator's tuning requirement of less than ±25 kHz deviation, although it actually proved to be a significant contributor to the oscillator's performance in this regard.

The residual flicker noise level of the SAW delay line was reduced through the use of a high power RF burn-in (+30 to +33 dBm incident on the device for 1 hour) to a level of approximately -130 dBc/Hz at 1 Hz carrier offset. A typical residual phase noise measurement on a SAW device is shown in Fig. 3. This turns out to be sufficiently low enough that the oscillator's phase noise at 100 Hz carrier offset is dominated by white FM ($1/f^2$). However, the SAW device's residual flicker noise is still at a level such that the oscillator's flicker FM noise at 1 Hz carrier offset is contributed by the SAW delay line device and not the loop amplifiers.

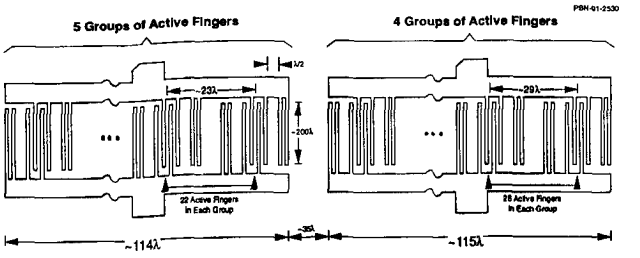
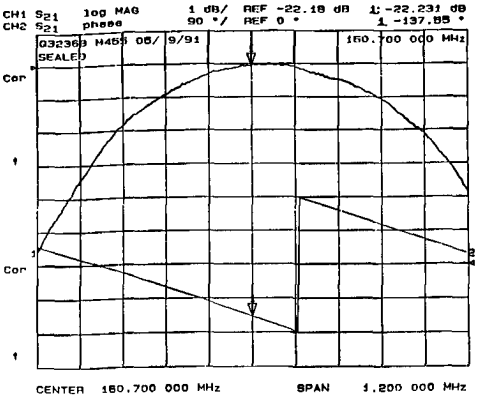
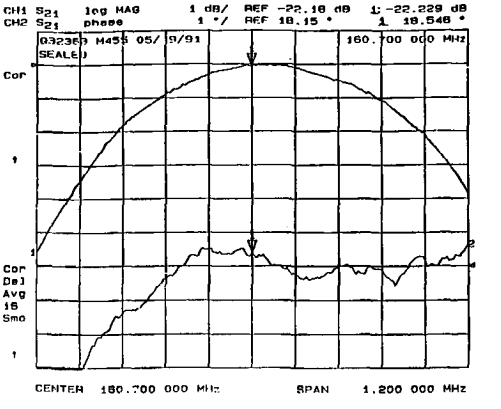


Figure 1. Basic features of SAW delay line device design.



(a)



(b)

Figure 2. (a) Measured insertion gain and phase shift for a typical AQP SAW delay line device. (b) Measured insertion gain and deviation from linear phase shift for a typical AQP SAW delay line device.

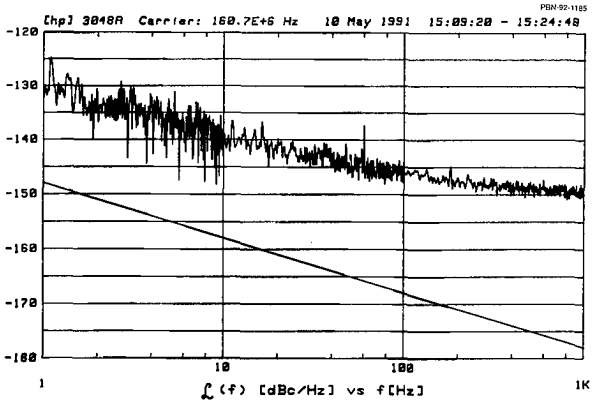


Figure 3. Residual phase noise measurement on a typical AQP SAW delay line device.

The delay line was packaged in a conventional All Quartz Package (AQP) as described previously [8] in order to achieve good long-term frequency stability [6] and low vibration sensitivity [9].

III. OSCILLATOR DESIGN

A. *Introduction*

Figure 4 shows a simplified block schematic diagram for the prototype oscillator. As indicated, each oscillator included: three loop amplifiers; an electronic phase shifter; a lumped element reactive power splitter (Wilkinson type); lumped element reactive phase adjusting circuits; an electronic temperature compensation circuit; an attenuator for loop gain adjustment; an attenuator for coupled output adjustment; a buffer amplifier; and a low pass output filter to suppress harmonics. The electronic phase shifter and electronic temperature compensation circuits were located between loop amplifiers G_1-G_2 and G_2-G_3, respectively, in order to minimize interaction between these two important circuits. Selection of the various electronic components proved to be a difficult task, since each component could potentially degrade the oscillator's phase noise spectrum if not properly chosen. However, the loop amplifiers, hyperabrupt varactors in the temperature compensation circuit, and SAW delay line itself were ultimately found to be the critical components. In addition, since a dc voltage regulator was needed, the output noise voltage of the regulator was also found to be critical in achieving the desired phase noise performance at 100 Hz carrier offset.

The temperature compensation circuit, electronic phase shifter circuit, voltage control port level shifting circuitry, and dc voltage regulator circuit are described in the following sub-sections.

B. *Temperature Compensation Circuit*

The specification required that for V_{CO} = 0v (the center of the electronic tuning range) the oscillator's fractional frequency

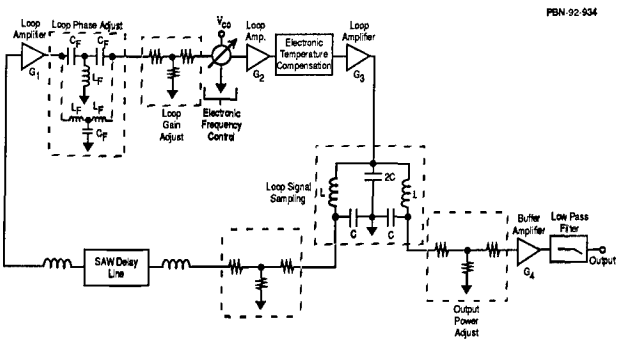


Figure 4. Simplified block circuit schematic diagram for the prototype oscillators.

stability be better than ±30 ppm for ambient temperatures in the range from -54°C to +71°. This is approximately a factor of 3 or 4 better than can be achieved for the free running oscillator. The basic analog temperature compensation scheme first demonstrated by Kinsman [7] was used, as illustrated in Fig. 5. The shunt varactor capacitance was imbedded in an L-C-L type phase shifting circuit in order to achieve another degree of adjustment through changing the inductor values. Figure 6

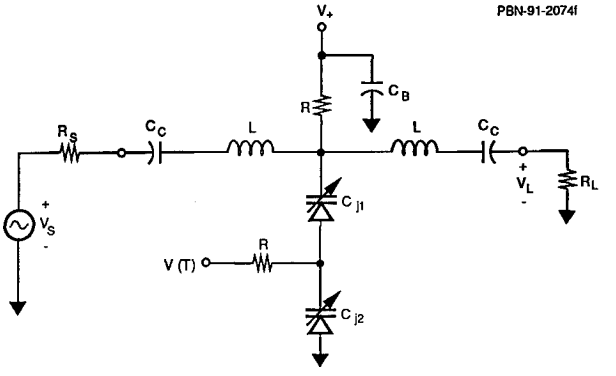


Figure 5. Basic analog electronic temperature compensation technique used, based on Kinsman [7].

shows specific details of the actual circuit used. An Analog Devices AD590 integrated circuit temperature sensor was selected for its good sensitivity and linearity. An OP-27 op-amp was used to buffer the temperature sensor output and drive the mid-point of the varactors in the electronic temperature compensation circuit. A temperature compensated zener reference diode (Z_2, +9.1v) was used to stabilize the bias voltage to the series connected hyperabrupt varactors. Experiments were performed to establish the proper bias set-up conditions for the varactors. Figure 7 illustrates the variation of transmission phase shift and insertion loss when the total dc bias voltage on

the two diodes was 7.4v, and the mid-point voltage corresponding to "turn-over" was 3.7v. While the phase variation is not perfectly parabolic, it is sufficient to achieve compensation of the oscillator over the -54°C to +71°C range, which nominally requires approximately 8 degrees of total phase shift. The hyperabrupt varactors were matched pairs of Frequency Sources type KV1502 devices with a typical junction capacitance of 40 pF at 3.5v bias. In the actual oscillator hardware, the total voltage was decreased slightly to 6.8v, and the mid-point (turn-over) voltage was 3.4v in order to achieve a slightly larger parabolicity. Figure 8 shows the measured fractional frequency variation for an oscillator incorporating the temperature compensation circuit. Again, the fractional frequency variation is a reasonable approximation to a parabola. This particular measurement was made at room temperature by varying the mid-point voltage and observing the resultant change in the oscillator's frequency.

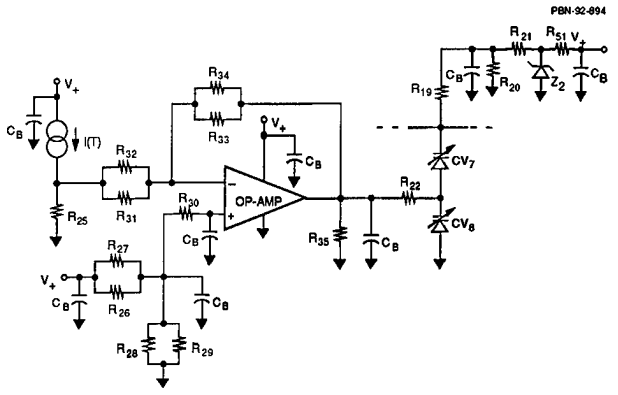


Figure 6. Schematic diagram of actual electronic temperature compensation circuit used.

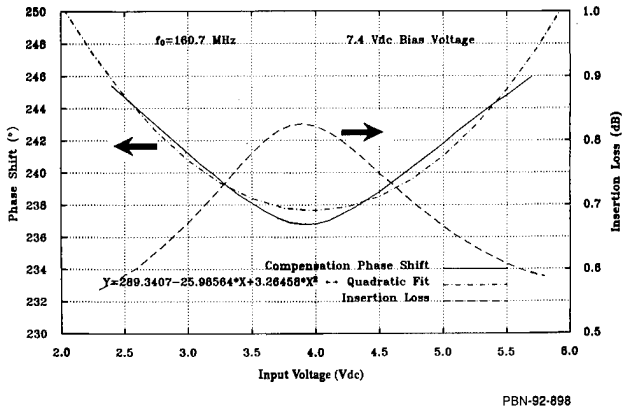


Figure 7. Measured variation of transmission phase shift and insertion loss for the electronic temperature compensation circuit.

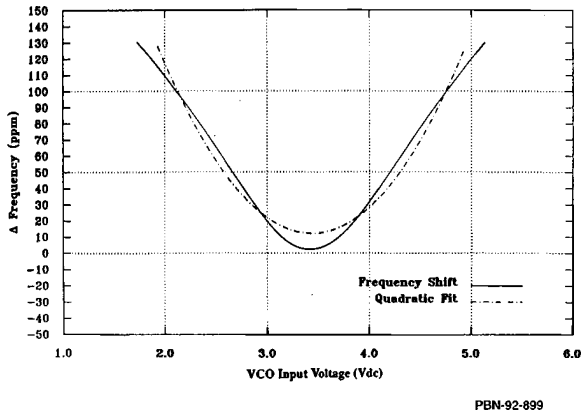


Figure 8. Measured variation of an oscillator's fractional frequency stability as a function of control voltage to the electronic temperature compensation circuit.

C. *Electronic Phase Shifter*

In order to achieve a total phase shift of 180°, the cascaded electronic phase shifter circuit shown in Fig. 9 was used. This design is an extension of our basic electronic phase shifter design which has been described previously [10]. Once again, Frequency Linear Tuning Varactors (FLTVARs, from Frequency Sources) are used in order to achieve a good linear relationship between tuning voltage and phase shift. The FLTVARs had a nominal junction capacitance of 20 pF at 5v bias. Figure 10 shows a typical plot of phase shift versus voltage for tuning voltages in the range from 0v to 12v. It is evident that the desired 180 degrees of phase shift may be readily achieved for a tuning voltage range of 0.0v to 8v, for example.

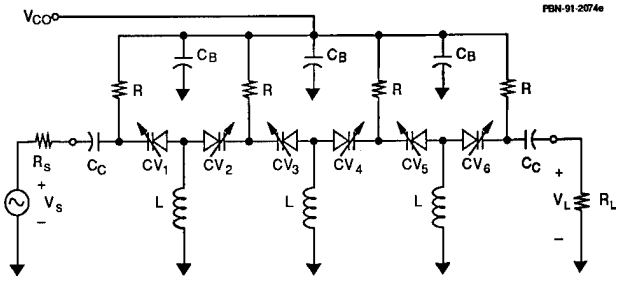


Figure 9. Cascaded three stage electronic phase shifter circuit used for electronic control of the oscillator's frequency.

D. *Control Port Level Shifting / Scaling Circuitry*

The oscillator's control voltage range is bipolar, i.e., 0v ± 5v, thereby dictating that an op-amp circuit be used to properly scale

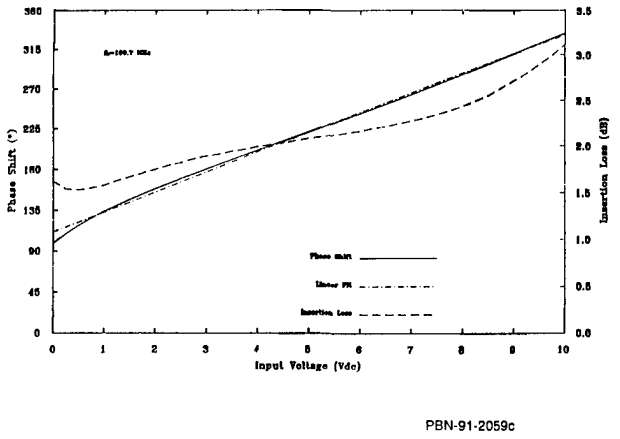


Figure 10. Typical plot of transmission phase shift versus control voltage for the cascaded three stage electronic phase shifter circuit shown in Fig. 9.

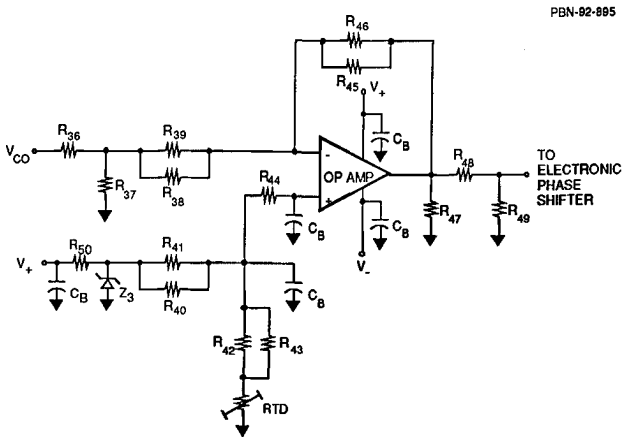


Figure 11. Control port electronic circuitry used for level shifting and scaling.

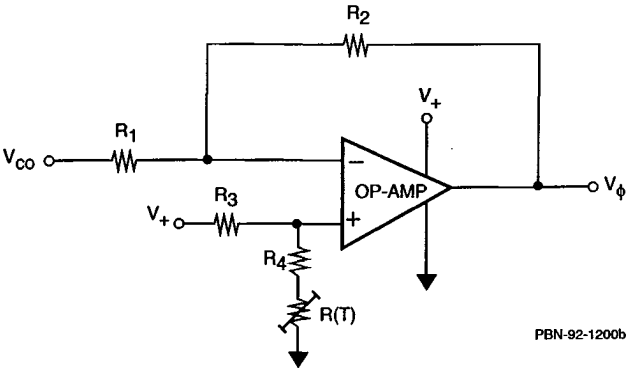
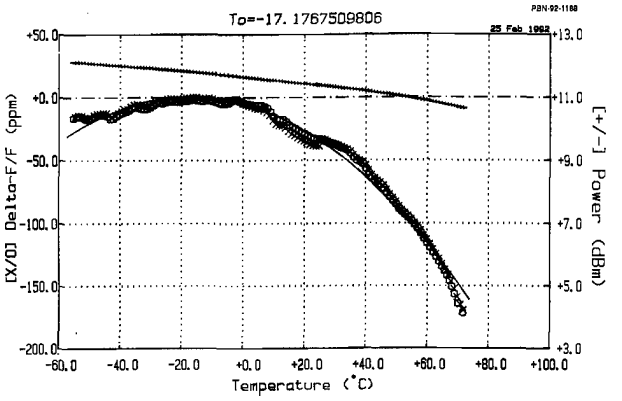


Figure 12. Basic control port circuit used for analysis purposes.

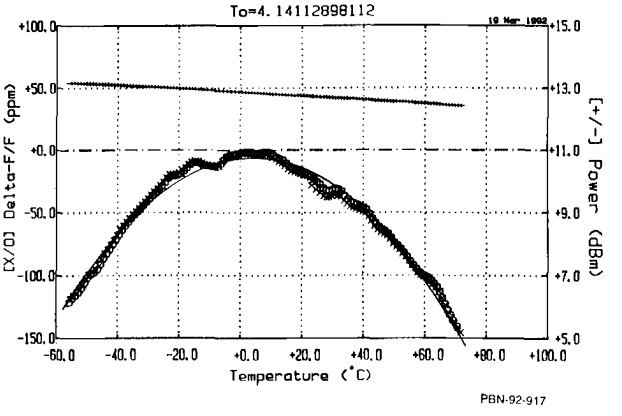
and interface the control voltage range to the 0v to 8v operating range of the electronic phase shifter circuit just described. Figure 11 shows the control port circuitry used in the prototype design. Once again, a temperature compensated zener reference diode (Z_3, +6.4v) was included to stabilize the reference voltage to the op-amp's non-inverting input. With a minimum output voltage of 0v required to properly drive the electronic phase shifter, it is necessary to provide both positive and negative bias voltages to the op-amp. This complication is discussed in the next sub-section. The variable resistor noted in Fig. 11 is a Resistance Temperature Detector (RTD). It was included in order to be able to adjust the oscillator's turn-over temperature. Our initial selection of 38° rotated Y-cut quartz for SAW device fabrication yielded devices with turn-over temperatures near 35°C. We had anticipated that the electronic circuitry in the feedback loop would lower the oscillator's turn-over temperature by approximately 30°C. The reality was such that the oscillator's turn-over temperature was near -50°C, a lowering of almost 85°C! The RTD, which possesses a very linear resistance versus temperature characteristic, was incorporated into the design as a means of electronically moving the oscillator's turn-over temperature back to nominally +8°C, the approximate center of the operating temperature range. By properly selecting the values of the resistors labeled R_3 and R_4 in the voltage divider shown in the simplified circuit schematic of Fig. 12, it is possible to raise the oscillator's turn-over temperature by as much as 50 to 75°C. This is illustrated in Figs. 13(a) and 13(b), where an oscillator's turn-over temperature has been shifted by approximately 25°C by simply making small adjustments to the values of R_3 and R_4 shown in Fig 12.

E. Voltage Regulator Circuitry

As indicated in the previous sub-section, the bipolar control voltage range in combination with the electronic phase shifter's operating voltage range initially dictated that both positive and negative voltages be used to bias the op-amp in the control port circuitry. The problem was that eventually the prototype oscillator was to be form-factored to operate in an existing system which only had +15v available! A voltage inverter (charge-pump) was used to generate a negative voltage and a dual voltage regulator was then used to provide stable ±11.5 dc voltages for the oscillator's circuitry. The charge-pump was a MAXIM637, while the dual regulator was an SG1502. The circuit schematic for the dc bias circuit is shown in Fig. 14. One problem associated with using the charge-pump is that it generates a very strong 40 to 50 kHz (and many harmonics thereof) ripple, which must be adequately filtered using L-C-L low pass filters, as noted in Fig. 14, in order to meet the oscillator's spurious requirement of -60 dBc (maximum). Figure 15(a) shows the typical level of spurious at the output of the inverter, while Figs. 15(b) and 15(c) show the extent of filtering achieved by the voltage regulator and its output L-C-L filters (1 mH - 1 μF - 1 mH) for the positive and negative



(a)



(b)

Figure 13. (a) Oscillator with its turn-over temperature electronically adjusted to approximately -17°C. (b) Oscillator with its turn-over temperature electronically adjusted to approximately +4°C.

output voltages, respectively. This degree of filtering is sufficient to drop the biased oscillator's output spurious well below -80 dBc. Finally, Figs. 16(a) and 16(b) show the measured noise voltages for the filtered positive and negative output voltages, respectively. Ultimately, the oscillator's inherent voltage pushing sensitivity to the positive supply voltage limited the phase noise performance of the oscillator at 100 Hz carrier offset. This is discussed further later in the paper, where a solution based upon only needing positive voltage for operation is briefly described.

F. *Summary*

Figure 17 shows a detailed circuit schematic for the entire oscillator, exclusive of the dc bias circuitry which was shown in Fig. 14. Where resistors are shown in parallel (e.g., R_{41} and R_{42}) this was done in order to maintain an ability to make bias adjustments with a resolution better than would have otherwise been possible with standard value, one percent tolerance resistors.

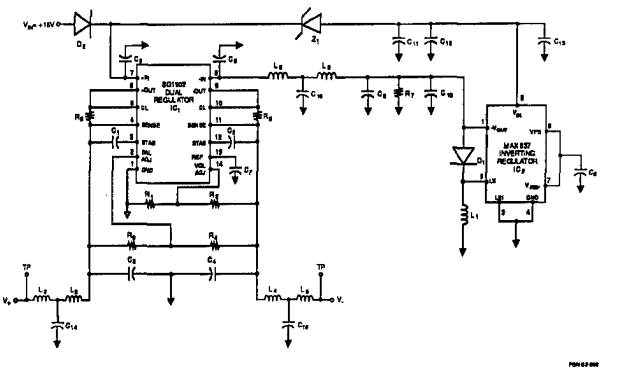


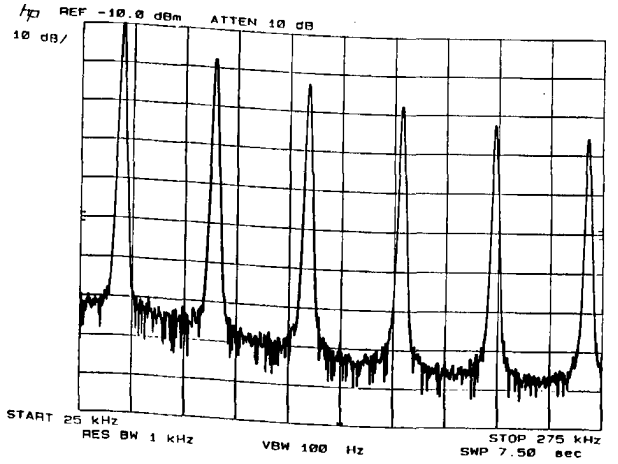
Figure 14. Circuit schematic for oscillator's dc circuitry, including MAX637 charge-pump and SG1502 dual regulator.

III. Oscillator Performance

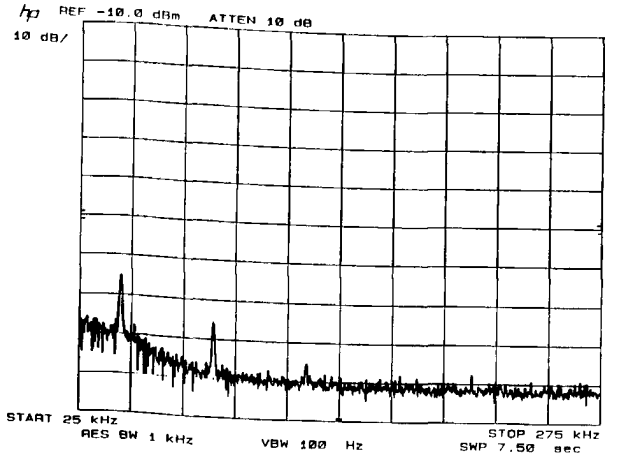
Figure 18 shows a photograph of a prototype oscillator in a 1.72"x1.72"x0.275" machined Kovar hybrid circuit package. The amplifiers are substrate carrier versions of the Avantek UTO-111. This particular amplifier was selected for the application because of its good dc to RF conversion efficiency, that is a power output at 3 dB compression of approximately +14 dBm when biased with +8.0v (9 mA).

The typical oscillator's frequency tuning characteristic and linearity (deviation from linear tuning) were excellent, as illustrated in Fig. 19(a) and 19(b), respectively. The maximum deviation from linear tuning was only ±8 or 9 kHz in the worst case. These measurements were performed at room temperature (+25°C), however, similar performance was observed over the entire operating temperature range of -54°C to +71°C.

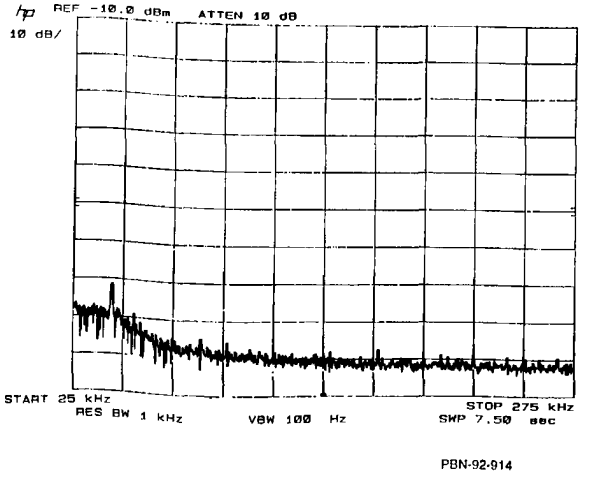
The oscillator's phase noise and spurious requirements are shown in Fig. 20. The most difficult phase noise specification to meet is at 100 Hz carrier offset, where a level of -80 dBc/Hz was needed. It should be noted that the phase noise requirement shown in Fig. 20 applies under vibration, although the 100 Hz carrier offset level is a considerable challenge even quiescently (i.e., non-vibrating). The phase noise specification corresponds to approximately 1.9° of integrated RMS phase jitter. Figures 21(a), 21(b), and 21(c) illustrate the typical phase noise performance achieved, for tuning voltages of -5v, 0v, and +5v, respectively. These measurements were again performed at +25°C with the oscillators operating off of external ±11.5v supplies.



(a)

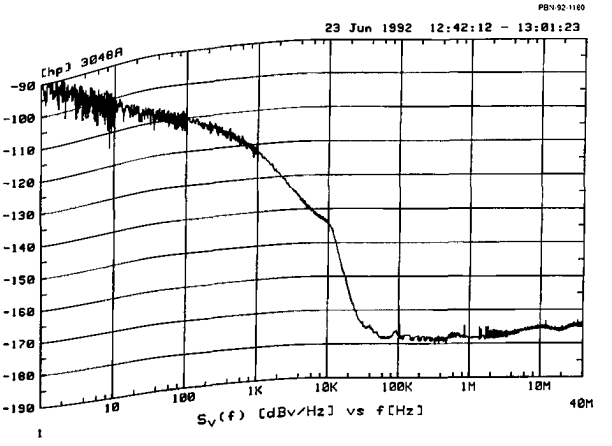


(b)

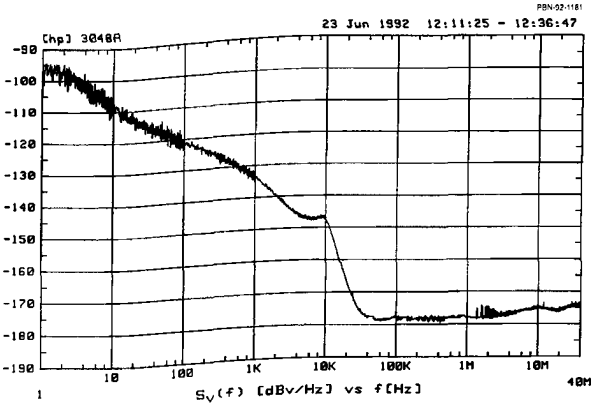


(c)

Figure 15. (a) Typical spurious level at the output of the MAX637 charge-pump integrated circuit. (b) Typical spurious level at the positive output of the SG1502 regulator after filtering (L-C-L). (c) Typical spurious level at the negative output of the SG1502 regulator after filtering (L-C-L).



(a)



(b)

Figure 16. (a) Typical measured voltage noise at the positive output of the SG1502 regulator after filtering (L-C-L). (b) Typical measured voltage noise at the negative output of the SG1502 regulator after filtering (L-C-L).

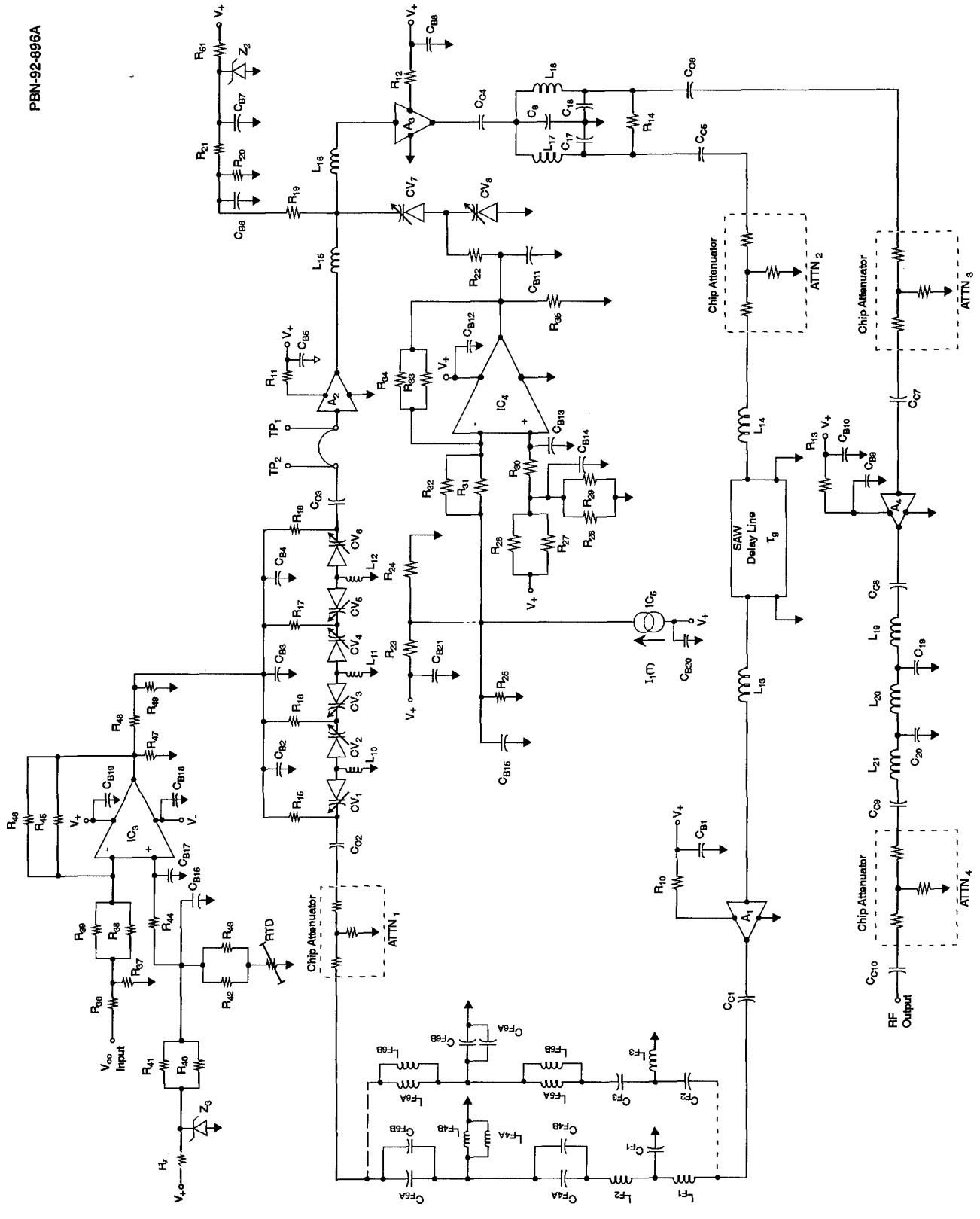
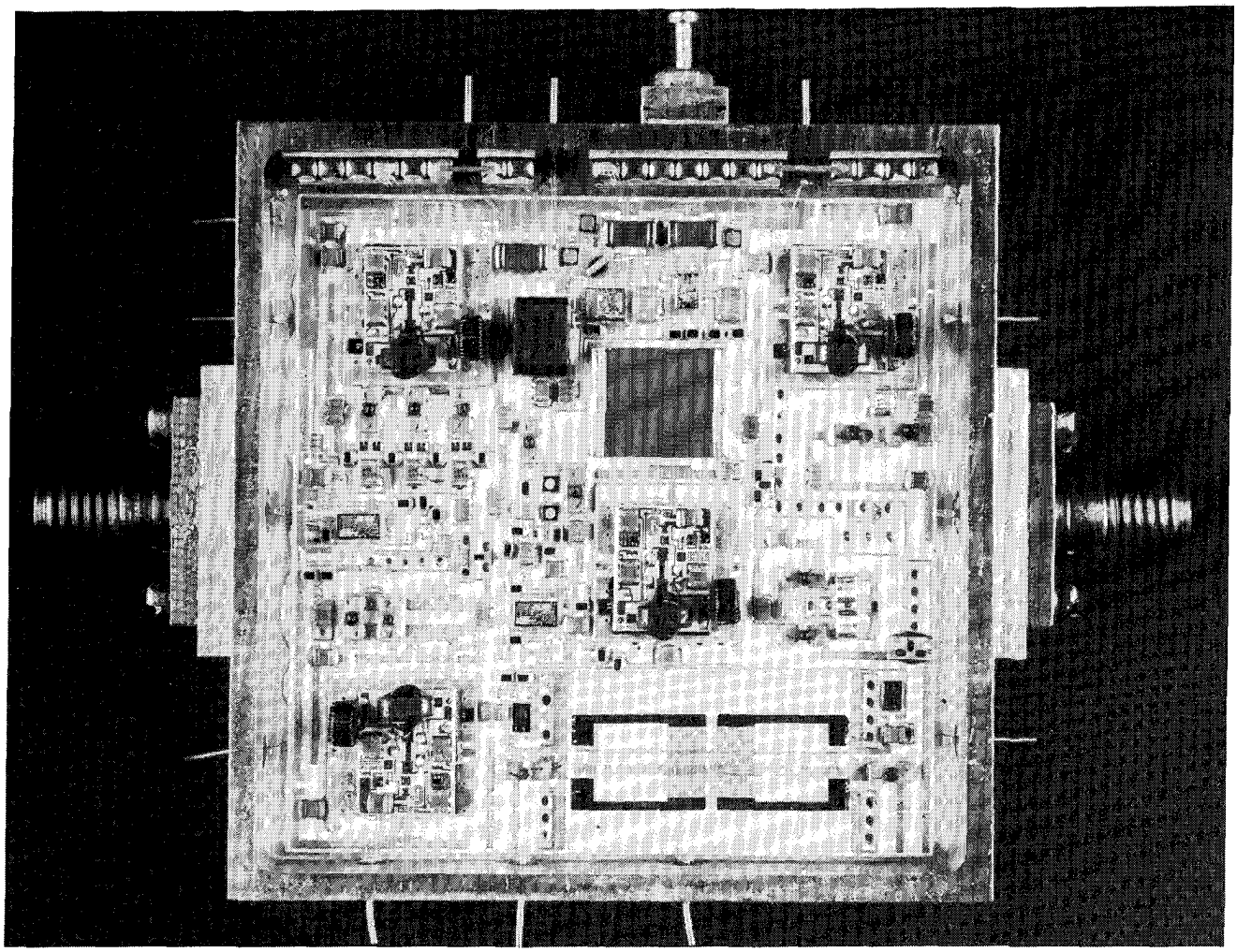
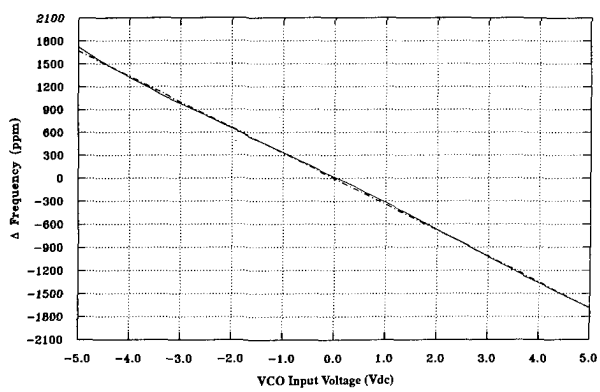


Figure 17. Detailed circuit schematic of the complete hybrid circuit SAW delay line oscillator, except for the dc circuitry which was shown in Fig. 14.



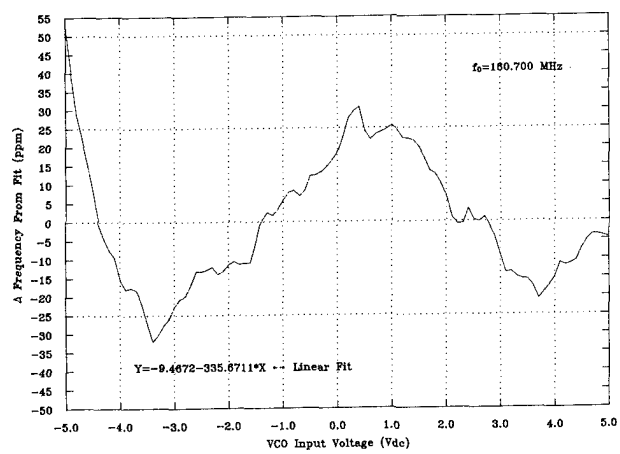
PBN-92-1186

Figure 18. Photograph of a prototype hybrid circuit AQP SAW delay line based oscillator. Package is 1.72"x1.72"x0.275", exclusive of connectors and pins.



PBN-92-900

(a)



(b)

Figure 19. (a) Typical fractional frequency variation as a function of control voltage at +25°C. (b) Typical deviation from linear tuning response as a function of control voltage at +25°C.

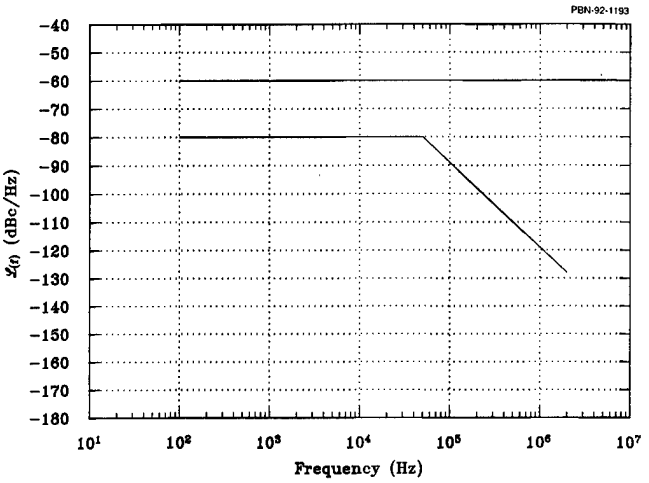
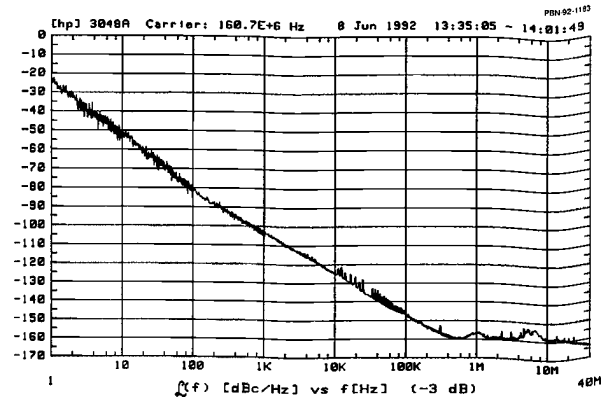


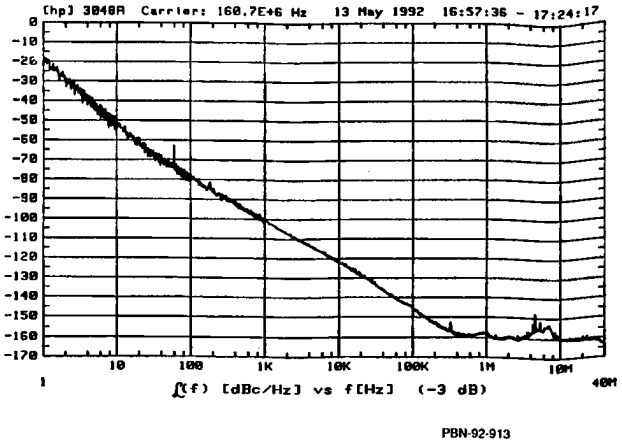
Figure 20. Phase noise (lower curve) and spurious (upper curve) specifications for hybrid circuit SAW oscillator.

During the course of oscillator development, we found that the hyperabrupt varactors could be a significant source of flicker noise, and in fact dominate the oscillator's close-to-carrier phase noise performance. This is illustrated in Fig. 22(a) where noise measurements were performed on an oscillator for two different varactor bias conditions: 1) a total voltage of +6.8v and a mid-point voltage of +3.4v, and 2) a total voltage of +12.0v and a mid-point voltage of +6.0v, the former bias condition corresponding exactly to the bias condition actually used in the oscillator's temperature compensation circuit set-up. From Fig. 22(a), it is readily apparent that the varactors are degrading the oscillator's flicker FM phase noise level by about 20 dB. Ultimately, we found that a thorough O_2 plasma cleaning followed by a UV/ozone cleaning of the varator chips prior to use seemed to eliminate the problem, although we don't know exactly what the actual source of the flicker noise was. The result of the cleaning procedure is illustrated in Fig. 22(b) for another oscillator incorporating varactor chips which had been cleaned, where very little difference is evident between the phase noise performance for the same two bias conditions just described. In fact, the lower curve in Fig. 22(b) actually corresponds to the varactors' operating bias condition in the temperature compensation circuit.

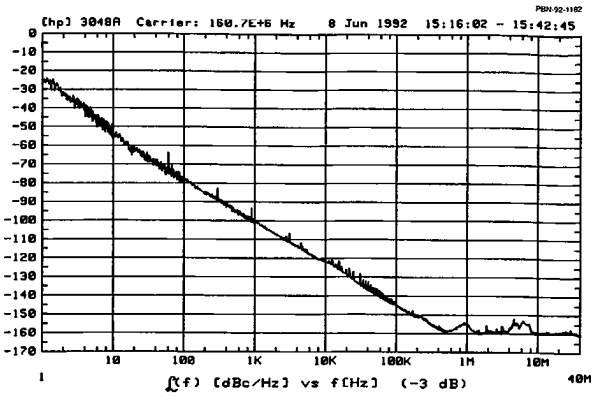
As mentioned previously, the oscillator's phase noise specification shown in Fig. 20 is to be met under vibration, with the random vibration spectra shown in Fig. 23(a) and Fig. 23(b) for the X- and Z-axes and the Y-axis, respectively. A straightforward calculation indicates that the oscillator's vibration sensitivity must be better than approximately $1x10^{-9}/g$ per axis in order to achieve this objective. Previous results suggest that all quartz packaged devices are indeed capable of performance at or below this desired level [9]. Figure 24(a) illustrates the measured magnitude of the vibration sensitivity vector, Γ, for a control voltage of 0v, while Figs. 24(b), 24(c), and 24(d) show the measured vibration sensitivities for each of the individual, mutually orthogonal axes. This result was typical for



(a)

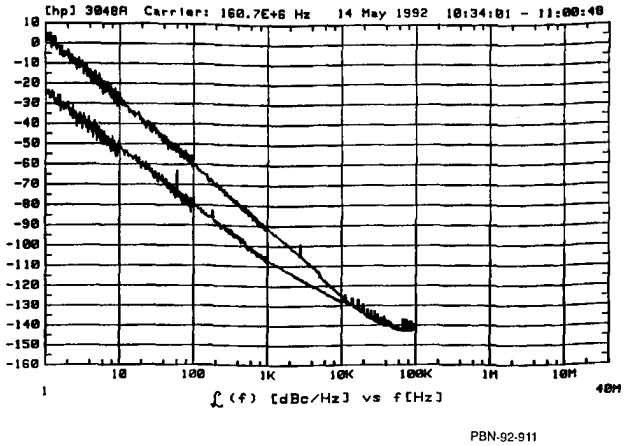


(b)

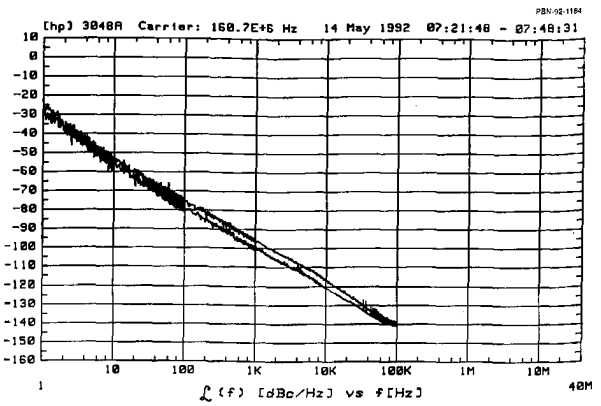


(c)

Figure 21. (a) Typical phase noise performance, $\mathcal{L}$(f), for one oscillator with V_{co} = 0v. (b) Typical phase noise performance, $\mathcal{L}$(f), for one oscillator with V_{co} = -5v. (c) Typical phase noise performance, $\mathcal{L}$(f), for one oscillator with V_{co} = +5v.

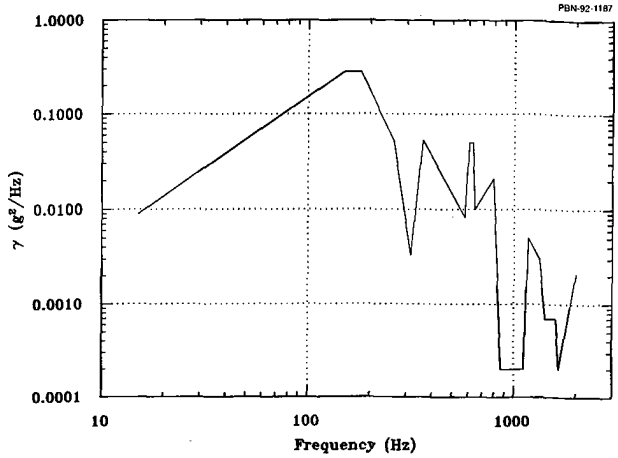


(a)

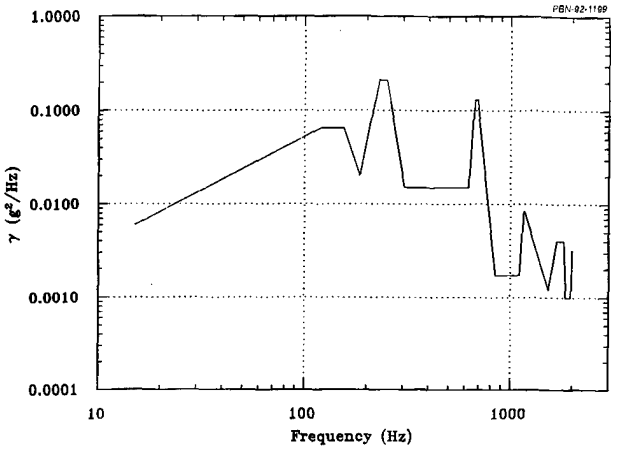


(b)

Figure 22. (a) Measured flicker FM phase noise degradation for an oscillator with "noisy" hyperabrupt varactors. (b) Measured flicker FM phase noise for an oscillator with "quiet" (cleaned) hyperabrupt varactors.



(a)



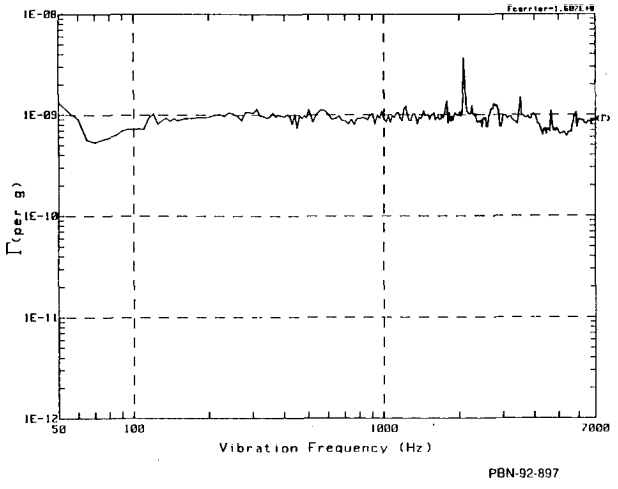
(b)

Figure 23. (a) Random vibration spectra for the X- and Z-axes. (b) Random vibration spectrum for the Y-axis.

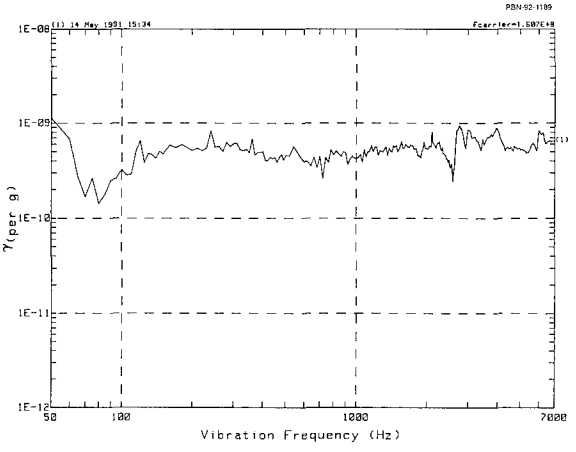
the five oscillators tested. In addition, measurements of γ_1 for vibration applied in a direction perpendicular to the plane of acoustic propagation were performed for a number of control voltages between -5v and +5v. These measurements are shown in Figs. 25(a) through 25(e) for control voltages of -5v, -2.5v, 0v, +2.5v, and +5.0v, respectively. While some variation with tuning voltage is evident, the oscillator's vibration sensitivity is still seen to be less than $1x10^{-9}/g$ for this, the generally dominant axis. The level of variation noted is consistent with recent observations [11-12], although we have not yet attempted to isolate its actual source.

Figure 26 shows a typical measurement of a temperature compensated oscillator's fractional frequency variation versus temperature. The fractional frequency variation falls within the desired ±30 ppm range. In addition, Fig. 26 also shows the measured variation of the oscillator's output power over the same -54°C to +71°C range. The degree of output power stability is exceptional, and is a tribute to the loop amplifiers' design which incorporates an active dc bias scheme.

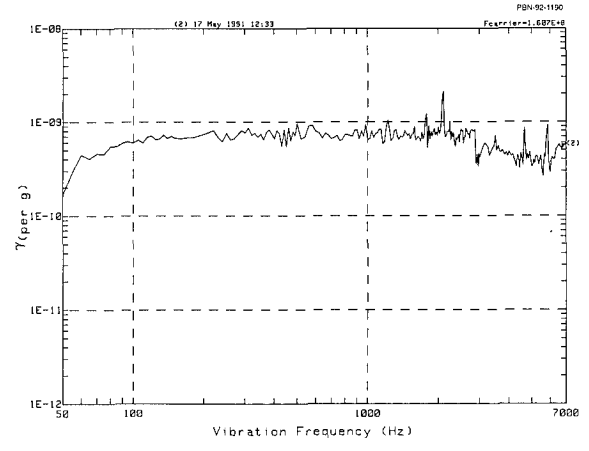
The typical oscillator's load pulling sensitivity was less than ±2 ppm for a 2:1 output VSWR, any phase. Its internal voltage pushing sensitivity for the +11.5v bias supply was found to be 6 ppm/volt, while its sensitivity to the -11.5v bias supply was not measureable due to the OP-27's excellent common mode voltage rejection. For single +15v supply operation, the oscillator's typical external voltage pushing sensitivity was less than ±0.1 ppm/volt. Finally, the output low pass filter adequately suppressed harmonics, with the worst harmonic level measured being -38 dBc (less than -30 dBc required).



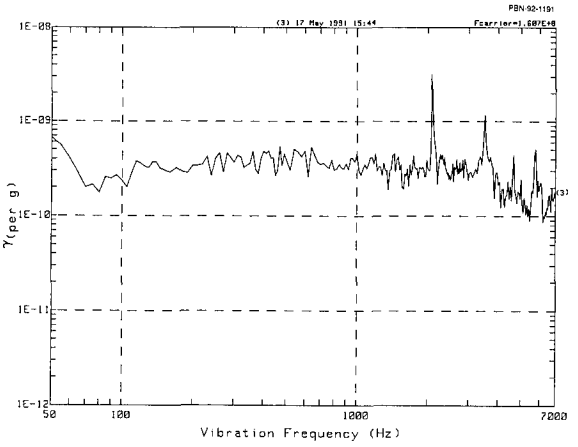
(a)



(b)



(c)



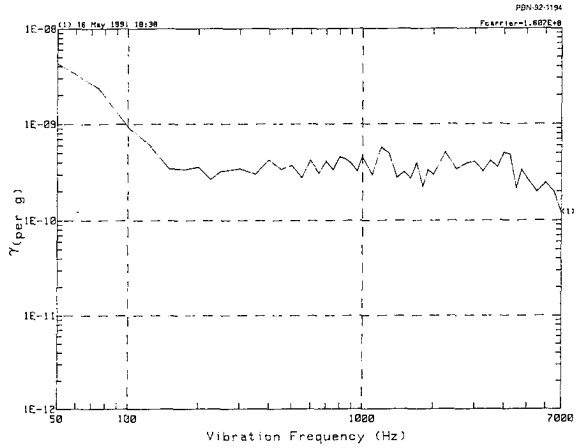
(d)

Figure 24. (a) Magnitude of the vibration sensitivity vector, Γ, for a typical oscillator with V_{co} = 0v. (b) Measured γ_1 for vibration applied in a direction perpendicular to the plane of acoustic propagation. (c) Measured γ_2 for vibration applied in the plane of acoustic propagation and in a direction perpendicular to the acoustic axis. (d) Measured γ_3 for vibration applied in the plane of acoustic propagation and in a direction parallel to the acoustic axis.

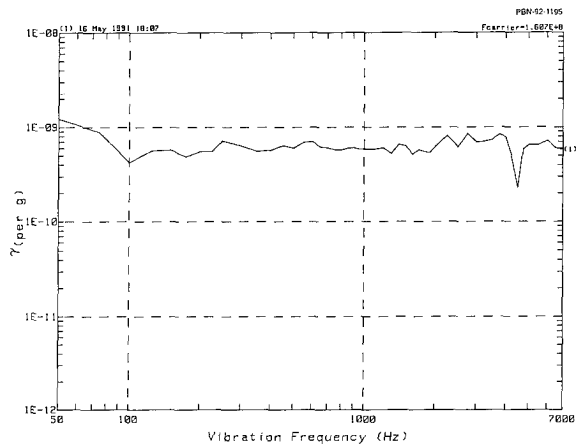
IV. Conclusions

The prototype 160.7 MHz AQP SAW delay line hybrid circuit oscillators just described have demonstrated state-of-the-art phase noise performance, while simultaneously maintaining extremely wide electronic tuning range (±275 kHz). In addition, excellent vibration sensitivity (< $1x10^{-9}/g$ per axis) was achieved. The Naval Air Warfare Center (NAWC), Indianapolis has assumed responsibilty for fabricating and testing fully form-factored prototype hardware for this application, using AQP SAW delay line devices fabricated by Raytheon's Research Division. Figure 27 shows a photograph of a form-factored prototype oscillator built by NAWC.

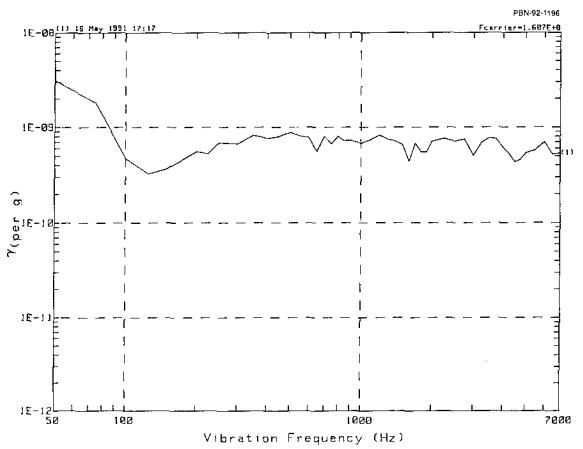
Finally, as noted previously, the dual regulator's output noise voltage (particularly the positive output voltage) was found to present a potential limit to the oscillator's phase noise performance, especially at 100 Hz carrier offset. This obstacle was eventually overcome by changing the operating voltage range for the electronic phase from 0v - 8v to 1.5v - 9.5v. This shift permitted the elimination of the negative supply voltage to the OP-27 in the control port since the op-amp could handle both positive and negative input voltages as long as the output voltage of the op-amp wasn't required to be less than +1.3v. This insures that the op-amp's output voltage isn't required to drop too close to ground. This circuit change doesn't degrade any aspect of the



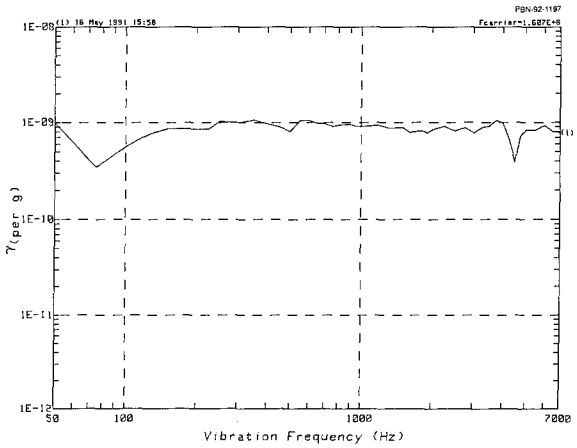
(a)



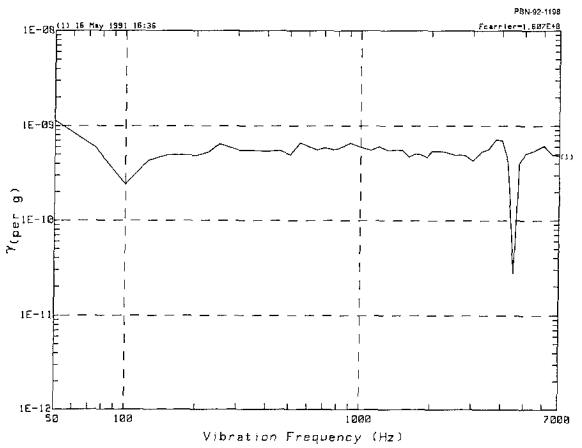
(b)



(c)



(d)



(e)

Figure 25. *Measurements of γ_1 (vibration applied in a direction perpendicular to the plane of acoustic propagation), for:* (a) V_{co} = -5v, (b) V_{co} = -2.5v, (c) V_{co} = 0v, (d) V_{co} = +2.5v, and (e) V_{co} = +5.0v.

oscillator's frequency tuning performance, and enabled us to eliminate the charge-pump circuit from the oscillator. In addition, since a negative bias voltage was no longer needed, the dual regulator was replaced with a 723 type regulator. The 723 type regulator is capable of superior output noise voltage performance, as illustrated in Fig. 28 (compare to Fig. 16). The circuit schematic for the 723 based regulator used is shown in Fig. 29.

These changes to the dc circuitry will be incorporated into future form-factored hybrid circuit SAW delay line oscillators shortly.

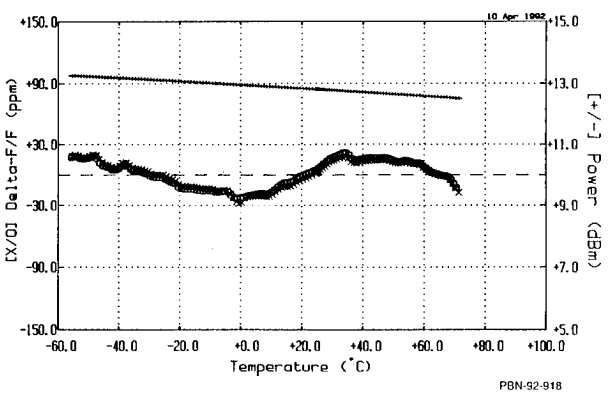


Figure 26. Typical fractional frequency stability (lower curve) and output power variation (upper curve) versus temperature for an electronically temperature compensated oscillator.

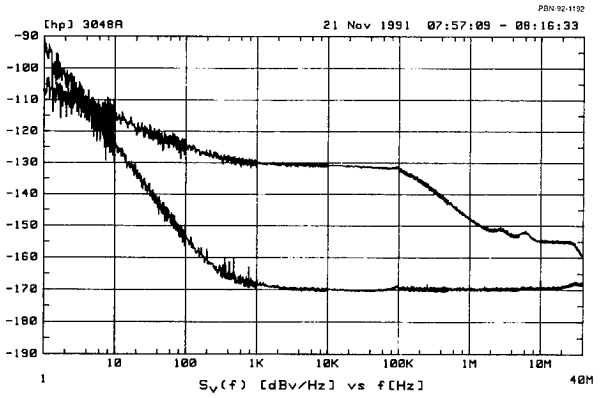


Figure 28. Output voltage noise for a typical 723 type voltage regulator circuit with filtering (lower curve) and without filtering (upper curve).

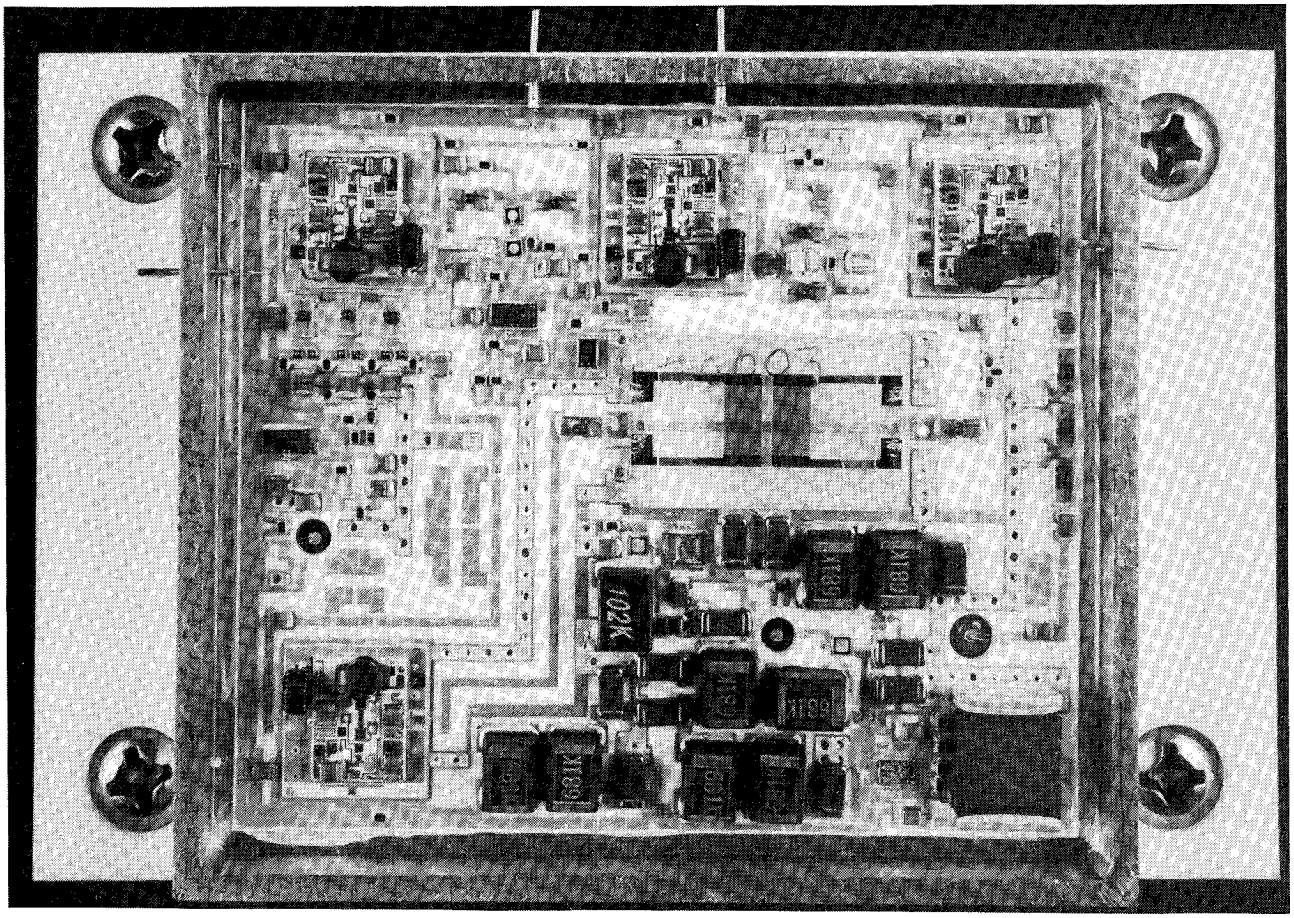


Figure 27. Photograph of a form-factored hybrid circuit prototype oscillator for an airborne system application.

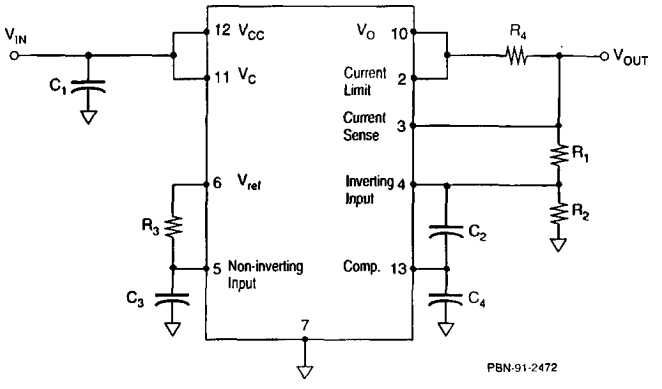


Figure 29. Circuit schematic for the 723 based regulator whose output voltage noise was shown in Fig. 28.

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, & 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] T. E. Parker and G. K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", *IEEE Transactions on Ultrasonics, Ferroelectrics, & Frequency Control*, vol. 35, no. 3, pp. 342-364, May 1988.

[7] R. G. Kinsman, "Temperature Compensation of Crystals with Parabolic Temperature Coefficients", in *Proceedings of the 32nd Annual Symposium on Frequency Control*, 1978, pp. 102-107.

[8] T. E. Parker, J. Callerame, and G. K. Montress, "A New All Quartz Package for SAW Devices", in *Proceedings of the 39th Annual Symposium on Frequency Control*, 1985, pp. 519-525.

[9] 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.

[10] G. K. Montress, "A High Performance, Hybrid Circuit Electronic Phase Shifter," in *Proceedings of RF Expo East*, 1989, pp. 231-236.

[11] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Acceleration Sensitivity Reduction in SAW and BAW Resonators and Oscillators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1990, vol. 1, pp. 573-576.

[12] J. A. Kosinski and A. Ballato, "Advances in Acceleration Sensitivity Measurement and Modeling", to be published in *Proceedings of the IEEE Frequency Control Symposium*, 1992.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

OPTIMAL AND SUBOPTIMAL SYNTHESIS OF SURFACE ACOUSTIC WAVE FILTERS

A.S.RUKHLENKO

Department of Radioelectronics
Minsk Radioengineering Institute
P. Brovka Street, 6, Minsk 220027 Republic of Belarus, CIS

Abstract

Optimal and suboptimal design techniques of surface acoustic wave (SAW) linear phase filters both based on the Remez exchange algorithm and the McClellan's computer program are considered.

An optimal synthesis provides the best fit to a design target and leads to an apodized interdigital transducer (IDT) of minimal length. The optimal solution properties are studied.

A suboptimal synthesis technique proposed allows to reduce considerably an amount of computations without significant sacrificing the approximation accuracy. Thus computer run time and storage are greatly saved if compared to an optimal synthesis. The detailed suboptimal theory and some practical design aspects are also discussed.

The design examples are presented which confirm the efficiency and the flexibility of the synthesis techniques proposed.

Introduction

Several SAW filter synthesis techniques based on the finite impulse response (FIR) digital filter theory [1] have been proposed in the past [2-15]. The most wide spread are 1) the windowing techniques [2-6], 2) the linear programming techniques [7-10], and 3) the Remez exchange algorithm techniques [11-15]. A comprehensive review of these may be found elsewhere [16].

But up to now there have been some problems in applying these techniques to a SAW filter design. That is why this is a common practice to use some simplifications, sometimes without sufficient foundations. For example, it is usually supposed that within a filter passband a contribution $F_1(\omega)$ of an unapodized IDT to the overall filter transfer function $F(\omega)=F_1(\omega)F_2(\omega)$ is negligible, i.e. $F_1(\omega)\approx 1$. Then the filter frequency response (FR) $F(\omega)\approx F_2(\omega)$ depends on the function $F_2(\omega)$ only, and consequently within the limits of the δ-function model [22,23] a SAW filter synthesis becomes equivalent to a FIR digital filter synthesis [1]. Unfortunately, this is the only case to apply the techniques above without any adaptations.

Further, while in the passband the roll-off of the overall function $F(\omega)$ due to the frequency response $F_1(\omega)$ might easily be compensated by a proper predistortion of the desired magnitude function $F_0(\omega)$, this is not the case in the filter stopband where the function $F_1(\omega)$ is usually sign-alternated.

It is yet shown in the paper how an original SAW filter FR approximation problem may be converted to an auxiliary one by a proper modifying of both: a desired magnitude function $F_0(\omega)$ as well as a weight function $W_0(\omega)$. An auxiliary approximation problem is solvable by means of standard linear Chebyshev approximation techniques using the Remez exchange algorithm by virtue of the McClellan's computer program [18] for example. In addition to a frequency response $F_1(\omega)$, both an element factor [24-26] and a multistrip frequency response [23] might also be accounted for if necessary.

An optimal synthesis provides uniquely the best fit to a design target and leads to an apodized IDT of minimal length. But its drawback is a considerable amount of computations due to a large number of optimized variables (OV) even if one uses an efficient McClellan's program. There have been some efforts to accelerate the algorithm convergence [20,21], but the results obtained are not sufficient for a real-time design.

To overcome this problem a suboptimal SAW filter synthesis technique also based on the McClellan's computer program [18] is proposed in the paper. While maintaining optimal synthesis generality and flexibility, the latter considerably reduces an OV number and hence the storage and the computation time. The detailed suboptimal synthesis theory and some practical design aspects will be discussed.

1. SAW Filter Optimal Synthesis

1.1. Optimal Approximation Problem Formulation and Solution

The SAW filter to be designed consists of two linear phase IDT. Frequency response $F_1(\omega)$ of one of them is supposed to be given a priori while the other's $F_2(\omega)$ is optimized providing a Chebyshev (minimax) approximation of the desired magnitude shape function $F_0(\omega)$. Another case, where the fil-

0-7803-0476-4/92 $3.00 © 1992 IEEE

ter consists of two apodized IDT to be optimized, is beyond the scope of this paper and is treated separately [12,27-29].

There are no constraints on a magnitude shape function $F_o(\omega)$ imposed. It may be symmetrical, non-symmetrical, multipassband etc. [15,17].

A weighted error function $\Delta F(\omega)$ can be written in the form

$$\Delta F(\omega)=W_o(\omega)[F_o(\omega)-F(\omega)] \quad (1)$$

where $W_o(\omega)>0$ is a positive-defined weight function and the function

$$F(\omega)=\xi(\omega)F_1(\omega)F_2(\omega) \quad (2)$$

describes a linear phase SAW filter FR, with the functions $F_i(\omega)$, i=1,2, being attributed to the IDT array factors and the function

$$\xi(\omega)=\omega\xi_1(\omega)\xi_2(\omega)\xi_3(\omega) \quad (3)$$

accounting for IDT element factors $\xi_1(\omega)$ and $\xi_2(\omega)$ [24-26] and multistrip coupler FR $\xi_3(\omega)$ [23] if necessary. Hereafter we suppose $\xi_3(\omega)=1$ for simplicity.

It follows from the symmetry relations for a linear phase IDT that the array factors $F_i(\omega)$, i=1,2, can take one of the four following forms:

$$F_i(\varphi)=\begin{cases}\sum_{k=0}^{n_i} a_k\cos k\varphi \\ \sum_{k=1}^{n_i} b_k\sin k\varphi\end{cases} \quad \text{for an odd number } N_i=2n_i+1 \quad (4)$$

or

$$F_i(\varphi)=\begin{cases}\sum_{k=1}^{n_i} a_k\cos(k-1/2)\varphi \\ \sum_{k=1}^{n_i} b_k\sin(k-1/2)\varphi\end{cases} \quad \text{for an even number } N_i=2n_i, \quad (5)$$

where $\varphi=\pi\omega/\omega_\pi$, $\omega_\pi=v/2p$ is a synchronism frequency, with v being a SAW velocity and p being an IDT period, and $N_{1,2}$ is a total number of the IDT acoustical sources (gaps or electrodes), with amplitudes $A_o=a_o$ and $A_{\pm k}=(a_k\pm b_k)/2$, $k\neq 0$, if numbered relatively to an IDT center.

In the quasistatic approximation [23] the element factors $\xi_i(\omega)$, i=1,2, are described by the expression [24-26]

$$\xi_i(\nu)=\frac{P_n(\cos\Delta)}{P_{-\nu}(-\cos\Delta)}\times\begin{cases}1 & \text{gap-weighted IDT} \\ \sin\pi\nu & \text{electrode-weighted IDT}\end{cases} \quad (6)$$

where

$\nu=\frac{\varphi}{2\pi}=\frac{\omega}{2\omega_\pi}$ dimensionless frequency variable;

$\Delta=\pi\eta$, $\eta=w/p$ being metallization ratio;

w, p electrode width and period (pitch);

$P_n(\cos\Delta)$ Legendre polynom of an order n;

$P_{-\nu}(-\cos\Delta)$ Legendre function;

$n=[\nu]$ spatial harmonic number $(n\le\nu<n+1)$.

The approximation problem can be stated as follows: given a desired magnitude function $F_o(\omega)$ and a weight function $W_o(\omega)$, one wishes to minimize the absolute weighted error function

$$\delta=\|\Delta F(\omega)\|=\max_{\omega\in\Omega_\pi}|\Delta F(\omega)| \quad (7)$$

within an approximation interval $\Omega_\pi=\{\omega\in[0,\omega_\pi]\}$ over the set of coefficients of the optimized function $F_2(\omega)$.

A feature of the approximation problem above is a multiplicative nature of the approximating function $F(\omega)$ given by the expression (2), with the function $F_1(\omega)$ sign-alternated in general case. Unfortunately, the McClellan's computer program [18] can not be directly applied to solve this problem, with the only exception of a special case $\xi(\omega)F_1(\omega)=1$.

Instead of the initial problem (1)-(7) let us consider an auxiliary one with an error function

$$\Delta\hat{F}(\omega)=sign\{F_1(\omega)\}\Delta F(\omega)=\hat{W}_o(\omega)[\hat{F}_o(\omega)-F_2(\omega)] \quad (8)$$

where

$$\hat{W}_o(\omega)=W_o(\omega)\,|\xi(\omega)F_1(\omega)| \quad (9)$$

$$\hat{F}_o(\omega)=F_o(\omega)/|\xi(\omega)F_1(\omega)| \quad (10)$$

One must be careful to omit in (8)-(10) those frequencies ω_i at which $\xi(\omega)F_1(\omega)=0$. The function $\xi(\omega)$ is usually monotone within Ω_π, the points ω_i being zeros of the sign-alternated function $F_1(\omega)$ only. At these frequencies an error function $\Delta F(\omega)$ takes the fixed values $\Delta F(\omega_i)=W_o(\omega_i)F_o(\omega_i)$, and the approximation might fail if the desired function $F_o(\omega_i)\neq 0$. Fortunately, if all the points ω_i are located in a filter stopband, where $F_o(\omega)=0$, then the error function $\Delta F(\omega_i)$ is also equal to zero. Therefore, there is no need to minimize an error at these points and their omitting does not influence an approximation accuracy.

Now an auxiliary approximation problem with an error function $\Delta\hat{F}_o(\omega)$ may be solved on a subset $\hat{\Omega}_\pi=\{\omega\in[0,\omega_\pi],\ \omega\neq\omega_i\}$ by any linear Chebyshev approximation technique [7-19]. The McClellan's computer program [18] can easily be applied, with the initial data changed according to the formulae (9) and (10).

An optimal solution obtained has some interesting properties that will be discussed below.

1.2. Optimal Solution Properties

Due to optimality and uniqueness properties an optimal solution does not depend on the optimization technique applied, would it be the Remez exchange algorithm or the linear programming.

A feature of an optimal solution is a behaviour of an error function $\Delta F(\omega)$ within a filter

stopband resulting from the relation (8) and from the Chebyshev alternation theorem [1]. According to this theorem, the error function $\Delta\hat{F}(\omega)$ and hence the error function $\Delta F(w)$ must exhibit on the subset $\hat{\Omega}_\pi$ at least n_2+1 equiripple extremuma, with n_2 being an order of the trigonometric polynomial $F_2(\omega)$. On the other hand the maximum number of the overall function $F(\omega)$ extremuma is defined by an order $n=n_1+n_2-1>n_2$ of the polynomial product $F_1(\omega)F_2(\omega)$. It is the difference between the extremuma number n and the alternation extremuma number n_2 that makes it possible for some extra extremuma between two neighbouring equiripple alternation ones to appear.

Indeed, unlike the usual alternation law, when two neighbour equiripple extremuma are always of the opposite signs, it follows from (8) that two extremuma 1 and 2 (Fig.1) must have the same sign if an odd order real zero ω_i of the function $F_1(\omega)$ is placed between them. This results in an extra extremum 3 of the opposite sign and of the lower amplitude to appear between extremuma 1 and 2 in the neighbourhood of the frequency ω_i. It is worth noting that even order real zeros or complex-valued roots of the function $F_1(\omega)$ do not violate the habitual alternation low.

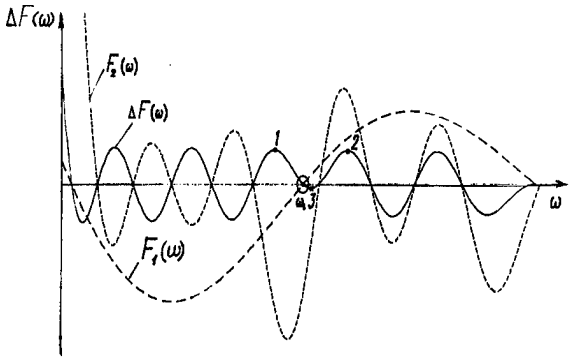
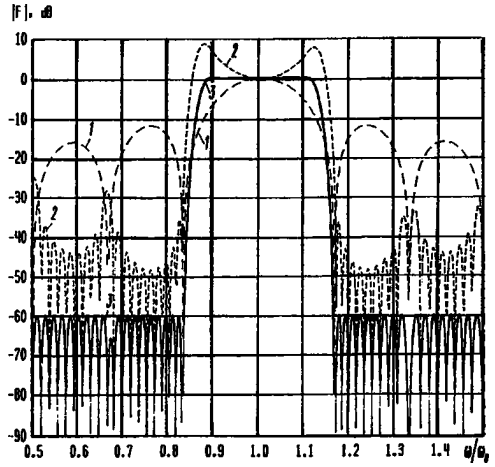


Fig. 1. Optimal Solution Error Function $\Delta F(\omega)$ and an Arrangement of Zeros and Extremuma of the Functions F1(w) and F2(w)

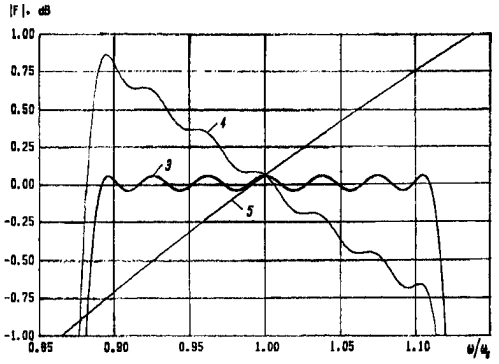
Besides, it is such a special arrangement of zeros and extremuma of the functions $F_1(\omega)$ and $F_2(\omega)$ in a filter stopband that ensures to a large degree a solution optimality, extremuma located in the neighbourhood of zeros and vice versa.

1.3. An Optimal Synthesis Design Example

The optimal synthesis technique was successfully applied to a SAW filter design. Fig. 2 shows an example of the optimal FR, with the curve 1 the FR $\sqrt{\omega}\,\xi_1(\omega)F_1(\omega)$ of the unapodized IDT, the curve 2 the FR $\sqrt{\omega}\,\xi_2(\omega)F_2(\omega)$ of the apodized IDT after optimization, and the curve 3 the overall optimal magnitude response $F(\omega)=\xi(\omega)F_1(\omega)F_2(\omega)$. For clarity, the polynomial product $F_1(\omega)F_2(\omega)$ (the curve 4) and the gap-weighted factor $\xi(\omega)=\omega\xi_1(\omega)\xi_2(\omega)$ (the curve 5) are also shown together with the filter passband ripple (the curve 3) in Fig. 2b. The function $\xi(\omega)$ for the metallization ratio $\eta=0,5$ was accounted for within a filter passband only. For convenience, the frequency characteristics are plotted versus a normalized frequency ω/ω_o where ω_o is a filter central frequency. An optimal solution was obtained using the McClellan's computer program [18] after a variable substitution (9), (10).



a) magnitude response



b) passband ripple

Fig. 2. SAW Filter Optimal Frequency Response

The design specifications are the following: the -3 dB passband width $\Delta\omega_o/\omega_o=25\%$, and the -3/-40 dB shape factor K=1,25. The IDT synchronism frequency is $\omega_\pi=2\omega_o$ that corresponds to IDT structures with splitted electrodes [30]. The IDT electrode numbers are $N_1=24$ and $N_2=200$ respectively.

As we can see from Fig. 2, the outofband attenuation is better than -60 dB and the passband peak-to-peak ripple is less than 0,1 dB.

The optimization was performed on the discrete frequency grid containing N_g=943 points with a discretization step $\Delta\omega_g=0,1\Delta\omega$, $\Delta\omega=2\omega_\pi/N_2$ being frequency sampling interval.

It took 110 iterations to obtain an optimal solution, the computation time being 32 minutes on a personal computer IBM PC/AT 286 with a math coprocessor.

2. SAW Filter Suboptimal Synthesis Technique

2.1. Suboptimal Approximation Problem Formulation and Solution

An optimal solution provides the best fit to a design target within a total approximation interval Ω_π and leads uniquely to an apodized IDT of minimal length. But the most serious drawback of an optimal synthesis is an excessive amount of computations. It is very desirable to find some way to reduce an optimized variables (OV) number n_2 and hence the computation time and the memory size needed. With this aim a suboptimal synthesis technique was elaborated which considerably reduces an OV number without significant sacrificing the approximation accuracy.

The key point of the suboptimal synthesis technique proposed is a splitting of the function $F_2(\omega)$ into two factors

$$F_2(\omega)=\bar{F}_2(\omega)\tilde{F}_2(\omega), \qquad (11)$$

with the function $\bar{F}_2(\omega)$ fixed and chosen a priori and the function $\tilde{F}_2(\omega)$ of the order $\tilde{n}_2 < n_2$ optimized within an approximation subinterval $\Omega\in\Omega_\pi$. Outside the subinterval Ω approximation accuracy depends mainly on the function $\bar{F}_2(\omega)$ which must secure a sufficient outofband attenuation. A synthesis technique of such a wideband window-type function $\bar{F}(\omega)$ will be discussed some later.

Rewriting the approximating function $F(\omega)$ in the form

$$F(\omega)=\xi(\omega)\bar{F}_1(\omega)\tilde{F}_2(\omega), \qquad (12)$$

where $\bar{F}_1(\omega)=F_1(\omega)\bar{F}_2(\omega)$, we note the function F(w) to be of the same structure (2), but with the function $F_1(\omega)$ replaced by the function $\bar{F}_1(\omega)$ and the optimized function $\tilde{F}_2(\omega)$ of the order $\tilde{n}_2$ replaced by the function $\tilde{F}_2(\omega)$ of the order $\tilde{n}_2 < n_2$. Therefore, a suboptimal approximation problem with the approximating function (12) can be converted to an auxiliary one and solved like an optimal one considered above, with an order $\tilde{n}_2$ of the optimized function $\tilde{F}_2(\omega)$ decreased.

The Fourier coefficients of the functions $F_2(\omega)$, $\bar{F}_2(\omega)$, and $\tilde{F}_2(\omega)$ are related via the convolution [1], that yields the relation

$$n_2=\bar{n}_2+\tilde{n}_2-1. \qquad (13)$$

The higher is the order $\bar{n}_2$ of the fixed function $\bar{F}_2(\omega)$, the lower is the order $\tilde{n}_2$ of the optimized function $\tilde{F}_2(\omega)$. Hence it appears desirable to increase the order $\bar{n}_2$ of the function $\bar{F}_2(\omega)$ until an approximation accuracy deteriorating. As a matter of fact a suboptimal solution is inevitably inferior if compared to an optimal one, but by a proper choice of the function $\bar{F}_2(\omega)$ the difference might be made negligible, with the low order function $\tilde{F}_2(\omega)$ optimized.

Thus the function $\bar{F}_2(\omega)$ performs here a twofold role: to decrease an OV number and to secure at the same time a sufficient approximation accuracy.

2.2. Window Function Construction

Let us suppose that the function $\bar{F}_2(\omega)$ to be constructed is completely defined by the Z-transform roots [1] $z_i=e^{j\varphi_i}$ at the points $\omega_i\notin\Omega_{op}$ allocated outside the optimization subinterval $\Omega_{opt}\in\Omega$ of the width $\Delta\omega_{opt}$, where $\omega_i=i\Delta\omega$, i=0,1,2, ..., with $\Delta\omega=2\omega_\pi/N_2$ being the frequency sampling interval (Fig.3). Such a window-type function $\bar{F}_2(\omega)$ is to force the frequency samples

$$F_2(\omega_i)=\bar{F}_2(\omega_i)\tilde{F}_2(\omega_i) \qquad (14)$$

to be zero outside the optimization subinterval Ω_{opt}, while the other frequency samples at the points $\omega_i\in\Omega_{opt}$ optimized. It has a bell-like magnitude response, its sidelobes rapidly decreasing outside Ω_{opt}.

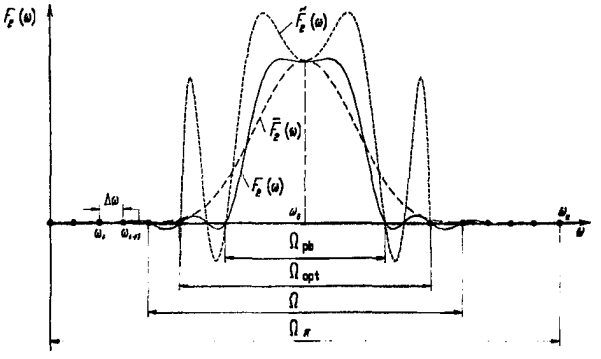
Using Z-transform properties we can write $\bar{F}_2(z)$ in the following form

$$\bar{F}_2(z)=\sum_{k=0}^{\bar{N}_2-1}\bar{A}_k z^k=\prod_{\omega_i\notin\Omega_{opt}}D_i(z)=\frac{1-z^{N_2}}{\prod_{\omega_i\in\Omega_{opt}}D_i(z)}, \qquad (15)$$

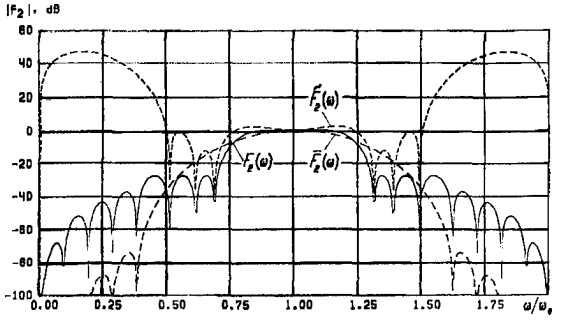
where

$$D_i(z)=\begin{cases}(z-z_i)(z-z_i^*), & z_i\neq\pm1;\\ (z-z_i), & z_i=+1 \text{ or } z_i=-1.\end{cases} \qquad (16)$$

The window function $\bar{F}_2(\varphi)$ is a Z-transform $\bar{F}_2(z)$ evaluated on the unit circle $z=e^{j\varphi}$.



a) linear scale magnitude response



b) Log magnitude response

Fig.3. Suboptimal Factorization of the Function F2(ω)

The coefficients $\bar{A}_k$ of the function

$$\bar{F}_2(\varphi)=\prod_{\omega_i \notin \Omega_{opt}} D_i(\varphi) \qquad (17)$$

of the order $\bar{N}_2$ could easily be calculated using the recurrent convolution of the root factors

$$D_i(\varphi)=\begin{cases} \cos\varphi-\cos\varphi_i, & \varphi_i \neq n\pi; \\ j\sin\varphi/2, & \varphi_i=2n\pi; \\ \cos\varphi/2, & \varphi_i=(2n+1)\pi, \end{cases} \qquad (18)$$

where $n=0,\pm1,\pm2,\ldots$. But usually there is no need to know the coefficients $\bar{A}_k$, and one can directly apply the formula (17) for calculations.

The wider is the bandwidth $\Delta\omega_{opt}$ of the optimization subinterval Ω_{opt}, the higher is a stopband attenuation of the function $\bar{F}_2(\omega)$ and the closer is a suboptimal solution to an optimal one. The number m of the frequency samples to be optimized is found from the simple relation

$$m = \frac{\Delta\omega_{opt}}{\Delta\omega}+1 = \frac{1}{2}N_2\frac{\Delta\omega_{opt}}{\omega_\pi}+1. \qquad (19)$$

For m > 10-15 one can use the following approximation for the OV number gain estimation

$$\frac{\tilde{n}_2}{n_2} \approx \frac{m}{n_2} \approx \frac{\Delta\omega_{opt}}{\omega_\pi}. \qquad (20)$$

In other words, the gain in the OV number is roughly proportional to the relative bandwidth $\Delta\omega_{opt}/\omega_\pi$ of the optimization subinterval Ω_{opt}.

The function $\bar{F}_2(\omega)$ may also be constructed by another way using a zero-extracting technique as opposed to a zero-inserting one above.

Indeed, an equivalent representation of the function $\bar{F}_2(\omega)$ follows from (15)

$$\bar{F}_2(\varphi) = \frac{\sin\frac{N_2}{2}\varphi}{\prod_{\omega_i \in \Omega_{opt}} D_i(\varphi)}, \qquad (21)$$

where an uncertainty at the singularity points $\varphi=\varphi_i$ can be avoided by using the Lopital's rule.

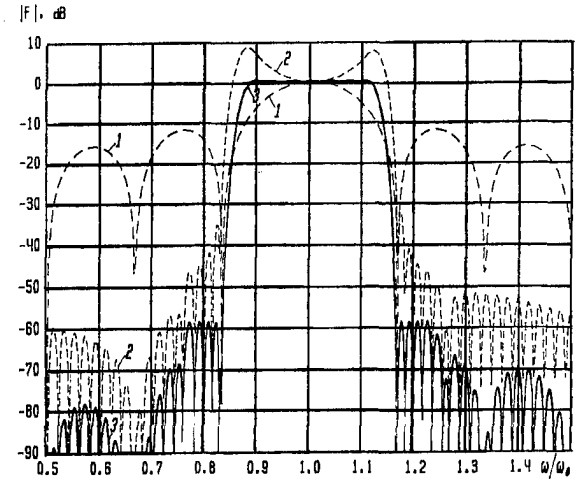
Here an amount of computations is proportional to $\tilde{n}_2$ instead of $\bar{n}_2$ in the formula (17). Hence the formula (21) is more preferable for $\tilde{n}_2 < \bar{n}_2$. Thus a rational choice of the function $\bar{F}_2(\omega)$ representation form allows somewhat to minimize an amount of computations in the function $\bar{F}_2(\omega)$ calculating.

As a rule of thumb we usually choose the optimization interval Ω_{opt} to be (6-10)$\Delta\omega$ wider than a filter passband Ω_{pb} (Fig.3). In turn an approximation subinterval Ω might be equal or (1-2)$\Delta\omega$ wider than optimization subinterval $\Omega_{opt} \in \Omega$. It is a rapid sidelobe attenuation of the window function on $\bar{F}_2(\omega)$ outside Ω_{opt} that secures an approximation accuracy to be not worse than that obtained within Ω, where the error function is minimized.

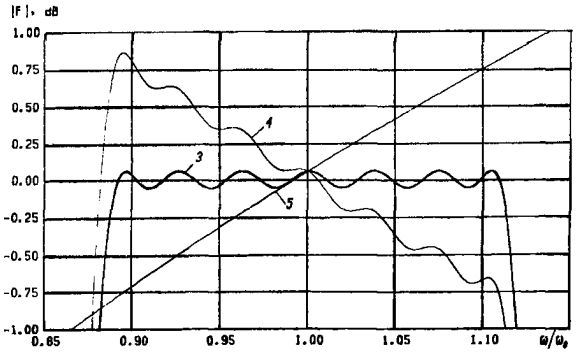
2.3. A Suboptimal Synthesis Design Example

For the comparison's sake it is convenient to use the same design example as for an optimal synthesis above. The suboptimal FR obtained is plotted in Fig.4. The initial specifications and the denominations are the same as for an optimal synthesis.

The suboptimal synthesis data are the following. The bandwidth of the optimization interval is $\Delta\omega_{opt}/\omega_{\pi}$=25%, with an approximation subinterval $\Omega \approx \Omega_{opt}$. The optimization was performed on the frequency grid containing N_g=209 points as opposed to the optimal synthesis, where the grid point number was as large as N_g=943.



a) magnitude response



b) passband ripple

Fig. 4. SAW Filter Suboptimal Frequency Response

The IDT electrode numbers are the same N_1=24 and N_2=200, but due to an a priori factorization of the function $F_2(\omega)$ the OV number was decreased from n_2=100 to $\tilde{n}_2$=25, with the number $m=\tilde{n}_2$ of the frequency samples optimized.

The stopband attenuation is -58,9 dB and the passband ripple is ≈0,11 dB. The detailed comparison of Fig. 2 and Fig. 4 shows that the solutions practically coincide within the subinterval Ω.

The difference in approximation accuracy is about 1,3 dB in a filter stopband and 0,015 dB in a filter passband that is quite negligible from the practical point of view.

It is the OV number gain $n_2/\tilde{n}_2$=25% in conjunction with a sufficient reduction of the frequency grid point number N_g that allows the computation time to be drastically reduced from 32 minutes to 15 seconds only, i.e. more than 125 times.

Conclusion

The optimal and suboptimal SAW filter design techniques have been considered above, both based on the McClellan's computer program [18].

Unfortunately, being of great theoretical importance, an optimal synthesis is rather impracticable one for a real-time design due to an excessive amount of computations.

While maintaining optimal synthesis generality and flexibility, the suboptimal synthesis technique allows to considerably reduce an OV number and hence the storage and the computation time nearly without sacrificing the approximation accuracy. Usually the difference between optimal and suboptimal approximations does not exceed 1-2 dB in a filter stopband and 0,01-0,05 dB within a passband that is more than acceptable for practical design purposes. Moreover, this slight discrepancy might easily be compensated by a small electrode number of an apodized IDT increasing if one wishes.

A feature of an suboptimal synthesis above is that an amount of computations depends mainly on the filter magnitude shape specifications but not on its central frequency. Consequently, for the most SAW filters the computation time is fairly small taking from some seconds to some minutes on a personal computer IBM PC/AT 286 with a math coprocessor. As a result a narrowband fast cut-off filter synthesis with an electrode number of several hundred and even thousand becomes possible due to dramatic OV number reducing.

It is the inherent efficiency of the Remez exchange algorithm in conjunction with a low order optimized function that makes the suboptimal synthesis technique proposed very attractive one for a SAW filter computer-aided design.

The design experience confirmed fast convergence, high computation speed, reliability, and flexibility of the suboptimal synthesis technique proposed, and good agreement between theory and experiment was obtained within the limits of the model and design constraints applied.

Acknowledgement

The author would like to thank Prof. V.M.Dashenkov for helpful discussions concerning the topics of the paper. Special thanks to Mr. V.S.Kiranov for his help in preparing the paper manuscript.

References

[1] L.R. Rabiner, B. Gold, Theory and Application of Digital Signal Processing. Englewood Cliffs, New-Jersey: Prentice Hall, Inc, 1975,ch.3.

[2] R.H. Tancrell, Analytic Design of Surface Wave Bandpass Filters, IEEE Transactions on Sonics and Ultrasonics, Vol. SU-21, No. 1, pp. 12-22, January 1974.

[3] C.D. Bishop, D.C. Malocha, Non-Iterative Design of SAW Bandpass Filters, IEEE Ultrasonics Symposium Proceedings, 1984, pp. 18-21.

[4] D.C. Malocha, C.D. Bishop, The Classical Truncated Cosine Series Functions with Applications to SAW Filters, IEEE Transactions Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-34, No. 1, pp. 75-85, January 1987.

[5] A.R. Reddy, Design of SAW Bandpass Filters Using New Window Functions, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-35, No. 1, pp. 50-56, January 1988.

[6] S.M. Richie, B.P. Abbott, D.C. Malocha, Description and Development of a SAW Filter CAD System, IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-36, No. 2, pp. 456-466, February 1988.

[7] K. Hohkava, S. Yoshikawa, SAW Filter Design Using Linear Programming Technique, Revue of the Electrical Communication Laboratories, Vol. 26, No. 5-6, pp. 755-766, May-June 1978.

[8] C.C.W. Ruppel, E. Ehrmann-Falkenau, H.R. Stocker, R. Velth, Optimum Design of SAW-Filters by Linear Programming, IEEE Ultrasonics Symposium, pp. 23-26, 1983.

[9] C.C.W. Ruppel, E. Ehrmann-Falkenau, H.R. Stocker, Generalized Design of Tap Weighted SAW-Filters by Optimization Techniques, Proceedings of the International Symposium on Circuits and Systems, pp. 1137-1140, 1985.

[10] V.M. Dashenkov, A.S. Rukhlenko, SAW Filter Synthesis Using ES Computer Mathematical Program Library, Radiotechnika i Elektronika, Minsk, Vyp. 13, pp.18-20, 1984 (in Russian).

[11] P.M. Jordan, B. Lewis, A Tolerance-Related Optimized Synthesis Scheme for the Design of of SAW Bandpass Filters with Arbitrary Amplitude and Phase Characteristics, IEEE Ultrasonics Symposium Proceedings, 1978, pp. 715-719.

[12] M. Morimoto, Y. Kobayashi, H. Hibino, An Optimal SAW Filter Design Using FIR Design Techniques, IEEE Ultrasonics Symposium Proceedings, 1980, pp. 298-301.

[13] V.M. Dashenkov, A.S. Rukhlenko, Optimal Design of Surface Acoustic Wave Filters, Izvestija Vuzov, Radioelektronika, Vol. 27, No. 7, pp. 76-78, July 1984 (in Russian).

[14] V.M. Dashenkov, A.S. Rukhlenko, Suboptimal Synthesis Technique of Surface Acoustic Wave Filters, Izvestija Vuzov, Radioelektronika, Vol. 28, No. 9, pp. 92-94, September 1985 (in Russian).

[15] P.M. Smith, C.K. Campbell, The Design of SAW Linear Phase Filters Using the Remez Exchange Algorithm, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-33, No. 3, pp. 318-323, May 1986.

[16] C.C.W. Ruppel, A.A. Sachs, F.J. Seifert, A Review of Optimization Algorithms for the Design of SAW Transducers, IEEE Ultrasonics Symposium Proceedings, 1991, pp. 1-11.

[17] F. Braun, A. Kinzl, H. Rothenbuhler, Chebyshev Approximation of Arbitrary Frequency Response for Nonrecursive Digital Filters with Li-Linear Phase, Electronics Letters, Vol. 9, No. 21, pp. 507-509, October 1973.

[18] J.H. McClellan, T.W. Parks, L.R. Rabiner, A Computer Program for Designing Optimum FIR Linear Phase Digital Filters, IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No. 6, pp. 506-526, December 1973.

[19] L.R. Rabiner, J.H. McClellan, T.W. Parks, FIR Digital Filter Design Techniques Using Weighted Chebyshev Approximation, Proc. IEEE, Vol. 63, No. 4, pp.595-610, April 1975.

[20] F. Bonzanigo, Some Improvements to the Design Programs for Equiripple FIR Filters, IEEE International Conference on Acoustics, Speech, and Signal Processing, 1982, pp. 274-277.

[21] A. Antoniou, Accelerated Procedure for the Design of Equiripple Nonrecursive Digital Filters, IEE Proceedings, Vol. 129, Pt. G, No.1, pp. 1-8, February 1982.

[22] H. Matthews, Editor, Surface Wave Filters: Design, Construction, and Use. New York-London-Sydney-Toronto: A Wiley - Interscience Publication, 1977, ch. 3.

[23] D.P. Morgan, Surface-Wave Devices for Signal Processing. Amsterdam-Oxford-New York-Tokyo: Elsevier, Tokyo, 1985, ch. 4, 5.

[24] B. Lewis, P.M. Jordan, R.F. Milsom, D.P. Morgan, Charge and Field Superposition Methods for Analysis of Generalized SAW Interdigital Transducers, IEEE Ultrasonics Symposium Proceedings, 1978, pp. 709-714.

[25] S. Datta, B.J. Hunsinger, Element Factor for Periodic Transducers, IEEE Transactions on Sonics and Ultrasonics, Vol. SU-27, No.1, pp. 42-44, January 1980.

[26] R.C. Peach, A General Approach to the Electrostatic Problem of the SAW Interdigital Transducer, IEEE Transactions on Sonics and Ultrasonics, Vol. SU-28, No. 2, pp. 96-105, March 1981.

[27] E. Ehrmann-Falkenau, H. R. Stocker, C.C.W.Ruppel, W.R. Mader, A Design for SAW Filters with Multistrip Coupler, IEEE Ultrasonics Symposium Proceedings, 1984, pp. 13-17.

[28] A.S. Rukhlenko, SAW Filter Factorizational Synthesis, Thesisy Dokladov XIII Vsesojuznoj Conferencii po Akustoelektronike i Kvantovoj Akustike, Chernovtsy, 1986, Pt. 2, p. 157 (in Russian).

[29] V. M. Dashenkov, A. S. Rukhlenko, I. G. Yur'evich, Synthesis of Minimum-Phase and Quasiminimum-Phase SAW filters, Akustoelektronnye Ustrojstva Obrabotki Informacii, Materialy Conferencii, Cherkassy, 1988, pp. 72-73 (in Russian).

[30] H. Engan, Surface Acoustic Wave Multielectrode Transducers, IEEE Transactions on Sonics and Ultrasonics, Vol. SU-22, No. 6, pp. 395-401, November, 1975.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ANALYSIS OF LEAKY SURFACE ACOUSTIC WAVES ON $LiTaO_3$ SUBSTRATE

H. Meier, P. Russer
Lehrstuhl für Hochfrequenztechnik der Technischen Universität München
Arcisstraße 21, 8000 München 2, Germany

Abstract – We have investigated leaky surface acoustic waves (LSAW) propagating in arbitrary directions of rotated Y-cuts on $LiTaO_3$ substrates. The inherent physical behavior of LSAW, consisting of coupled modes involving terms decaying into the substrate and a term representing a radiating bulk wave, is fully explained. The high propagation velocity of LSAW, and hence the fundamental mode operation, may lead to a further advance of SAW technology in commercial GHz-systems, especially in low-loss SAW filter applications. But nevertheless, to further reduce the insertion loss, all relevant effects influencing the device performance, e.g. diffraction spreading losses, and spurious bulk waves, have to be considered as well. Thus, we have analyzed the well-known Y-rotated $LiTaO_3$ substrate with wave propagation in the crystalline X-direction, whereas also deviations from this specific propagation direction are taken into account. Phase velocity and attenuation constant are computed for different rotation angles, as well as for different propagation directions and the measured results are shown to compare with those obtained from the theoretical analysis.

INTRODUCTION

High electromechanical coupling and low temperature sensitivity are generally required in piezoelectric materials for their practical use. Ferroelectric crystals, such as lithium niobate ($LiNbO_3$) and lithium tantalate ($LiTaO_3$), have received much attention because of their large coupling coefficients. Extensive investigations have been done to find optimum crystal cuts and propagation directions in order to achieve effective coupling on these substrates. It has been shown [1] that, aside from the well known SAW, another type of surface wave can exist on rotated Y-cuts of $LiNbO_3$ substrate for certain rotation angles. This wave, called leaky surface acoustic wave (LSAW), consists of coupled modes involving terms decaying into the substrate and a term representing a bulk wave radiating into the solid [2].

Consequently, the behavior of LSAW on rotated Y-cuts of $LiTaO_3$ has been analyzed and it has been stated that a variety of acoustic modes exist, depending strongly on the electrical boundary conditions on the substrate surface[3]. Due to the high propagation velocity, and hence the fundamental mode operation, LSAW devices may play a key role for the break-through of SAW technology in commercial GHz-systems [4][5][6].

In order to encourage this process we have investigated the properties of LSAW, propagating in the crystalline X-direction of Y-rotated $LiTaO_3$ substrate, whereas also deviations from the X-axis propagation direction are taken into account. Our analysis is based only on the metallized surface case because, for this boundary condition, the LSAW is the dominating wave [3]. The mass- and stress-loading effect, caused by an isotropic and metallic film with finite thickness, is exactly taken into consideration by analyzing the layered configuration. Furthermore, also thin film induced effects, e.g., partial oxidation of an aluminum film due to the fabrication process, can be investigated[7][8].

THEORY

The elastic wave propagation in a piezoelectric and anisotropic medium is governed by the well-known set of four coupled wave equations for the three components of elastic displacement u_i and electric potential ϕ. The coordinate system to be used is illustrated in Fig. 1, from which it can be seen that the $x_3 = 0$ plane is the interface between the layer and the substrate.

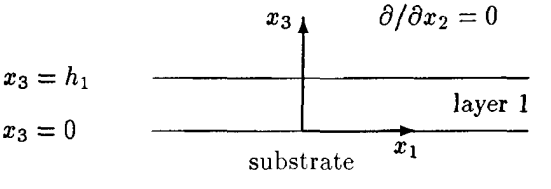


Figure 1: Coordinate system

Supposed that propagation is along the x_1-axis of a

0-7803-0476-4/92 $3.00 © 1992 IEEE

cartesian coordinate system (x_1, x_2, x_3) with the substrate being in the lower halfspace $x_3 < 0$. Then the wave solution in the substrate can be assumed to be linear combinations of partial waves as follows:

$$u_i = \sum_{n=1}^{4} W_n A_{in} exp[jk(\alpha_n x_3 + (1 + j\delta)x_1 - vt)] \quad (1)$$

$$\phi = \sum_{n=1}^{4} W_n A_{4n} exp[jk(\alpha_n x_3 + (1 + j\delta)x_1 - vt)] \quad (2)$$

where W_n represents the weighting factors and the A_{in} are the normalized partial amplitudes. The acoustic wave number is k, α_n are the decay constants in the direction of depth, and δ is the attenuation constant in the propagation direction regarded to LSAW solutions. Because we are only seeking for LSAW solutions, the layer is required to be "thin", i.e., wave propagation is in principle in the substrate with a small perturbation owing to the film.

Substitution of Eqs.(1) and (2) into the coupled wave equations for the substrate gives a polynomial equation of eighth order for the decay constants α_n with complex polynomial coefficients. With common surface waves these coefficients are real and in general four lower halfplane roots will be selected. Note that with complex coefficients the roots are no longer restricted to be exactly conjugate pairs. Especially for LSAW one of the terms α_n must represent the bulk term with a negative real part corresponding to a propagation vector tilted down into the substrate and a relatively small positive imaginary part [2]. Furthermore, the remaining three α_n should be chosen in the usual way in order to have the energy density concentrated within a few wavelengths beneath the surface.

The layer is assumed to be a highly conducting metallic film. Thus, the simplified equations of elastic motions for the isotropic case can be used. Due to the finite layer thickness, a linear combination of six partial waves for the three particle displacements u_i in the layer is retained, having the same form as in Eqs.(1).

The elastic displacement components and the normal components of stress must be continuous across the interface plane $x_3 = 0$, the normal stress components must vanish at the plane $x_3 = h_1$ and the electric potential is forced to be zero at the interface $x_3 = 0$. Finally, the unknown weighting factors W_n may be calculated by substituting the assumed partial wave solutions in the substrate and the metallic film into the ten boundary condition equations. In order to have nontrivial solutions, successive values of the phase velocity v and the attenuation constant δ are chosen until the boundary condition determinant is made equal to zero within some numerical accuracy.

Results and Discussion

1. Ideal metallized surface

There is still some uncertainty whether a rotation angle of 36° will be optimal particularly when a minimal attenuation constant in the propagation direction is needed. Thus, we have calculated the LSAW parameters for different rotation angles (Euler-angle μ). Generally it can be concluded that, only for a certain range of μ (25° - 43°), a LSAW solution exists which completely fulfills the boundary conditions. The absolute value of the weighting factor W_1, representing the bulk term in the partial wave solution, also varies over the angular range of interest. But W_1 is always much less than the other three weighting coefficients representing the usual surface wave solution. For different rotation angles μ^* the decay constant δ as well as the weighting factor W_1 are minimal, showing the close connection between these two factors. But nevertheless, this specific angles μ^* are still a function of the material parameters used for the $LiTaO_3$ substrate. We have calculated the field amplitudes and the energy flow characteristics for three rotation angles, yielding LSAW solutions with different attenuation constants. The rotation angles and the accompanied calculated phase velocities and attenuation constants are listed in Table I.

Table I:

rotation angle μ [°]	phase velocity v [m/s]	attenuation δ [dB/λ]
30.0	4094.2	$2.2 * 10^{-2}$
36.0	4108.5	$2.7 * 10^{-4}$
39.0	4112.4	$2.6 * 10^{-3}$

As can be seen in Fig. 2a,c,e the LSAW solutions behave like surface waves with polarization predominantly in the surface plane, similarly to Bleustein-Gulyaev waves, but with a penetration depth comparable to normal surface waves. On the other hand, the relatively small bulk wave part has a quasi saggital-plane polarization with slightly increasing u_3-component, corresponding to an increasing weighting factor W_1. This small bulk wave term in the partial wave solution has a phase velocity vector tilted towards the substrate at an angle $\varphi = 28.5°$ referred to the substrate surface for 36° YX $LiTaO_3$. Using simple phase matching arguments and the slowness curves of the bulk waves on 36° YX $LiTaO_3$ (Fig. 3), it can be concluded that the

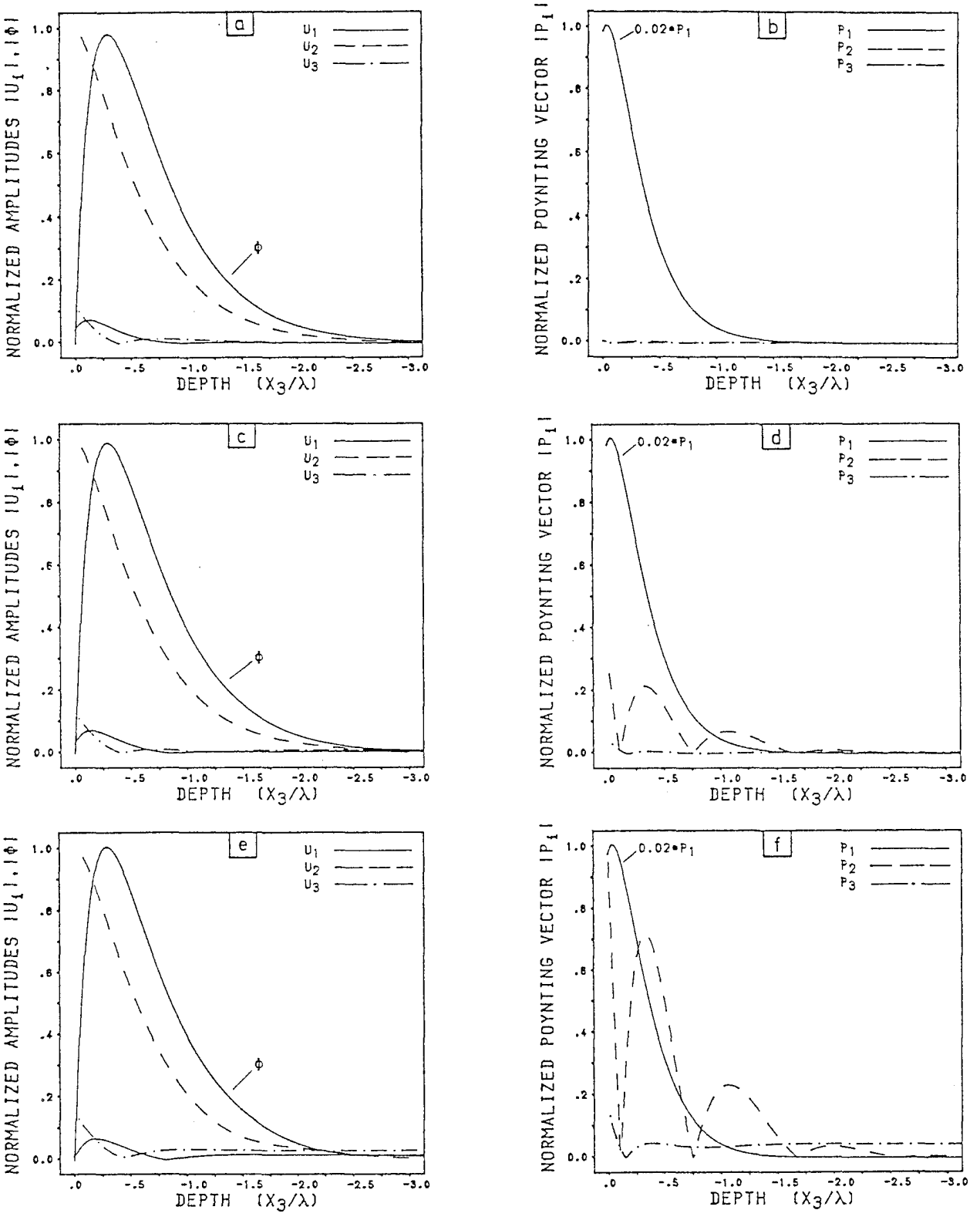


Figure 2: Field amplitudes and energy flow versus depth x_3/λ for metallized surface case and different rotation angles μ; Fig.2 a,b $\mu = 36.0°$; Fig.2 c,d $\mu = 30.0°$; Fig.2 e,f $\mu = 39.0°$;

radiating bulk wave is a slow shear wave (SSW) having a quasi-saggital plane polarization.

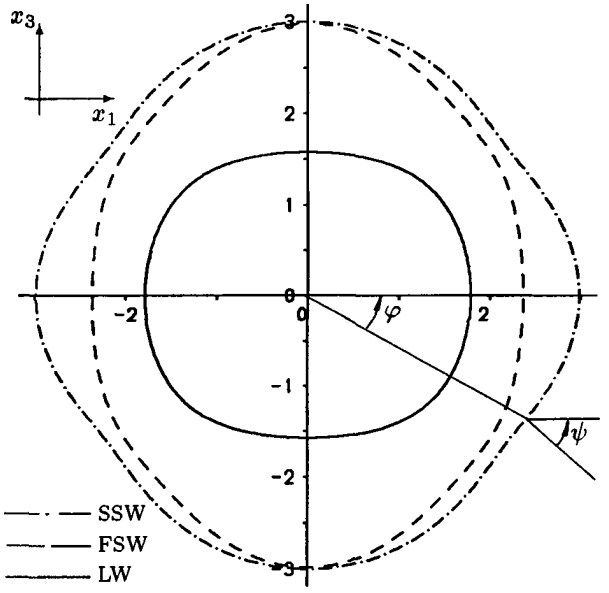


Figure 3: Slowness curves of the bulk waves on 36YX $LiTaO_3$; $S = 10^4/V$

With regard to power flow characteristics, it is common to define a complex acoustical Poynting vector **P**.

$$P_i = -\frac{1}{2} Re\{ j\omega (T_{ij} u_j^* - \phi D_i^*)\} \qquad (3)$$

The P_i are the time averaged real power flow components across an area normal to x_i (i=1,2,3). T_{ij} is the strain tensor component and D_i the electric displacement component. For all normal surface waves there are no components perpendicular to the surface.
Referring to LSAW, energy flow must be defined with caution according to the energy leakage into the substrate. Taking this into account, we only calculate the real parts of the complex acoustical Poynting vector as a function of the penetration depth. The normalized real components P_i are plotted in Fig. 2, with the dominating component P_1 being multiplied by a factor 1/50, in order to emphasize that P_3 is responsible for the leaky behavior. As a result, Fig. 2b,d,f shows that with increasing attenuation, the energy leakage will also become greater. Fig. 4 gives the power flow angle ψ into the substrate. Near the surface, energy flow is large and essentially parallel to the surface as for common surface waves. However, for greater depths the bulk wave term is the only one left in the solution and the energy flow is that of the SSW with an angle of $\psi \approx 38°$, which is still constant even for greater penetration depths $x_3/\lambda > 300$ [2]. This result agrees well with calculations of bulk wave radiation by a line source on 36° YX $LiTaO_3$. Additionally, there is also a strongly radiated fast shear wave (FSW) with a power flow angle nearly in the same direction as that obtained for the SSW [9][10].

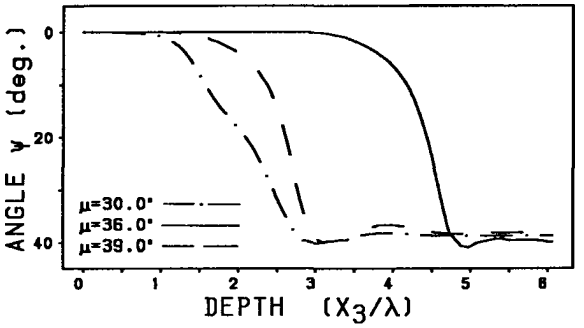


Figure 4: Power flow angle ψ

2. Real metallized surface
For a finite layer thickness the phase velocity will become dispersive. Owing to this fact, the phase velocity vector φ of the radiating SSW will also change slowly, in order to fulfill the phase matching condition. Field amplitudes are affected by the finite film height, too. But nevertheless, the well-known SH-type polarization and the low penetration depth are still retained.

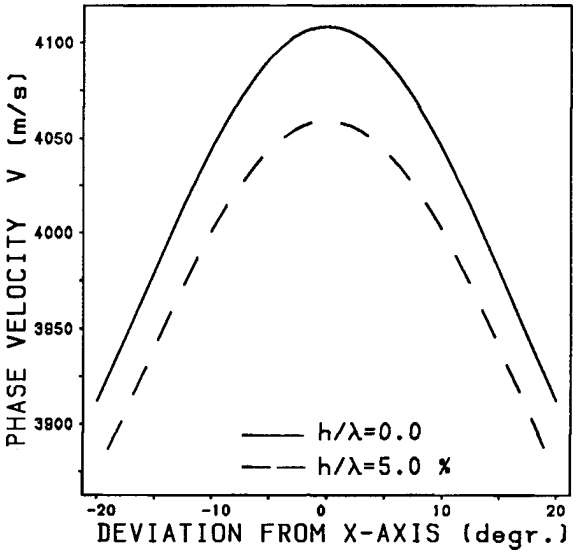


Figure 5: Phase velocities for different propagation directions and metallization heights; 36YX $LiTaO_3$

To study the influence of a thin metallic layer, we have investigated a pure aluminum film on 36°YX

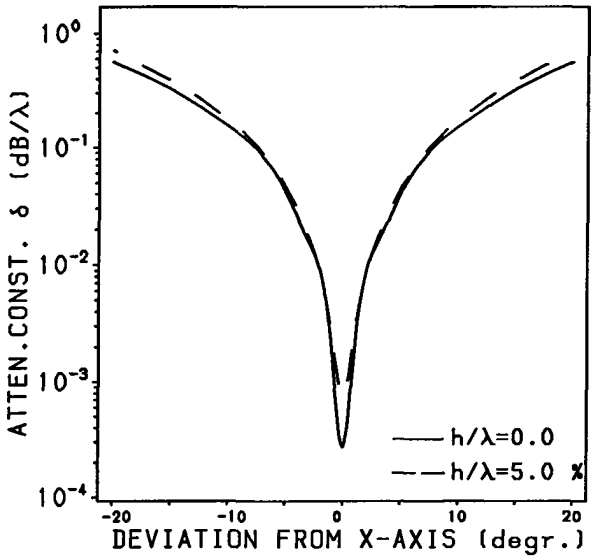


Figure 6: Attenuation for different propagation directions and metallization heights; 36YX $LiTaO_3$

$LiTaO_3$ substrate. In Fig. 5 and Fig. 6 phase velocity and attenuation constant of LSAW are plotted as a function of the deviation angle, whereas also two different metallization heights ($h/\lambda = 0$ ideal metallized surface; $h/\lambda = 5\,\%$ pure aluminum film) are taken into consideration. The inherent attenuation of the LSAW is seen to increase markedly for small deviations from the X-axis propagation direction (Fig. 6). Furthermore, the directions of power flow and phase velocity are no longer parallel, i.e., beam spreading losses occur.
Experimental values were determined by impulse response measurements using simple delay line configurations with different aluminum films (h/λ : 0.5–8 %) deposited on 36 °YX $LiTaO_3$. Group velocity and attenuation constant have been measured with the launching transducer orientated in the crystalline X-direction. The receiving interdigital transducers have been circularly arranged in an appropriate distance to measure the far field. Thus, also diffraction spreading effects, caused by the finite aperture of the transducers, are experimentally determined. A comparison of calculated and experimental results is illustrated in Fig. 7. The influence of an aluminum oxide film, which may be formed during the fabrication process, is clearly seen, i.e., for small layer thicknesses phase velocity is "stiffened" and for greater film thicknesses the mass loading effect is dominant. All calculations have been done by using improved material constants for $LiTaO_3$ [11]. Measuring group velocities and attenuation constants for small deviations ($-5\,°to\,5\,°$) from the crystalline X-axis was strongly affected by radiated and reflected spurious bulk wave signals. But nevertheless, the experimentally determined values are in good aggreement with those obtained from the theoretical analysis. The influence of imprecise material parameters, of both the substrate and the metallic film, as well as the accurate determination of group velocity and attenuation constant, is still a matter of further investigation.

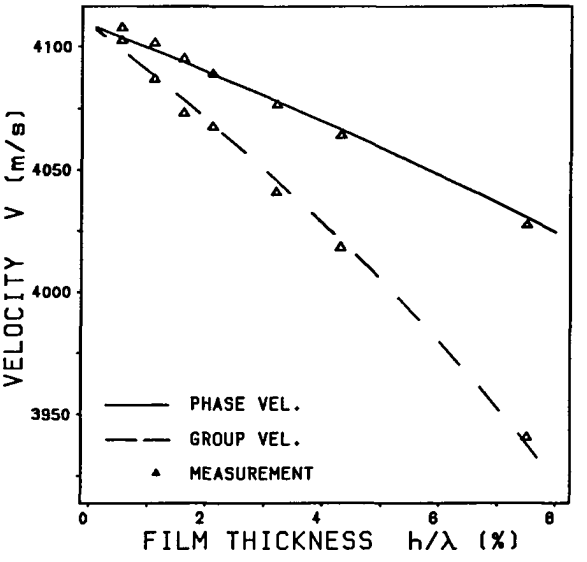


Figure 7: Dispersion curves for 36YX $LiTaO_3$

Conclusion

We have investigated, both theoretically and experimentally, the properties of LSAW propagating in the crystalline X-direction on Y-rotated $LiTaO_3$ substrate, whereas also small deviations from this specific propagation direction are taken into account. The effect caused by a thin metallic film was considered by analyzing the layered configuration. The inherent physical behavior of LSAW was fully explained and field amplitudes, as well as power flow components, have been computed. Thus, we were able to determine the radiating bulk wave behavior.
Group velocity and attenuation constant have been measured for small deviations from the X-axis propagation direction. It could be shown that the measured results agree well with those obtained from the theoretical analysis.

Acknowledgement

This work was supported by the Corporate Research and Development Center, Siemens AG München, Germany.

REFERENCES

[1] K.Yamanouchi,K.Shibayama, *Propagation and Amplification of Rayleigh Waves and Piezoelectric Leaky Surface Waves in $LiNbO_3$* , Journal of Applied Physics, Vol.43, pp. 856-862, 1972

[2] T.C.Lim,G.W.Farnell, *Character of Pseudo Surface Waves on Anisotropic Crystals*, Journal of the Acoustical Society of America, Vol.45, pp. 845-851, 1969

[3] K.Hashimoto,M.Yamanouchi,H.Kogo,*Experimental Verification of SSBW and Leaky SAW Propagating on Rotated Y-Cuts of $LiNbO_3$ and $LiTaO_3$* , IEEE Ultrasonics Symposium Proc., pp. 345-349, 1983

[4] H.Meier,R.Weigel,K.Anemogiannis,P.Russer, *SAW Microstrip Antenna Duplexer for Radio Communication Transceivers in the GHz Range*, 21st European Microwave Conference Proc., pp. 398-403, 1991

[5] K.Anemogiannis,S.Berek,H.Zottl, *GHz Low Loss SAW Filters Employing Standard Fabrication Process*, IEEE Ultrasonics Symposium Proc., pp. 129-133, 1990

[6] W.Ruile,H.Meier,P.Russer, *Field Theoretical Analysis for Improved CAD Modeling of SAW Filters in Mobile Communication Systems*, to be presented at ISSSE Paris, France, September 1-4, 1992

[7] H.Meier,P.Russer, *Propagation Characteristics of Leaky Surface Acoustic Waves for Two Thin Metal Layers on $LiTaO_3$* , presented at IEEE Ultrasonics Symposium, Lake Buena Vista, USA, Dezember 8-11, 1991

[8] B.K.Sinha,S.Locke, *Thin Film Induced Effects on SAW Devices* , IEEE Ultrasonics Symposium Proc., pp. 289-294, 1987

[9] Y.Zhang, *Surface Generated Bulk Acoustic Waves by Interdigital Transducers in Strongly Coupling Materials*, URSI International Symposium Signals, Systems and Electronics Conference Proc., pp.458-461, 1989

[10] R.Dill,C.Kappacher, *Analysis of the Complete Spectrum of Acoustic Waves on 36°rot YX $LiTaO_3$ by Means of Wigner Distribution*, IEEE Ultrasonics Symposium Proc., pp. 139-143, 1990

[11] G.Kovacs,M.Anhorn,H.E.Engan,G.Visintini,C.C.W. Ruppel, *Improved Material Constants for $LiNbO_3$ and $LiTaO_3$* , IEEE Ultrasonics Symposium Proc., pp. 435-438, 1990

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PARAMETER EXTRACTION OF SAW RESONATOR EQUIVALENT CIRCUIT PARAMETERS AND PACKAGE PARASITICS

MARK CAVIN, NANCY EISENHAUER AND DONALD C. MALOCHA
Electrical and Computer Engineering Department, University of Central Florida
Orlando, FL 32816-2450

Abstract

This paper will focus on surface acoustic wave (SAW) resonator and package parasitic parameter extraction. A previous approach used multiple applications of the EIA-512 standard to extract several of the package parasitics. The results were good over a reasonable bandwidth near resonance, but diverged over the wide bandwidths.[1] This was primarily due to the effects of mutual inductance between the input and output ports in two port SAW resonators. This paper will present a new approach which extracts the effects of the mutual inductance and transducer mutual coupling capacitance. The approach first isolates the package parasitics, and then extracts the actual resonator parameters. Two port S-parameter data is taken over the narrowband near resonance and wideband, normally from 100 KHz to 3 GHz. The approach and analysis then uses a series of network and parameter transformations in order to extract all the required circuit parameters from the measured data.

The theoretical approach and analysis will be presented for the SAW resonator parameter extraction. A number of commercially available devices will be used to show model verification, both narrowband as well as broadband. Data comparison of several approaches to the parameter extraction will also be presented.

I. Introduction

The motivation for this work is to develop a high frequency model for extraction of SAW two port resonator parameters. Because SAW resonators typically operate in the frequency band between 100 MHz and 2 GHz, the effects of packaging parasitics can not be ignored when attempting to find the resonator's equivalent electrical circuit model. The typical resonator model used for bulk and one port SAW resonators is shown in Figure 1.[2,3] This models the motional arm as a series R-L-C in parallel with a static capacitor. The static capacitance is due to the electrode plates for BAW resonators or due to

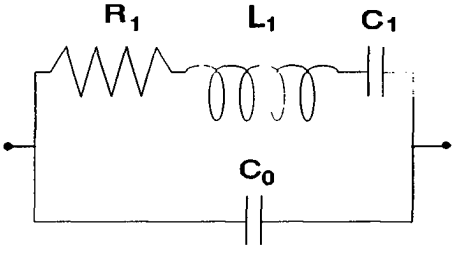


Figure 1. BAW and SAW one port resonator model.

the transducer for the SAW resonator. The typical equivalent circuit model for a two port SAW resonator is shown in Figure 2. The motional arm is represented by the series R-L-C network and the two shunt capacitors represents the input/output transducers. The capacitor, C_{12}, which represents the very small coupling capacitance between transducers, is sometimes included in the model. These models have been found to be accurate over narrowbands near the resonator center frequency.

At higher frequencies it is necessary to de-embed the package parasitics in order to obtain the actual resonator parameters. A packaged SAW resonator will typically have, on both the input and output ports, a parallel capacitance associated with the header pin feedthrough, followed by a series lead inductance due to the pin

0-7803-0476-4/92 $3.00 © 1992 IEEE

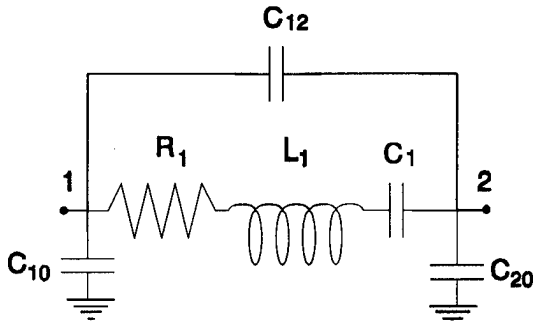


Figure 2. Two port SAW resonator model without package parasitics.

and bond wire inductance, and a series transducer capacitance to ground. In addition, there will be a small coupling capacitance between the input and output transducers and mutual inductance between the input and output inductances. The equivalent circuit model based on a physically packaged, single mode, two port SAW resonator is shown in Figure 3. This is the equivalent circuit model which will be analyzed and presented.

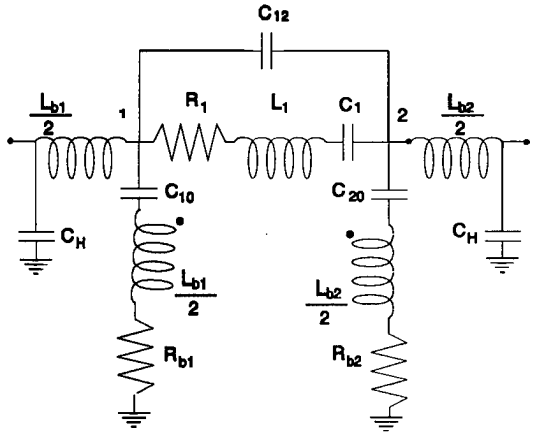


Figure 3. Equivalent circuit model for a typical packaged two port SAW resonator.

A number of different two port SAW resonators were tested. The resonators were provided by RF Monolithics, Dallas and Sawtek, Orlando, from their standard manufacturing line. Different devices were tested at various center frequencies. The basic equivalent circuit model appeared valid for all devices tested. Only a sample of the results are presented here which are believed to be representative of the general case. Devices were tested using a calibrated HP 8753 automatic network analyzer to obtain the frequency dependent S-parameters.[4]

II. Model Analysis

It is now necessary to determine the actual model parameters. It should be noted that the frequency dependent conductance and capacitance of the SAW transducers have been ignored in order to obtain a lumped element equivalent circuit model. A lumped element model has several advantages for oscillator design and for inclusion in other circuit analysis computer aided design (CAD) programs, as mentioned previously [5].

The first element to be determined is the header pin capacitance, C_H. It is assumed that this element can be measured accurately using an impedance analyzer or network analyzer. It has been experimentally found that this capacitance is fairly constant, from header to header, for a given type and manufacturer. This element is assumed a known, given constant value which has been accurately measured.

The bond wire inductance in each port has been split in two. This represents the bond wire and pin inductance to the transducer and then the bond wire and pin inductance to the ground pin. The finite lead resistance has been included as a single resistance R_B. Far from the SAW resonance, the motional elements form an open circuit and can be ignored. Further, knowing that C_{12} is a very small value, usually on the order of femto-farads, the lead inductance can be combined at each port. The simplified circuit now becomes a simple pi-network composed of resistors, inductors and capacitors, as shown in Figure 4.

A. Package Parameter Extraction

First the transducer and circuit elements associated with the package will be extracted. It is assumed that the frequency dependent data will be obtained suitably far removed from the SAW resonance such that the SAW motional elements can be ignored, as shown in Figure 4. One problem which arises in the analysis is the mutual inductance term. The network described in Figure 4 is based on physically identifiable elements due to the device and package, however, the mutual

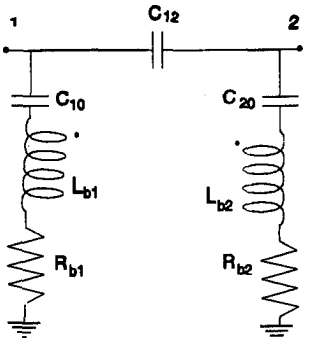


Figure 4. Simplified equivalent circuit model for two port packaged SAW resonator at frequencies removed from all SAW resonances.

inductance term does not easily lend itself for analysis. The approach used is to convert the Pi-network to a T-network, introduce the mutual inductance, and then extract the Pi-network equivalent circuit parameters.

The equivalent Pi and T network block diagrams are shown in Figure 5. The elements of the Pi and T networks can be written as

$$Z_A + Z_C = \frac{Y_2+Y_3}{\overline{Y^2}}$$

$$Z_B + Z_C = \frac{Y_1+Y_2}{\overline{Y^2}}$$

$$Z_C = \frac{Y_2}{\overline{Y^2}}$$

where $\overline{Y^2} = Y_1Y_2 + Y_1Y_3 + Y_2Y_3$. At this point the mutual inductance term has not yet been added. Let $Z_C \rightarrow Z_C + j\omega M$ where M is the mutual inductance term. This couples the input and output ports via the mutual inductance and is straight forward to analyze. Further, since $Y_2 = j\omega C_{12}$ and C_{12} is very small, then

$$\overline{Y^2} \approx Y_1Y_3$$

The equations can be further simplified to

$$Z_A = 1/Y_1 - j\omega M = Z_1 - j\omega M$$

$$Z_B = 1/Y_3 - j\omega M = Z_3 - j\omega M$$

$$Z_c = Y_2/Y_1Y_3 + j\omega M = Z_1Z_3/Z_2 + j\omega M$$

The form of the impedances of the T-element network for Z_A and Z_B are simple series R-L-C with the addition of the mutual inductance term and these are directly related to the series R-L-C network arms in the Pi network of Figure 5. These impedances can be written as

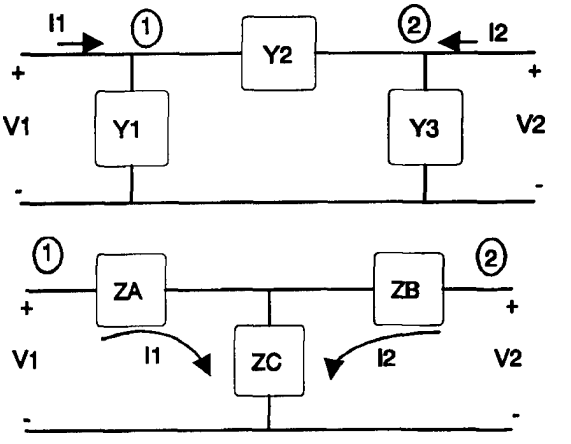


Figure 5. Pi and T network block

$$Z_A = R_{b1} + j\omega(L_{b1} - M) - 1/j\omega C_{10}$$

$$Z_B = R_{b2} + j\omega(L_{b2} - M) - 1/j\omega C_{20}$$

where $R_A=R_{b1}$, $L_A=L_{b1}$-M and $C_A=C_{10}$ and similar expressions for Z_B. The reactance of Z_C can be expanded, ignoring small and third order frequency terms as

$$Im[Z_C] = \omega\{C_{12}R_1R_2 + C_{12}(L_1/C_3 + L_3/C_1) + M\} - C_{12}/(\omega C_1C_3)$$

where the first term corresponds to L_C and the second term corresponds to C_C. Therefore, the T-network elements look like series R-L-C networks. Further, it has been generally found that the two SAW resonator ports are nearly identical, which simplifies the equations. For identical ports 1 and 2,

$Y_1 = Y_2$ and $Z_A = Z_b$.

Solving the Pi-network elements in terms of the T-network elements yields

$R_b = R_A$, $C_{10} = C_A$, $C_{12} = C_A^2/C_C$, $L_b = (L_A+L_C)/2$,

and $M = L_C - 2(L_A+L_C)(C_1/C_3)$.

Once the package elements have been determined, each element can be de-embeded from the Y or Z parameters until the only remaining elements are the SAW motional parameters.

B. Implementation

A computer program was written to perform the resonator model extraction using the methods developed in the preceding sections. Inputs to the program are wideband and narrowband two port S-parameter data files. The program first calculates the parasitic model from

the wideband data and then subtracts these effects from the narrowband data before extracting resonant arm values. Table I lists the process steps used to extract the packaged SAW resonator parameters. Once the package elements are extracted from the data, the conventional technique, similar to EIA-512, can be used to extract the actual SAW device resonator motional arm parameters.[1]

- Convert narrowband S-parameters to Y-parameters.
- Subtract admittance of C_H from Y_{11},Y_{22}.
- Convert Y-parameters to Z-parameters.
- Find Z_A, Z_B, and Z_C from package resonances.
- Subtract impedance of L_b and R_b from Z_{11},Z_{22}.
- Subtract impedance of M from Z_{12},Z_{21}.
- Convert Z-parameters to Y-parameters.
- Subtract admittance of C_{12} from $-Y_{12}$,$-Y_{21}$.

Table I. Extraction Procedure of Package Elements

Having found and de-embeded the package elements, the modified Y-parameters describe the SAW resonator motional elements and can be suitably extraccted.

III. Example of a 905 MHz SAW Resonator

A 905 MHz SAW resonator was modeled using the parameter extraction software and results of ten measurements of a sample device are shown in Table II. The resonant arm values were obtained after the average parasitic network from the ten runs was subtracted. As can be seen the percent errors in the extracted values are very low, and the resonant frequency is extracted with a standard deviation to mean ratio of approximately 3 parts per million. No particular care was taken in handling the devices or controlling temperature.

Figures 6 through 12 show comparisons between the model and the resonator measurement. Figure 6 shows wideband S_{12} comparisons and improvements of the developed model over simpler models

#	Ct pf	C12 fF	Lb nH	Mt nH	R ohm
1	0.847	10.9	1.97	0.864	4.7
2	0.849	11.1	1.97	0.879	13.2
3	0.875	11.1	1.91	0.837	12.7
4	0.866	11.1	1.93	0.849	16.9
5	0.844	10.7	1.98	0.888	9.8
6	0.876	11.2	1.91	0.862	12.3
7	0.86	11	1.94	0.851	8.6
8	0.858	11.2	1.96	0.889	10.6
9	0.899	10.9	1.87	0.86	1.8
10	0.87	10.7	2.03	1.06	10
Mean	0.8644	10.99	1.947	0.8839	10.06
S.d.	0.0157937	0.175784	0.0426732	0.0608481	4.0998049
S.d./mean	0.0182713	0.0159949	0.0219174	0.0688405	0.4075353
%	1.8271251	1.5994901	2.1917399	6.8840468	40.753528
ppm	18271.251	15994.901	21917.399	68840.468	407535.28

#	R5 ohm	L5 x 10^-4 H	C5 x 10^-16 F	fr Hz	Q
1	202.1	2.051606	1.506671	905240319	5773.9292
2	207.6	2.098436	1.473076	905231468	5749.2069
3	207.6	2.125072	1.454613	905231227	5822.1815
4	208.1	2.168308	1.425617	905228388	5926.3456
5	208.1	2.103242	1.469715	905229915	5748.5191
6	206.6	2.109979	1.465017	905231557	5808.8133
7	206.5	2.129476	1.451595	905234249	5865.3453
8	206.1	2.103148	1.469776	905231359	5804.0527
9	204.6	2.084518	1.482911	905231620	5794.816
10	205.8	2.08902	1.479716	905231379	5773.4677
Mean	206.31	2.1062805	1.4678707	905232148	5806.6677
S.d.	1.7552493	0.0293149	0.0203477	3057.075	51.976432
S.d./mean	0.0085078	0.0139179	0.0138621	3.377E-06	0.0089512
%	0.8507824	1.3917869	1.3862063	0.0003377	0.8951163
ppm	8507.8245	13917.869	13862.063	3.3771171	8951.1635

Table II. Results of Parameter Extraction Program for 905 MHz

including only transducer capacitance and series resonant arms.[5,6] The full model and the measured data have very good agreement over very wideband frequencies. Figure 7 shows wideband S_{11} and S_{21} comparisons. In addition, the phase comparisons shown in Figures 8 of S_{21} and Figure 9 of S_{11} show excellent agreement. Figure 10 shows narrowband S21 phase comparisons. When the resonator is used in a feedback oscillator, the insertion phase of the resonator will determine the frequency of oscillation. The net loop phase requirement of an integer multiple of 2 will make the resonator insertion phase at oscillation equal to 2 - the net

insertion phase of all components in the loop including the package, loop amplifier, interconnects, power dividers etc. It is very desirable to predict the device phase versus frequency as well as knowing it's sensitivity. Figure 11 shows narrowband insertion loss and return loss comparisons and Figure 12 shows narrowband S_{11} phase comparisons. Again, agreement is good between the model and the measurements.

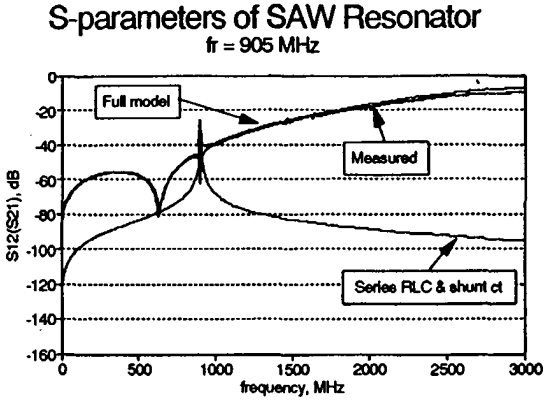


Figure 6. Wideband S_{21} magnitude comparison of measured data, the present model, and the simple series RLC model for a 905 MHz SAW resonator.

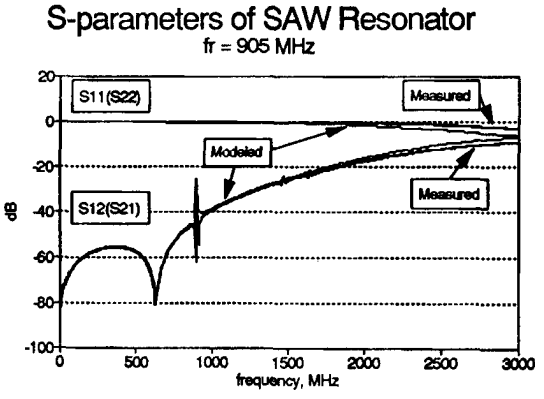


Figure 7. Wideband S_{11} and S_{21} comparison of measured data and the present model.

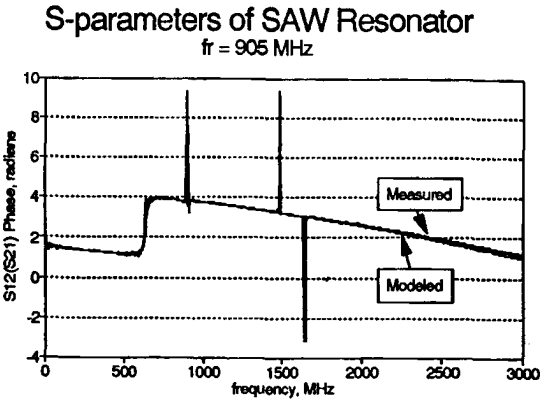


Figure 8. Phase comparison of S_{21} measured data and modeled results.

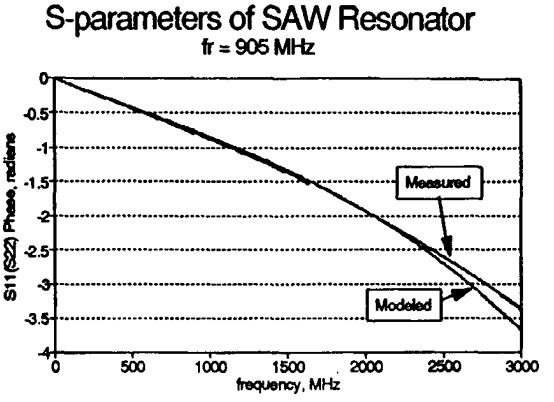


Figure 9. Wideband S_{11} measured and modeled phase data.

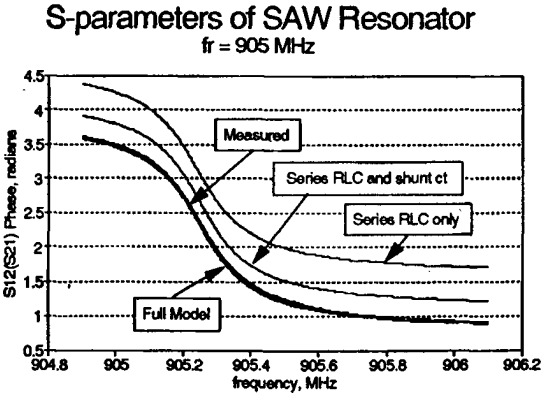


Figure 10. Narrowband S_{21} phase plots of measured versus various modeled results.

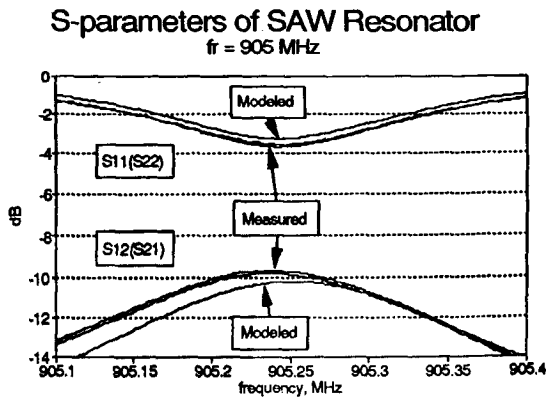


Figure 11. Narrowband S_{21} and S_{11} measured data and predicted model results.

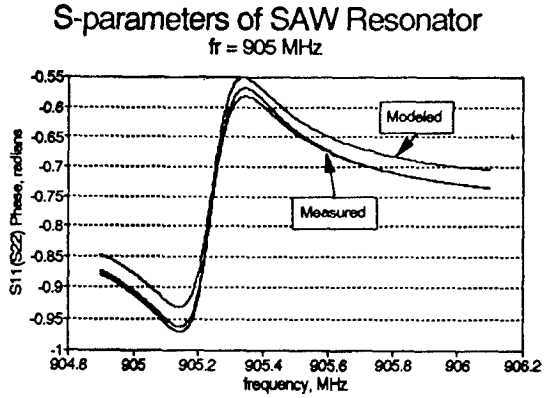


Figure 12. Narrowband S_{11} measured data and modeled phase response.

IV. Example of a 675 MHz SAW Resonator

In order to further test the modeling approach, a series of commercially available SAW resonators were tested. The results of extracting both the package elements as well as the SAW resonator parameters is shown in Table III. As seen, very good correlation between the extracted package elements as well as the resonator motional parameters is found from device to device. The relative frequency is the difference between the nominal center frequency of 675 MHz and the actual measured value. The last column of Table III given in parts-per-million (ppm) is the frequency deviation for a given device measured five times. The average deviation over all 5 devices (25 runs) shows a slightly smaller

Dev	C01	C12	Lb	M	Rbl	L1	C1	R1	Relative	ppm
	pF	fF	nH	nH	ohm	xE-4 H	xE-16 F	ohm	Freq(Hz)	
1	1.754	11.64	2.214	0.859	11.32	1.516	3.667	54.36	123898	3.80
						1.542	3.605	54.82	124053	
						1.513	3.672	54.23	127066	
						1.501	3.703	53.49	120948	
						1.532	3.628	54.34	127060	
2	1.782	11.84	2.21	0.882	8.88	1.477	3.763	56.24	124047	3.35
						1.511	3.677	57.45	124102	
						1.513	3.672	57.50	128983	
						1.512	3.676	57.64	124011	
						1.553	3.578	58.79	126973	
5	1.75	11.06	2.176	0.872	11.1	1.616	3.44	58.50	120929	6.57
						1.594	3.487	57.23	123994	
						1.603	3.466	57.80	128889	
						1.608	3.455	57.99	123871	
						1.611	3.45	57.97	132006	
6	1.724	11.68	2.144	0.779	10.94	1.603	3.468	57.38	118005	5.76
						1.591	3.493	57.13	117997	
						1.578	3.521	56.42	121004	
						1.596	3.482	57.25	118058	
						1.638	3.393	58.57	126970	
8	1.742	11.72	2.21	0.838	11.5	1.553	3.578	55.33	124092	3.07
						1.558	3.566	55.51	124015	
						1.556	3.572	55.53	126991	
						1.528	3.637	54.38	121132	
						1.561	3.561	55.16	124036	
mea	1.75	11.59	2.191	0.846	10.75	1.559	3.568	56.4	124125	
s.d.	0.021	0.304	0.03	0.041	1.066	0.043	0.099	1.59	3542	
ppm	12049	26277	1E+05	48404	99154	27850	27877	28222	5.25	

Table II. Parameter Extraction of 675 MHz Two Port SAW Resonators.

frequency standard deviation value (5.25 ppm) than on some of the single devices under test. This is most probably due to the fact that no control on handling or temperature was conducted. This hypothesis was tested by measuring Device #5 with and without careful controls. When measured with minimum handling and by attempting to limit temperature fluctuations over a measurement cycle, the frequency deviation over 5 runs decreased from 6.5 ppm to less than .1 ppm. This gives some bound on what might be typically found when measuring in a manufacturing environment.

V. Conclusion

This work has presented an approach for modeling both the wideband and narrowband responses of two port SAW resonators. Experimental data has been presented and has shown good agreement between several devices tested. This lumped element modeling approach is useful to extract the SAW motional parameters and can also be

used in conventional lumped element modeling programs.

Acknowledgements

The authors would like to thank the Electronic Industries Association for their present and past support, and Motorola, Inc. for past support by a Motorola University Partnership in Research Grant. In addition, we would like to thank Sawtek, Inc. and RF Monolithics, Inc. for providing test devices.

References

1. EIA, "Standard Methods for Measurement of the Equivalent Electrical Parameters of Quartz Crystal Units, 1kHz to 1 GHz," ANSI/EIA-512-1-1990.
2. Smith, W.L., "Electronic Industries Association Standard 512." Some Further Discussions and Comments, Proc. 7th Quartz Devices Conference, pp13-51, 1985.
3. Hafner, E., "The Piezoelectric Crystal Unit-Definitions and Methods of Measurement", Proceedings of the IEEE, vol 57, #2, Feb 1969, pp 179-201.
4. Jones, Tom, "Design and Development of an RF Data Acquisition System," Thesis: University of Central Florida, 1990.
5. Malocha, D.C., Ng, H., And Fletcher M., "Quartz Resonator Model Measurement and Sensitivity Study," RF Design, August 1989, pp38-50.
6. Horine, B.H. And Malocha, D.C., "Equivalent Circuit Parameter Extraction of SAW Resonators," Proceedings of 1990 Ultrasonics Symposium, Vol. 1, pp477-482.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

MICROWAVE OSCILLATORS STABILIZED WITH SURFACE TRANSVERSE WAVE RESONANT DEVICES

IVAN D. AVRAMOV

Institute of Solid State Physics
72 Trakia Blvd.
1784 Sofia
Republic of Bulgaria

Abstract

Although surface transverse waves (STW) were thought to be very attractive for stable oscillator applications over the last 15 years, not much has been published on STW oscillators with a performance comparable or better than that of the well established SAW oscillators. The reason for this fact is that, before 1987, designers of SAW oscillators believed that it is impossible to build a good STW resonator.

This paper is an attempt to show that STW resonant devices are not only very well suited for stable oscillator applications but have some unique features offering greater design flexibility than their SAW counterparts. Different designs of single and multimode resonators and resonator filters are presented and their performance in fixed frequency and voltage controlled oscillators in the range of 750 MHz - 2 GHz is discussed. Data from frequency trimming, wide-band frequency modulation (FM), injection locking, temperature stability and results from phase noise and short-term stability measurements are presented.

Introduction

Since the time when shear horizontal (SH) surface waves were first coupled to bulk shear transverse waves on piezoelectric substrates by means of interdigital transducers (IDT) [1] these waves have drawn considerable attention. Browning and Lewis called them shallow bulk acoustic waves (SBAW) or surface skimming bulk waves (SSBW) and demonstarted a variety of devices such as filters and delay lines [2]. By that time their most attractive feature was the 1.6 times higher propagation velocity which inspired many engineers to try and build microwave oscillators. Although successful temperature stable 5 GHz SBAW delay line oscillators were demonstrated in 1983 [3], all the efforts to design a high Q SBAW resonator failed. It seemed to be impossible to build an efficient reflector for a wave diffracting into the bulk.

The first successful high-Q SH-resonator on temperature stable quartz was demonstarted by Bagwell and Bray in 1987 [4]. By designing a periodic metal strip waveguide they succeeded in confining the SH-wave close to the surface where it behaves very similar to a SAW and propagates with negligible loss under the metal strip grating. The authors demonstrated a single mode quartz device operating at 1.7 GHz with an unloaded Q of 5600 which is as high as 92 % of the material limit. In their resonator they successfully applied the phenomenon of energy trapping that has been investigated in earlier work [5], [6], [7] in respect to SH-waves. These horizontally polarized trapped shear surface waves were called surface transverse waves (STW) [6], [7] or guided Bleustein-Gulyaev waves (BGW) [8]. As it will be shown below, these waves are very well suited not only for high-Q single and multimode resonator designs but also for low-Q low-loss resonant devices which can be used in stable voltage controlled oscillators (VCO) with excellent wideband FM and injection locking (IL) capabilities and respectable temperature stability and phase noise performance.

This paper presents experimental data from several designs of STW resonant devices on rotated Y-cut quartz, all using a simple unweighted synchronous IDT [9] metal strip two-port resonator configuration. Since such devices allow a control over the resonant Q in very wide limits with retaining low insertion loss they offer a great flexibility in the design of a variety of stable fixed frequency and tunable microwave oscillators.

Propagation of SH-waves under metal strip gratings

Two types of SH-waves are of practical importance to the design of resonant devices. These are STW and SSBW. If the SH-wave is excited and detected by two interdigital transducers (IDT) separated by a free surface from each other, as shown in Fig. 1, then the SH-wave is a SSBW. In

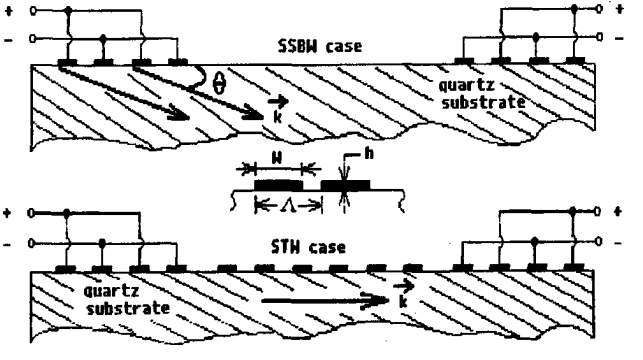


Fig. 1. SH-wave propagation under periodic metal strip gratings.

0-7803-0476-4/92 $3.00 © 1992 IEEE

this case the power flow is radiated under the angle θ in respect to the surface and diffracts into the bulk of the crystal. This results in an increased insertion loss especially when the IDT are separated by a large distance from each other [2]. If a metal strip grating with a period equal to the IDT one is deposited between the IDT it slows the wave down, converting it from a SSBW into a STW. In this case the slowing effect of the grating traps the wave close to the surface preventing it from diffraction into the bulk. The k-vector becomes parallel to the surface (see Fig. 1) drastically reducing the insertion loss. The slowing effect of the grating is two-fold: First the trapping mechanism is similar to that of a Love wave where addition of a slow smooth surface layer slows down the SSBW creating thereby a decay of the wave function into the depth of the substrate. The second slowing effect is due to multiple reflections from the individual grating elements that create a stopband in the transfer function as shown in Fig.2 [6]. The close to stopband behavior

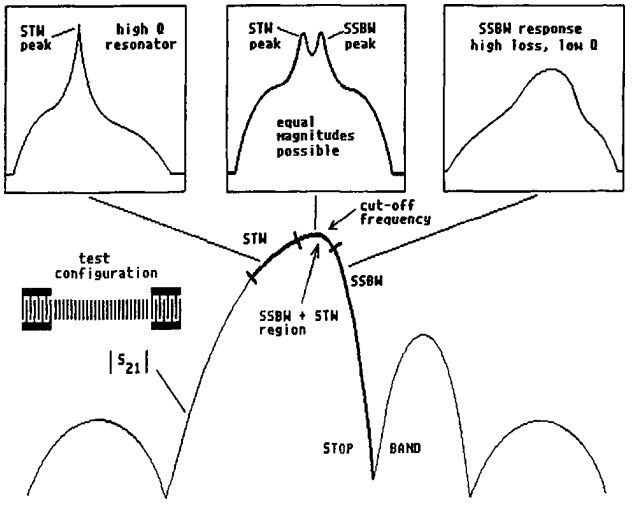


Fig. 2. Transfer function of a test device consisting of two IDT separated by a metal strip grating.

of the transfer function has three characteristic regions which are essential for the design of SH-wave resonant devices. Far below the stopband the SH-wave is very well trapped to the surface and is therefore a STW. We will call this region "STW-region". If a Fabri-Perot resonator is designed in such manner that its resonance is positioned in this region the device would have a low-loss and a high Q. The second characteristic part of the transfer function is on both sides of the so called "cut-off frequency" [8] that is the frequency at which the stopband behavior sets on. In 1980 it was shown by D. L. Lee that here both waves, the STW and the SSBW, can exist and their magnitudes and difference in propagation velocities will strongly depend on the metallization and the number of active finger pairs in the IDT [8]. In this part, which we will call "STW/SSBW-region" a variety of finger pairs and metal thickness combinations can support both waves with equal magnitudes and a controlled frequency displacement, a phenomenon which was successfully applied to the design of Chebyshev type combined mode resonator filters (CMRF) [10]. Finally, the third region, which we will call "SSBW-region", is on the left stopband slope where the SH-wave diffracts into the bulk as a SSBW. If the resonant peak is positioned in this region it is difficult to achieve low-loss and high Q, especially if the stopband center is approached. If we work in the upper part of this region, however, a trade-off between insertion loss and resonant Q can be established such that low-Q resonators for linear wideband FM applications can be designed.

The next section will describe different resonant devices, fabricated on rotated Y-cut quartz that have been designed for operation in the three regions discussed above.

Design of SH-Wave Resonant Devices

Generally there are two mechanisms of positioning the resonance in the desired region. The first mechanism is the metallization. As shown in [11] thicker metallization will slow the wave down and therefore move the resonance in the STW region while thinner metal will cause a decay of the wave function into the bulk and the resonance will be moved in the STW\SSBW- or the SSBW-region.

The second mechanism is based on a control of the resonant peak position via altering the distance L between both IDT in the two-port resonator configuration by small fractions of the acoustic wavelength λ. This can be done by coupling of modes (COM) analysis. Even though not ideal (COM) analysis was found to be a powerful tool for modeling the responses of SH-wave resonant devices. Generally it is necessary to adjust the COM parameters according to the region where one wishes to position one or more resonances. Clearly for all three regions the coupling coefficient $\varkappa$, the grating loss coefficient α and the acoustic mismatch parameter ε [12] will be different. In the STW-region, where the wave is a surface wave in nature, COM analysis gives excellent results below the stopband center frequency. There is a discrepancy between theory and experiment above the stopband where the theoretically predicted ripples do not occur in the measured responses. This is so because the wave energy strongly diffracts into the bulk in this region and there is no standing wave behavior any more. This discrepancy is valid for all three regions. If one intends to work in the STW/SSBW-region it is necessary to check in what relation the STW and SSBW amplitudes and difference in propagation velocities would be [8], [13]. In this case the overall frequency response can be obtained by a superposition of two responses: a STW one and a SSBW one, both with different values of the COM parameters and a frequency displacement calculated beforehand.

The devices presented here were modeled with a COM-analysis software, initially developed for calculating the responses of SAW unidirectional transducers and two-port resonators [12]. Here it was modified as discussed above. The COM-parameters were adjusted by comparing the theoretically predicted responses with those obtained from test devices.

A Single Mode STW Resonator Using a SAW Design

The device in Fig. 3 is identical with a two-port configuration widely used in metal-strip single mode SAW resonators for CATV converters [14]. The only difference is in the metal

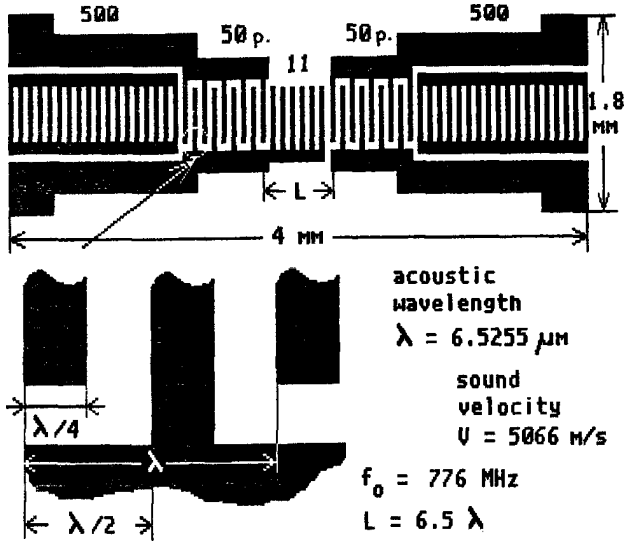


Fig. 3. Configuration of a STW resonator using a SAW design.

thickness and the corrected distance L which is λ/8 longer than the optimum value of nλ/2 - λ/8 for SAW resonators as determined in [14]. Here λ is the acoustic wavelength and n is an integer. The correction of λ/8 in the STW design makes the main resonance slightly unsymmetric by moving the resonant peak towards a lower frequency, deeply in the STW-region as shown in Fig. 4a which is the COM-response of the STW device. The resonance is well below the stopband according to the behavior of the synchronous IDT configuration [9]. Fig. 4b is the measured frequency and phase response of the device. It has an unloaded Q of 5400 and the sidelobe suppression is 14 dB. This data was obtained with a metal thickness h/λ=1.6% which, as shown in [18], yields best results for devices of this type. The comparison of the responses in Fig. 4a and b shows a fairly good agreement between theory and experiment below the stopband. The presence of ripples in the COM-response for frequencies above the stopband can be explained with the fact that for SAW the grating loss coefficient α is assumed to be constant over the entire frequency range. These ripples can be removed from the COM-response by introducing a frequency dependent loss coefficient which increases rapidly when the stopband is approached.

Fig. 5 shows the measured frequency and phase response of a 390 MHz SAW resonator using the configuration in Fig. 3 for comparison. Here L = 6.5λ and h/λ = 2% [14]. Despite the low aperture (40λ) for both devices there is a strong presence of higher order transverse modes in the response of the SAW device.

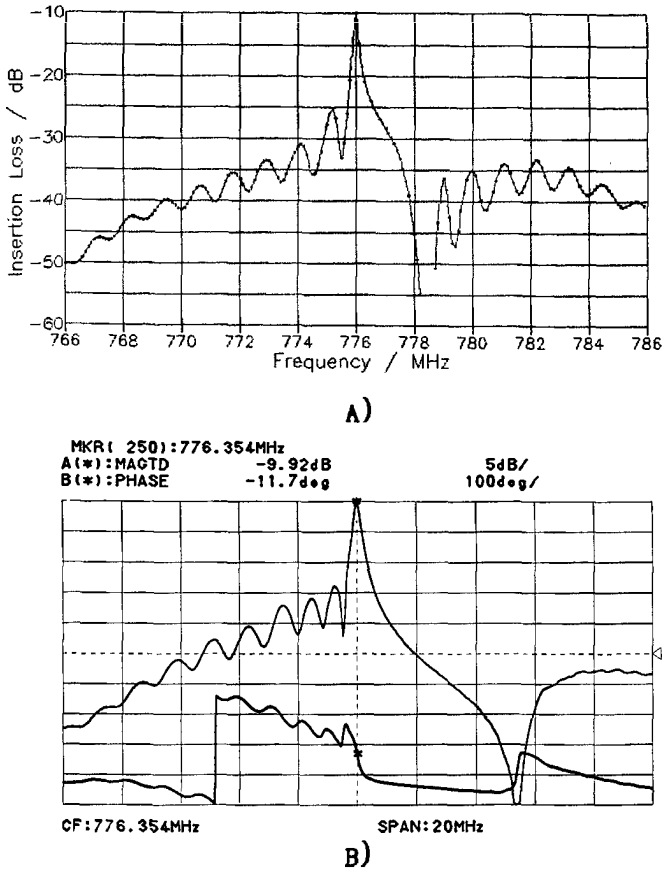


Fig. 4. Characteristics of the device from Fig. 3.
a) predicted frequency response
b) measured frequency and phase response.

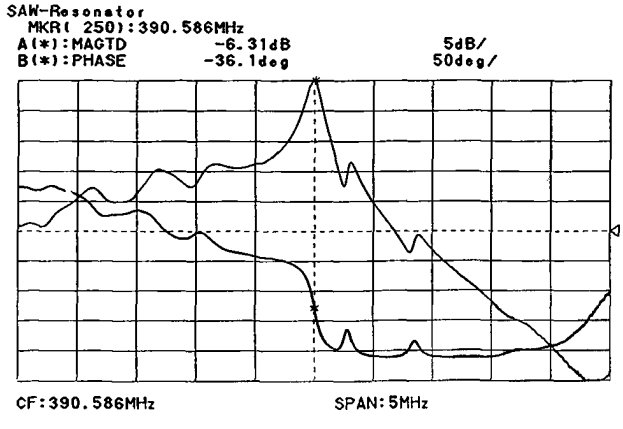


Fig. 5. Frequency and phase response of a SAW device using the configuration in Fig. 3

Extended Cavity Single Mode High-Q STW Resonators

The device described in the previous section has a moderate Q and occupies a small substrate area. Even though the unloaded Q can be increased up to about 6000 using dry etching frequency trimming techniques, as it will be shown below, for

many applications higher Q values may be required. STW resonators offer an elegant way of increasing the resonant Q with retaining single mode operation. This can be achieved because of the fact that STW have a smaller penetration depth into the metal strip reflector gratings (not into the bulk !) compared to SAW. Thus STW resonators allow an increase of the distance L between the transducers without deteriorating their single mode operation. Fig. 6 is a schematic of such a device. Here the resonant peak is still in the STW-region, as evident from the COM-response in Fig. 7a but, in order to keep the wave close to the surface, a

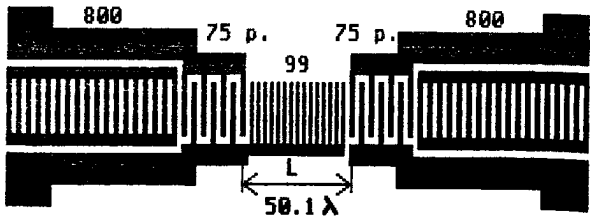
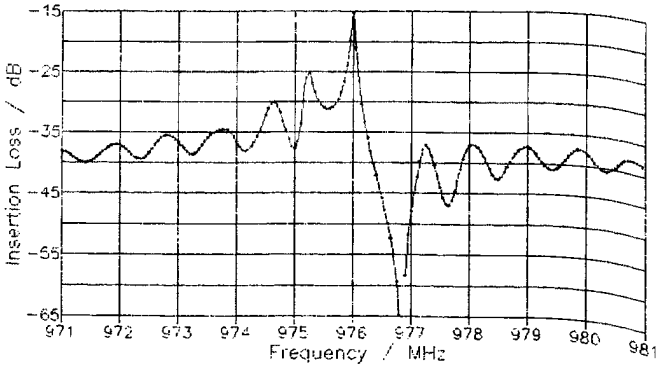
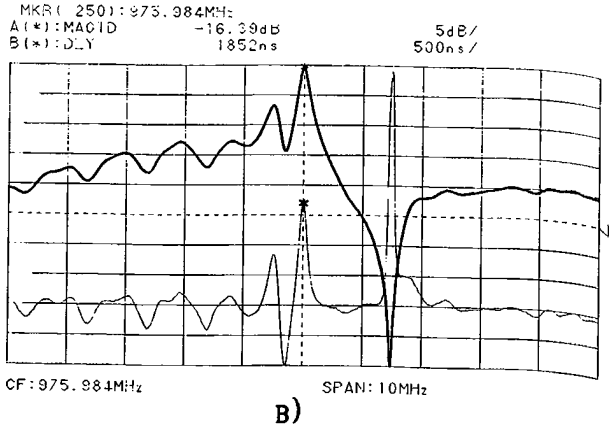


Fig. 6. Configuration of a single mode extended cavity STW resonator.



A)



B)

Fig. 7. Characteristics of the extended cavity single mode high-Q STW resonator.
a) predicted COM-response
b) measured frequency and group delay responses.

grating of 99 strips with a period slightly lower than the transducer/reflector one has been added in the extended part of the cavity. The operation of this grating is explained in [7]. While being transparent for the travelling wave it still keeps the wave energy confined to the surface which is essential for a low-loss device operation. Fig. 7b shows its measured frequency and group delay responses. At 976 MHz the device operates with a loaded group delay of 1.85 μs which yields a loaded Q of 5670 and an unloaded Q of 6700 that is about 62 % of the material limit. The first left side lobe is suppressed by only 7 dB but this was not found to have any effect on the stability of the oscillator using this device.

A Multimode High-Q STW Resonator

It is well known that the best SAW oscillators built to date utilize multimode SAW resonators [26]. These devices allow not only quality factors close to the material limit but also operation at high RF-power levels since they occupy a large substrate area. STW also seem to perform very well in similar multimode designs. Fig. 8. is a schematic of a multimode STW resonator. Its predicted and measured frequency responses are shown in Fig. 9a and 9b respectively. By placing

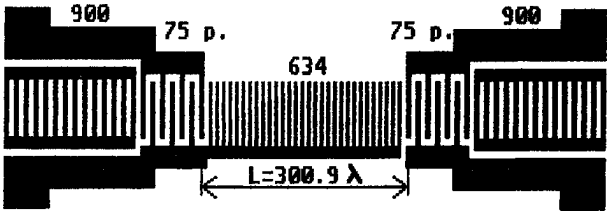
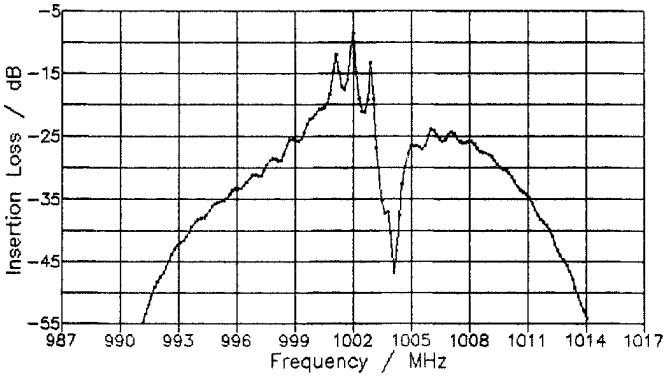


Fig. 8. Configuration of a three-mode long cavity high-Q STW resonator.

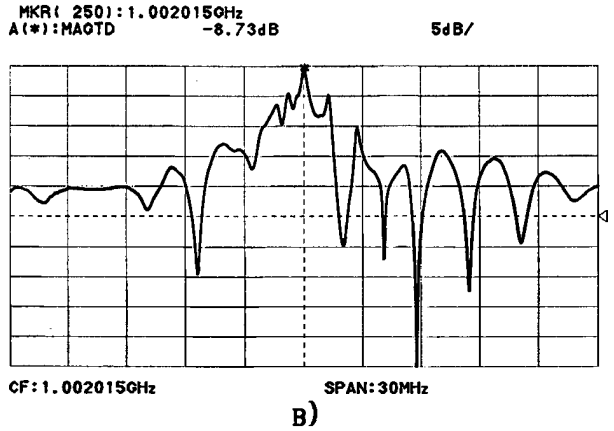
a waveguide grating of about 600 strips in the cavity the latter was designed to support 3 resonant modes in the STW- and STW/SSBW-regions. As evident from Fig. 9a and 9b COM-analysis predicts the resonant and stopband behavior fairly well. Fig. 10 presents the narrowband frequency and phase responses of the device in detail. At the main resonance there is a smooth phase transition over 0 deg. while at the left and right resonances the phase is reversed by 180 deg. Thus in an oscillator loop the device would provide stable single mode operation only at the main resonance, since at the adjacent resonances the oscillator would be out of phase. There is a slight distortion between the resonant peaks which can be explained with interaction of STW and SSBW modes that exist in the STW/SSBW-region. This distortion is well beyond the 1 dB bandwidth of the device and was not found to have any effect on the oscillator stability.

The best 1 GHz device of this design, characterized in Fig. 11, has an insertion loss of 7 dB, a loaded Q of 4000 and an unloaded Q of 7200 which is about 70% of the material limit. This result shows how efficient the center waveguide grating can work even with very long cavities. The maximum RF-power at which the degradation peak

stress should occur was calculated according to [4] (see below) and has the respectable value of about 2.6 W. In fact the device was run for several hours at a power of 2 W. No measurable performance degradation was observed.



A)



B)

Fig. 9. a) predicted and b) measured frequency frequency responses of the device from Fig. 8.

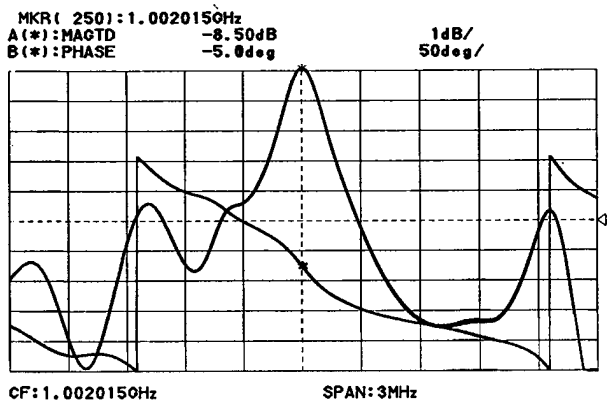


Fig. 10. Detailed frequency and phase responses of the device from Fig. 8.

It is the author's belief that these multimode devices could become the basis for designing precision GHz range oscillators in the near future.

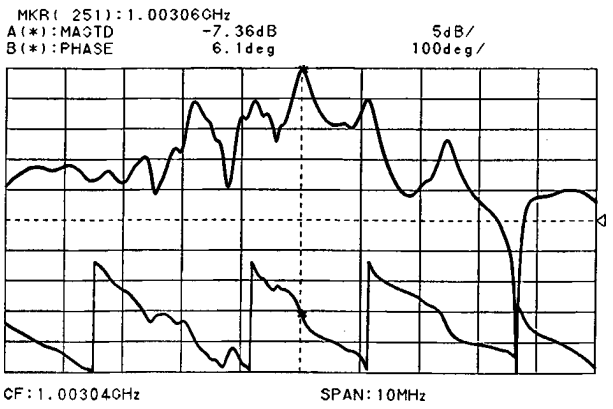


Fig. 11. Frequency and phase responses of the best multimode device fabricated.

Resonant SH-Wave Devices for Voltage Controlled Oscillator (VCO) Applications

For VCO applications the STW device should have a larger bandwidth, a lower group delay or a decreased phase steepness and a single phase transition within its passband in order to allow smooth tuning and stable single mode operation. If the VCO has to be frequency modulated the maximum modulation rate or the ability of the VCO to transmit FM data will strongly depend on the loaded group delay τ_{gl} of the resonant device. In the author's opinion SH-wave devices are ideally suited for VCO and FM applications since they allow a control over the resonant Q and the group delay in wide limits without seriously affecting the insertion loss. Thus a trade-off between VCO tuning range and its ability to handle high rate FM data on one hand, and phase noise performance on the other hand, can be established.

There are two ways of designing resonant devices with an increased bandwidth for VCO and FM applications. First, by placing the resonance in the upper part of the SSBW-region (see Fig. 2.), using thinner metallization, it is possible to achieve loaded Q values of less than 1500. Second, by designing Chebyshev type filters utilizing a controlled excitation of STW- and SSBW modes in the STW/SSBW-region. These devices can have 3 dB bandwidths of more than 0.2 % of the center frequency and an insertion loss well below 10 dB at 1 GHz. Some examples will be discussed next.

Chebyshev Type Combined Mode Resonator Filters

These devices, first reported in [10], are a good example for simultaneous excitation of STW and SSBW modes in the STW/SSBW-region. Fig. 12 is a schematic of such a device. Also here a short center waveguide grating, as the one in Fig. 6 and 8 keeps the insertion loss low. The frequency and group delay responses of such a device are shown in Fig. 13. Its passband response is formed by two STW cavity modes and a SSBW mode with a predicted

magnitude, excited right in the middle between the STW ones. The design of these devices is described in [13] in detail. With proper selection of the

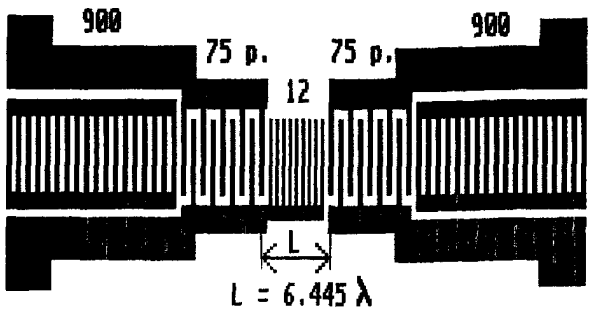


Fig. 12. Schematic of a Chebyshev type CMRF.

number of active fingers in the IDT and careful control over the metallization these filters can be designed to have 3 dB bandwidths of 0.2% of the center frequency, a bandpass ripple within 1 dB and a symmetric group delay response (see Fig. 12).

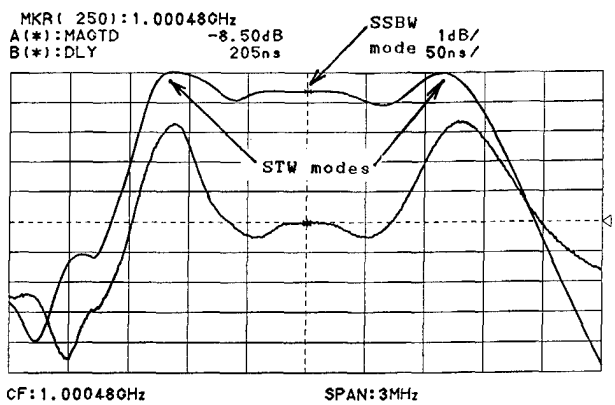


Fig. 13. Typical frequency and group delay responses of a 1 GHz CMRF.

The group delay can achieve extremely low values, of 200 ns and even less. In the author's opinion such low group delay values, so essential for wideband FM and injection locking applications, are very difficult to achieve with other techniques on quartz. Another attractive feature of the CMRF, evident from Fig. 14, is its smooth phase response of about 200 deg. over the entire 3 dB bandwidth which is more than twice as much as the phase shift of SAW or STW resonators. Thus, if the device is used in a low-cost oscillator, no additional phase shifters, that can be rather bulky, are necessary to adjust the phase condition for oscillation. In most cases reversing the bond leads makes the oscillator jump from the left to the right STW mode where the phase slope is maximum and the phase noise is well suppressed. In a VCO this phase response allows smooth tuning with retaining stable single mode operation over the entire 3 dB bandwidth. Such a 1 GHz VCO [15] was tuned in ten 200 KHz steps and frequency modulated at each step. Very linear wideband FM with modulation bandwidths of 150 KHz was obtained [16]. In another application the VCO tuning range was used to build a temperature compensated oscillator. Using simple analog and digital CMOS circuitry a stability of ± 1.4 ppm was obtained in the temperature range of

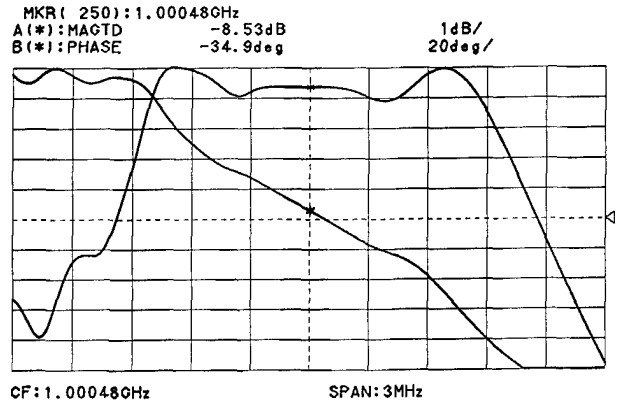
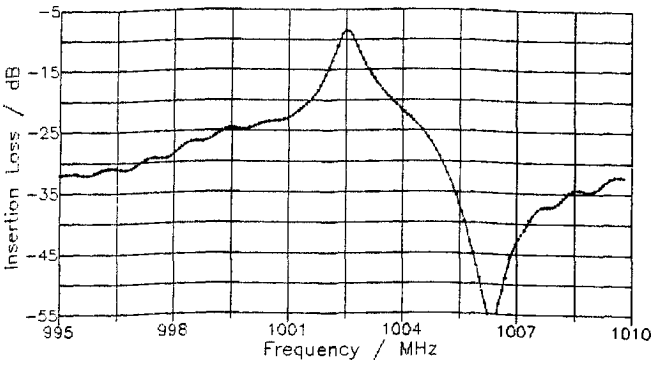


Fig. 14. Frequency and phase response of a CMRF.

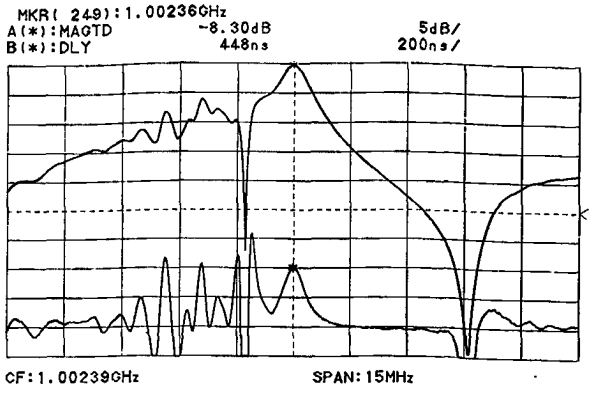
(-45 ... +75)deg. C [17]. In this application only 20% of the entire available tuning range was used for the compensation. The rest can be used for temperature compensation at multiple frequencies. Further work regarding this problem is on the way.

Low-Q SH-Wave Resonators. The CMRF described above can easily be modified into a low-Q single mode resonator using thinner metallization. The CMRF was fabricated with an optimum metal thickness $h/\lambda = 1.5\%$ at a metallization fill factor of 0.4 [13]. The device, characterized in Fig. 15a and b was fabricated with the same photolithographic mask as the CMRF (see Fig. 12) but with a metal thickness of 1.2% which, according to [18] and [19], should move the resonant response by about 2000 ppm higher and position it in the SSBW region where a considerable diffraction into the bulk strongly reduces the out-of-band ripple. The COM response in Fig. 15a, obtained with an increased value of the grating loss coefficient, is in a good agreement with the experimental frequency response in Fig. 15b. The distortion on the measured frequency response below resonance is in the STW/SSBW-region and is due to interaction of STW and SSBW modes. The insertion loss of 8.3 dB is still low, first because it operates in the upper part of the SSBW-region and second, because of the large number of IDT fingers and a small distance between the IDT.

A first attempt was made to build also a 2 GHz low-Q resonator using electron beam lithography (EBL) equipment. Its frequency and phase responses are shown in Fig. 16. Unfortunately, in the evaporator used, it was difficult to control very small thicknesses precisely and in fact this device was built with a much thinner metal than the optimum of 40 nm. As a result the frequency was about 6 MHz higher than expected and the resonance occurred deeply in the SSBW-region, not far from the stopband center frequency. This and the increased electrode resistance as a result of the thinner metallization are the reasons for the increased insertion loss of more than 24 dB.



A)



B)

Fig. 15. Characteristics of a low-Q resonator
a) COM response
b) Frequency and group delay response.

As evident from Fig 15b and Fig. 16 these low-Q devices have a symmetric group delay response and a very smooth S-shaped phase response and are very suitable for linear FM applications such as cordless microphones and portable communications. A frequency modulated VCO stabilized with the device from Fig. 15 was found to be able to easily handle modulating frequencies of 20 KHz.

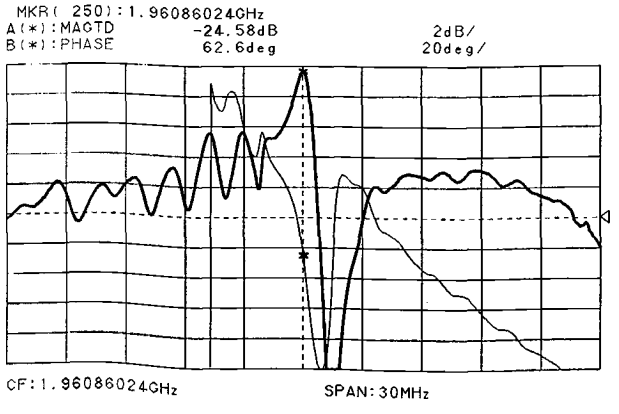
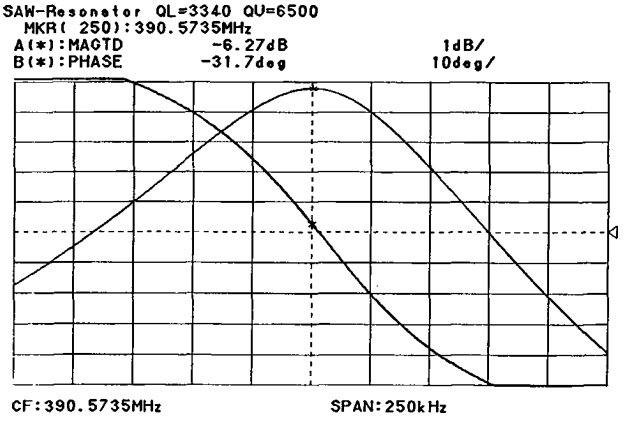


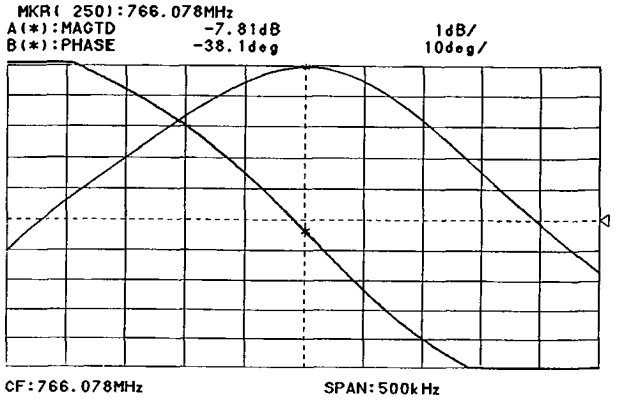
Fig. 16. Frequency and phase responses of a 2 GHz EBL device.

Equivalent and Characteristic Parameters of STW Resonators

As evident from Fig. 17 there is basically no difference in the close to resonance behavior of a SAW and a STW resonator. Therefore the same equivalent circuit used for SAW two-port resonators



A)



B)

Fig. 17. Detailed frequency and phase responses in the vicinity of resonance of:
a) a 390 MHz SAW resonator and
b) a 766 MHz STW resonator.

can be used to describe the behavior of STW resonators. This circuit is shown in Fig. 18. It contains the static IDT capacitances C_t, and the

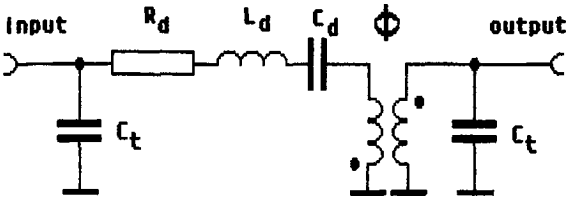


Fig. 18. Equivalent electric circuit of SAW and STW resonators.

resonator dynamic parameters R_d, L_d and C_d that can be derived from the measured loaded quality factor Q_L, the resonant frequency f_r, the insertion loss IL and the unloaded Q-factor Q_U using following equations [14]:

$$Q_L = f_r / \Delta f_{3dB} \quad (1)$$

$$Q_U = Q_L / [1 - 10^{-(IL/20)}] \quad (2)$$

$$R_d = Q_L (R_g + R_l) / (Q_U - Q_L) \quad (3)$$

$$L_d = Q_U R_d / (2\pi f_r) \quad (4)$$

$$C_d = 1 / [(2\pi f_r)^2 L_d] \quad (5)$$

where Δf_{3dB} is the loaded 3 dB bandwidth and R_g and R_l (usually equal to 50 Ω) are the source and load impedances of the measurement system. The static capacitances C_t can be measured directly with a capacitance meter or if the number of IDT fingerpairs N and the acoustic aperture W are known, the C_t values can be calculated as:

$$C_t = C_2 N W \quad (6)$$

where C_2 is the capacitance of one finger pair per unity length.

The transformer in Fig. 18 indicates the possibility of adding an additional phase shift of 180 deg. by reversing the device's bond leads. This may be necessary to meet the phase condition of oscillation.

The resonance behavior of a two-port STW resonator is evident also from Fig. 19 which shows

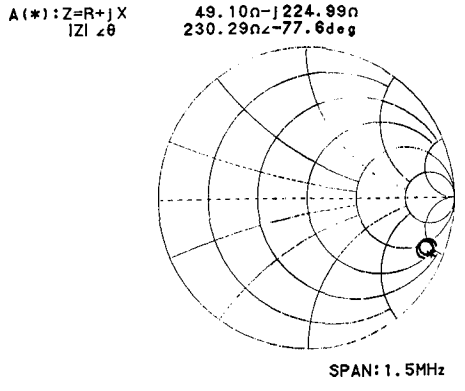


Fig. 19. Characteristic input or output impedance of the STW device from Fig. 17.

its characteristic input or output impedance as a loop on the Smith chart. This loop is located in the capacitive part of the chart which is due to the static IDT capacitances C_t.

Another useful equation gives the relationship between Q_L and the resonator loaded group delay τ_{gl} as follows:

$$\tau_{gl} = Q_L / \pi f_r \quad (6)$$

One of the most attractive features of STW resonators is their ability to handle high levels of RF power. As shown in [4] the maximum RF power P_{max}, a STW resonator can dissipate, can be calculated using following equation:

$$P_{max} = \left(\frac{|T_{12}|_p}{2050} \right)^2 \frac{L_e D W}{Q_U \lambda} \quad (7)$$

where $|T_{12}|_p$ is the highest component of the shear stress that can occur in the STW device and here it is assumed to be as high as the degradation peak stress value of $2.8*10^8$ N/m^2 for SAW resonators with Al-Cu metallization, D is the penetration depth into the bulk which is assumed to be 0.75λ [4] and W is the acoustic aperture. The effective resonant cavity length L_e can be calculated from the cavity mode distance Δf_{cm} between two adjacent standing wave maxima and can be measured as a distance between the ripples below resonance in the frequency response (see for example Fig. 4b). Then L_e is calculated as:

$$L_e = (f_r / \Delta f_{cm}) \lambda \quad (8)$$

Equations (1) ... (8) characterize the resonant and power handling behavior of STW two-port resonant devices fairly well and have the advantage that all characteristic values can be derived only from three measured parameters: f_r, Δf_{3dB} and IL. These can be measured with sufficient accuracy in a 50 Ω environment using a low-cost scalar network analyzer or level meter, a frequency counter and a high resolution signal generator. Characteristic parameter values of the 776 MHz STW device from Fig. 4b compared with those from the 390 MHz SAW device in Fig. 5 and 17a are given in Table 1. Both devices use the same SAW design. It is evident that the SAW resonator has about 3.5 dB lower insertion loss which results in a slightly higher unloaded Q than the STW device but the latter can handle more than 30 times higher RF power levels and would therefore perform much better in an oscillator in respect to the thermal noise floor. In respect to loaded Q the STW resonator is superior to the SAW device despite the twice as high operating frequency. Finally the absence of higher order transverse modes which can present a real problem in SAW resonators with larger apertures and the 1.6 times higher operating frequency for the same lithographic resolution give a definite priority to the performance of STW resonators in the UHF range.

Parameter	776 MHz STW device	390 MHz SAW device
Insertion Loss	9.9 dB	6.3 dB
Loaded Q	3650	3340
Unloaded Q	5370	6500
Loaded gr. delay	1.47 μs	2.72 μs
Sidelobe suppr.	14 dB	13 dB
RF power handling	1.7 W	0.05 W
IDT capacitance	1 pF	1.3 pF
Dynamic resistance	212 Ω	106 Ω
Dynamic inductance	233 μH	280 μH
Dynam. capacitance	180 fF	59 fF

Table 1. Comparison of characteristic parameters of SAW and STW resonators using a SAW design.

Frequency Trimming of STW resonators

In most cases high Q resonators have to be trimmed in frequency in order to meet a given specification requirement. Frequency trimming experiments using heavy Xe^+ ion bombardment were performed in the laboratories of Dr. M. Baklanov at the Institute of Semiconductor Physics in Novossibirsk, Russia. The Xe^+ reactor used allows a very fine control over the ion current, from 0.01 $\mu A/cm^2$ to 100 $\mu A/cm^2$ which allows first a rough and then a fine frequency trimming with an accuracy of 1 ppm if in situ monitoring provisions are made. The trimming mechanism is based on mechanical sputtering of the quartz between the metal electrodes of the devices. The trimming process is performed in an oxygen atmosphere which makes the process selectiv. The Al-Cu electrodes were found to be etched at a rate about 8 times lower than quartz.

Frequency trimming was performed on the 776 MHz single mode devices and the 1 GHz three mode resonators characterized in Fig. 4 and 9. Some results are presented in Fig. 20 which shows the frequency shift and unloaded Q behavior of 776 MHz devices with trimming time. The ion current was 0.213 $\mu A/cm^2$ in this experiment. In the beginning of the etching process there is a rapid change in the resonant frequency by about 500 ppm which is due to the abrupt decrease in STW propagation velocity because of the grooves that occur between the electrodes. This results in an abrupt change of the reflection coefficient which was found to decrease the insertion loss by about 2 dB and increase the unloaded device Q by about 20 %. Further downtrimming results in a linear frequency shift and a relatively constant unloaded Q with trimming time. In another experiment we found that this behavior remained unchanged even after a downtrimming by 3500 ppm for the single mode devices. Further downtrimming results first in a decrease of the device loaded Q and then in a

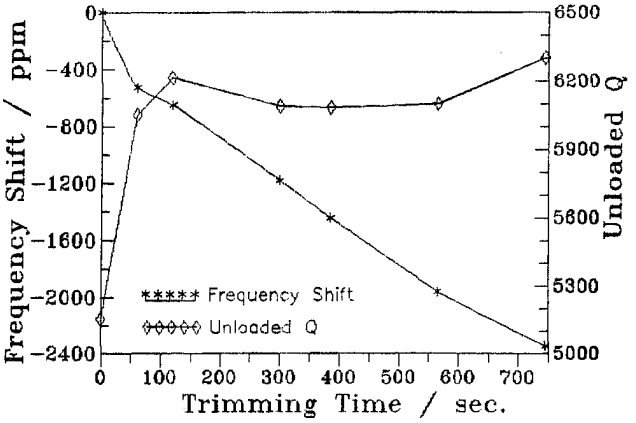


Fig. 20 Frequency trimming of 776 MHz single mode STW resonators.

strong deterioration of the overall performance. This is illustrated in Fig 21. The rightmost curve is the frequency response of the device before trimming. The medium curve was measured after a downshift by 4.2 MHz (5400 ppm). Here the insertion loss is still the same but a decrease in

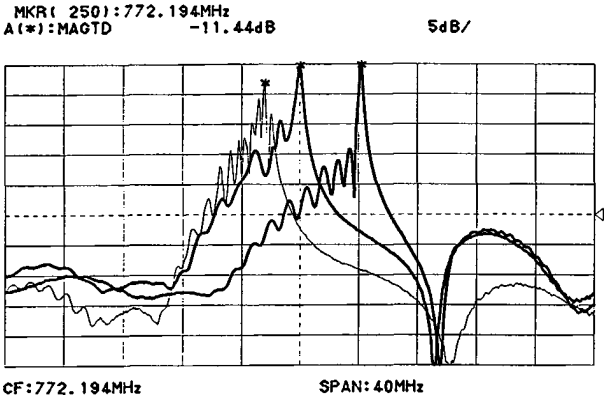


Fig. 21 Frequency response behavior of STW single mode resonators with downtrimming:
right curve: before trimming
medium curve: after 4.2 MHz downtrimming
left curve: after 6.5 MHz downtrimming.

loaded Q and sidelobe suppression is evident. The left curve corresponds to a downshift of 6.5 MHz (8400 ppm). We found that the total etched groove depth for this frequency shift was 185 nm which is almost twice as large as the metal thickness.

Multimode devices were found to withstand about 1000 ppm frequency downshift before deteriorating their performance. A substantial improvement of their parameters was also observed after the first trimming. It is the author's belief that Xe^+ heavy ion bombardment is very suitable for frequency trimming and improving the parameters of STW resonators.

Microwave Oscillators Using SH-Wave Resonant Devices

Although many authors believed that STW oscillators may offer some advantages in respect to temperature stability, aging and phase noise performance in the lower GHz range [2], [3], [4] so far, except for the investigations of Bagwell and Bray [4] and a recent paper describing a 2 GHz STW Collpits oscillator [20] not much has been published on microwave oscillators using SH-wave resonators. This section will review some results obtained with SH-wave resonator oscillators at the Institute of Solid State Physics in Sofia.

The basic feedback oscillator circuit used in all studies is shown in Fig. 22. The RC-amplifier

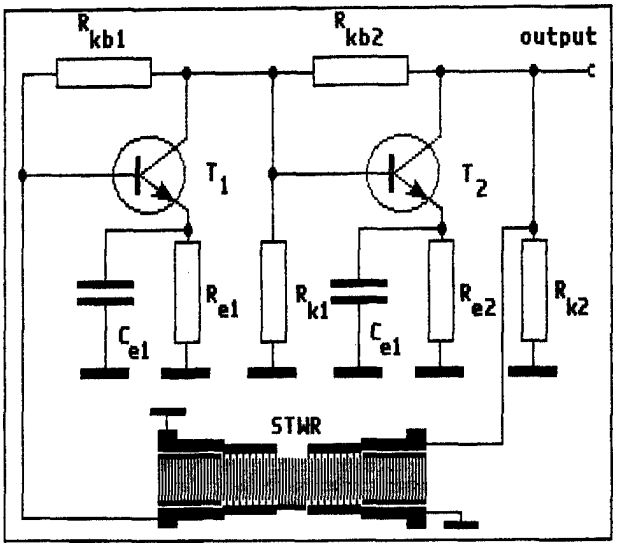


Fig. 22. Basic STW feedback oscillator circuit.

used in this circuit was initially developed for delay line oscillators but was found to behave well also in resonator oscillators in the frequency range 0.3 - 2 GHz. Even though STW resonators are low-loss devices and in most cases one transistor would be sufficient to compensate for the device loss, the two-stage amplifier in Fig. 22 has some important advantages for the design of low-noise oscillators. First, by using a low collector current in the first transistor stage and a high collector current in the second stage, the noise figure of the amplifier can be improved by 2-3 dB compared to a single stage amplifier with the same output power. Second, each transistor stage has a separate voltage and frequency dependent current feedback realized with the resistors R_{kb1} and R_{kb2} and the RC groups R_{e1}, C_{e1} and R_{e2}, C_{e2} respectively. This complex feedback allows a control over the amplifier's input and output admittances such that optimum matching conditions to the SH-wave device can be established. For example in a fixed frequency oscillator this feedback allows capacitive input and output impedances to be adjusted so that a trade-off between device loaded Q and insertion loss in the loop is established. If a wide tuning range VCO or an injection locked oscillator is desired then the amplifier's input and output admittances should be made inductive in order to tune out the IDT capacitances C_{t1} and C_{t2} and decrease the loaded resonant Q. This case is illustrated in Fig. 23 which shows the computer optimized input and output short circuit admittances of the amplifier Y11 and Y22 in the frequency range of 0.4 - 1.2 GHz. Their

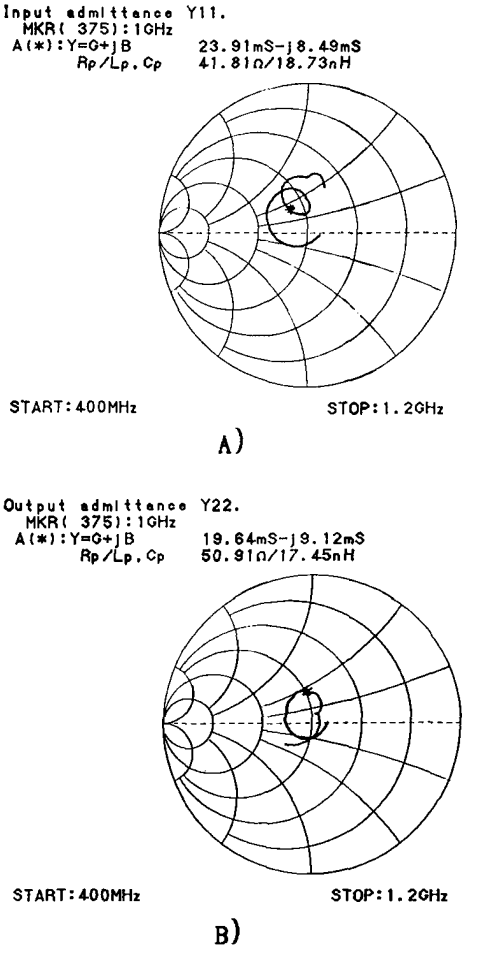


Fig. 23 a) Input and b) Output short circuit admittances for inductive matching.

inductive components have been optimized in such manner that an IDT capacitance of 1.4 pF is tuned out completely at 1 GHz. It is important to note that the admittance curves are dominantly in the inductive part of the Smith Chart which means that the amplifier can provide inductive matching over a wide frequency range without using any external inductors or transmission lines. While the input and output admittances of the amplifier strongly depend on the values of the emitter feedback elements the amplifier's phase and the gain can be controlled in wide limits by the feedback resistors R_{kb1} and R_{kb2} without seriously affecting the matching conditions. Thus the whole oscillator circuit can be optimized only by altering the values of the feedback elements without the necessity of using external components other than

the device bond leads. These are included in the oscillator analysis. A CAD software for optimization of SAW oscillators using this amplifier circuit has been developed at the Institute of Solid State Physics in Sofia. It calculates the oscillator eigenfrequency, the gain compression and the transmission function around the loop and allows a quick circuit optimization. Details are given in [21] and [22].

Another advantage of the amplifier in Fig. 22 is the emitter degeneration realized by the resistors R_{e1} and R_{e2} which is known to reduce the harmonic content and decrease noise downmixing. 1 GHz STW resonator oscillators with 13 dBm of output power were found to suppress the harmonics by more than 40 dB without additional filtering.

The frequency and phase responses of the amplifier are shown in Fig. 24. A constant gain of

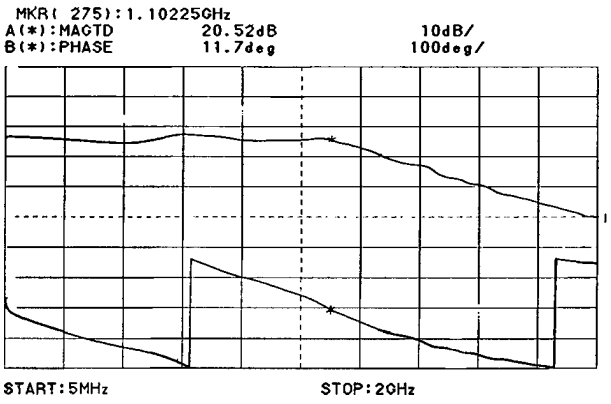


Fig. 24. Frequency and phase responses of the broadband amplifier from Fig. 22

10 - 20 dB can be adjusted over a frequency range of up to 1.2 GHz using 7.5 GHz transistors. A gain of 26 dB at 2 GHz was achieved using bipolar transistors with a transition frequency of 12 GHz. The circuit has a very low phase slope and behaves well with temperature. STW resonator oscillators using this amplifier were found to operate without problems in a temperature range of (-120 ... +120) deg. C. Finally it occupies a very small area and is convenient for realization in hybrid and especially surface mounted devices (SMD) technology.

Frequency Tuning of SH-Wave Oscillators

As shown in the beginning of this study SH-wave devices can be designed to have very low Q values with retaining an insertion loss of less than 10 dB. CMRF based 1 GHz VCO with a tuning range of 1.8 MHz were found to have a phase noise suppression of more than 95 dBc/Hz at an offset of 1 KHz away from carrier [15]. Using low Q SH-wave resonators, which allow a control over the resonant Q in very wide limits, from several thousand to values slightly above 1000, a trade-off between tuning range and phase noise performance can be achieved. Electronic tuning of a VCO stabilized with such a device can be realized with the circuit in Fig. 25. It contains two single transistor

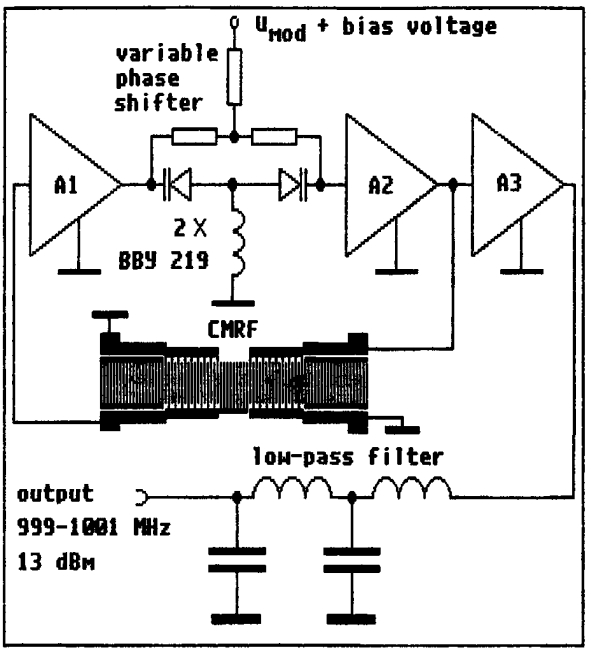


Fig. 25 Schematic of a varactor tuned VCO.

amplifier stages A1 and A2 as the ones in Fig. 22, a varactor tuned CLC variable phase shifter (VPS), known for its tuning linearity and ability to behave well at high RF levels, the SH-wave device in the loop, a buffer bipolar transistor amplifier A3 for reducing load pulling and a low-pass filter for suppression of harmonics that are mainly generated in the buffer amplifier. With careful selection of the varactors and the value of the inductor in the VPS a phase shift of 90 deg. can be obtained from a single VPS. The tuning characteristic of a 1 GHz VCO stabilized with a low Q resonator and tuned over its 3 dB bandwidth is shown in Fig. 26. Smooth tuning over 900 KHz is achieved with a tuning voltage of 0 ... 25 V. Fig. 27 illustrates the tuning ability of a VCO using a

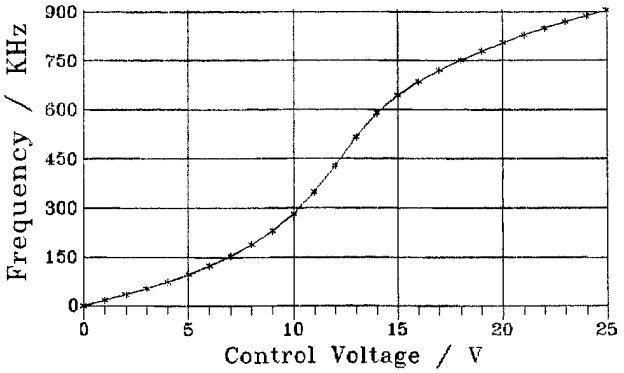


Fig. 26 Tuning characteristic of a 1 GHz VCO stabilized with a low-Q resonator.

1 GHz CMRF with a 3 dB bandwith of 2.5 MHz. Here the tuning range of 2.4 MHz was obtained by

cascading two identical VPS as the one in Fig. 25. The overall linearity is not very good because the CMRF used here had a rather distorted group delay response but for many applications, as temperature compensated oscillators [17] or frequency agile sources a better linearity may not be necessary.

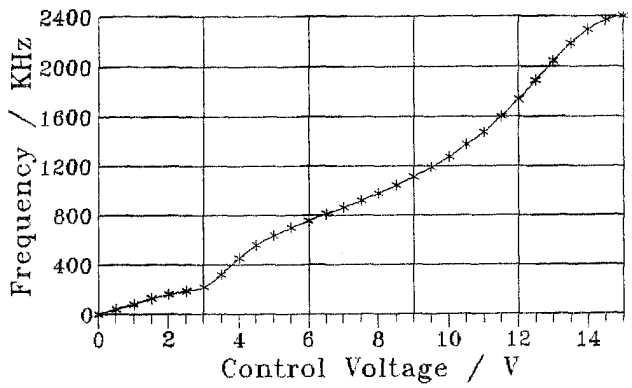


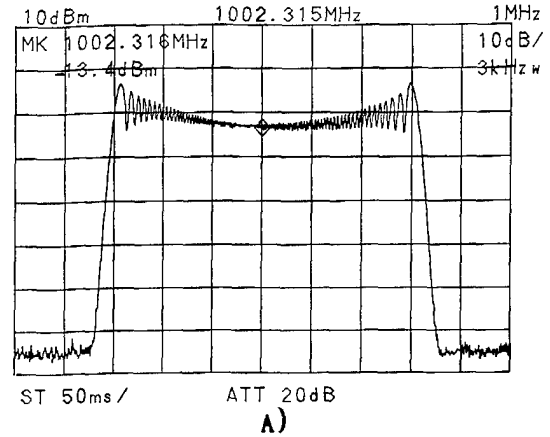
Fig. 27. Tuning characteristic of a CMRF based VCO.

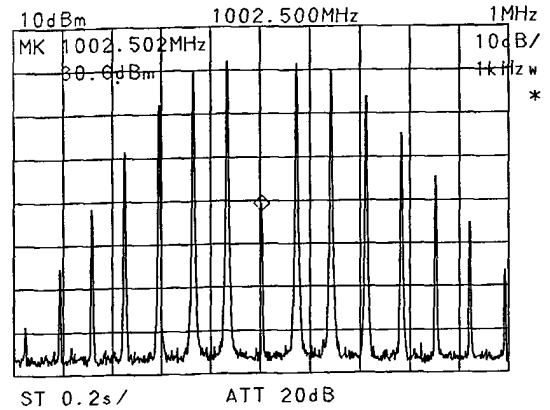
Frequency Modulation of SH-Wave Oscillators

UHF SAW resonator oscillators usually operate with loaded group delays of several microseconds and are very difficult to modulate because energy storage effects in the SAW device usually limit the FM bandwidth to a few KHz or even less which makes high rate data transmission impossible. The rapid development of personal communications and local area radio networks requires low noise temperature stable FM sources in the frequency range of 0.8 - 3 GHz with reduced size, weight and cost which are capable of transmitting data at a rate of several hundred Kbits/s up to a few Mbits/s [23]. SH-wave stabilized microwave FM sources seem to be a very good solution for such purposes since the control over resonant group delay allows a reasonable compromise between FM bandwidth and phase noise performance.

Recently a CMRF based 1 GHz VCO with a phase noise suppression of 97 dBc/Hz at 1 KHz away from carrier was found to be able to transmit wideband FM data at a rate of more than 200 Kbits/s [15]. Further experiments were performed with a low-Q SH-wave resonator based VCO which operated with a group delay of 350 ns in the loop. Its tuning characteristic was very similar to the one in Fig. 26. Figures 28 a, b and c show wideband sine wave FM spectra obtained with this oscillator at a frequency deviation of ± 300 KHz which yields a modulation bandwidth of about 600 KHz. The spectrum in Fig. 28a, measured at a modulating frequency of 1 KHz, is symmetric which indicates a very low distortion. In fact a nonlinear distortion of less than 0.1 % was measured with a distortion analyzer after detection but since this value was very small it was not clear if the distortion occurred in the VCO or in the detector.

The spectrum in Fig. 28b was measured with a modulating frequency of 71 KHz at which the carrier disappeared for the last time when the frequency





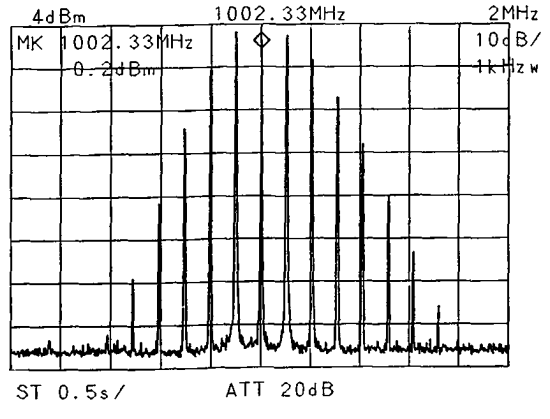


Fig. 28. FM spectra at modulating frequencies of: a) 1 KHz , b) 71 KHz and c) 103 KHz.

was increased. Here the modulation index is greater than one. Further increase of the modulating frequency caused a decrease of the frequency deviation due to energy storage in the

resonator. The frequency at which the deviation decreased by 3 dB [15] was 103 KHz. This case is illustrated in Fig. 28c. Here the carrier becomes equal to both adjacent sidebands and the modulation index becomes smaller than one. Without storage effects in the resonator this case would occur at a modulation frequency of 300 KHz.

Even though the loaded group delay is the main limiting factor of the maximum FM data rate the latter seems to depend to a lower extent also on the frequency deviation, the gain compression and the loop power. Further investigations on how these factors limit the maximum data rate are on the way.

SH-wave based VCO were found to be able to transmit FM data at rates as high as 2 Mbits/s at a modulation index smaller than 1. It is the author's belief that these unique FM features could find a variety of future applications in wireless data transmission networks.

Injection Locking of SH-Wave Oscillators

Injection locking is often used to improve the medium and long term stability and the close to carrier phase noise performance of microwave oscillators. If a low-power signal of a higher harmonic, obtained from a stable crystal oscillator, is injected into the loop of an unstable oscillator running at a frequency close to that of the injected signal then the frequency of the unstable oscillator tracks the frequency of the injection source within a synchronization bandwidth B. According to Adler's equation B depends on the the loop power P_{loop} and injection power P_{inj} in the point of injection, and the loaded group delay as follows [24]:

$$B = \frac{1}{\pi \tau_{gl}} \sqrt{\frac{P_{inj}}{P_{loop}}} \qquad (9)$$

Here again the resonant group delay in the loop plays the major role in determining the oscillator's injection locking behavior.

Experiments were performed with a CMRF based VCO using the circuit in Fig. 25. A signal from a noisy synthesizer with quartz stability was injected at the input of the amplifier A1 where the loop power had its lowest value. Then the synchronization bandwidth was measured as a function of the injected power. The experiment was performed 4 times at 4 different frequencies within the VCO tuning range, each corresponding to a different value of the group delay. These frequencies were adjusted on the free running oscillator with the variable phase shifter and the loop power in the point of injection was measured before locking. The results are shown in Fig. 29. Even though the loop power varied by 10 dB at the different free running frequencies the synchronization bandwidth is mainly determined by the group delay.

Fig. 30 is the spectrum of the injection locked VCO compared with the spectra of the synchronizing source and the free running VCO. Since the injection locked oscillator performs like a noise filter a substantial improvement of the

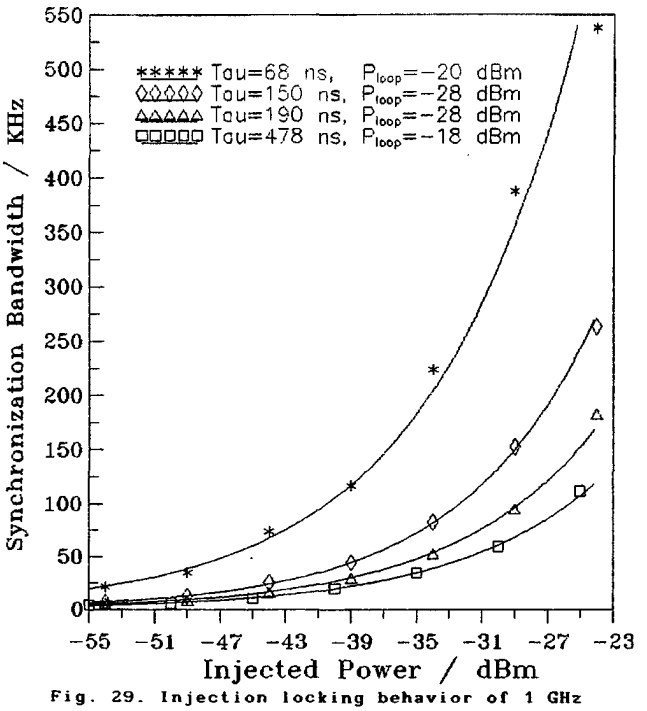


Fig. 29. Injection locking behavior of 1 GHz CMRF based VCO.

signal to noise ratio is observed. Here the output power of the injection locked VCO was 9 dBm while the injected power was only -40 dBm. Thus such an injection locked SH-wave VCO can successfully be used for low-noise amplification of multiplied quartz frequencies in the lower GHz range.

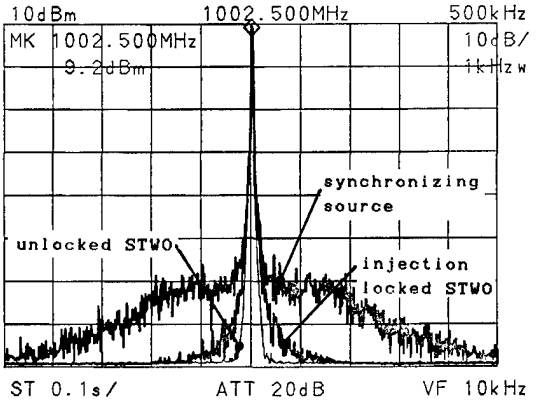


Fig. 30. Spectra of the free running and the injection locked VCO compared with the spectrum of the synchronizing source.

Temperature Stability of SH-Wave Oscillators

So far it is known that the temperature behavior of STW resonator oscillators on 36 deg. rotated Y-cut quartz is similar to that of AT-cut bulk wave resonators [4] and depends on

metallization [7]. More systematic results regarding the dependence of the resonant frequency on temperature in metal strip STW resonators, to the best of the author's knowledge, have not been published so far.

Here we will present some experimental results obtained from fixed frequency oscillators stabilized with 766 MHz STW resonators and from a VCO using a 1 GHz CMRF. The 766 MHz devices used in this study were fabricated on different quartz cuts close to the AT cut and with different metal thicknesses. The 1 GHz CMRF was realized with a metal thickness h/λ = 1.6% on 36°40' rotated Y-cut quartz. Fig. 31 is the temperature behavior of 766 MHz resonator oscillators at a constant metallization (h/λ = 1.6%) but on different quartz cuts. The curves are close to parabolic although below the turn-over temperature some unsymmetry indicates that an inversion point exists at very low temperatures. Assuming that for small changes of the cut orientation the shift of the turn-over point with cut angle is linear, from the results in Fig. 31 we can say that 1 deg. change in orientation yields about 20 K shift of the turn-over temperature. Thus STW resonators seem to be about three times more sensitive to cut orientations than their SAW counterparts [26].

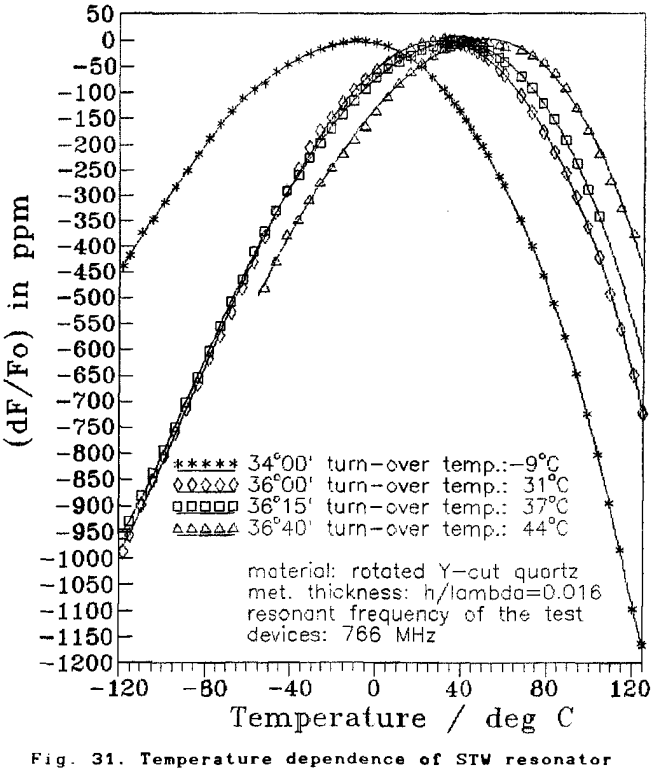


Fig. 31. Temperature dependence of STW resonator oscillators with cut orientation.

The curves in Fig. 32 were obtained with 766 MHz resonators all fabricated on the same quartz cut but with different metal thicknesses. Here the AT-cut cubic behavior is clearly visible especially at thick metallization. The overall temperature instability of the oscillator with a metal thickness h/λ = 2% in the range of (-100 ... +35) deg. C does not exceed ± 60 ppm. With a proper cut correction this curve can easily be shifted towards

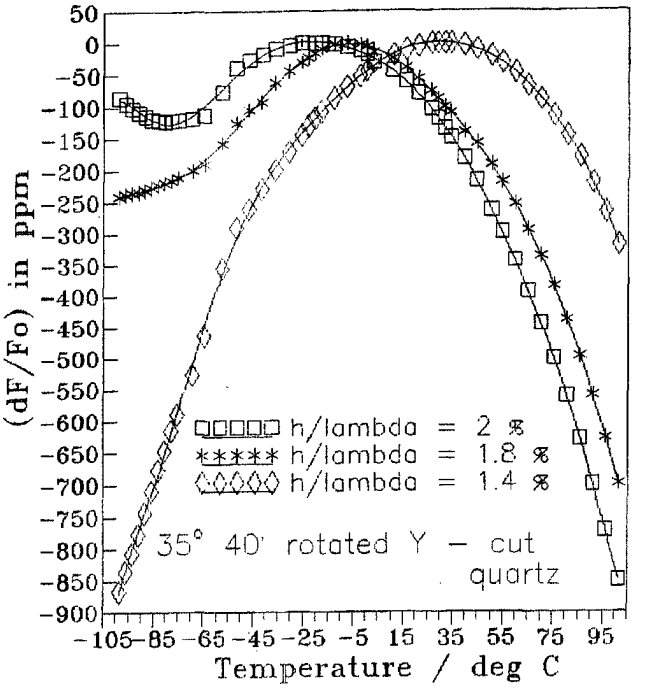


Fig. 32. Influence of the metallization on the temperature behavior of STW resonators.

higher frequencies in order to achieve maximum stability for example at room temperature but so far it is not clear whether maximum device Q can be achieved with such a thick metallization.

The temperature stability of the CMRF based 1 GHz VCO was measured at three characteristic frequencies. The two curves in Fig. 33 correspond to a frequency adjusted at the SSBW mode and a second one adjusted at one of the STW modes in Fig. 13. Although there is not much difference in the behavior of both curves the turn-over temperatures were found to differ by about 7 K from each other. Then the VCO frequency was adjusted at room temperature between the SSBW mode and one of the STW modes in Fig. 13. An unexpected temperature behavior, illustrated in Fig. 34 was observed at this frequency. Even though the overall stability is not very different from the one in Fig. 32 the frequency shift with temperature is rather distorted. Similar behavior is observed in bulk wave resonant devices in which two bulk modes of different origin interact with each other [25]. If the reason for the distortion in Fig. 33 is the same as in bulk wave devices then this is another proof for the different origin of the modes forming the frequency response of the CMRF in Fig. 13. Further work is needed to investigate and explain this interesting phenomenon in STW/SSBW combined mode resonator filters.

The results from Figures 31, 32, 33 and 34 show that the overall temperature stability of SH-wave resonant devices depends not only on the cut orientation but to a large extent also on the

metallization. As it was shown in the previous sections metallization is the key factor for achieving high device performance. Therefore first the resonant device has to be optimized in respect

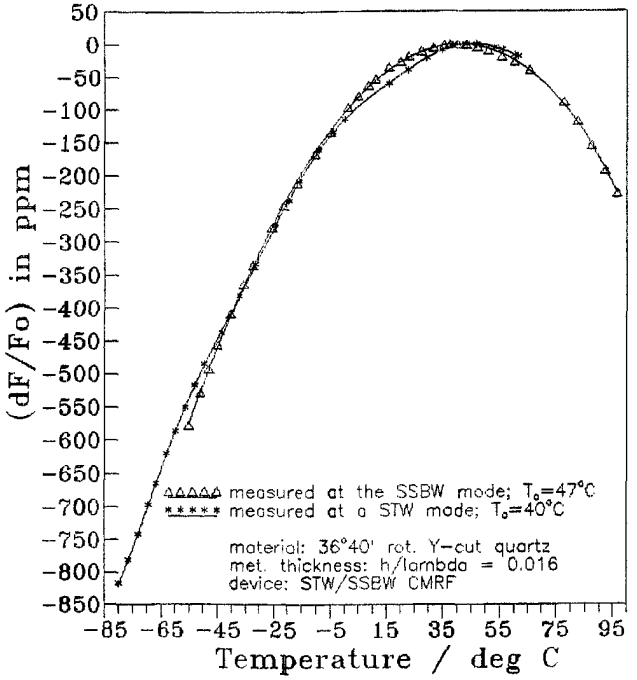


Fig. 33. Temperature behavior of SSBW and STW modes in a CMRF.

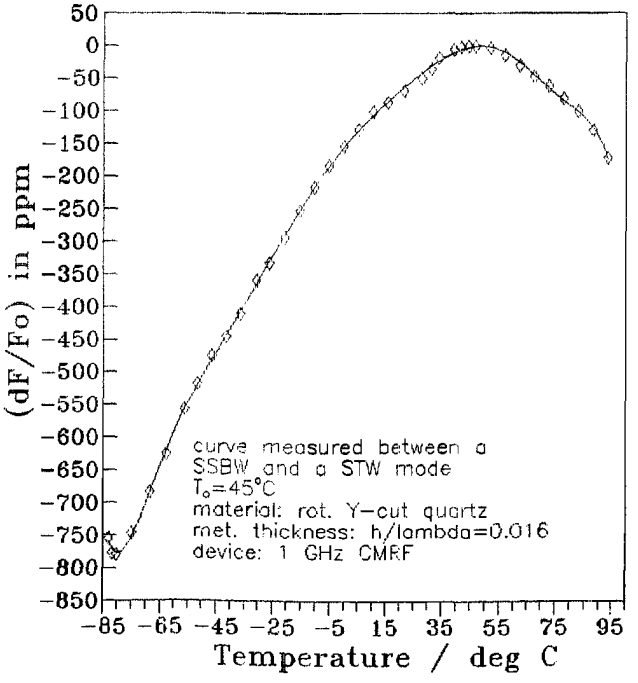


Fig. 34. Temperature dependence of the CMRF at a frequency between the SSBW- and one STW mode

to its electric parameters and then its temperature behavior has to be adjusted with a proper selection of the cut orientation. Metal thicknesses of about 1.6 % which seem to be close to optimum for most SH-wave devices considered in this study yield a close to parabolic temperature behavior with an overall stability comparable to that of SAW resonators on rotated Y-cuts of quartz [26]. Thicker metallization seems to bring a substantial improvement of the overall temperature stability by making the dependence cubic and similar to the AT-cut behavior of bulk wave resonators. Further work is necessary to establish a reasonable compromise between electrical performance and temperature stability of SH-wave resonant devices.

Phase Noise Performance and Short Term Stability of STW Resonator Oscillators

In their 1987 study Bagwell and Bray demonstrated phase noise performance of burned-in metal strip STW resonators comparable to that of state-of-the-art SAW resonators [4]. Here some results of phase noise and short term stability measurements on fixed frequency oscillators stabilized with single and multimode STW resonators of the designs in Fig. 3 and Fig. 8 will be presented. Close to carrier measurements were performed with the HP 5390 A Frequency Stability Analyzer and the HP 8662 A Synthesizer as a reference source.

Fig. 35 is the close to carrier phase noise performance of a 767 MHz fixed frequency oscillator using a single mode device of the design in Fig. 3.

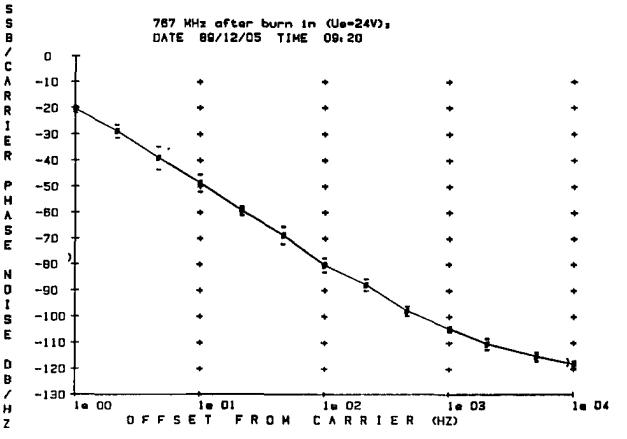


Fig. 35. Close to carrier phase noise of a 767 MHz fixed frequency STW oscillator.

The STW device operated at a loaded Q of about 3000 while the loop power was 24 dBm. The measurement was performed after a continuous operation of the oscillator for a period of 5 days. Since the loop power was high it can be assumed that the STW device was exposed to a self burn-in process. The result is a very smooth phase noise behavior with a typical decrease of 30 dB/decade in the flicker frequency region [26] up to about 1 KHz offset from carrier. The suppression of 20 dBc/Hz at 1 Hz away from carrier for an unsealed STW device which was not exposed to high temperature bakeout, UV-cleaning, high vacuum packaging and other

processes which are essential to the oscillator overall stability [26], is a respectable result.

Fig. 36 presents results from Allan-variance measurements on the same oscillator. The flicker frequency floor begins at a measurement time of about 1 ms, in agreement to the results from Fig. 35, and remains constant at an average value of about 2×10^{-10}/s up to measurement times of 1 s.

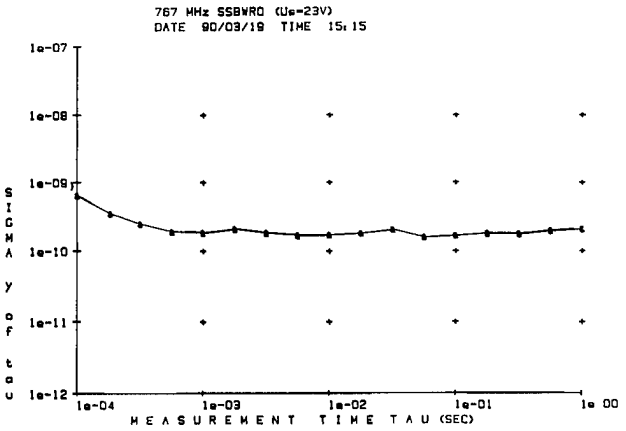


Fig. 36. Results from Allan-variance measurements on the 767 MHz STW oscillator.

The same measurements were repeated with a 1015 MHz STW resonator oscillator operating at a loop power of 10 dBm and with a buffer amplifier at the oscillator output. The STW device was of the multimode design in Fig. 8 and the resonant Q in the loop was about 4000. The close to carrier phase noise suppression of this oscillator, presented in Fig. 37, is by about 5 dB worse than

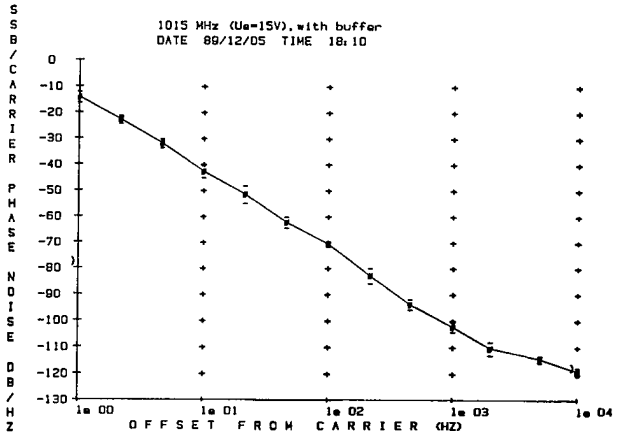


Fig. 37. Close to carrier phase noise of the 1015 MHz STW oscillator.

in the 767 MHz one despite the higher resonant Q. One reason for that could be the fact that the device was not burned-in and the second reason may be due to the buffer amplifier which was found to generate a high harmonic content. The latter may have caused some noise downmixing which deteriorates the overall phase noise performance. Still a flicker frequency noise floor of about 2.5×10^{-10}/s is a good result for a 1 GHz oscillator.

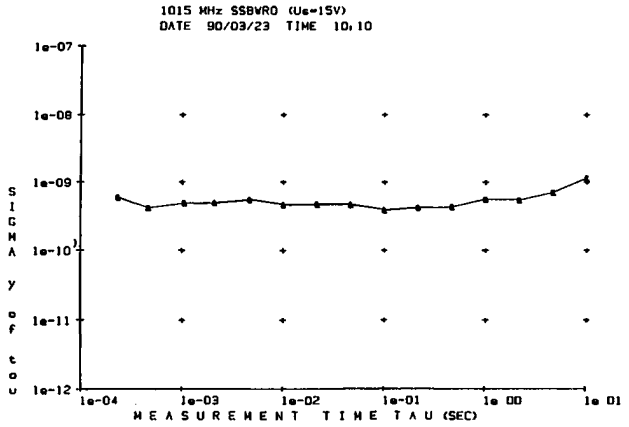


Fig. 38. Results from Allan-variance measurements on the 1015 MHz STW oscillator

Table 2 summarizes results from short term stability measurements on both STW oscillators discussed above compared with results obtained from a SAW oscillator using the 390 MHz device from Fig. 5. It is evident that, despite the much

Meas. time	Short term stability		
	767 MHz STW oscill.	1015 MHz STW oscill.	390 MHz SAW oscill.
0.1 ms	6.4×10^{-10}	4.4×10^{-10}	-
1 ms	2.2×10^{-10}	2.3×10^{-10}	-
10 ms	1.7×10^{-10}	2.5×10^{-10}	1.94×10^{-10}
100 ms	1.5×10^{-10}	2.4×10^{-10}	1.9×10^{-10}
1 s	1.8×10^{-10}	2.7×10^{-10}	3.2×10^{-10}
10 s	2.0×10^{-10}	2.9×10^{-10}	6.6×10^{-10}

Table 2. Numeric results from Allan-variance on STW and SAW resonator oscillators.

higher operating frequency the STW oscillators have a better short term stability in the flicker frequency region than the 390 MHz SAW oscillator.

Fig. 39 is the spectrum of a 1.96 GHz STW oscillator stabilized with the device characterized in Fig. 16. The relatively high device loss was easily compensated for by an amplifier of the type in Fig. 22 which was built with 12 GHz bipolar transistors. As evident from Fig. 39 the oscillator generates a power of about 5 dBm and has

a very clean spectrum.

The results presented in this section and in [15] indicate that STW oscillators in the UHF and the lower GHz range have a respectable phase noise performance and behave even better in the flicker frequency region than SAW oscillators stabilized

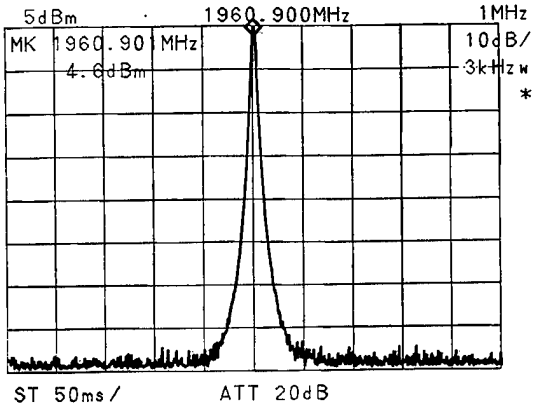


Fig. 39. Output spectrum of a 1.96 GHz STW resonator oscillator.

with metal strip resonators. Since STW resonators have a much higher power handling ability than their SAW counterparts STW oscillators are expected to yield lower thermal noise floors than SAW oscillators of a similar design. In the author's opinion STW oscillators will very soon achieve phase noise performance comparable or better than that of state-of-the-art SAW oscillators.

Summary and Conclusions

This paper has presented results from a study on SH-wave resonant devices and microwave oscillators stabilized with them. Several different designs of single and multimode high-Q and low-Q metal strip resonators and combined mode resonator filters in the frequency range of 750 - 2000 MHz were presented and discussed. Results from stable fixed frequency and tunable oscillators using these devices were compared with those obtained from a SAW resonator oscillator of a similar design. Aspects of frequency trimming of STW resonators, frequency modulation, injection locking, temperature stability and phase noise behavior of fixed frequency and voltage controlled SH-wave oscillators were discussed.

SH-wave resonant devices are more difficult to design than their SAW counterparts especially if the required specifications are tight. They are also more sensitive to fabrication tolerances such as metallization and cut orientation. But if we learn how to design them the efforts will pay off. STW resonators can achieve unloaded Q values as high as 90 % of the material limit at microwave frequencies and can handle orders of magnitude higher amounts of RF power than SAW resonators. Phase noise values comparable with those of state-of-the-art SAW oscillators have already been demonstrated. On the other hand they demonstrate some unique features that cannot be achieved with SAW devices. These features come from the fact that the resonant Q and the group delay can be controlled in very wide limits without a significant increase in insertion loss. This phenomenon offers a great flexibility in the design of voltage controlled oscillators with a wide tuning range and unique properties in respect to wideband frequency modulation and injection locking. In order to fully explore the advantages of SH-wave resonant devices further work regarding temperature stability, the influence of frequency trimming on the temperature behavior, phase noise and aging is necessary.

It is the author's belief that the SH-wave resonator oscillator will become a serious competitor of its SAW counterpart in the near future and will find a variety of applications in wireless data transmission networks, personal communications and local area radio networks.

Acknowledgements

A lot of people have contributed to this study. The author wishes to gratefully acknowledge Prof. I. B. Yakovkin, Drs. A. S. Kozlov and A. L. Asseev from the Institute of Semiconductor Physics in Novossibirsk, Russia for the opportunity to have the SH-wave devices fabricated in their laboratories. Thanks to Yu. Nastaushev, V. Sandirev, M. Dolgov, V. Fishchuk, N. Matveyenkova, T. Belenkova, N. Filipova, L. Korzhova, M. Gubina, S. Sudarikov, and G. Strelnikova from the same institute for their patience in fabricating the devices. Special thanks also to S. Denisenko and M. Kachanova for their efforts in realizing the EBL devices and Dr. V. Aliev from the same institute who performed the frequency trimming. The author expresses his appreciation to Dr. P. M. Smith and P. J. Edmonson from McMaster University in Hamilton, Ontario, Canada for their allowance to use the COM-analysis software and to D. Mihailov and A. Stratieva from the Institute of Solid State Physics in Sofia, Bulgaria for expert preparing the quartz substrates. Finally special thanks to T. Thienel from the Technical University in Dresden, Germany for his help in the design of the printed circuit boards for the microwave oscillators and M. Knittel from the same University for performing the short term stability measurements on the SAW oscillator.

References

[1]. K. H. Yen, K. L. Wang, and R. S. Kagiwada, "Efficient Bulk Wave Excitation on ST-cut Quartz", Electronics letters, 13, (1), 1977, pp. 37-38.

[2]. T. I. Browning and M. F. Lewis, "New Family of Bulk Acoustic Wave Devices Employing Interdigital Transducers", Electronics Letters, 13, (5), March 1977, pp. 128-130.

[3]. K. F. Lau, K. H. Yen, A. M. Kong, K. V. Rousseau, "SBAW Oscillators at 3 to 5 GHz

Range", IEEE Proc. 1980 Ultrasonics Symposium, pp. 263-266.

[4]. T. L. Bagwell and R. C. Bray, "Novel Surface Transverse Wave Resonators with Low Loss and High Q", IEEE 1987 Proc. Ultrasonics Symposium, pp. 319-324.

[5]. B. A. Auld, J. J. Gagnepain, and M. Tan, "Horizontal Shear Waves on Corrugated Surfaces", Electronics Letters, 12, (24), November 1976, pp. 650-651.

[6]. D. F. Thompson and B. A. Auld, "Surface Transverse Wave Propagation Under Metal Strip Gratings", IEEE Proc. 1986 Ultrasonics Symposium, pp. 261-266.

[7]. B. A. Auld and D. F. Thompson, "Temperature Compensation of Surface Transverse Waves for Stable Oscillator Applications", IEEE Proc. 1987 Ultrasonics Symposium, pp. 305-312.

[8]. D. L. Lee, "S-Band SSBW Delay Lines for Oscillator Applications", IEEE Proc. 1980 Ultrasonics Symposium, pp. 245-250.

[9]. P. S. Cross, W. R. Shreve and T. S. Tan, "Synchronous IDT SAW Resonators with Q Above 10,000", IEEE 1979 Proc. Ultrasonics Symposium, pp. 824-829.

[10]. I. D. Avramov, "1 GHz Low Loss Coupled Resonator Filter Using Surface Skimming Bulk Waves and Bleustein-Gulyaev Waves", Electronics Letters, 5, (27), February 1991, pp. 414-415.

[11]. C. A. Flory and A. E. Baer, "Surface Transverse Wave Mode Analysis and Coupling to Interdigital Transducers", IEEE Proc. 1987 Ultrasonics Symposium, pp. 313-318.

[12]. C. K. Campbell and C. B. Saw, "Analysis and Design of Low-Loss SAW Filters Using Single-Phase Unidirectional Transducers", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC-34, No. 3, May 1987, pp. 357-367.

[13]. I. D. Avramov and A. S. Kozlov, "Design of Combined Mode Resonator Filters Based on Surface Transverse and Surface Skimming Bulk Waves", Proc. 10th European Conf. on Circuit Theory and Design ECCTD-91, 2-6 Sept. 1991, Lyngby, Denmark, pp. 729-738.

[14]. T. Morita, M. Tanaka, Y. Watanabe, K. Ono and Y. Nakazawa, "UHF Two-port SAW Resonator", TOYO's Technical Bulletin, No. 37, 1987, pp. 1-11.

[15]. I. D. Avramov, "1 GHz Voltage Controlled Oscillator Using a SSBW/BGW Combined Mode Resonator Filter", Proc. 45th Annual Symp. on Frequency Control, 1991, pp. 230-238.

[16]. I. D. Avramov, "Surface Transverse Wave Based 1 GHz FM Transceiver for Local Area Radio Networks", Proc. 5th Conference "Acoustoelectronics'91", 10-13 Sept., 1991, Varna, Bulgaria, pp. 60-69.

[17]. M. Taslakov, I. Avramov, "A Temperature Compensated 1 GHz STW Oscillator", IEEE Proc. 1991 Ultrasonics Symposium, (in press).

[18]. I.D. Avramov, "High Q Metal Strip Resonators Using a SAW Design", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Vol. UFFC-37, No.6, pp. 530-534.

[19]. E. Bigler, B. A. Auld, E. Ritz and E. Sang, "Design and Test of Surface Transverse Wave (STW) Resonators on Quartz", Proc. 5th Conference "Acoustoelectronics'91", 10-13 Sept. 1991, Varna, Bulgaria, pp. 8-22.

[20]. D. F. Thompson and A. R. Northam, "High Frequency, Low Loss Fundamental Mode STW Resonators Used for Stable Oscillator Design", IEEE Proc. 1991 Ultrasonics Symposium, (in press).

[21]. I. D. Avramov, "Aspekte der Elektrischen Optimierung von Einzelmoden- und Multimodenoszillatoren auf der Basis von Verzoegerungsleitungen mit Akustischen Oberflaechenwellen", Dissertation zur Erlangung des akademischen Grades Dr.-Ing., Technische Universitaet Dresden, Maerz, 1988.

[22]. I. D. Awramow, "An Active Electrical RC-Circuit for Stable SAW Oscillators in the UHF Range", Proceedings of the Conference on Electronic and Piezoelectronic Components and Circuits EPCC-1988, July 5-7, 1988, Liberec, Czechoslovakia, pp.3-10.

[23]. M. Hikita, T. Tabuchi, N. Shibagaki, T. Akagi, Y. Ishida, "New High-Performance and Low-Loss SAW Filters Used in Ultra Wideband Cellular Radio Systems", IEEE Proc. 1991 Ultrasonics Symposium, (in press).

[24]. R. Adler, "A Study of Locking Phenomena in Oscillators", Proc. IRE, Vol. 34, No. 6, June 1946, pp. 351-357.

[25]. L. Spassov, Private consultation, 1991.

[26]. 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.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

AN X-BAND, 2.5 WATT CONTINUOUS WAVE DIELECTRIC RESONATOR OSCILLATOR FOR FUTURE MILITARY SYSTEMS

MUHAMMAD MIZAN

U.S. Army LABCOM, Electronics Technology & Devices Laboratory, SLCET-MP-C, Fort Monmouth, N.J. 07703-5601.

ABSTRACT

A 9 GHz Dielectric Resonator Oscillator (DRO) generating a continuous wave (CW) power output of 2.5 watts at room temperature has been designed and fabricated by incorporating a high power GaAs FET amplifier and a Murata Erie dielectric resonator (DR) in a feedback loop oscillator configuration. The DC to RF conversion efficiency of this DRO is about 11%. The oscillator exhibited a frequency stability of less than 130 ppm over the temperature range of -50°C to + 50°C. Measured single-sideband phase noise levels were -105 dBc/Hz and -135 dBc/Hz at 10 kHz and 100 kHz carrier offset frequencies, respectively. The output power varied from +35 dBm (3.2 watts) at -50°C to +33 dBm (2 watt) at +50°C. The DRO output power is then fed into a single-stage GaAs FET amplifier, resulting in a total power output of 6.5 watts at X-band. The overall efficiency of the 6.5 watt DRO unit is about 12%. The pulse power capability of the DRO unit was also investigated. The unit can deliver 10 watts of pulse power for a pulse duration of one microsecond or less.

INTRODUCTION

The high power, high efficiency, light weight, small size, temperature stable, DRO reported will aid in satisfying the frequency stability and the output power requirements of an Army man-portable radar transponder. The man-portable radar transponder is deployed by long-range Special Operation Forces (SOF) patrols for use as a forward air controller/strike aircraft rendezvous aid. When the transponder is interrogated by an incoming aircraft, it responds either at X- or Ku-band frequency.

The X-band, 6.5 watt DRO unit contains only two GaAs FET amplifiers. One of which is used in the oscillator circuit and the other one is used for amplifying the DRO output signal. Circuits employing fewer active devices are very cost effective, easy to reproduce, and are more reliable because of less parts count. An overall block diagram of the high power DRO unit is shown in Figure 1.

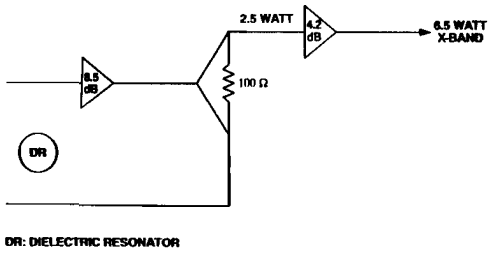


Figure 1. DRO Unit Block Diagram

The peak power capability of the DRO unit was studied in detail for the transponder application. A minimum power output of 10 watts was measured over a wide temperature range. This power is needed to drive an output power stage that contain four 10 watt GaAs FET amplifiers in parallel to produce 25 watts of X-band peak power over the military temperature range. The 25 watt pulsed power, when fed into a microstrip patch antenna having 9 dB gain, will correspond to more than 200 watts of effective radiated power (ERP). The transponder range requirement of 100 kilometer is easily achieveable with the stated propagating energy. Link equations were developed and solved to calculate the required transmit power. Some of the parameters used in the link equations are transponder and interrogator antenna gain, path loss, receiver sensitivity, multiple path loss, etc. The transponder transmitter power for X- and Ku-band were found to be 18 watts and 13 watts, respectively.

The sophisticated state-of-the-art power DRO discussed in this paper will provide enhanced system performance for the transponder, and can

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

be configured for use in future systems to meet advanced DoD requirements for smart munitions, EW survivable all-weather surveillance, and target acquisition systems. A Ku-band, high power DRO is still under construction. The goal is to produce twice the power at half the frequency and use a ISIS diode frequency doubler circuit to obtain a Ku-band signal.

OSCILLATOR DESIGN

The feedback loop oscillator is designed by combining a two-port GaAs FET amplifier with a two-port transmission-mode dielectric resonator. The amplifier and the DR enclosed in a metal cavity are treated separately. Dielectric resonator oscillators and acoustic resonator based oscillators designed this way have been shown to have excellent phase noise and frequency stability [1-5]. The residual phase noise of the oscillator's components should be evaluated to avoid overall phase noise degredation due to low quality/faulty RF components. The absolute phase noise of the oscillator, within the resonator's bandwidth, can be calculated by knowing the amplifier/resonator residual noise and the loaded Q of the resonator [6-7]. The resonator used in the oscillator design was procured from Murata Erie. It has a dielectric constant of 24.2 and a temperature coefficient of zero parts per million per degree centrigrade. The dielectric resonator had an unloaded Q of 22,000 at X-band. The dimension of the resonator are 7.07 mm in diameter and 3.14 mm thick.

CAVITY DESIGN

The cavity dimensions are chosen such that the $TE_{01\delta}$ mode of the resonator is well separated from the cavity modes. The amount of coupling and the positioning of the DR in the cavity are very critical in obtaining an optimum performance. A spacer made of low dielectric constant material is used for mounting the DR inside the cavity. Improper mounting will degrade the loaded Q, and increase vibration sensitivity. The cavity design is shown in Figure 2.

The cavity is made of brass and is cylindrical in shape. The cavity mode in the absence of the resonator is analyzed using the cylindrical cavity resonant frequency formula. The calculated modes are then varified by network analysis measurements. Figure 3 shows the air filled cavity modes. These mode are identified as TE_{111}, TM_{010} and TM_{011} operating at 8.765 GHz, 8.985 GHz and 10.415 GHz respectively. These modes were excited by a 50 ohm microstrip line, located at the bottom of the cavity. Figure 4 shows the mode spectra in the presence of the DR.

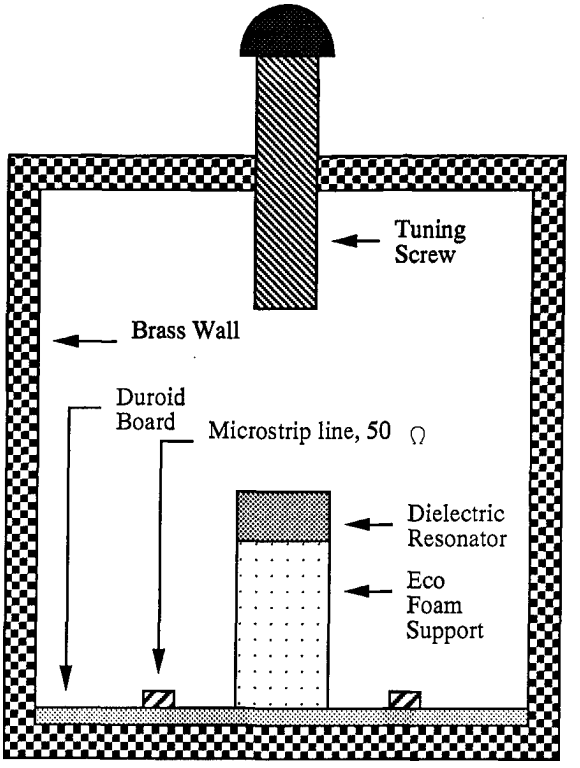


Figure 2. Two Port Cavity Design

As expected, the cavity modes had shifted lower in frequency due to the presence of high dielectric constant material, (i.e. the DR). Figure 4 also shows the mode separation between the $TE_{01\delta}$ mode and the cavity modes.

The resonance frequency of the DR is tuned via fringing fields perturbations by a metallic screw. The cavity modes are also very sensitive to the tuning screw. Figure 5 illustrates the behavior of the cavity modes as the screw is plunged into the cavity from the top. The TM_{011} mode is extremely sensitive to the tuning screw and completely overlaps the $TE_{01\delta}$ mode at some tuning position. Figure 5 suggests that one has to be very careful with the tuning screw, carelessness could result in operating on an undesired cavity mode.

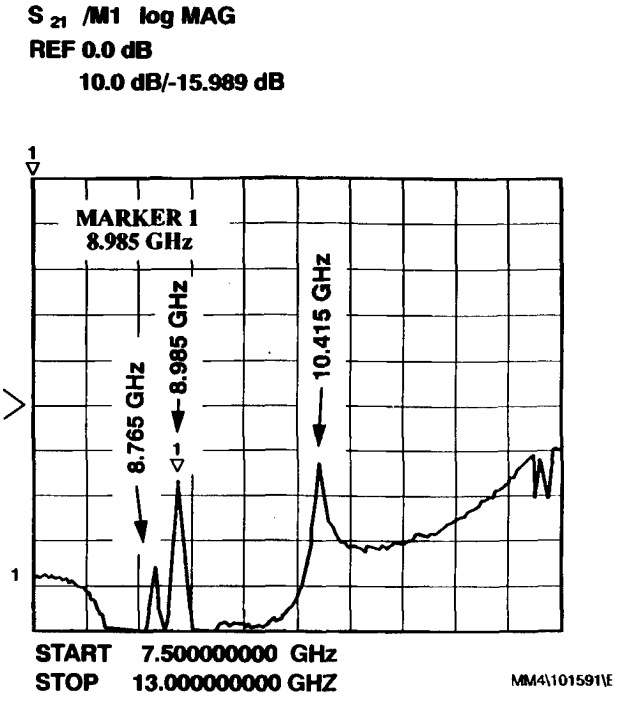


Figure 3. Cavity Modes Excited by Microstrip Line

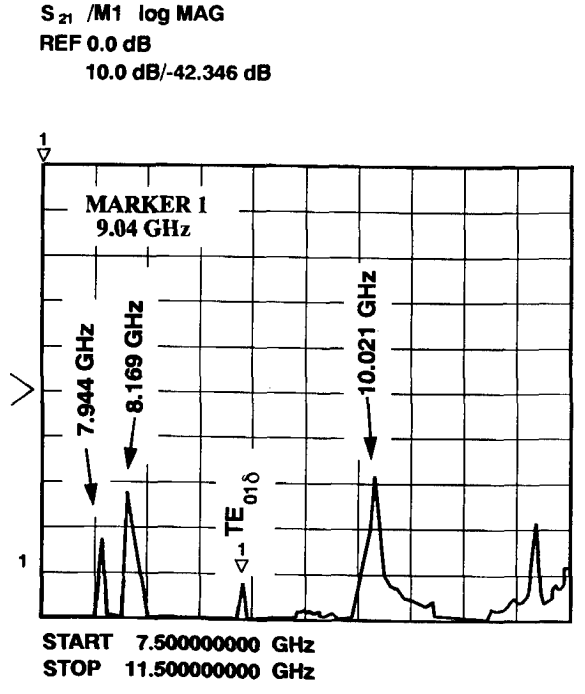


Figure 4. Cavity Modes and the $TE_{01\partial}$ Mode of the DR

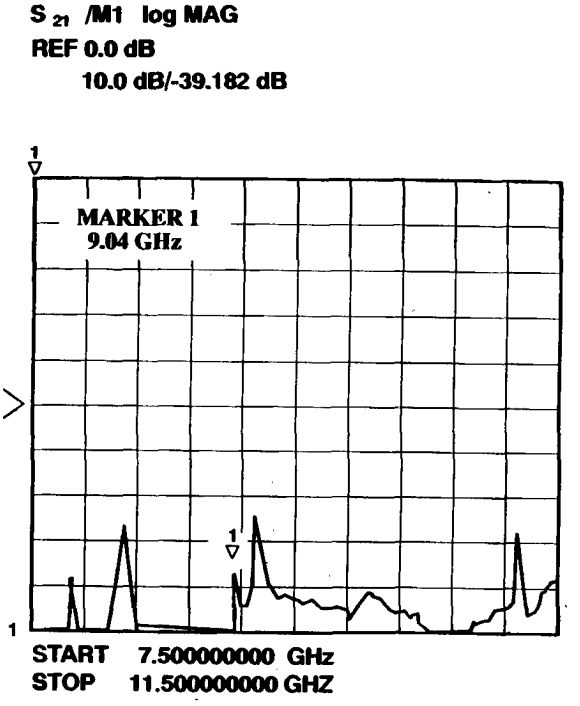


Figure 5. Behavior of Cavity Modes as a Function of Tuning Screw

The result of loaded Q measurement is shown in Figure 6. A 3 dB bandwidth of 1.49984 MHz at a center frequency of 9.041GHz corresponds to a Q of 6,028. The insertion loss was about 4 dB. A loaded Q of 15,000 was measured with 20 dB

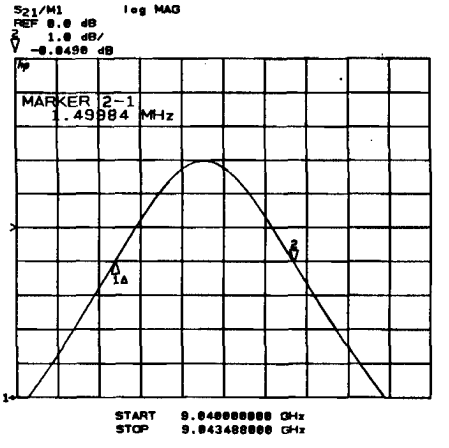


Figure 6. Measured 3 dB Bandwidth of the Resonator. Corresponding loaded Q is 6,028 with 4 dB insertion loss.

insertion loss. Three loop amplifiers would have been required to over come this insertion loss.

AMPLIFIER DESIGN

The amplifier is designed with a commercially available, internally matched, Fujitsu FLM 8596-8C GaAs field effect transistor. The 1 dB gain compression point of the amplifier is about 8 watts. The amplifier has bandwidth of approximately 1 GHz with maximum gain at 9.041 GHz, which is the puck frequency. The residual phase noise of the high power GaAs FET amplifier was calculated to be -130dBc/Hz at 100 Hz offset frequency. Bias network and the tuning stubs were fabricated on a 10 mil thick RT 5880 duroid board. Tuning was needed to achieve maximum gain at the puck frequency. Design precautions were taken to avoid any bias circuit oscillation.

MEASURED RESULTS

The microstrip circuit board layout of the DRO is shown in Figure 7. The amplifier, the Cavity and the Wilkinson power divider were designed and tested individually for optimum performance.

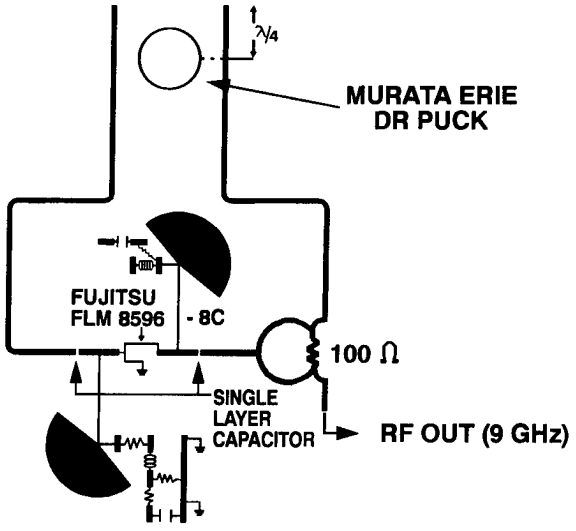


Figure 7. Dielectric Resonator Stabilized Single-Stage Feedback-Loop Microstrip Circuit Board Layout

The DRO's output power, center frequency and the spectral purity was measured by a HP 8566B spectrum analyzer. The test setup and the measured data are shown in Figures 8 and 9, respectively. A 20 dB pad was used between the DRO output port and the spectrum analyzer. The output power is 33.9 dBm (2.5 watt). (12.90 dBm from the plot of Figure 8 + 20dB pad + 1dB cable loss =33.90 dBm). The oscillator has no spurious oscillation and the second harmonic power level is 35 dB below the carrier power level.

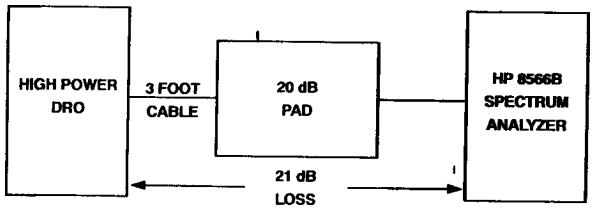


Figure 8. Power Output Measurement Test Setup

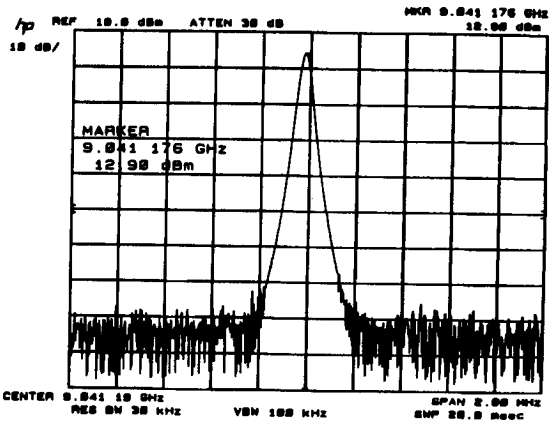


Figure 9. Measured DRO Output Power. (12.9 dBm+ 20 dB Pad+1 dB Cable loss=33.90 dBm)

The center frequency and the output power variation with temperature from -50°C to +50°C are given in Figures 10 and 11, respectively. The measurement started after stabilizing the environmental chamber at -50°C for 1 hour. The total frequency drift is less than 130 ppm and is well within the transponder specification. The frequency vs temperature characteristic was reinvestigated under low output power (20 dBm)

condition. The result of this measurement is shown in Figure 10. The highest frequency drift observed in this case is about 80 ppm. This is a factor of 1.6 improvement over the previous data. Non-linearities in the active device may be responsible for the additional degredation observed under high power condition. The output power varied from +35 dBm (3.2 watts) at -50°C to +33 dBm (2 watt s) at +50°C. The room temperature power is 33.9 dBm (2.5 watts).

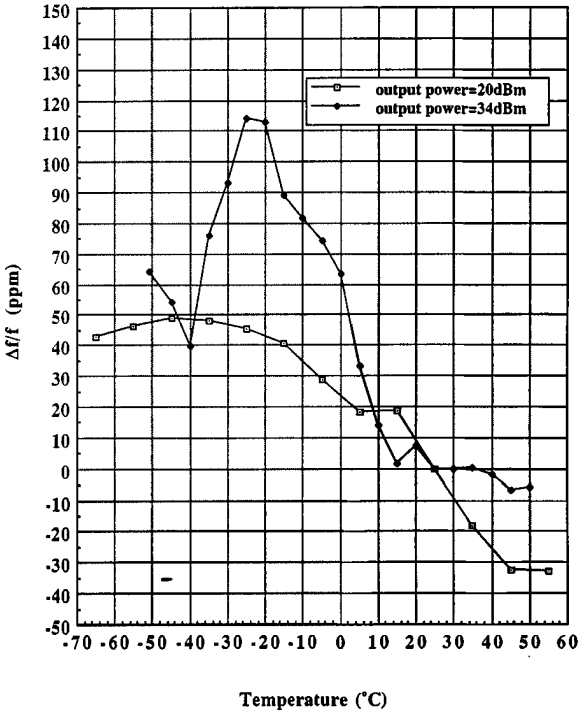


Figure 10. Frequency vs Temperature Plot

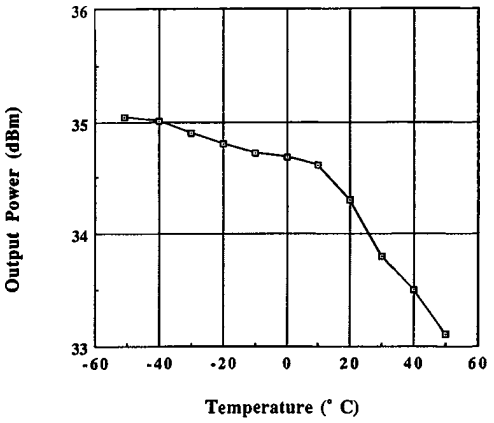


Figure 11. High Power DRO Output Power vs Temperature Data

The absolute phase noise performance of the high power DRO was measured with a HP 3047A noise measurement system. The test setup is illustrated in the graph of Figure 12. The DRO test signal was downconverted to 151 MHz signal by mixing it with a low noise high overtone bulk acoustic resonator (HBAR) oscillator. The 151 MHz signal is then phase locked to an HP 8662A frequency synthesizer driven by an external 10 MHz VCXO. This arrangement was necessary due to a lack of a second source with VCO capability. The single sideband phase noise plot is shown in Figure 13. The synthesizer noise starts to show up at 100 kHz offset frequency. The spurs between 100 kHz and 1 MHz are due to mixer products. The spurs between 60 Hz and 1 kHz are due to AC line spurs.

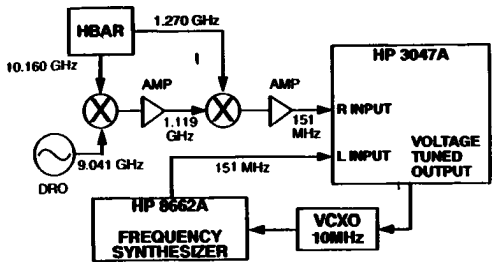


Figure 12. Phase Noise Measurement Test Setup

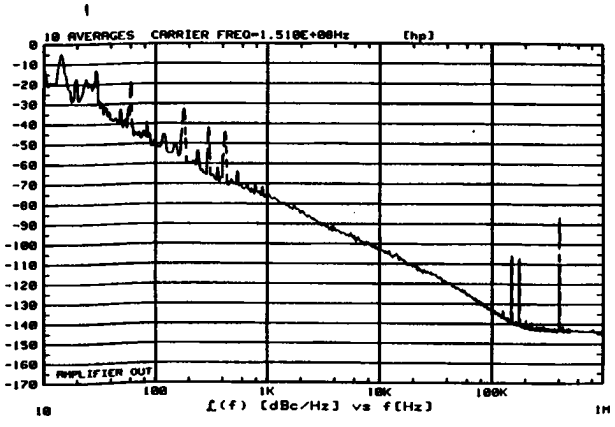


Figure 13. Absolute Phase Noise of High Power DRO

Finally the output power was measured again after adding the single-stage amplifier at the output of the DRO. The measurement test setup is exactly the same as the one described in Figure 8. The result of this measurement is shown in Figure 14. The total output power is 38.2 dBm (6.5 watt).

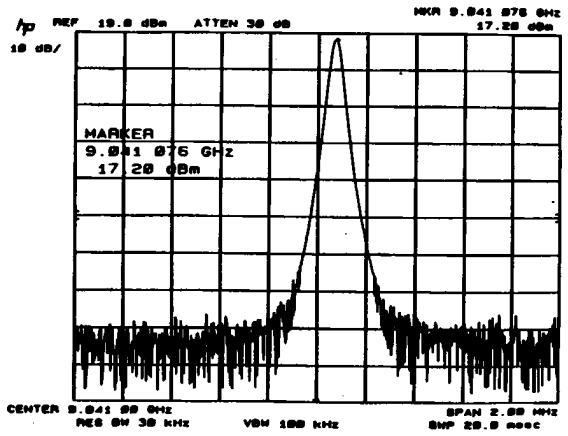


Figure 14. Output Power Measured for the DRO Unit (17.2 dBm+21 dB Line loss=38.2 dBm)

CONCLUSIONS

A detailed study of a 2.5 watt, X-band DRO has been reported in this paper. The spectral purity and the frequency stability are considered excellent for such a high power DRO. The overall unit provides 6.5 watt CW output power at 9 GHz and has 11.82 % DC to RF conversion efficiency. The unit is small and draws 4.23 amp at 10 VDC. It can also be used for pulse power application.

REFERENCES

[1] M. Mizan and R. McGowan, "Ultra-Low Noise, 8.3 GHz Dielectric Resonator Oscillator", Proceeding of 45th Annual Symposium on Frequency Control 1991, pp. 693-699 IEEE Catalog No. 91CH2965-2.

[2] G. K. Montress and T. E. Parker, "Design Techniques for Achieving State-of-the-Art Oscillator Performance", Proceeding of 44th Annual Symposium on Frequency Control 1990, pp. 522-535. IEEE Catalog No. 90CH2818-3.

[3] M. Mizan, R. McGowan, T. Lukaszek and A. Ballato, "Determination of the Limiting Factors in the Absolute Phase Noise of an L-Band Dielectric Resonator Oscillator", Proceeding of 45th Annual Symposium on Frequency Control 1991, pp. 687-692 IEEE Catalog No. 91CH2965-2.

[4] J. Dick and Jon Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric ring resonator for Ultra-Low Noise", Proceeding of 43rd Annual Symposium on Frequency Control 1989, pp. 107-114. IEEE Catalog No. 89CH2690-6.

[5] R. A. Jelen and M. M. Driscoll, "Multiple Frequency Source for Air Defense Radar", R&D Technical Report SLCET-TR-87-0718-F, September 1990.

[6] T. E. Parker, "Characteristics and sources of Noise in Stable Oscillators", Proceeding of 41st Annual Symposium on Frequency Control 1987, pp. 99-110. IEEE Catalog No. 87CH2427-3

[7] F. L. Walls and A. E. Wainwright, "Measurement of the Short-Term Stability of Quartz Crystal Resonators and the Implications for Crystal Oscillator Design and Applications", IEEE Trans. Instrum. Meas., vol. IM-24, pp. 15-20, Mar. 1975.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PIECE-WISE LINEAR MODELING OF SOLID-STATE VARACTOR-TUNED MICROWAVE OSCILLATORS

Ali Behagi, Associate Professor, Penn State Harrisburg,
Middletown, PA. 17057
Stephen Turner, MAXTECH, Inc., State College, PA. 16802

Abstract

In this paper, the methodology for the piece-wise linear modeling of a varactor-tuned microwave oscillator is developed. The methodology is then applied to the X-band VCOs designed for this purpose and the component values of the lumped models are evaluated. The measured and calculated values are compared with the results obtained from computer simulations.

I-Introduction

A varactor-tuned oscillator uses the voltage-controlled capacitance of a varactor diode for electronic tuning. A simplified block diagram of a negative-resistance varactor-tuned microwave oscillator is shown in Fig. 1.

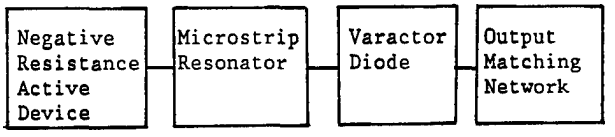


Fig.1-Varactor-Tuned Oscillator Block Diagram

The negative-resistance active device consists of the transistor with its terminating network which provides the negative resistance necessary for oscillation. The microstrip resonator is the network that determines the frequency of oscillation. The output matching network provides the necessary impedance match between the oscillator output impedance and the standard 50 Ohm load.

At microwave frequencies, the parasitic capacitances of the transistor provide some or all the feedback needed for oscillation and any transient excitation due to noise in the circuit will initiate the oscillation.[1]

II-A Negative-Resistance X-band VCO Design

The negative-resistance design basically consists of selecting a transistor in an oscillator topology that provides the required output power. In this paper, a common-collector topology with a packaged transistor, NE64587, is utilized. Following Niehenke's Design [2], an emitter inductance to ground, consisting of a short length of 50 Ohm microstrip line terminated by a chip capacitor, was used as the terminating network. The impedance looking at the base of the transistor was inductive at the X-band frequencies and the magnitude of its reflection coefficient remained relatively constant for the 5 to 8 percent tuning bandwidth achieved using silicon or gallium arsenide varactor diodes. Due to the inductive impedance of the transistor, a length of 50 Ohm microstrip line was used to parallel resonate the transistor. Both Si and GaAs varactors, in abrupt and hyperabrupt types, were used for the tuning purpose; but only the case of the VCO with silicon abrupt varactor is reported in this paper. A small tuning inductance was added to series resonate the varactor at the oscillator center frequency. A quarter-wave transformer, consisting of two sections of microstrip lines, was used to match the output impedance of the oscillator to the standard 50 Ohm load. A simplified layout of the VCO is given in Fig. 2.

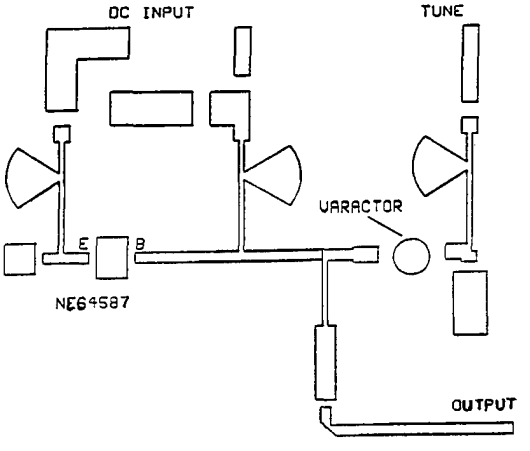


Fig. 2 Simplified Microstrip Layout of VCO

0-7803-0476-4/92 $3.00 © 1992 IEEE

III-The VCO Equivalent Circuit

The equivalent circuit of the VCO within the tuning range of the oscillator can be represented by the following RLC circuit:

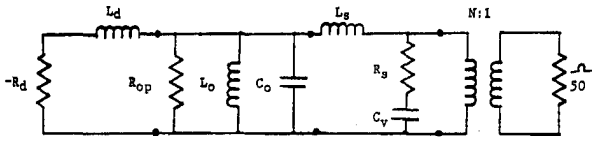


Fig. 3-The VCO equivalent circuit

The microstrip resonator, represented by R_o, L_o and C_o is assumed to have fixed components while the transistor and the varactor, which are non-linear active devises, are assumed to have frequency or voltage dependence variable components. The negative resistance $-R_d$ and the inductance L_d represent the active device with its terminating network while R_s, L_s and C_v represent the varactor with its series resistor and inductor. The coupling of the VCO to the 50 Ohm output is represented by an impedance matching transformer with the turns ratio of N to 1.

Combining the transistor and the resonator inductances into a single inductance L_t and transforming the varactor into an equivalent parallel RLC network, the VCO can then be represented by the following parallel RLC circuit.

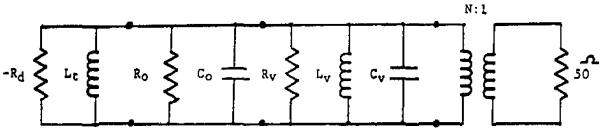


Fig. 4 A Simplified VCO Equivalent Circuit

In the following sections, a methodology for the component evaluation of the model is presented. It is clear from Fig. 4 that the frequency of oscillation at a given tuning voltage Vn is given by:

$$f_n = \frac{1}{2\Pi\sqrt{L_n(C_o+C_{vn})}} \qquad (1)$$

Where the index n denotes that the voltage V_n changes with n = 1,2,... and L_n is the parallel combination of L_{tn} and L_{vn}.

IV-Piece-wise Linear Modeling Based on Varactor Specifications

Clearly due to frequency dependency and non-linearity of the active elements used in the VCO design, a fixed lumped element representation is not appropriate. However, a piece-wise linear model, which represents the VCO at several critical frequencies, is more realistic. In this paper, the voltage and capacitance data provided by the varactor manufacturers are used as the basis for VCO piece-wise linear modeling.

Varactor diodes are fabricated using silicon or gallium arsenide and they are characterized as having either an abrupt or a hyperabrupt junction. The capacitance variations measured for two types of junction profiles[3] are shown in Fig. 5.

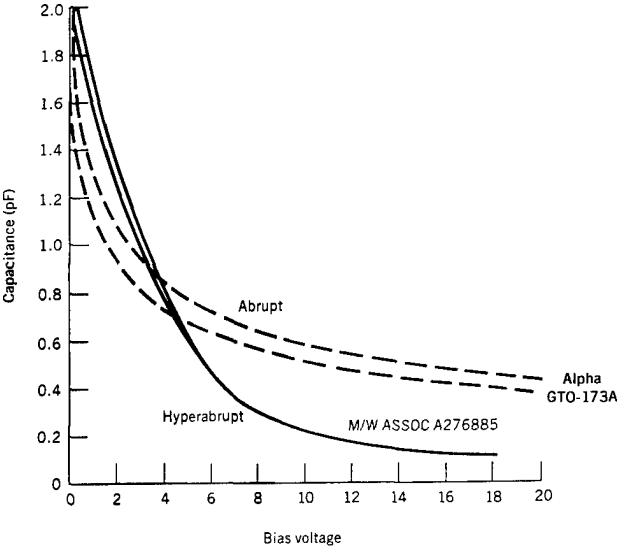


Fig. 5. Capacitance variation for varactor diodes

Most manufacturers specify the capacitance of the varactors at 0, 4 and 30 volts. They also specify the varactor Q factor at 50 MHz. For piece-wise linear modeling, in addition to the varactor data, it is required that the VCO frequencies and the external and resonator Q factors be measured at the three specified varactor bias voltages. It is also required that the oscillator frequency and the resonator Q factors be measured when the varactor is removed from the VCO.

V-Determination of the elements of the model

The individual elements of the piece-wise linear model can be evaluated as follows:

(a) determination of C_o

From equation (1), assuming that L_n remains essentially constant within the tuning range, the fractional tuning ratio referenced to f_1, may be expressed as:

$$t=\frac{f_3-f_1}{f_1}\cong\sqrt{\frac{C_o+C_{v1}}{C_o+C_{v3}}}-1 \qquad (2)$$

The resonator capacitance C_o can then be calculated from:

$$C_o=(\frac{r-1}{(2+t)\,t}-1)\,C_{v3} \qquad (3)$$

where r is the varactor capacitance ratio at bias voltages V_1 and V_3 and t is given by equation (2).

(b) Calculation of the varactor equivalent resistance and inductance:

In varactor-tuned oscillators, the primary factor limiting the tuning range is the resistive loading effects of the varactor diode. The series resistance of the varactor diode is given by:

$$R_{sn}=\frac{X_c}{Q_{vn}}=\frac{1}{2\Pi f_n C_{vn} Q_{vn}} \qquad (4)$$

Where Q_{vn} is the varactor Q factor at the VCO frequency. The equivalent parallel resistance of the varactor diode can be approximated[4] by:

$$R_{vn}=R_{sn}Q_{vn}^2(C_{v1}/C_{vn})^2 \qquad (5)$$

and the varactor Q factor, which is normally given at 50 MHz and 4 volts bias voltage, is calculated from:

$$Q_{vn}=(50MHz/f_nMHz)\,Q_{50} \qquad (6)$$

Now by assuming that the varactor circuit also resonates at the VCO frequency, its inductance is obtained from the following equation:

$$L_{vn}=\frac{1}{(2\Pi f_n)^2 C_{vn}} \qquad (7)$$

(c) Determination of transistor equivalent inductance

From the specified varactor bias voltages and the corresponding VCO frequencies, the transistor equivalent inductance L_t can be obtained from equation (8) by calculating L_n and L_{vn} from equations (1) and (7) respectively:

$$L_{tn}=\frac{L_{vn}L_n}{L_{vn}-L_n} \qquad (8)$$

(d) Evaluation of equivalent load resistance and transformer turn ratio

From an external Q factor measurement[5], the equivalent load resistance can be determined from equation (9):

$$R_{1n}=Q_{En}(2\Pi f_n L_n) \qquad (9)$$

Where Q_{En} is the VCO external Q factor measured at the bias voltage V_n. The corresponding transformer turn ratio is then calculated from:

$$N_n=\sqrt{\frac{R_{1n}}{50}} \qquad (10)$$

(e) Calculation of the resonator equivalent resistance

From the measurements of the VCO resonator Q factor,[3] the equivalent series resistance of the resonator can be calculated from equation (11):

$$R_{os}=\frac{X_{C_o}}{Q_o}=\frac{1}{Q_o(2\Pi f_o C_o)} \qquad (11)$$

Where f_o is the oscillator frequency when the varactor is removed and Q_o is the resonator Q of the microstrip resonator. The equivalent parallel resistance of the resonator then becomes:

$$R_{op}=Q_o^2 R_{os} \qquad (12)$$

(f) Calculation of the transistor negative resistance

Once the load, the varactor and the resonator equivalent resistors are calculated, the transistor negative resistance can be obtained by utilizing the oscillation condition which results in:

$$R_{dn}=R_{1n}\|R_{vn}\|R_o \qquad (13)$$

This completes the calculation of all the pertinent VCO components at the three specified tuning voltages.

VI-Experimental Results

Four identical X-band VCOs using four different type varactors were designed according to the procedure discussed in section II of the paper. Only the experimental results on the silicon abrupt VCO is reported here. The varactor data (Metellics MVS34069-T55) and the VCO frequency and Q factor measurements are given in Table 1.

Table 1. Varactor and VCO Measurements

Parameter	Specified or Measured Data			Units
Varactor Bias Voltage	V_1=0	V_2=4	V_3=30	Volts
Varactor Capacitanee	C_{v1}=2.38	C_{v2}=1.15	C_{v3}=.559	pF
VCO frequency	f_1=9.05	f_2=9.176	f_3=9.699	GHz
VCO External Q	Q_{E1}=46	Q_{E2}=48	Q_{E3}=54	-
VCO Resonator Q	Q_o=32	Q_o=32	Q_o=32	-

From Table 1, the tuning range and the capacitance ratio are:

t=0.0717 and r= 4.257

By removing the varactor, the frequency of oscillation was measured to be f_o = 8.6 GHz. The lumped elements, representing the VCO piece-wise linear model at the three specified varactor tuning voltages (0, 4 and 30 volts), were calculated using equations developed in section V. The results are given in Table 2.

Table 2 - Calculated elements of the piece-wise linear model

Model Elements	For n=1	For n=2	For n=3	Units
Varactor Voltage	V_1=0	V_2=4	V_3=30	volts
Resonator Capacitance (C_o)	11.7	11.7	11.7	pF
Resonator Series Resistance (R_{os})	.0495	.0495	.0495	Ohm
Resonator parallel Resistance (R_{op})	50.7	50.7	50.7	Ohm
Varactor series Resistance (R_{vs})	.739	1.5	2.93	Ohm
Varactor parallel Resistance (R_{vp})	47	411	3398	Ohm
Varactor parallel inductance (L_{vp})	.13	.26	.48	nH
Varactor specified Capacitance (C_{vn})	2.38	1.15	.559	pF
Transistor Inductance (L_{tn})	26.48	25.7	23.05	pH
Equivalent Load Resistance (R_{tn})	57.5	64.7	72.4	Ohm
Transformer Turn Ratio (N_n)	1.07	1.13	1.20	-

VII-Simulation of the model

Both series and parallel representations of the piece-wise linear model were simulated on the computer and the results reported in Table 3 and Fig. 6. The actual VCO frequency vs tuning voltage is also shown in Fig. 7. The comparison shows that the piece-wise linear model produces less than 1% error within the application range of the model while any of the three fixed element models results in a worst case error of more than 3% within the VCO tuning bandwidth.

Table 3. Comparison of Simulation Results.

Models	VCO freq. (MHz) 0V Bias	VCO freq. (MHz) 4V Bias	VCO freq. (MHz) 30V Bias
Piece-wise linear model	9040	9180	9690
Model at 0v	9040	9470	9690
Model at 4v	8770	9180	9400
Model at 30v	9040	9470	9690
Actual measurement	9050	9176	9699

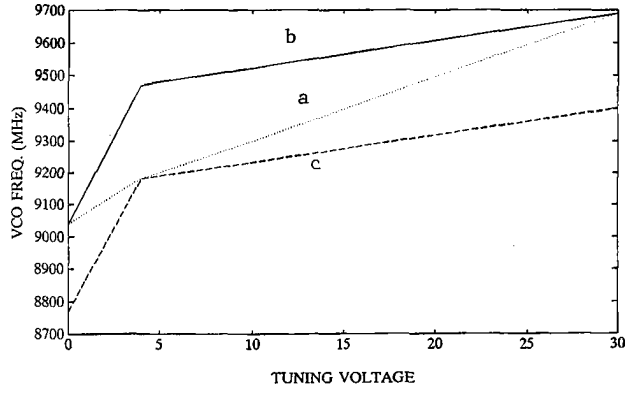


Figure 6. Simulation Results a)Piece-wise Linear Model b)Model at 0/30 Volts c)Model at 4 Volts

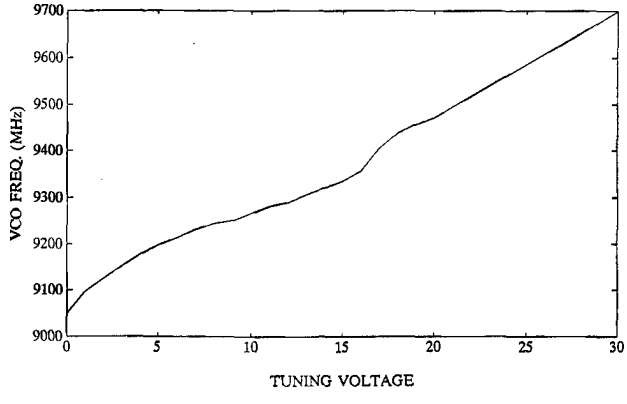


Fig. 7 VCO Frequency vs. Tuning Voltage

VII-SUMMARY

Negative-resistance microwave oscillators use a semiconductor active device which is linear at the initial stage of the oscillation but it becomes non-linear once the oscillation amplitude grows large. Additionally, a varactor-tuned microwave oscillator utilizes a non-linear varactor diode for the tuning purpose which adds to the complexity and non-linearity of the oscillator. As a result, a lumped linear model representing the VCO within its tuning bandwidth will produce large errors while a piece-wise linear model, which has variable component values at different tuning voltages, will greatly reduce the percentage of error.

In this paper, based on varactor specifications and some external measurements, a piece-linear model has been developed which matches the VCO response at several points within the VCO tuning bandwidth. A computer simulation of the piece-wise linear model shows that the model reproduces the VCO response with more accuracy.

References

[1] G. Gonzalez, Microwave Transistor Amplifiers, Prentice-Hall, 1984, chapter 5.

[2] E. C. Niehenke, R. D. Hess, A Microwave Low-Noise X-Band Voltage-Controlled Oscillator, IEEE Transactions on Microwave Theory and Techniques, Vol. 27, No. 12, Dec. 1979, pp 1075-1079.

[3] E. A. Wolff, R. Kaul, Microwave Engineering and Systems Applications, John Wiley & Sons, 1988, p. 355.

[4] S. F. Paik, Q Degradation in Varactor-Tuned Oscillators, IEEE Transactions on Microwave Theory and Techniques, May 1974, pp. 578-579.

[5] A. Behagi, Accurate Calculation and Measurement of Solid State Microwave Oscillator Q factor, Journal of Engineering Technology, vol. 9, No. 1, Spring 1992, pp. 24-26.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

Ultra-Stable Microwave and Millimeter Wave Photonic Oscillators[1]

Ronald T. Logan Jr. and Lute Maleki

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, California 91109

Abstract

Novel photonic realizations of ultra-stable microwave and millimeter-wave sources are discussed. The phase noise performance of an "all-photonic" microwave and millimeter wave oscillator based on a mode-locked semiconductor laser stabilized by a fiber-optic delay-line is investigated.

Introduction

Previously, a dramatic improvement in the phase stability of a widely tunable microwave signal generator was demonstrated using a microwave fiber optic delay-line discriminator [1]. In that work, a 45 dB reduction in the close-to-carrier (< 1 kHz) phase noise of a cavity-tuned microwave oscillator was achieved at 7.8 GHz using a 6.2 km fiber optic delay line in excellent agreement with theoretical predictions. It was shown that the amount of improvement possible was limited by the additive close-to-carrier phase noise of the laser diode transmitter in the microwave fiber optic delay line, and that further improvement was possible with improved fiber optic link phase noise performance.

In the present work, a novel implementation of a delay-line stabilized oscillator using a self-mode-locked semiconductor laser is investigated. The mode-locked laser has much-reduced additive RF phase noise, compared to the previous laser diode transmitters, yielding improved performance for the stabilized oscillator output. In addition, the conventional microwave oscillator of the previous implementation is eliminated in the new scheme, replaced by the self-oscillating mode-locked laser diode to yield an "all-photonic" microwave/millimeter-wave oscillator. Also, an inherent RF frequency-multiplication is obtained for the detected mode-locked laser output.

The phase noise performance of the self-mode-locked photonic oscillator is calculated using the previously derived theory [1], based on RF phase noise measurements of mode-locked semiconductor laser diodes.

Phase Noise Performance Limitations

To motivate the investigation of the photonic oscillator, we briefly revisit the considerations affecting the phase noise performance of the fiber optic delay line stabilized oscillator. The block diagram of the fiber optic delay line stabilized microwave oscillator described previously [1] is as shown in Figure 1.

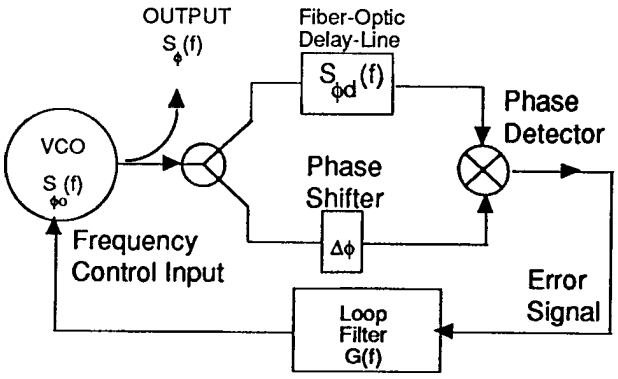


Figure 1. Block diagram of fiber optic delay line stabilized oscillator.

The phase noise for the stabilized output in the limit of large loop gain is well approximated by [1]

$$S_\phi(f) = \frac{1}{(2\pi\tau)^2}\,\frac{S_{\phi d}(f)}{f^2} \qquad (1)$$

where τ = nL/c is the delay time for the microwave signal in the optical fiber of length L and group

[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.

0-7803-0476-4/92 $3.00 © 1992 IEEE

index n, c is the velocity of light in vacuum, f is the offset frequency from the carrier, and $S_{\phi d}(f)$ is the additive double-sideband phase noise power spectral density of the delay line.
First test results at 7.8 GHz agree well with the predictions of this theory, for frequency offsets $f << 1/\tau$. The phase noise of the stabilized output is thus dependent on the additive phase noise of the fiber optic delay line.

Mode-locked Semiconductor Lasers as Fiber Optic Transmitters with Reduced RF Phase Noise

The primary source of additive RF phase noise in fiber optic links employing injection-modulated semiconductor lasers has been identified as up-conversion of low-frequency intensity fluctuations to the vicinity of the modulation signal in the laser diode transmitter [2,3]. These low-frequency intensity fluctuations typically exhibit a 1/f frequency dependence, with a corner frequency below 1 MHz. A major source of the 1/f intensity noise of semiconductor lasers has been shown theoretically [4] to be caused by competition between longitudinal optical modes of the laser. Experiments with a mode-locked semiconductor laser [5] have shown a reduction in low-frequency intensity noise compared to the free-running case, presumably due to the reduction of mode-competition when the modes are phase locked.

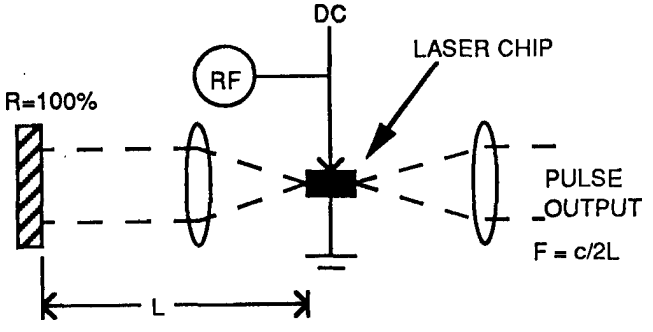


Figure 2. Block diagram of actively mode-locked external cavity semiconductor laser.

The longitudinal optical mode spacing of a semiconductor laser is determined by the laser cavity length, and is typically several hundred GHz. The mode spacing can be brought into the microwave regime by anti-reflection coating one of the laser chip facets and using an external mirror to increase the cavity length. When the laser gain is modulated at the same frequency as the longitudinal optical mode spacing, the laser modes become strongly coupled and phase-locked. Active mode-locking can be achieved by applying a sinusoidal drive signal to the laser at the longitudinal optical mode spacing frequency of the laser cavity $F_{cav} = c/2L$, as illustrated in Figure 2. The output of a mode-locked semiconductor laser is a train of ultra-short optical pulses, typically 1-10 picoseconds in duration. Detection of this optical signal with a fast photodiode yields an electrical signal with a frequency spectrum that contains the fundamental mode-locking frequency and many harmonics that typically extend to well beyond 100 GHz.

A mode-locked semiconductor laser can thus be used as a source of microwave and millimeter-wave signals [6-8], and its optical output makes it inherently suited to the fiber-optic delay-line stabilization scheme. The residual pulse-to-pulse timing jitter added by the laser will degrade the phase stability of the detected fundamental and harmonics. Recent work [9] has demonstrated that the absolute level of timing jitter for an actively mode-locked semiconductor laser is typically dominated by the phase stability of the electrical mode-locking drive signal. The residual RMS jitter of the actively mode-locked laser has been observed to be less than 50 fs [9]. This represents an extremely low level of jitter, thus making the mode-locked semiconductor laser an attractive fiber optic transmitter, and RF-to-millimeter-wave frequency multiplier.

As a typical example, consider a mode-locked laser with a pulse width of 5 picoseconds in a 6-inch external cavity with a longitudinal mode spacing of 1 GHz. The mode-locked laser intensity detected by a high-speed photodiode and viewed on a microwave spectrum analyzer will consist of the fundamental frequency of 1 GHz and harmonics extending beyond 50 GHz. The harmonics can be amplified and used as millimeter-wave signals that are phase coherent with the mode-lock drive signal.

When using the mode-locked laser as a fiber optic transmitter or frequency multiplier, it is desirable to know the fundamental level of phase instability added by the laser. For a directly modulated semiconductor laser, the low-frequency intensity noise is up-converted to the vicinity of the modulation frequency [3], and comprises the fundamental limit to the phase stability of a microwave intensity modulation. We propose that

a similar effect limits the pulse-to-pulse stability of the actively mode-locked laser, since mode-locking can be considered as a form of narrow-band modulation. However, there is a crucial difference between the two situations, as discussed below.

A fortuitous consequence of mode-locking is a reduction of low-frequency intensity noise of the laser, compared to the free-running case (i.e., no mode-locking). In recent experiments [5], the low-frequency intensity noise of a single longitudinal mode of a monolithic waveguide/laser structure with a longitudinal mode spacing of 300 GHz was studied. The individual modes were separated using a spectrometer, and the low-frequency intensity noise of a single longitudinal mode observed in both the free-running and mode-locked conditions. It was observed that the 1/f low frequency intensity noise of the free-running laser was 10 times higher than that of the mode-locked laser. The reduction of noise is attributed to the reduction of competition between the longitudinal laser modes when they are locked together. The remaining level of 1/f low-frequency intensity noise is less well-understood, however. It is most likely due to 1/f carrier and absorption fluctuations in the bulk semiconductor material and contacts [10,11]. In a semiconductor laser, these fluctuations in carrier density modulate the optical intensity, producing a 1/f intensity fluctuation.

Based on the preceding assumptions, an estimate of the timing jitter of the actively mode-locked laser requires knowledge of the level of low-frequency intensity noise due to the bulk 1/f-noise processes of the semiconductor material. For the calculation of the fundamental level of timing jitter, it will be assumed that the low-frequency intensity noise of the mode-locked laser, to first order, is equal to the intensity noise below threshold. In making this conjecture, it is assumed that mode competition noise is negligible in a below-threshold laser and in a mode-locked laser, so the 1/f noise of the bulk semiconductor dominates in both instances.

The phase noise of the fundamental frequency component of the mode-locked laser is estimated using the noise translation factor formalism [3]. The noise to signal ratio in the vicinity of the microwave modulation signal at frequency ω, due to relative intensity noise RIN(Ω) at a low frequency Ω, is given by

$$\left(\frac{N}{S}\right)_{\omega} = RIN(\Omega)\ T(\omega) \qquad (2)$$

where the noise translation factor $T(\omega)$ is

$$T(\omega) = \left(\frac{[g(\omega) + \gamma\epsilon P_o]\ [P_o + \beta + h(\omega)]}{\sqrt{2}f(\omega)}\right) \qquad (3)$$

and

$$g(\omega) = i\omega + \gamma(1 - N_o) + \gamma\epsilon P_o \qquad (4)$$

$$h(\omega) = i\omega + 1 + P_o \qquad (5)$$

$$f(\omega) = h(\omega)g(\omega) + \gamma N_o(P_o + \beta) \qquad (6)$$

In these expressions, P_o is the average laser power, N_o is the steady-state carrier density, ϵ is the gain compression factor, γ is the ratio of spontaneous carrier to photon lifetime, and β is the spontaneous emission factor. At frequencies greater than the relaxation oscillation resonance, $T(\omega)$ approaches an asymptotic value of -3 dB [3].

The external cavity actively mode-locked laser can be considered as a laser with a narrow-band modulation response at a frequency equal to the cavity separation. The effect of the external cavity is to decrease the cavity mode frequency separation $F_{cav} = c/2nL$ from hundreds of GHz to several GHz. The effect of the external cavity enhances the modulation response of the laser at F_{cav}, which can be higher than the relaxation oscillation frequency [12]. From the preceding analysis, the noise translation factor $T(\omega) = -3$ dB for modulation frequencies greater than the relaxation oscillation frequency. Thus, the low-frequency phase noise of the mode-locked laser fundamental beat note will be 3 dB less than the low-frequency relative intensity noise below threshold, as discussed previously.

Measurements of the low-frequency intensity noise of various other laser structures [13] were used to obtain a rough value for the level of the low-frequency intensity noise below threshold. The 1/f noise below threshold at 1 kHz varies from approximately -126 dB/Hz for a buried-crescent heterostructure laser to -142 dB/Hz for a transverse-junction-stripe laser.

Picking the worst-case noise of -126 dB/Hz as the noise of the buried crescent yields a value of -129 dB/Hz for the intensity noise at 1 kHz from the fundamental beat frequency component of the mode-locked laser. It can be shown that the single-sideband RF phase noise, L(f), measured in dBc/Hz (dB below carrier in a 1 Hz bandwidth), is 3 dB less than the intensity noise level [14]. Thus, the RF phase noise is predicted to be -132 dBc/Hz at 1 kHz from the fundamental, with a 1/f slope, yielding -102 dBc/Hz at 1 Hz offset. This is below the measured

value of -95 dBc/Hz [9], which was equal to the noise of the measurement system. If the best value for the low-frequency intensity noise of -142 dB/Hz is assumed, the phase noise at 1 Hz offset is calculated to be -118 dBc/Hz. A more sensitive measurement apparatus is under construction to test these predictions.

All-Photonic Delay Line Oscillator

The output of the mode-locked laser may be detected, amplified, and used to provide the laser RF input current in a self-mode-locked configuration [15]. Using this configuration, the mode-locked semiconductor laser replaces the electronic VCO and fiber optic transmitter of Figure 1, as shown in Figure 3.

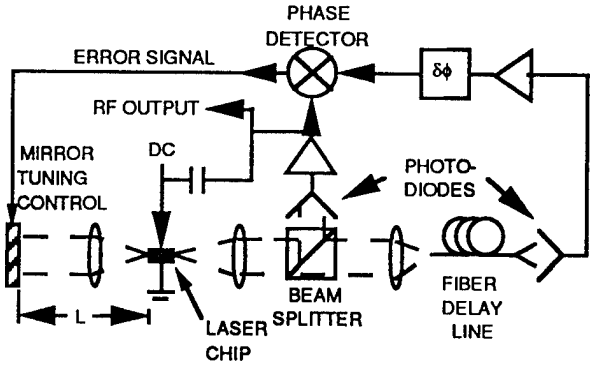


Figure 3. Block diagram of all-photonic microwave oscillator using self-mode-locked semiconductor laser.

The output of the fiber optic delay line discriminator then provides an error signal that drives a tuning input to the mode-locked semiconductor laser, thus stabilizing the repetition rate of the pulses, and hence the detected RF signal. Tuning can be achieved by physically changing the external cavity length with a piezo-electric mirror translation stage, or by changing the index of refraction in a segment of the laser diode via a DC-biased tuning segment [8].

This results in an "all-photonic" microwave oscillator, where the oscillating element is now a photonic device, and the stabilization is provided by the high-Q optical delay line. Since the output spectrum of the mode-locked laser contains many harmonics of the fundamental, this system comprises both an oscillator and a frequency multiplier.

The phase noise performance of this configuration will be limited by the phase noise of the mode-locked laser. Taking the measured residual phase noise value for the actively mode-locked semiconductor laser of -95 dBc/f at 1 Hz from a 5 GHz carrier [9], when stabilized with a 6 kilometer optical fiber delay-line discriminator, the phase noise of the photonic oscillator is calculated using Equation 1 to be $L(f) = -20.5$ dBc/f^3 at 1 Hz from a 5 GHz carrier. If the more optimistic value of -118 dBc/Hz is assumed, the photonic oscillator performance is calculated to be -43.5 dBc/f^3 at 1 Hz offset. This compares very favorably with the performance of state-of-the-art microwave oscillators [16]. Further research is underway to experimentally confirm these calculations and to determine the actual RF phase noise of mode-locked laser diodes.

Conclusions

The first demonstration of a microwave fiber optic delay-line stabilized oscillator [1] had phase noise performance limited by the laser diode transmitter in the microwave fiber optic delay line. In this work, a novel "all-photonic" implementation of a delay-line stabilized oscillator was presented, employing a mode-locked laser diode. The use a mode-locked laser offers lower phase noise than previous fiber optic transmitters, and inherent RF frequency multiplication. The expected phase noise performance was calculated using the previously derived theory, and shown to compare favorably with the performance of state-of-the-art microwave oscillators.

References

[1] R.T. Logan Jr., L. Maleki, M. Shadaram, "Stabilization of Oscillator Phase Using a Fiber Optic Delay Line," Proc. of 45th IEEE Symposium on Frequency Control, 508 (1991).

[2] R.T. Logan Jr., G.F. Lutes, L.E. Primas, and L. Maleki, "Design of a fiber optic transmitter for microwave analog signal transmission with high phase stability," Department of Defense Fiber Optics Conference '90, McLean VA, March 20-23, 1990, pp. 359-365.

[3] K.Y. Lau and H. Blauvelt, "Effect of low-frequency intensity noise on high-frequency direct modulation of semiconductor injection lasers," Applied Physics Letters, Vol. 52, No. 9, pp. 694-696,

29 February 1988.

[4] C.B. Su, J. Schlafer, R.B. Lauer, "Explanation of low-frequency relative intensity noise in semiconductor lasers," Applied Physics Letters, Vol. 57, No. 9, pp. 849-851, 27 August 1990.

[5] Y.K. Chen, M.C. Wu, T. Tanbun-Ek, R.A. Logan, M.A. Chin, "Multicolor single-wavelength sources generated by a monolithic colliding-pulse mode-locked quantum well laser," IEEE Transactions Photonics Technology Letters, Vol. 3, No. 11, November 1991.

[6] K.Y. Lau, IEEE Journal of Quantum Electronics **QE-26**, 250 (1990).

[7] C. Harder, J.S. Smith, K.Y. Lau, A. Yariv, Applied Physics Letters **42**, 772 (1983).

[8] D.J. Derickson, R.J. Helkey, A. Mar, J.G. Wasserbauer, and J.E. Bowers, IEEE Symposium on Microwave Theory and Techniques (MTT-S), 1992.

[9] D.J. Derickson, A. Mar, J.E. Bowers, "Residual and Absolute Timing Jitter in Actively Mode-Locked Semiconductor Lasers," Electronics Letters, Vol. 26, No. 24, pp. 2026-2027, 22 November 1990.

[10] G. Tenchio, "1/f noise of continuous-wave semiconductor lasers," Electronics Letters, Vol. 13, No. 20, pp. 614-616, 29 September 1977.

[11] R.J. Lang, K.J. Vahala, A. Yariv, "The effect of spatially dependent temperature and carrier fluctuations on noise in semiconductor lasers," IEEE Journal of Quantum Electronics, Vol. QE-21, No. 5, pp. 443-451, May 1985.

[12] K.Y. Lau, "Short pulse and high-frequency signal generation in semiconductor injection lasers," IEEE Journal of Lightwave Technology, Vol. 7, No. 2, pp. 400-419, February 1989.

[13] A. Dandridge and H.F. Taylor, "Correlation of low-frequency intensity and frequency fluctuations in GaAlAs lasers," IEEE Journal of Quantum Electronics, Vol. QE-18, No. 10, pp. 1738-1750, October 1982.

[14] K.Y. Lau, "Microwave phase stability of directly modulated semiconductor injection lasers," Applied Physics Letters, Vol. 52, No. 17, 25 April 1988.

[15] D.J. Derickson, R.J. Helkey, A. Mar, P.A. Morton, and J.E. Bowers, "Self-mode-locking of a semiconductor laser using positive feedback," Applied Physics Letters, Vol. 56, No. 1, pp. 7-9, 1 January 1990.

[16] G.J. Dick, "Microwave oscillators for superior short term stability and ultra-low phase noise," this conference.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

TEMPERATURE DYNAMIC MODEL AND TEMPERATURE DYNAMIC COMPENSATION OF CRYSTAL OSCILLATORS

Anatoly V. Kosykh , Boris P. Ionov

Departament of Radioengineering Omsk Polytechnical Institute
11, Mira Prosp., Omsk, 644050, Russia

Alexander M. Vasiliev

Microsistem Inc., Moscow, Russia

SUMMARY

Until now in thermocompensated crystal oscillators (TCXO) the compensating action was estimated as a function of one variable - environmental temperature. At the same time experimental data show that TCXO frequency stability is essentially dependent on temperature process dynamic parameters. The more drastic is the temperature variation the more is decompensating effect.

In this work the origins of this phenomenon are analysed and the ways are shown of reducing its effects. A new in principle approach is suggested according to which the compensating action is formed as a function of two wariables: the temperature and the rate of its variation.

A theoretic substantiation of this method is given. Structural circuits of devices perfoming the temperature dynamic compensation are suggested. The experimental investigations of TCXO with microprocessor based temperature dynamic compensation system confirmed the possibility of essential TCXO frequency stability improvement in non-stationary thermal conditions. In a system using AT - cut resonator and a separate Y - cut thermosensor the application of the temperature dynamic compensation resulted in ten times actual stability rise as compared to the conventional digital thermocompensation procedures. Even with dual-mode SC-cut resonators the application of the temperature dynamic compensation proved to be useful.

INTRODUCTION

Temperature compensated crystal oscillators (TCXO) are widely used in self-contained miniature equipment especially in cases when high stability is needed alongside with economical efficiency and short operational readiness time. It is characteristic of such devices operational conditions that rapid environmental temperature variations are possible while with small mass and absense of heat isolation the smoothing effect of the case is lowered. In such conditions the crystal resonator, thermosensor and other elements of the oscillator are under the actions of fast alternating temperature fields which results with deteriorated thermocompensation. The phehomenon of TCXO stability deterioration under transient thermal effects is well known but at present the problem became especially acute. The matter is in digital control methods applied now which ensure high accuracy of compensation function. As a result the other components of frquency nonstability became more evident, those with analoque compensation system imperfections. Hence, to create a high stability TCXO it is not enough only to step up the compensation accuracy under slow (quasistatic) temperature variations. It is quite necessary to take steps of lowering the effects of temperature shocs and sharp temperature variations.

The above mentioned effects had been described earlier [1 and oth.] but it is our opinion that the phenomenon of TCXO dunamic discompensation has not yet been fully investigated. Furthermore, the engineering solutions are absent which could ensure the increase of temperature-dynamic frequency stability with practicaly any kind of TCXO construction, using only pure circuit procedures.

In this work origins of temperature dynamic unstability are discussed in detail. A temperature dynamic TCXO model and means of temperature dynamic frequency compensation are suggested. Theoretic analysis of suggested structures effectiveness and experimental investigations results are given.

TEMPERATURE DYNAMIC MODEL OF T C X O

The TCXO model used now for compensation action description only as an environmental temperature function cannot explain thermocompensation deterioration under nonstationary thermal conditions and show the ways of stability increase. Therefore a new model is needed based on the analysis of temperature dynamic phenomena in oscillator.

First of all the terminology should be defined which will be used to describe the non-reproducibility of the F-T curve form after temperature effects: temperature and frequency hysteresis and temperature dunamic F-T curve distortion. In both cases after temperature cycle is over the F-T cur-

0-7803-0476-4/92 $3.00 © 1992 IEEE

ve assumes a shape of a loop though different factors are responsible for the departure from stationary state. Terminology assumed in [2] will be used here and the definitions of terms are given below.

F-T curve hysteresis - is a departure of an oscillator (or a resonator) frequency from the value corresponding to the static F-T curve and depending on the previous temperature procedures with the oscillator. F-T curve hysteresis is defined by processes running in TCXO crystal resonator and is related to plastic deformations occuring in points of crystal plate contact with holder and leads. This parameter is of structural origin. Its effect could be lowered by means of thermotraining, by proper choice of points and means of piezoelement fastening to the holder etc. The most effective solution is utilization of modern designs of crystal plates eliminating the possibility of plastic deformations in points of fastening (BVA, QAS and other). Frequency departures due to hysteresis are recovered very slow after the temperature actions are over.

TCXO F-T -curve temperature dynamic distortions are provoked with instability of thermal action on the oscillator. The distortions are of reversible nature and after the thermal action is stabilised they disappear rather soon. There are several components of TCXO F-T -curve temperature dynamic distortios.

The first component is a temperature dynamic distortion of the proper crystal resonator primary F-T curve. This distortion arise as a result of mechanical stresses occuring in the piezocrystal volume and in the region of electrodes and holders fastening when the warm-up of the resonator is irregular. The deformations observed here are of elastic nature and the frequency departure values depend on mechanic stresses amounts and the distribution law in the volume of the piezocrystal and on the resonator strain sensitivity. For the first time this phenomenon was investigated in detail by a team of autors in the USSR in mid-60-ies [3,4]. They introduced the notion "temperature-dynamic frequency ratio". Later more profound and theoretical investigation of this phenomenon was carried out by R. Holland, [5,6], A. Ballato [7,8], G. Theobald and other [9,10].It was shown, in particular, that crystal resonator F-T curve depends not only on the temperature proper but also on the rate of temperature variation and is described with the following expression:

$$\delta f = \sum_{i=0}^{3} C_i (T_a - T_o)^i + [\tilde{a} + \hat{a}(T_a - T_o)]\dot{T}_a, \quad (1)$$

where δf - crustal resonator relative frequency departure;
T_a - temperature of the crystal oscillator;
C_i - static F-T -curve coefficients;
T_o - reference temperature;
$\tilde{a}$, $\hat{a}$ - dynamic coefficients of shift and rotation respectively;
$\dot{T}_a = dT_a/dt$ - temperature variation rate.

It follows from (1) that under unstable temperature conditions crystal resonator F-T -curve is distorted by shifting upwards (or downwards) along the frequency axis and rotation around a certain point and the amounts of these shifts are proportional to the rate of temperature variations and depend on $\tilde{a}$ and $\hat{a}$ coefficients. The numerical values of a and a coefficients depend on the type of resonator crystal cut and on its design.

To minimize frequency dynamic departures desensitized cuts were developed: TD-cut(yxbl 23° 25' /35°) in USSR [11]; SC-cut -in USA [12] and others. As the form constancy of TCXO F-T - curve is of prime importance the use of dynamically compensated cuts is the most preferable.

The second component of temperature frequency unstability has not yet been properly investigated. It is specific only for TCXO. The nature of this component is as follows: when the temperature changes are fast the compensating action lags behind or advances the needed value because the ways of heat conduction to TCXO circuit parts are quite different and in unstable temperature conditions the circuit elements have different temperatures. But oscillator tuning is usually performed under quasistatic conditions when the temperatures of different elements are assumed to be equal. Since the compensating action is formed as a function of thermal sensor signal we can say that in this case the decisive role will belong to temperature dynamic difference between the sensor and other circuits elements having temperature - frequency coefficient. For majority of TCXOs the most significant are the following dynamic temperature differences (in order of diminishing effectiveness),

- between the thermal sensor and a crystal resonator;
- between the thermal sensor and a reference voltage source;
- beetween the thermal sensor and an active oscillator circuit.

As a first approximation for a TCXO built according to a traditional circuit (Fig.1) a temperature-dynamic model could be suggested having the view of Fig.2. In this model an oscillator case is simulated with an aperiodic thermoinertial link (1) with thermal time constant τ_s . Thermal sensor (3), reference voltage source (4) and oscillator cirquit (6) are represented with serial connected aperiodic inertial links having their corresponding thermal time constants τ_s ,τ_v ,τ_{osc} and with uninertial nonlinear links, describing the dependence of their output parameter on temperature as a power series. A crystal resonator is represented with an

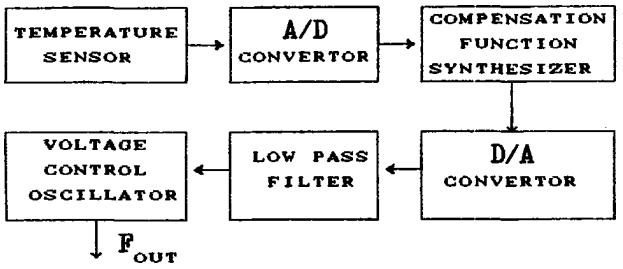


Fig.1. DTCXO Block Diagram

inertial link (having the time constant) and with in-series connected network simulating the equation (1). TCXO output frequency is determined in this model as a sum of crystal resonator frequency F_Q and frequency corrections defined by compensation system and oscillator circuit temperature coefficients ΔF_C, ΔF_{osc}

$$F_{OUT} = F_Q + \Delta F_C + F_{osc} \qquad (2)$$

Compensating function synthesizer (CFS) (5) transmission function Z(X) under quasistationary thermal conditions is determined in such a way that at any temperature F_{OUT} = const. Compensation system time delay is allowed for in a time constant of a filter (7). If a crystal thermal sensor is used in a TCXO the model is complicated as under non-stationary conditions F-T -curve of the sensor will be described with an expression similar to (1).

T C X O MODEL PARAMETRS DETERMINATION

In order to have the possibility of TCXO frequency departures estimation in different temperature conditions it is necessary to have numerical values of temperature dynamic model parameters (Fig.2). High acurracy theoretic determination of these values is impossible because of numerous factors acting, therefore the numerical data needed canbe obtained only throught experimental procedures. Separate units static transmission coefficients can be easyly represented in analytical form by means of experimental data approximation using least squares procedures while thermal time constants and dynamic coefficients determination is rather difficult.

Crystal resonator and thermosensor time constant are difficult to determine because there is no possibility to control their temperature. Only output signals F_Q (for resonator) and X_S (for thermosensor) are non-linearly related to internal temperatures T_Q and T_S and so are accessable for measurements. Sufficiently accurate thermal time constant determination can be perfomed by means of semigrafical procedure using the object response to positive or negative temperature jumps [13]. Tran-

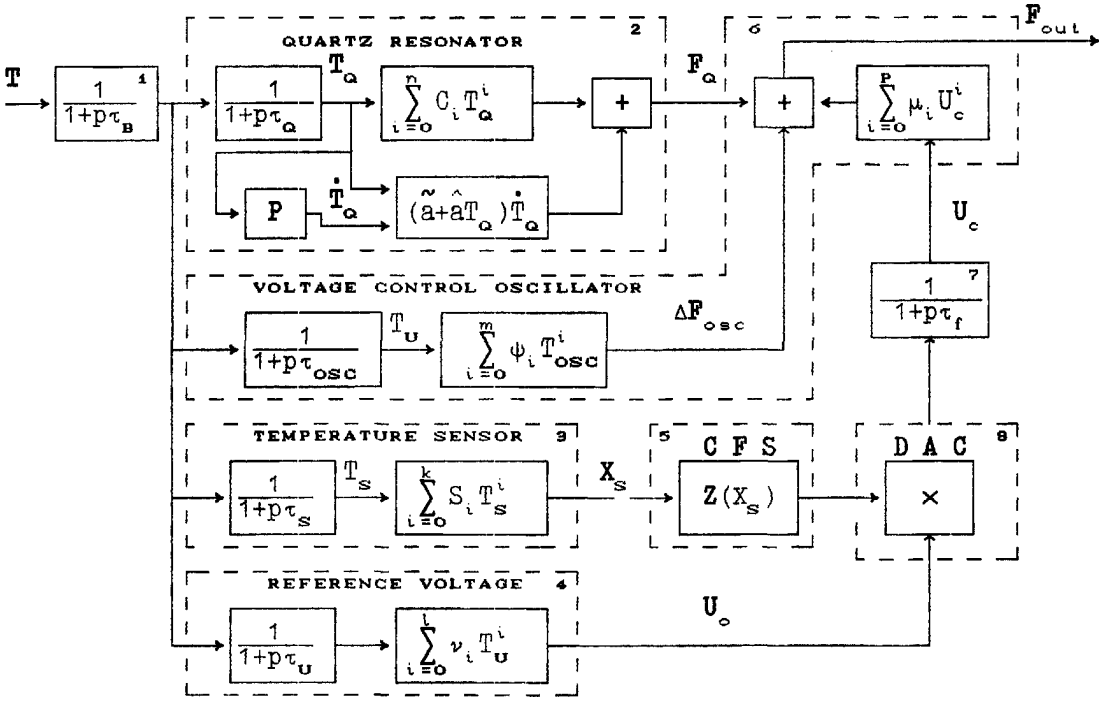


Fig.2. DTCXO Temperature Dynamic Model Block Diagram

sient response origins should be coincided in time as it is shown in Fig.3. If a thermosensor or a resonator has no dynamic sensitivity (thermoresistor, p-n junction, SC-cut resonator), thermal time constant can be determined using the formula

$$\tau = t^*/\ln 2 ,$$

where t^* - time passed beginning from the transient process origin up till the moment of curves crossing.

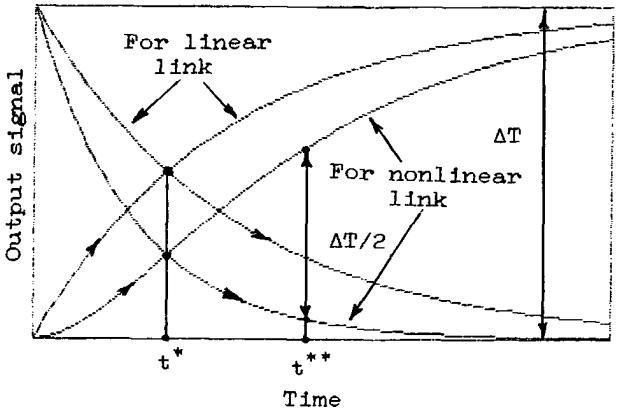


Fig. 3. Method of Thermal Time Constant Determination

In the cases when F-T -curve dynamic coefficients make an essential contribution to general frequency departure (for AT-cut resonators) the procedure becomes more complicated. The effect of temperature dynamic frequency departures may be accounted by means of additional measurement: time interval t^{**} beginning from process origin up to the point where difference between curves values equals the half of the jump T amplitude. In this case

$$\tau = (t^{**} - t^*)/\ln 2.$$

The results of time constants experimental investigations for different "resonator-thermosensor" couples are given in Fig.4. These structures were not optimized for implementation in TCXO but were really used in different stages of our investigation. It is evident that resonator and thermosensor time constants values may differ even in the case when they are placed in the same enclosure.

MODEL DYNAMIC COEFFICIENTS DEFINITION

As it was indicated above TCXO dynamic decompensation is defined by temperature dynamic sensitivity of a crystal resonator ($\tilde{a}$ and $\hat{a}$ coefficients) and by thermal inertia differences between a reference channel and a thermosensing channel. The dynamic temperature difference arising between the resonator and the thermosensor is defined by "resonator - thermoresistor" system quality and may be conveniently characterised with a coefficient

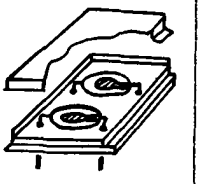
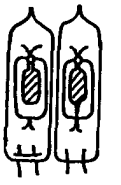
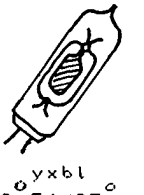
AT-cut Y-cut	AT-cut Y-cut	Dual-mode resonator
		
τ_a = 25 s τ_s = 27 s $\tilde{a}$=-1X10 s/K $\hat{a} \approx 0$ b = -3 s	τ_a = 66 s τ_s = 59 s $\tilde{a}$=-1.5X10 s/K $\hat{a}$=-1,3X10 s/K^2 b = -6 s	τ_a = 76 s τ_s = 75 s $\tilde{a}$=-4X10 s/K $\hat{a}$=-1,6X10s/K^2 b ≈ 0
a	b	c

Fig.4. "Resonator-thermosensor" Couples parameters

$\tilde{b} = \tau_a - \tau_s$. In cases when a crystal thermosensor is used its own F-T -curve dynamic coefficient $\tilde{a}$ begins acting. Here $\tilde{b} = \tau_a - \tau_s - \tilde{a}_s / S_s$, where S_s - thermosensor F-T curve slope.

Dynamic coefficients values may be defined by F-T curve transformation nature analysis under different rates of temperature variations. For this purpose a special thermal chamber capable to withstand the rate of cooling and heating specified. In coordinate axes "resonator frequency - thermosensor indications" F-T- curves family of a non-compensated oscillator under different temperature variation rates has the view as in Fig.5 and is described in terms of the following expression

$$\delta f = \sum_{i=0}^{n} C_i (T_s + \tilde{b}\dot{T})^i + [\tilde{a} + \hat{a}(T_s + \tilde{b}\dot{T}_s - T_d)]\dot{T} \quad (3)$$

Here T_{e1} and T_{e2} - temperatures of the lower and the higher (along temperature range) extremes of

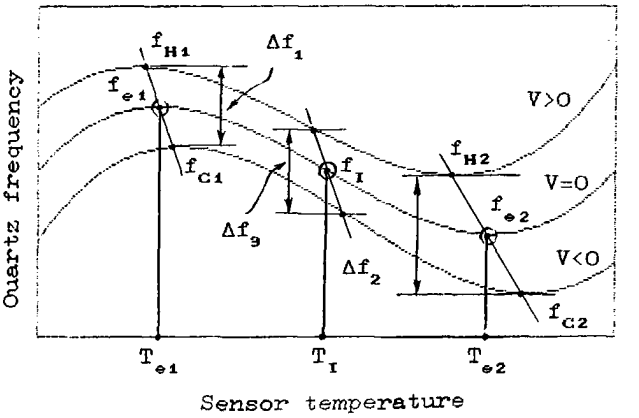


Fig.5. Method of Dynamic Coefficients $\tilde{a}$, $\hat{a}$ and $\tilde{b}$ Determination

F-T curve; T_I - temperature corresponding to the F-T curve inflection (the point where the second derivative of the frequency on temerature is zero).

Let us consider the vicinities of F-T curve two extremes. As F-T curve slope is low in the extremes region the effect of $\tilde{b}$ coefficient may be neglected and all the frequency departures under various temperature variation rates will be defined only by temperature dynamic coefficients $\tilde{a}$ and $\hat{a}$. In this case the resonator frequency at the extreme point under heating and cooling will differ from that value in stationary state

$$f_{Hi} = f_{ei} + [\tilde{a} + \hat{a}(T_{ei} - T_d)]V_H$$

$$f_{Ci} = f_{ei} + [\tilde{a} + \hat{a}(T_{ei} - T_d)]V_C$$

where V_H and V_C - heating and cooling rates. Dynamic frequency difference in the point of the extreme will have the following view:

$$\Delta f_i = f_{Hi} - f_{Ci} = \tilde{a}(V_H - V_C) + \hat{a}(T_{ei} - T_d)(V_H - V_C) \quad (4)$$

In similar way - for the second extreme

$$\Delta f_2 = \tilde{a}(V_H - V_C) + \hat{a}(T_{e2} - T_d)(V_H - V_C)$$

If an equality is provided $V_H = V_C = V$,

$$\Delta f_1 - \Delta f_2 = 2\tilde{a}V(T_{e1} - T_{e2})$$

where from

$$\hat{a} = (\Delta f - \Delta f) / 2V(T_{e1} - T_{e2}) \quad (5)$$

Having expressed $\tilde{a}$ according to (4), we will have

$$\tilde{a} = \frac{\Delta f_i}{2V} - \hat{a}(T_{ei} - T_d) \quad (6)$$

In order to determine temperature shift dynamic coefficient b F-T curves family in vicinity of T temperature will be considered. F-T curve in this point is linear with slope S_I, while the highest powers of the series in expression (3) may be neglected. The frequency difference under cooling and heating in this point will be as follows

$$\Delta f_3 = 2V[S_I\tilde{b} + \tilde{a} + \hat{a}(T_I - T_d)]$$

where from

$$\tilde{b} = \frac{\Delta f_3 - 2\tilde{a}V - 2\hat{a}V(T_I - T_d)}{2 S_I V}. \quad (7)$$

Thus, after analysing the F-T curve form dynamic transformation in the extremes and bend region according to the expression (5), (6) and (7) the coefficients $\tilde{a}$, $\hat{a}$ and $\tilde{b}$ can be determined. It is noteworthy that the coefficient b found using such a procedure is more significant than thermal time constants difference as all the factors responsible for F-T curve shift along the temperature axis are taken into account.

T C X O MODEL PARAMETERS EFFECT ON DYNAMIC TEMPERATURE FREQUENCY UNSTABILITY

TCXO output frequency can be calculated using equation (2) under any temperature effects if model (Fig.2) parameters are known. Direct substitution of nonlinear and inertial links transmission coefficients into (2) results in very awkward expression wich is very difficult to analyse. Much simple and easy tob visualise is to carry out the simulation of TCXO responses to different temperature effects using a numerical procedure on an electronic computer. For this purpose a special DYNO-LAN program was developed capable to change easyly any model parameter and to represent TCXO F-T curve plot when the temperature variation law is specified.

Though there is theoretical possibility to determine any reaction to no matter how complex action we confined during calculations by two types of temperature process:

- linear temperature variation up and down with the rate specified;
- sinusoidal temperature variation with cycle closed.

In both cases external temperature variation is in the operational range limits and besides, under sinusoidal action a maximum value of temperature derivative with respect to time is specified.

The following objectives were to be attained during numerical investigation.

1. To find out the nature of dynamic decompensation dependence on the following factors:
   - temperature variation law;
   - thermal time constants values correlation between a resonator and a thermosensor;
   - the form of resonator F-T curve;
   - case thermal time constant;
   - compensation system output filter electrical time constant;
   - reference voltage source temperature coefficient;
   - resonator dynamic coefficients values.
2. To develope specification for separate TCXO units with intention to ensure the F-T curve disposition in required limits under real thermal conditions.

Model parameter values obtained in experiments were used in calculations and the variation of those values in certain limits was used in order to check their sensitivity. Calculation results are given in Fig.6 - Fig.11. The parameters indicated in Fig.6 - Fig.11 correspond to structural circuits Fig.13 designations.

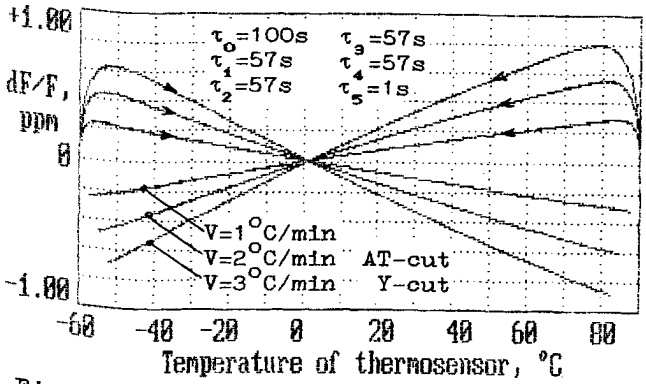


Fig.6. DTCXO frequency dynamic unstability vs. temperature simulation: coefficient $\tilde{a}$ and $\hat{a}$ influence

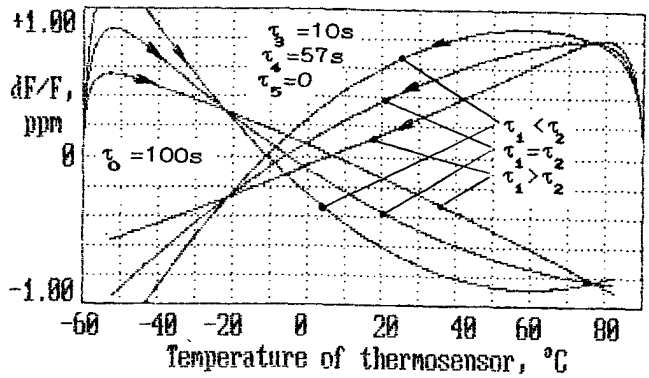


Fig.9. DTCXO frequency dynamic unstability vs. temperature simulation: The difference between τ_a and τ_s influence

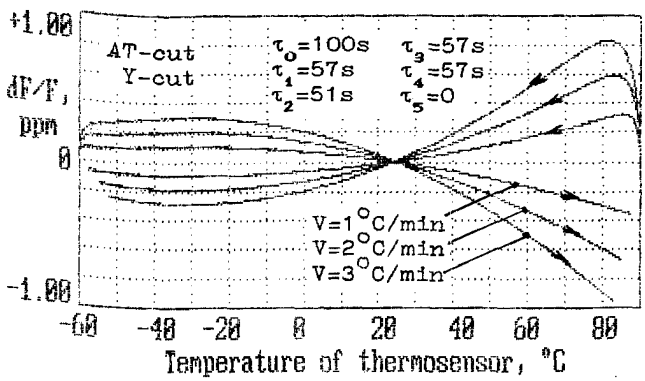


Fig.7. DTCXO frequency dynamic unstability vs. temperature simulation: $\tau_a > \tau_s$

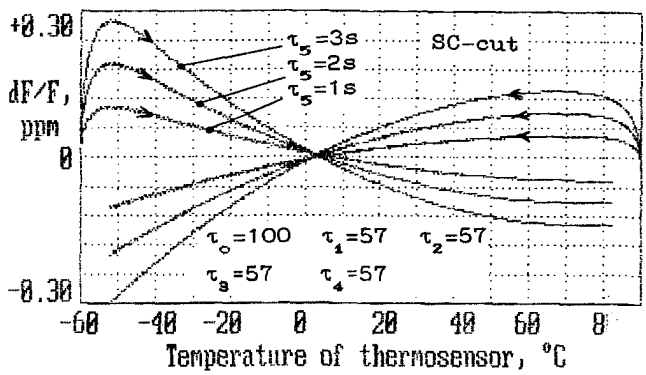


Fig.10.DTCXO frequency dynamic unstability vs. temperature simulation: The filter electrical time constant influence

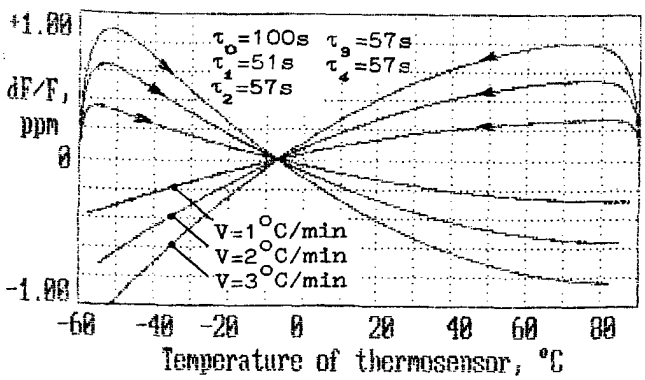


Fig.8. DTCXO frequency dynamic unstability vs. temperature simulation: $\tau_a < \tau_s$

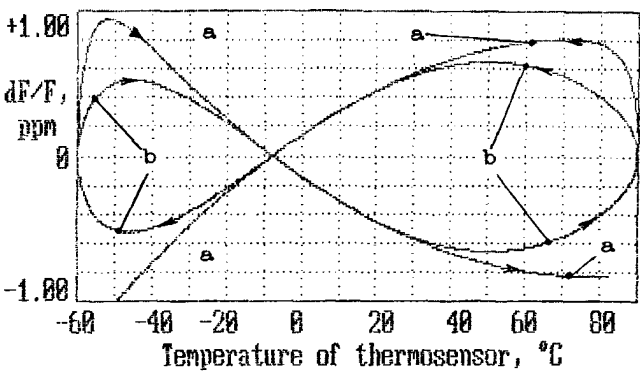


Fig.11.DTCXO frequency dynamic unstability vs. temperature simulation: Linear (a) or Sinusoidal (b) temperature effects

TEMPERATURE DYNAMIC COMPENSATION

As a result of crystal oscillator termocompensation with an initial F-T curve C(T) the final frequency unstability will be described by the following expression

$$\delta f = C(T_Q) - Q(T_S) \qquad (8)$$

where $Q(T_S)$ – frequency departures function under the effect of thermocompensation system;

T_Q – resonator temperature;

T_S – sensor temperature.

The form of $Q(T_S)$ function is affected with non-linearities of a thermosensor and the oscillator frequency control law, and also with the transfer function $Z(X_S)$ of the compensation function synthesizer. The usual task of the thermocompensation procedure comes to such a form $Z(X_S)$ selection that would ensure minimum decompensation function (8) in the whole temperature range (with due regard for the oscillator F-T curve real form C(T), frequency control characteristics $\mu(V)$ and thermosensor nonlinearities S(T) (See Fig.2). In existing TCXOs the form of a transfer function of a compensation function $Q(T_S)$ synthesiser is determined in quazistationary temperature conditions when dynamic effects are not active while in real condition the F-T curve of an ideal static compensated TCXO will be decompensated:

$$f(T,\dot{T}) = \sum_{i=0}^{n} C_i (T_S + \tilde{b}\dot{T} - T_o)^i - \sum_{i=0}^{n} C_i (T_S - T_o)^i + [\tilde{a} + \hat{a}(T_S + \tilde{b}\dot{T} - T_o)]\dot{T} \qquad (9)$$

The form of this function under different coefficient values combination has been investigated in above section rater in detail.

At the bottom of the TCXO dynamic decompensation is that the first term of the expression (8) is a function of two variables: T – temperature and $\dot{T}$ – its derivative, while the secont term (compensation function) – is a function of only one variable: T_S – temperature. To lessen the dynamic termocompensation error the compensation function should also be considered and plotted as a function of two variables. It is exactly the essence of a new in principle approach to the thermocompensation problem. For the implementation of the idea a sensor of temperature variation rate should be introduced into the thermocompensating device, capable to control the special channel operation in temperature dynamic compensation system.

From an analysis of TCXO model (Fig.2) it follows that this channel should modify the zero and the first terms of static compensation function decomposition in proportion to a temperature variation rate and to coefficients $\tilde{a}$ and $\hat{a}$ values. Furthemore, it should induce the argument modification of the compensation function synthesizer in proportion to $\tilde{b}$ value and to a temperature variation rate. The above algorithm can be realized by means of a device given in Fig.12 [14]. Since to form the compensation function for this case rather complex actions areneeded it would be sound practice to realize thecompensation function synthesizer to be microprocessor based. By the way, the rate of temperature variation determination in this case can be carried out by algorithmic procedure using one of the known algorithms of thermosensor signal differentiation.The circuit will not differ from the standard one (Fig.1).

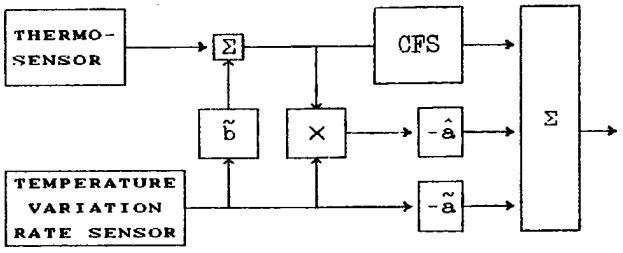


Fig.12. Dynamic Temperature compensation Device Block Diagram

For the sake of theoretic verification of the temperature dynamic compensation idea the TCXO structural circuit with dynamic thermocompensating device (Fig. 13) was analysed using the program DYNO-LAN. The results (Fig.14 – Fig.15) show that if the parameters of channel are properly chosen, the procedure ensures the essential thermocompensation improvement. Since the treatment of only the first derivative of temperature process is provided in the model, the system does not ensure full compensation of sharp temperature effects: in this case the temperature variation law compries the highest derivatives. When the temperature variation is linear in steady-state operation mode the thermocompensation is essentially ideal.

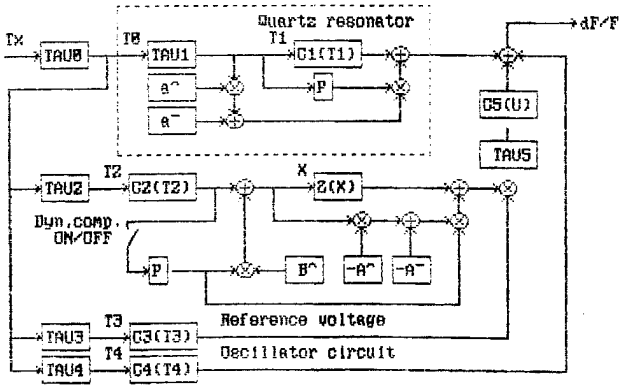


Fig.13. Temperature Dynamic Compensated DTCXO Model Block Diagram

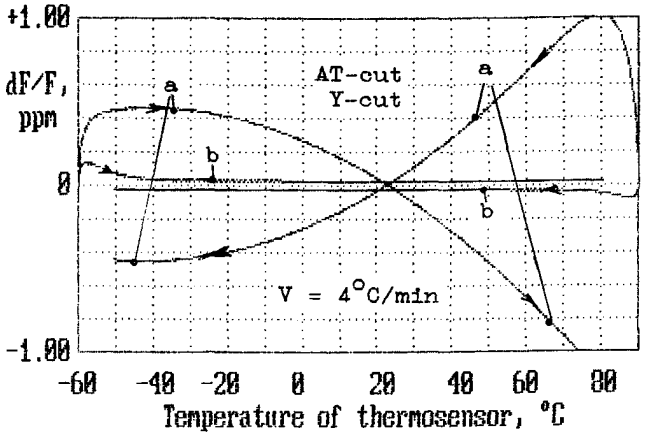


Fig.14. DTCXO frequency dynamic unstability vs. temperature simulation: Dynamic compensation system OFF (a) or ON (b)

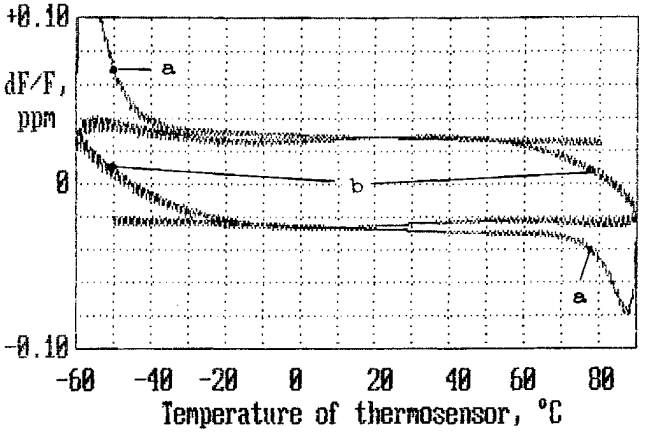


Fig.15. Dynamic compensated DTCXO under linear (a) and sinusoidal (b) thermal effects

EXPERIMENTAL RESULTS

A TCXO with temperature dynamic compensation system was developed for experimental testing of suggested solutions. AT-cut crystal resonators were used and a separately disposed Y-cut crystal thermosensor (the case of the worst initial temperature dynamic stability) and also TD-cut dual mode resonators (owing to the influence of a filter, an oscillator and other components the initial dynamic temperature unstability in this case is lower but not zero). In Fig.16 F-T curves of the oscillator are given with separate resonators and the dynamic compensation system ON and OFF, while the rate of temperature variation is 2 °C/min.

In Fig.17 similar results are shown for a TCXO with dual-mode resonator under the rate of temperature variation of 6 °C/min . The results obtained confirm that the temperature dynamic frequency stability of any design TCXO can be increased with pure circuit procedures.

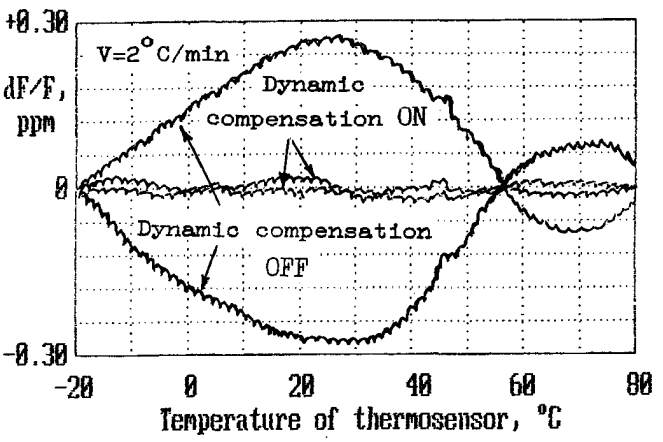


Fig.16. Experimental F-T curve of DTCXO using AT-cut and Y-cut Resonators

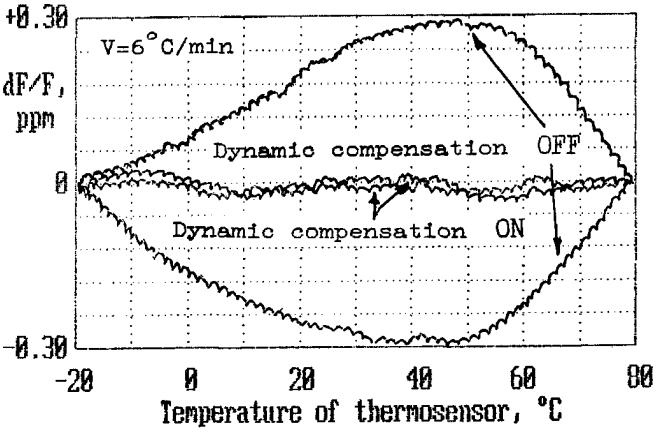


Fig.17. Experimental F-T curve of DTCXO using Dual-Mode Quartz Resonator

CONCLUSION

The investigations described allow to draw the conclusions given below.

1. The TCXO response to a non-stationary thermal effect can be predicted, if the parameters of its temperature dynamic model and the temperature variation law are known.
2. In order to reduce dynamic frequency departures every effort should be applied to "spacial coincidence" and matching of thermal time constants of those TCXO elements affecting the frequency through the temperature variations.
3. The temperature dynamic frequency departures can be reduced by using a special system of temperature dynamic compensation where the compensation action is formed as a function of at least two variables: the temperature and the rate of the variation.

4. If the second and higher derivatives dominate in the law of the temperature variation, the quality of temperature dynamic compensation is deterioratedas the higher derivatives are not accounted for in suggested system.
5. It is reasonable to fulfill the implementation of the temperature dynamic compensation system on the base of microprocessor.
6. The algorithms of temperature dynamic compensation need further development especially concerning the time spent on the high accuracy measurement of the temperature proper and its derivatives.

ACKNOWLEDGEMENTS

The authors express their acknowledgements to Mr. A.N. Lepetaev, a colleague for computer data treatment and especially for DYNO-LAN program developed for this investigation.

REFERENCES

1. R.L. Filler." Measurement and Analysis of Thermal Histeresis in Resonators and TCXOs". Proc. 42nd Ann. Symp. of Frequency Contror, pp. 380-388, 1988.
2. B. Parzen. Design of Crystal and Other Harmonic Oscillators. New York: A Willey - Interscience publication, 1983, pp. 110-113.
3. Л.Е. Ивлев, В.П. Багаев, Э.М. Фромберг, "О температурно-динамическом коэффициенте частоты кварцевых резонаторов", Вопросы радиоэлектроники. Сер.3, Детали и компоненты (USSR). 1965, вып. 2, стр. 50-59.
4. Л.Е. Ивлев, "Температурно-динамический коэффициент частоты и температурно-динамические характеристики кварцевых резонаторов". Электронная техника. Сер. 9 (USSR),1967, вып.4, стр. 20-29.
5. R. Holland, "Nonuniformly Heated Anisotropic plates. Pt.1.",IEEE Trans. Sonics and Ultrasonics, 1974, July, vol. SU-21, pp. 592-598.
6. R. Holland, "Nonuniformly Heated Anisotropic plates. Pt.2."1974 Ultrasonics Symposium Proc., IEEE, Cat.74, CHO 896, ISU, pp. 592-598.
7. A. Ballato."Static and Dynamic Frequency-Temperature Behavior of Singly and Doubly Rotated Ovencontrolled Quartz Resonators." Proc. 32nd Ann. Symp. of Frequency Control, 1978.
8. A. Ballato. "Static and Dynamic Behavior of Quartz Resonators.", IEEE Trans. Sonics and Ultrasonics. 1979, N4, pp. 299-306.
9. G. Theobald, G. Marineau, R. Pretot, I.I. Gagnepain. "Dynamic Thermal Behavior of Quartz Resonators". Proc. 33rd Ann. Symp. of Frequency Control, 1979, pp. 239-246.
10. J.P. Valentin, G. Theobald, I.I. Gagnepain "Frequency Shift Arising from In-plane Temperature Gradient Distribution in Quartz Plates" Proc. 38th Ann. Symp. of Frequency Control, 1984, pp. 157-163.
11. Invention sertificate N 243977 (USSR). Устройство для стабилизации частоты генераторов / А.Н. Дикиджи, Л.Ш. Дикиджи, Л.Е. Ивлев, В.С. Теренько. Открытия. Изобретения. 1969, N 17.
12. J.A. Kusters. "The SC-cut Crustal - an Overviev". Proc. 1981 Ultrasonics Symposium, pp.402-409.
13. Invention sertificate N 1332528 (USSR). Способ определения показателя тепловой инерции термопреобразователя./ Б.П. Ионов. Открытия. Изобретения. 1987, N 26.
14. Invention sertificate N 1197035 (USSR). Устройство термокомпенсации кварцевого генератора./ А.В. Косых, В.П. Ионов, Е.Г. Гросфельд. Открытия. Изобретения. 1985, N 45.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

Frequency Temperature Behavior and Its Hysteresis of S-TCXOs Using New Shape GT Cut Quartz Crystal Resonators

Hirofumi KAWASHIMA and Kenji SUNAGA

Seiko Electronic Components Co., Ltd.
Tochigi-shi, Japan 328

Abstract

This paper describes frequency temperature behavior and its hysteresis of self-temperature compensated crystal oscillators (S-TCXOs) using new shape GT cut quartz crystal resonators. The object of this paper is to clarify the frequency temperature behavior, its hysteresis and aging characteristics of the S-TCXOs.

First, a S-TCXO is successfully obtained with frequency deviation of +30/-150ppb at best over a wide temperature range of -30 to +85 ℃, needless to say, without any external temperature compensation such as thermister and so on.

Second, the thermal hysteresis of S-TCXOs is extremely small for the excursion of fall-temperature to rise-temperature, but occurs somewhat for the excursion of rise-temperature to fall-temperature. The variation quantity for each S-TCXO is, however, less than 0.38ppm in frequency deviation for the temperature excursion of -45 to +90 and -45 ℃, whose excursion shows the outstanding hysteresis.

Finally, aging quantity after the pre-aging for about two weeks is also found to be 0.33ppm at a mean value of eight samples for one year and two months, and after the pre-aging for about six months, the quantity is 65ppb at a mean value for eight months.

§ 1. Introduction

Recently, a temperature compensated crystal oscillator(TCXO) that is used as source of a frequency standard of principally handy-type communication equipment and consumer products demands excellent frequency temperature behavior, and simultaneously, a further miniaturized oscillator is required with low consumption power. As an oscillator satisfying this requirement, it was already reported[1], [2] by one of the authors that a self-temperature compensated crystal oscillator employing a new shape GT cut quartz crystal resonator(S-TCXO) [3]-[7] is very available. In addition, many papers were reported on thermal hysteresis of resonators[8]-[12].

The object of this paper is to clarify the frequency temperature behavior, its hysteresis and aging characteristics for the S-TCXOs. First, it is shown that a S-TCXO employing a new shape GT cut quartz crystal resonator is successfully obtained with frequency deviation of +30/-150ppb at best in a temperature range of -30 to +85 ℃ without any external temperature compensation and is extremely excellent in frequency temperature behavior. Second, the thermal hysteresis characteristics of five S-TCXOs with frequency deviation of less than 0.7ppm in a temperature range of -30 to +85℃ are shown. An examination of the thermal hysteresis is also performed for the temperature excursions of (a) to (d):

(a) +25℃→+90℃→−45℃→+25℃
(b) +25℃→−45℃→+90℃→+25℃
(c) −45℃→+90℃→−45℃
(d) +90℃→−45℃→+90℃

As a result, the thermal hysteresis occurs somewhat for the excursion of rise-temperature to fall-temperature, but its variation quantity is very small and the present S-TCXOs are found to have hardly problem on the thermal hysteresis, even when the frequency deviation of ±2.5ppm is demanded in the temperature range of -30 to +85℃.

0-7803-0476-4/92 $3.00 © 1992 IEEE

Finally, the aging characteristics are examined for eight samples. The aging quantity after the pre-aging for about two weeks is less than 0.5ppm in frequency deviation for one year and two months and after the pre-aging for six months, it is less than 0.15ppm for eight months. Thus, the present S-TCXOs are found to have extremely small frequency aging. A detailed description will be performed below.

§ 2. New Shape GT Cut Quartz Crystal Resonator and Its Supporting Means

Since resonator shape and its supporting means of a GT cut quartz crystal resonator to be treated in this paper may influence the thermal hysteresis characteristics, they are first described.

Figure 1 represents a new shape GT cut quartz crystal resonator and its supporting means. As described detailedly in the reference [6], the present GT cut quartz crystal resonator consists of a vibrational portion and two supporting portions. In addition, the resonator is mounted on a ceramic pedestal at both ends of the supporting portions by adhesive. The pedestal is also fixed on two lead wires, as a result, it has high shock-resistance.

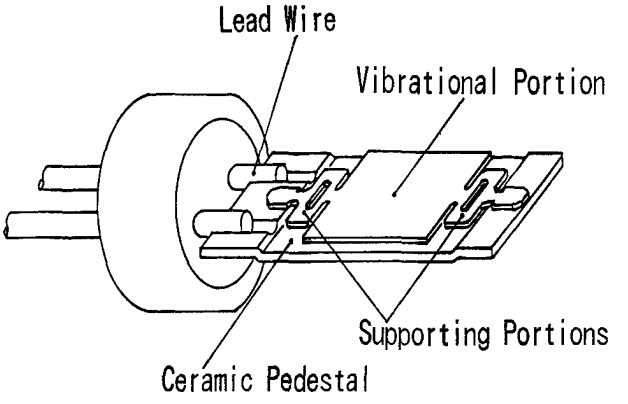


Fig. 1 New shape GT cut quartz crystal resonator and its supporting means.

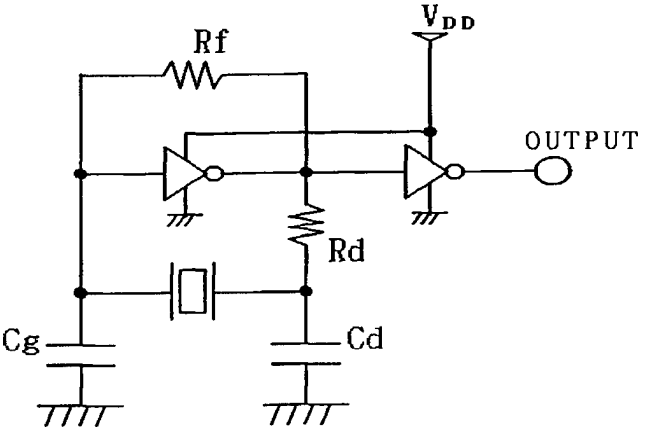


Fig. 2 Circuit diagram of a S-TCXO.

§ 3. Circuit Diagram and Oscillation Frequency

Figure 2 represents a circuit diagram of a S-TCXO. The circuit consists of a CMOS inverter, drain capacitance Cd, gate capacitance Cg, feedback resistance Rf, drain output resistance Rd, voltage source V_{DD} and a new shape GT cut quartz crystal resonator, and a signal is output through a buffer. Next, an oscillation frequency is derived from an equivalent circuit of the present circuit diagram, but the derivative procedure is explained simply because the detailed procedure was already-described in the references[1], [2].

A CMOS inverter for a small signal can be expressed by voltage source μVg (μ:amplification ratio, Vg:gate voltage) and drain resistance rd, as is well-known. Also, the impedance of a new shape GT cut quartz crystal resonator which is a coupling resonator can be expressed by a summation of real part Rei(resistance part, i=1, 2, i=1:Principal vibration, i=2:Subvibration) and imaginary part Xe(reactance part). Therefore, the circuit in Fig. 2 is replaced by an electrical equivalent circuit. From the circuit, when an angular frequency is taken as ω, Xe is given as

$$Xe = \frac{1}{\omega C_L} \cdot \frac{1}{1 - \dfrac{1+\mu}{Rf\ (rd+Rd)\ \omega^2\ CgCd}} \qquad (1)$$

$$\frac{1}{C_L} = \frac{1}{Cg} + \frac{1}{Cd}\left(1 + \frac{Rei}{rd+Rd}\right) \qquad (2)$$

Moreover, Xe for a coupling between principal-

and sub-vibrations of the GT cut resonator which is taken as a capacitance coupling, was already-calculated in the reference[1]. The result is, therefore, employed here, also, Rei of equation(2) is calculated under the condition that a coupling between principal-and sub-vibrations is very weak. Consequently, Xe and Rei are, respectively, given as

$$Xe=-\frac{1}{\omega Co}\cdot\frac{\left(\frac{\omega^2}{\omega_1^2}-1\right)\left(\frac{\omega^2}{\omega_2^2}-1\right)-\frac{C_1C_2}{Cm^2}}{\left[\left(\frac{\omega^2}{\omega_1^2}-1\right)\left(\frac{\omega^2}{\omega_2^2}-1\right)-\frac{C_1C_2}{Cm^2}-\left\{\frac{C_2}{Co}\left(\frac{\omega^2}{\omega_1^2}-1+\frac{C_1}{Cm}\right)+\frac{C_1}{Co}\left(\frac{\omega^2}{\omega_2^2}-1+\frac{C_2}{Cm}\right)\right\}\right]} \quad (3)$$

$$Rei=Ri\left(1+\frac{C_0}{C_L}\right)^2 \qquad (i=1,2) \quad (4)$$

where ω_1, ω_2: angular frequencies of principal-and sub-vibrations uncoupled

C_0 : shunt capacitance

Cm : coupling capacitance

C_1, C_2 : motional capacitance of principal-and sub-vibrations

R_1, R_2 : series resistance of principal-and sub-vibrations

Accordingly, substituting equ. (3) into equ. (1), the oscillation frequencies $f_\pm$ are given as

$$f_\pm^2=\frac{1}{2}\left[(f_1'^2+f_2'^2)\pm\sqrt{(f_1'^2-f_2'^2)^2+4Ks^2f_1^2f_2^2}\right] \quad (5)$$

where,

$$f_1'^2=\left(1+\frac{1}{r_1}\cdot\frac{a}{1+a-h}\right)f_1^2 \quad (\because f_1=\omega_1/2\pi)$$

$$f_2'^2=\left(1+\frac{1}{r_2}\cdot\frac{a}{1+a-h}\right)f_2^2 \quad (\because f_2=\omega_2/2\pi)$$

$$Ks^2=Km^2\left(1+\frac{b}{1+a-h}\right)^2$$

$$h=\frac{1+\mu}{Rf(rd+Rd)\ \omega^2CgCd}$$

$$a=Co/C_L \qquad r_1=Co/C_1$$

$$Km^2=\frac{C_1C_2}{Cm^2}$$

$$b=Cm/C_L \qquad r_2=Co/C_2$$

Equation (5) gives two oscillation frequencies, and shows an oscillation frequency f_+ of principal vibration for a positive sign in front of the root and also an oscillation frequency f_- of sub-vibration for a negative sign. In addition, substituting equ. (4) into equ. (2) and taking $(C_0/Cd)^2 \ll 1$ into account, load capacitance C_L is rewritten as

$$\frac{1}{C_L}=\frac{\frac{1}{Cg}+\frac{1}{Cd}\left(1+\frac{Ri}{rd+Rd}\right)}{1-\frac{Co}{Cd}\cdot\frac{2Ri}{rd+Rd}} \quad (6)$$

As is apparent from equ. (5), it is easily predicted and understood that frequency temperature behavior varies with C_L, because $f_\pm$ is given as a function of Ks and also the Ks is given as a function of the C_L. By determination of an optimum value C_L, a S-TCXO extremely excellent in frequency temperature behavior is obtained as will be described later.

§ 4. Results and Discussion

According to the following procedure, the results are explained and discussed. First, electrical equivalent circuit parameters of principal-and sub-vibrations for new shape GT cut quartz crystal resonators mounted on ceramic pedestals that are used in this study, are shown. Second, frequency temperature behavior of S-TCXOs with the aforementioned resonators is shown and the S-TCXOs are found to have a frequency deviation of less than 0.7ppm in a temperature range of -30 to +85℃. Third, the thermal hysteresis characteristics for the obtained S-TCXOs are examined for the temperature excursions of four sequence. Also, the thermal hysteresis of a novel GT cut resonator with no pedestal is compared with that of the GT resonators with the pedestal. Finally, long term aging characteristics one year are shown over.

4.1 Electrical Equivalent Circuit Parameters

Table 1 represents electrical equivalent circuit parameters of principal-and sub-vibrations for new shape GT cut quartz crystal resonators that are used in the present S-TCXOs. Samples No. 1 to No. 5 are employed for thermal hysteresis characteristics and samples No. 6 to No. 13 are employed for aging characteristics.
As is obvious from Table 1, the various values are sufficiently satisfactory as a resonator.
Here, capacitance ratio r_i is defined as $r_i=C_0/C_i$ (i=1, 2).

Table 1 Values of electrical equivalent circuit parameters used in the S-TCXOs.

No.		f (MHz)	$R_{1,2}$ (Ω)	$L_{1,2}$ (H)	$C_{1,2}$ (fF)	C_0 (pF)	$r_{1,2}$	$Q_{1,2}$ ($\times 10^3$)
1	Prin.	2.0997	59.9	1.00	5.82	2.43	418	220
	Sub.	1.8941	1460	2.76	2.56		950	22
2	Prin.	2.0997	60.8	0.97	5.92	2.45	414	210
	Sub.	1.8942	625	2.68	2.63		930	51
3	Prin.	2.0997	82.5	0.98	5.86	2.43	414	157
	Sub.	1.8941	365	2.74	2.58		943	89
4	Prin.	2.0997	59.5	0.98	5.86	2.45	418	217
	Sub.	1.8941	1250	2.75	2.57		953	26
5	Prin.	2.0997	65.2	0.97	5.92	2.44	412	196
	Sub.	1.8941	565	2.67	2.64		924	56
6	Prin.	2.0997	49.5	0.99	5.83	2.48	425	264
	Sub.	1.8928	222	2.68	2.64		939	143
7	Prin.	2.0997	51.0	0.99	5.80	2.44	420	256
	Sub.	1.8934	196	2.79	2.54		961	169
8	Prin.	2.0997	58.1	0.96	5.96	2.46	413	218
	Sub.	1.8937	595	2.77	2.55		965	55
9	Prin.	2.0997	58.1	0.96	5.96	2.49	418	218
	Sub.	1.8936	1235	2.59	2.73		912	25
10	Prin.	2.0997	49.0	0.95	5.96	2.46	413	256
	Sub.	1.8927	235	2.53	2.80		879	128
11	Prin.	2.0997	52.5	0.98	5.86	2.43	415	246
	Sub.	1.8942	354	2.75	2.57		946	92
12	Prin.	2.0997	59.8	0.97	5.92	2.45	414	214
	Sub.	1.8942	736	2.67	2.64		928	43
13	Prin.	2.0997	68.5	0.98	5.86	2.43	415	189
	Sub.	1.8940	451	2.73	2.59		938	72

4.2 Frequency Temperature Behavior

Figure 3 represents the frequency temperature behavior of five S-TCXOs employing new shape GT cut quartz crystal resonators. Load capacitance C_L is so selected that the first order temperature coefficient α_+ of principal vibration reaches zero approximately at room temperature. As a result, the frequency deviation $\triangle f/f$ is easily found to be less than 0.7ppm in the temperature range of -30 to +85℃. Specifically, the frequency deviation for sample No.1 is +30/-150ppb in the foregoing temperature range. This is because the present GT cut resonator has a value of the second order temperature coefficient $\beta_+=0$ approximately and also a very small absolute value of the third order temperature coefficient γ_+ such as 10^{-12} to 10^{-13}. This value is one hundredth smaller than that of an AT cut quartz crystal resonator which is widely used. Where temperature

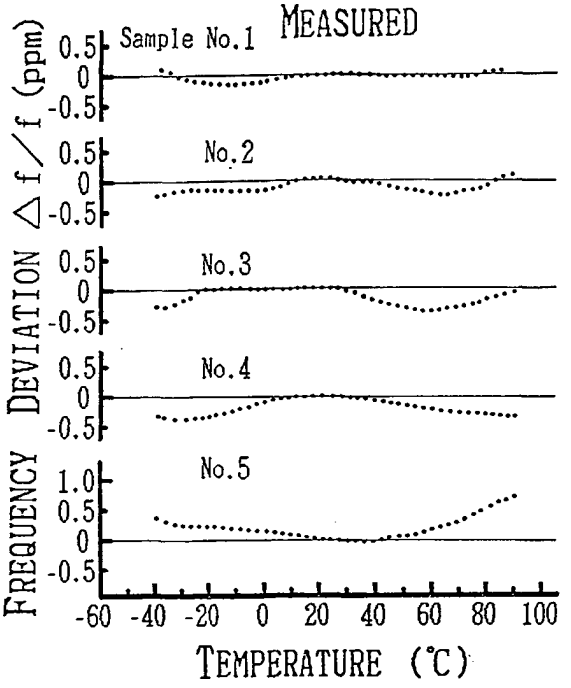


Fig. 3 Frequency temperature behavior of five S-TCXOs.

measurements were performed at the excursion of low temperature (-30 ℃) to high temperature (+85℃).

4.3 Thermal Hysteresis Characteristics

Frequency hysteresis versus temperature is next shown. The thermal hysteresis is examined for four temperature excursions of (a) to (d) as is described below, and also measured for frequency-stable time of five minutes versus temperature change of 3℃ and for shorter frequency-stable time of two minutes.

(a) +25℃→+90℃→−45℃→+25℃

(b) +25℃→−45℃→+90℃→+25℃

(c) −45℃→+90℃→−45℃

(d) +90℃→−45℃→+90℃

Figure 4 represents the thermal hysteresis characteristics of five S-TCXOs measured at an average temperature speed of 0.6℃/minute (3℃/5minutes) for each temperature excursion.
For the excursion(a), the samples show each the thermal hysteresis of 0.13 to 0.2ppm at maximum at the excursion of rise-temperature(+25℃→+90℃) to fall-temperature(+90℃→+25℃). However, the thermal hysteresis is hardly observed for the excursion of fall-temperature

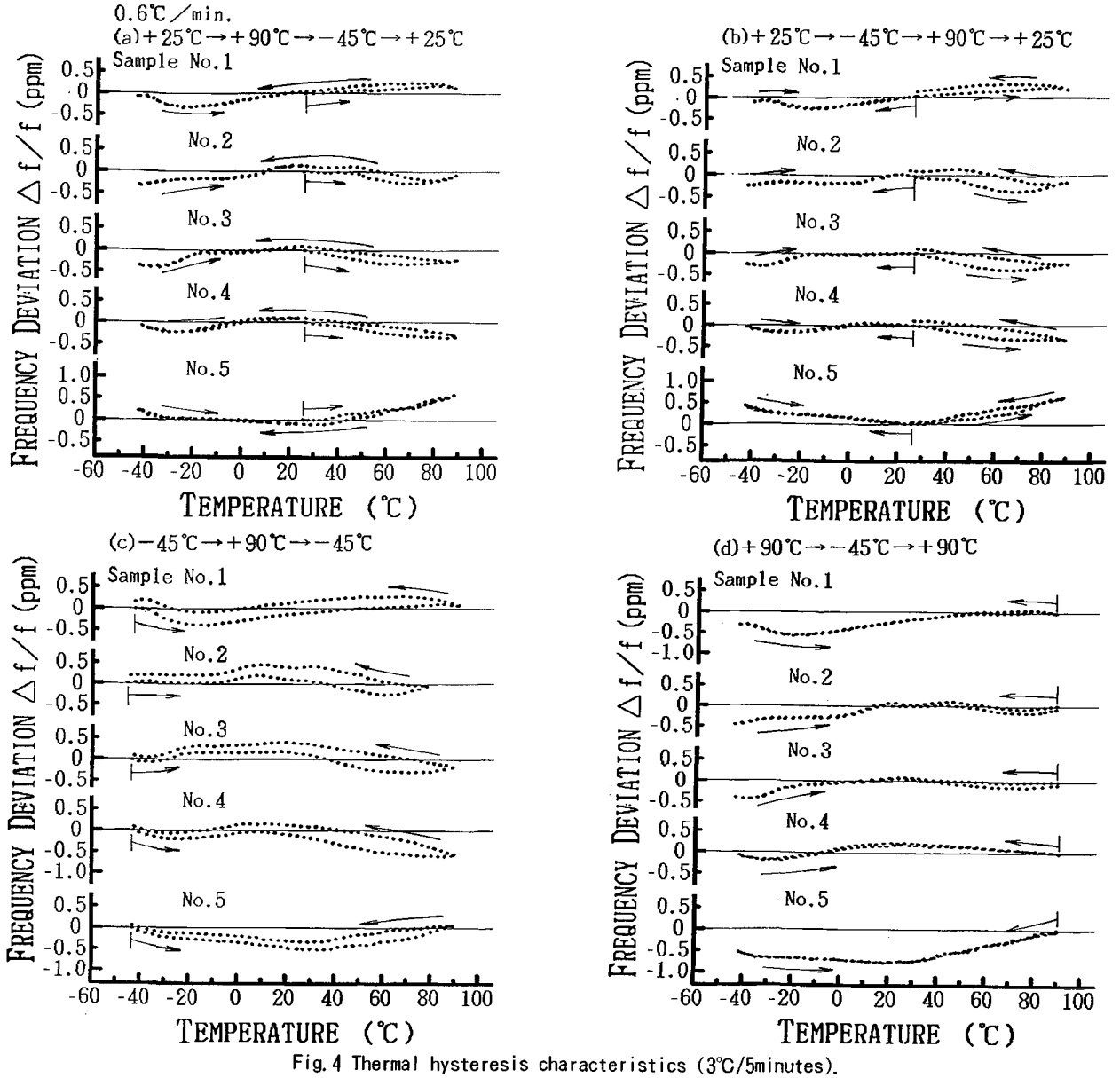


Fig. 4 Thermal hysteresis characteristics (3℃/5minutes).

(+25℃→−45℃) to rise-temperature (−45℃→+25℃). Next, for the excursion(b), the thermal hysteresis is hardly observed, similar to the excursion(a), for the excursion of fall-temperature(+25℃→−45℃) to rise-temperature (−45℃→+25℃), but the samples show each the thermal hysteresis of 0.15 to 0.31ppm at maximum for the excursion of rise-temperature(+25℃→+90℃) to fall-temperature(+90℃→+25℃).

From the results of (a) and (b), the thermal hysteresis is found to occur somewhat for the excursion of rise-temperature to fall-temperature. It is, therefore, predicted that somewhat large thermal hysteresis may occur in a wider temperature range of rise-temperature to fall-temperature. The excursion(c) is an example for the wider temperature excursion (−45℃→+90℃→−45℃) of rise-temperature to fall-temperature. Each sample shows the somewhat large thermal hysteresis of 0.25 to 0.38ppm at maximum as compared with the aforementioned temperature excursion of (a) and (b). To the contrary, it is predicted from (a) and (b) that the thermal hysteresis hardly occurs even in a wider temperature range for the excursion of fall-temperature to rise-temperature.

The excursion (d) is an example for the wider temperature excursion(+90℃→−45℃→+90℃) of fall-temperature to rise-temperature.

The samples show each the thermal hysteresis of 0.01 to 0.1ppm at maximum and the hysteresis becomes extremely small as predicted. Thus,

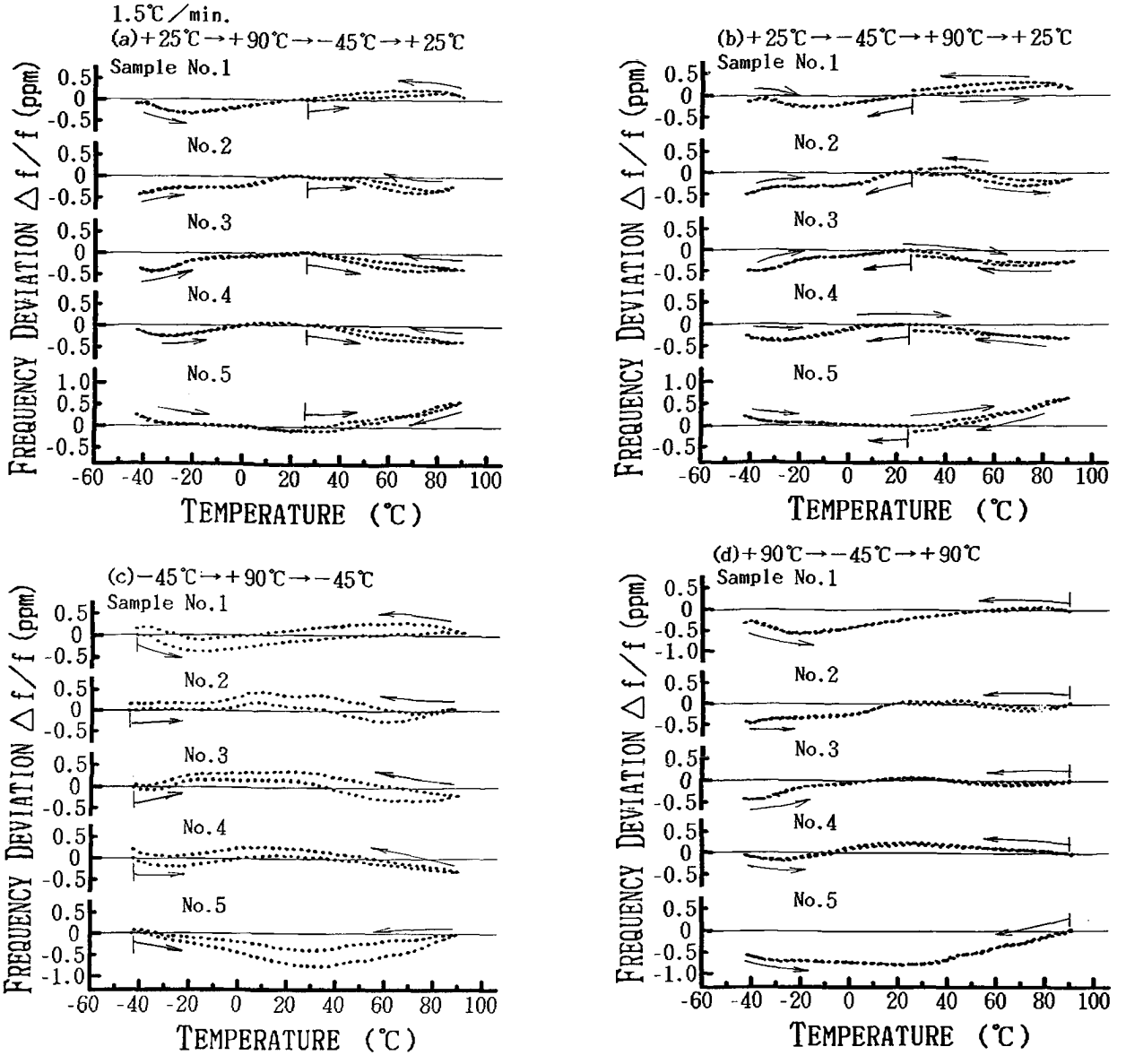


Fig. 5 Thermal hysteresis characteristics (3℃/2minutes).

frequency hysteresis versus temperature of the S-TCXOs employing new shape GT cut quartz crystal resonators is somewhat dependent on the temperature excursion, but the excellence of the S-TCXOs is due to the fact that the frequency deviation is less than 0.7ppm in a wide temperature range of -30 to +85℃ in frequency temperature behavior and 0.38ppm at maximum at the temperature excursion of -45 to +90 and -45℃ for the thermal hysteresis characteristics. In addition, the reason why causes thermal hysteresis slightly dependent on the temperature excursion is not clarified, but it is conceivable that the pedestal whereon a quartz crystal resonator is mounted, influences the resonator.

Figure 5 represents the thermal hysteresis characteristics of five S-TCXOs measured at an average speed of 1.5℃/minute(3℃/2minutes) for each temperature excursion. Namely, this is the case that the average speed is faster than that of the examples in Fig. 4. The appearance where thermal hysteresis occurs for the excursions of (a) to (d) shows the same tendency as that of Fig. 4. Further, for each excursion of (a), (b), (c) and (d), each sample has the thermal hysteresis of 0.13 to 0.18ppm, 0.12 to 0.21ppm, 0.25 to 0.38ppm and 0.02 to 0.08ppm at maximum, respectively. These results are by no means inferior to those in Fig. 4. This implies that thermal hysteresis is not dependent upon a

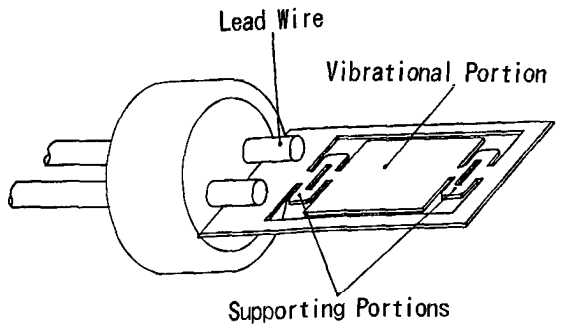


Fig. 6 A novel GT resonator with no ceramic pedestal.

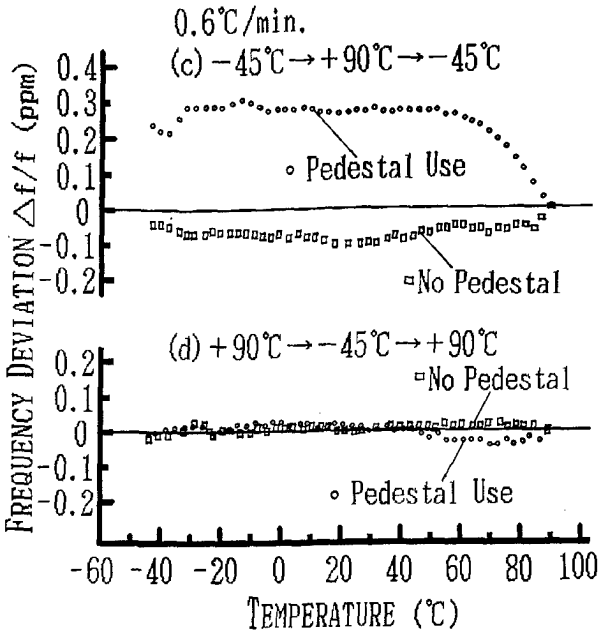


Fig. 7 Comparison of the largeness of thermal hysteresis for pedestal use and no pedestal.

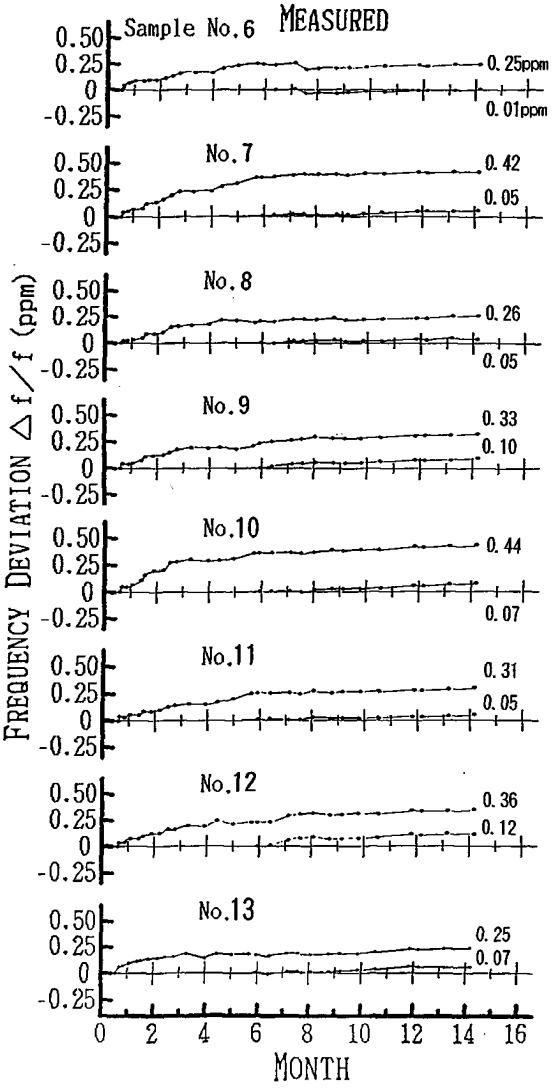


Fig. 8 Aging characteristics.

temperature speed when the speed is so taken that the temperature in an oven gets stable.

A ceramic pedestal whereon a new shape GT cut resonator is mounted, may influence the thermal hysteresis, therefore, in order to examine its influence, a novel GT resonator with no ceramic pedestal is designed and its hysteresis characteristics are examined.

Figure 6 represent a novel GT resonator of 2.1MHz and R_1=93Ω with no ceramic pedestal. The resonator comprises a vibrational portion and two supporting portions, the supporting portions are also connected to the surrounding frame, and the resonator is mounted on two lead wires.

Figure 7 represents the thermal hysteresis for pedestal use and no pedestal at the temperature excursions of (c) and (d). As is apparent from Fig. 7, the resonator with no pedestal shows the hysteresis smaller than that with the pedestal. The frequency deviation is 310ppb at maximum for pedestal use, while it is 95ppb at maximum for no pedestal. In case of the excursion (d), the frequency deviation by the hysteresis gets small, namely, it is 100ppb for pedestal use and 10ppb for no pedestal at maximum. Thus, the thermal hysteresis for the present S-TCXO was found to be very dependent on a means for mounting a GT resonator, but the reason is not known why the hysteresis takes place and also is dependent on the temperature excursion.

4.4 Aging Characteristics

Figure 8 represents aging characteristics of S-TCXOs using new shape GT cut quartz crystal resonators. Aging quantity is shown after the pre-aging of about two weeks and six months.
The aging quantity for one year and two months is from 0.25 to 0.44ppm(mean:0.33ppm) after the pre-aging for about two weeks and also the quantity for about eight months is from 0.01 to 0.12ppm (mean:0.065ppm) after that for about six months. Like this, the S-TCXOs are excellent in frequency stability for a long term.
Particularly, the frequency deviation for eight months is, as previously mentioned, 0.065ppm at the mean value after the pre-aging for six months, thus, frequency deviation for the S-TCXOs is hardly observed. Hereafter, a further long term aging will be examined.

§ 5. Conclusions

In this paper, a study was performed with a view to clarifying the frequency temperature behavior, its thermal hysteresis and aging characteristics of S-TCXOs using new shape GT cut quartz crystal resonators.

First, a circuit diagram of a S-TCXO was shown and by using new shape GT cut quartz crystal resonators, the frequency deviation was found to be less than 0.7ppm over a wide temperature range of -30 to +85℃ without any external temperature compensation. Especially, the S-TCXO has a extremely small frequency deviation of +30/-150ppb at best in the foregoing temperature range.

Second, the frequency thermal hysteresis characteristics versus temperature of the S-TCXOs were examined in a wide temperature range of -45 to +90℃. As a result, the thermal hysteresis is hardly observed for the excursion of fall-temperature to rise-temperature, but it is observed somewhat for the excursion of rise-temperature to fall-temperature, however, the variation quantity is so small as to be satisfactory as a resonator.

Finally, frequency deviation was examined for the aging term of about one year and two months. The aging quantity is 0.33ppm at the mean value after the pre-aging for about two weeks and 0.065ppm at the mean value after that for about six months. Thus, the S-TCXOs were found to be successfully obtained with extremely small aging.

As the coming subjects, an examination will be performed on the reason of thermal hysteresis occurance for the excursion of rise-temperature to fall-temperature and further long term aging.

References

[1] H. Kawashima: " An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using a Coupling Quartz Crystal Resonator" , Proc. the 44th Annual Symp. on Frequency Control, pp.585-592(1990).

[2] 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).

[3] H. Kawashima: "Variational analysis of GT cut quartz crystal resonators with the supporting portions at the ends" , Proc. the 40th Annual Symp. on Frequency Control, pp.193-200(1986).

[4] H. Kawashima: " A vibration analysis of a GT cut quartz crystal resonator taking account of dissipation term" , Trans. IEICE(A), Japan, J72-A, 4, pp.633-639 (April 1989).

[5] H. Kawashima: "An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New Shape GT Cut Quartz Crystal Resonator Formed by an Etching Method" , IEEE Ultrasonics Symp. , Proc. pp.465-470 (1989).

[6] H. Kawashima, H. Sato and O. Ochiai: " New frequency temperature characteristics of miniaturized GT cut quartz resonators " , Proc. the 34th Annual Symp. on Frequency Control, pp.131-139(1980).

[7] H. Kawashima and M. Matsuyama : " An Analysis of Electrical Equivalent Circuit Parameters for a GT Cut Quartz Crystal Resonator with the Supporting Portions at Both Ends" , Trans. IEICE(A), Japan, J71-A, 6, pp.1217-1224(June 1988).

[8] D.L.Hammond, C.A.Adams and A.Benjaminson : "Hysteresis Effects in Quartz Resonators" , Proc. the 22nd Annual Symp. on Frequency Control, pp.55-66 (1968).

[9] H.J.Forster: "Thermal Hysteresis of AT + SC-Cut Quartz Crystal Resonators" , Proc. the 36th Annual Symp. on Frequency Control, pp.140-147 (1982).

[10] R.L.Filler : " Measurement and Analysis of Thermal Hysteresis in Resonators and TCXO's" , Proc. the 42nd Annual Symp. on Frequency Control, pp.380-388 (1988).

[11] J.A.Kusters and J.R.Vig: "Thermal Hysteresis in Quartz Resonators-A Review " , Proc. the 44th Annual Symp. on Frequency Control, pp.165-175 (1990).

[12] R.L.Filler : "Thermal Hysteresis in Quartz Crystal Resonators and Oscillators " , Proc. the 44th Annual Symp. on Frequency Control, pp.176-184 (1990).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

TWO-MODE QUARTZ RESONATORS FOR DIGITAL TEMPERATURE COMPENSATED QUARTZ OSCILLATORS

Igor V. Abramson

Omsk Research Institute of Device-making,
231 Maslenikov St., Omsk, 644063, Russia

Abstract: Goal of the resonator development program described in this paper was to design dual-mode quartz resonators being free from anomalies on the both fundamental and thermosensing modes and having appropriate electrical parameters, maximum posible frequency vs. transient temperature stability, low aging and frequency hysteresis.

Well-behaved dual mode resonator of yxbl/23,25/34° cut with properties being close to propeties of SC-cut has been developed. Plano-plano plate resonator of rectangular contour are shown to be well-behaved on the both fundamental C-mode, at 10 MHz, and on B-mode, at 10,9 MHz, over, at least, a -50 to +70 °C temperature range. Such resonators are shown to be suitable for DTCXO of $1 \cdot 10^{-7}$ stability.

Introduction

The appearance of digital methods of thermocompensation provides considerable improvement of parameters of TCXO, first of all, temperature-frequency stability, which in DTCXO may be better than $3 \cdot 10^{-7}$ in wide temperature range [1]. The noval design of the oscillator demands developing the appropriate design of the resonator, which must meet following requirements: availability of temperature sensing mode in the crystal, high temperature-dynamic and temperature-frequency stability, minimum posible aging and hysteresis of frequency.

As it follows from previous studies [2], the SC-cut resonators exited by thickness electric field meet the requirements most fully. The resonators have extremely low sensitivity to temperature-dynamic influence and quite smooth temperature-frequency dependance. Also they can be excited on two modes, one of which is high stability C-mode and another is temperature sensing B-mode with frequency versus temperature slope of $-30 \cdot 10^{-6} 1/^{o}C$.

Unfortunatly developing these resonators meets the obstacles due to presence of dips of activity on the B-mode while ambient temperature is changing [3] as well as that B-mode activity exceeds the activity of C-mode.

The subject of the work was to develop dual-mode resonators operating on the fundamental C-mode at 10 MHz, and on the B-mode at 10,9 MHz, which have no anomalies in the modes behavior and possess the parametors to be suitable for DTCXO.

The modes activity relationship

That characteristic of the dual-mode resonators is very important to provide normal operating of two-mode oscillator. Usually the mode resistance relationship is to be 1.5 ÷ 3.0 within operating temperature range. In fact, the relationship in SC-cut plate with strong "energy trapping" is 0.6÷0.8 due to B-mode is of larger coupling factor than C-mode.

Indeed, B and C-modes motional resistance relationship may be defined as following :

$$\frac{r_b}{r_c} = \frac{f_c * C_c * Q_c}{f_b * C_b * Q_b} \qquad (1)$$

where f_b, f_c, r_b, r_b, C_b, C_c, Q_b, Q_c are the frequency, motional resistance, capacitance and Q-factor of B and C-modes respectively.

The motional capacitance is calculated with the well-known equation:

$$C_c = \frac{8 \cdot C_o \cdot k_c^2}{\pi} \cdot F_c , \qquad (2)$$

$$C_b = \frac{8 \cdot C_o \cdot k_c^2}{\pi} \cdot F_b , \qquad (3)$$

where C_o is a static capacitance of the resonator being of the same value for the both modes; k_c and k_b are coupling fac-

0-7803-0476-4/92 $3.00 © 1992 IEEE

tor, which for this cut are equaled to 4.52% and 4.85%, respectively [4]; F_c, F_b are coefficients depended on "energy trapping" in the plate.

Taking into account that in this resonator F_c is approximetly equal to F_b and also that f=1.9f, the expression for the resistance relationship takes the form:

$$\frac{r_b}{r_c} = 0.8 \frac{Q_c}{Q_b} \qquad (4)$$

Thus, the given resistance relationship in this design of the resonator can be obtained only by controlling the Q-factor relationship. An example of such controlling is described in [5], where to suppress the B-mode the circular plano-convex SC-cut plate is fixed at two points, placed on the line being parallel to Z' crystallografic axis. There is used that displacement of the plate particles on B-mode vibration is in direction being close to Z' that results pin more appreciable energy dissipation of B-mode vibrations in the places of mounting and hence results in its resistance increasing.

It is evident the weaker the "energy trapping" in the plate the stronger the described effect. This effect is most appreciable in the plano-plano plate, where "energy trapping" is provided by the film electrodes only. This idea is supported by experiments with the yxbl/23°25'/34°-cut rectangular plates of contour sizes 7x13 mm, which is mounted at two points, located on the short sides (Fig. 1).

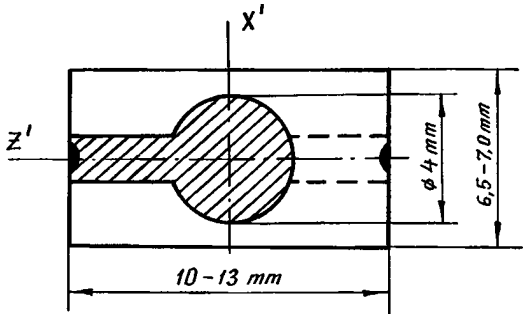


Fig. 1. Design of the dual-mode plate.

When the plate length is along Z', the relationship r_b/r_c =1.5 ÷ 3,5,while at the length being along X' - r_b / r_c= 0.6 ÷ 0.9.

Also, the plate length and the electrode thickness influence on the modes activities relationship was studied. As it can be seen from curves shown in Fig. 2, while the plate length decreases the B-mode resistance increases at almost constant resistance of C-mode. Moreover, when mass loading of the electrodes increases, the modes activity and r_c/r_b somewhat increase owing to the "energy trapping" rising.

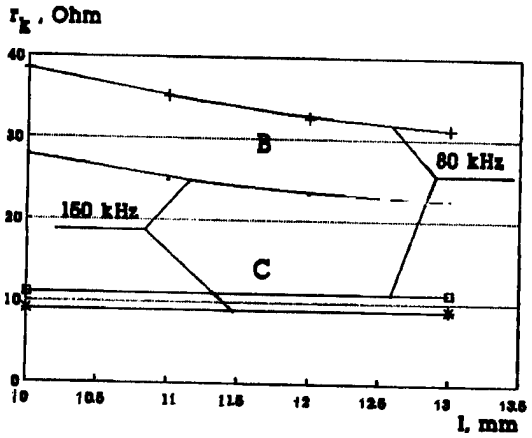


Fig. 2. B and C mode resistance vs. plate length and mass loading

Thus, varing the electrode thickness and the plate length, the modes activity relationship can be obtained to be suitable for dual-mode oscillators.

Dips of activity

The dips of activity is the main reason that the dual-mode resonators are not used in DTCXO. Unlike plano-convex plates, in which dips of activity take place on the B-mode only [3], in the plano-plano plate with weak "energy trapping" they occur on the both modes. Fig.3 shows the typical behavior of motional resistance of yxbl/23°25'/34°-cut resonator on C and B modes in wide temperature range. Frequency of fundamental C-mode was 10 MHz, and of temperature sensing B-mode - 10,9 MHz.

It is seen that deep activity dips down to 200-300 Ohms take place on the both modes that results in these resonators are not suitable for the DTCXO.

The nature of the activity dips in double-rotated cut plates is difficult to determine. But the most like cause is coupling the main resonance with spurious modes at the frequencies of these vibrations coincide.

On the base of AT-cut resonator designing experience it has been supposed that the most like cause of the activity

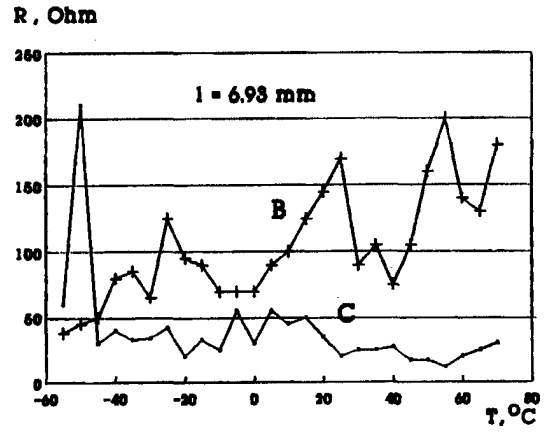


Fig. 3. Typical B and C modes resistance vs. temperature behavior at unoptimal width of the plate.

dips of the resonator with weak "energy trapping" is couple of main thickness-shear vibrations with flexuable or contour-shear modes. It is evident that the coupling with these vibrations may be changed by varying the plate contour size which controls their frequencies but doesn't influence on the B and C-mode frequency.

The behavior of 100 resonators of different plate width from 6.5 to 7.0 mm vs. temperature was studied on B and C modes. The statistic treatment of experimental data has been made. Fig. 4 shows part of units without dips on B and C modes (within given temperature range) in total number of tested resonators vs. the plate width.

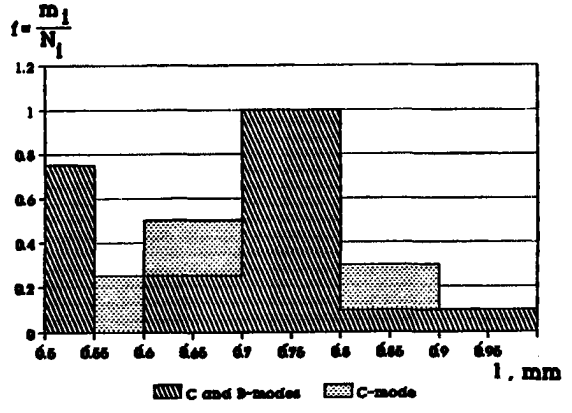


Fig. 4. Statistic data of well-behaved dual-mode resonators vs. plate width.

From the analysis of the experimental data, the following has been found.

The B and C-modes behavior within temperature range depends on the plate's width.

There are optimal values of the width for each of the modes, at which the activity dips are absent.

There are optimal width values, at which the activity dips on the both modes are absent. For the studied values of the plate width, the optimal sizes are located in area from 6.70 to 6.80 mm.

The optimal sizes remain the same at the different plate's lenth at least from 8 to 13 mm.

The typical resistance vs. ambient temperature behavior of the "optimal" resonators are shown in Fig. 5. It is seen neither dips no significant resistance fluctuatin are on the both modes within temperature range from -50° to $+70^{\circ}$C.

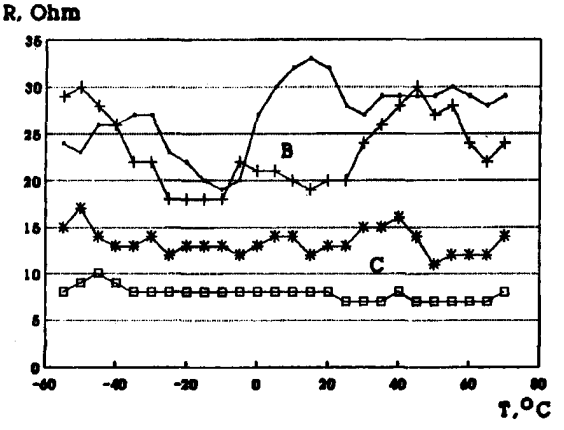


Fig. 5. Typical resistance vs. temperature of the "optimal" resonators.

Basic characteristics of two-mode resonators

Temperature-frequency characteristic

The temperature-frequency dependence of yxbl/23,25'/34-cut resonator has the same as SC-cut cubic paraboloid form but with the inflection point at 110°C. To minimize the frequency nonstability of the resonator within operating temperature range the lower turnover temperature T_o is placed close to the temperature range middle. Fig. 6 shows the form of these temperature-frequency dependence with the extremum at 10°C. It is seen, that the temperature nonstability in the range $(-50 \div +70)^{\circ}$C is about $50 \cdot 10^{-6}$, that approximately corresponds to the AT-cut resonator frequency nonstability.

The T_o versus the plate orientation slope in this temperature range is about 2 $^{\circ}$/min. It is somewhat less then in the

AT-out plates that makes the resonator convenient to produce.

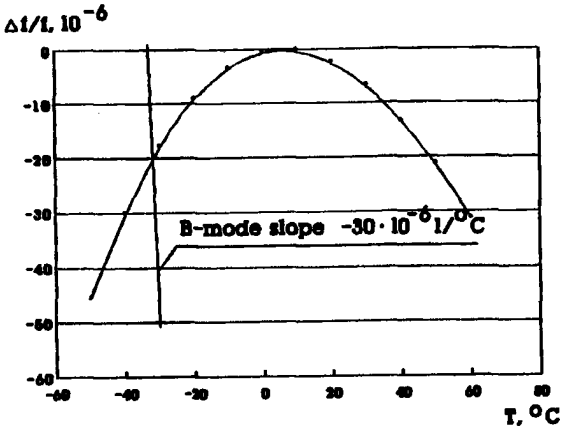


Fig. 6. The B and C modes frequency vs. temperature.

The B-mode frequency versus temperature dependence displied in Fig. 6, has almost linear form with slope of $-30 \cdot 10^{-6}$ $1/^{\circ}C$. So significant slope provides high accuracy of the plate temperature determination that results in high stability of DTCXO despite the resonator and oscillator aging.

Temperature dynamic charasteristics

Temperature dynamic characteristic describes behavior of resonators under the ambient temperature fast changing. Since the temperature nonstability of DTCXO is reduced significantly the temperature-dynamic nonstability becomes one of the main part in total nonstability of the oscillator.

It's convinuent to determine the temperature-dynamic nonstability with temperature-dynamic coefficient of frequency, which is equiled to:

$$\bar{a} = (\Delta F / F)_{max} / (\Delta T/ t)_{max} , \quad (5)$$

where $(\Delta F/F)$ - maximum dynamic frequency shift under ambient trmperature changing with maximum speed of $(\Delta T/ t)_{max}$.

This coefficient for noval dual-mode resonators has been obtained.This characteristic was studied by following means. The resonator was placed into temperature chamber with temperature equaled the lower turnover temperature of the resonator.The resonator was excited in two-mode oscillator and the modes frequency were measured. Then the chamber frequency was increasing for $1^{\circ}C$ with maximum speed of 2.5 $^{\circ}$/min. During the temperature rising the transient frequency vs.time characteristics of B and C modes were measuring.

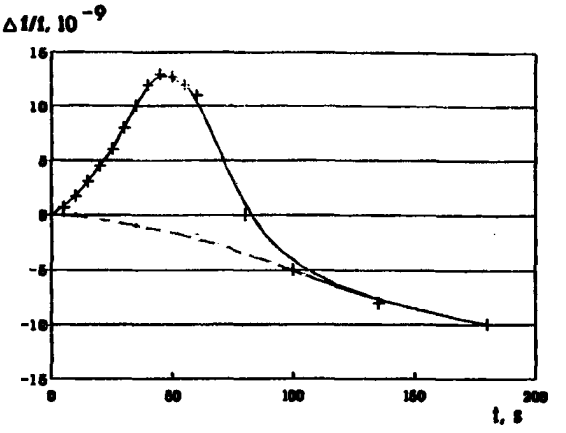


Fig. 7. The C-mode frequency vs.transient temperature.

The transient C-mode characteristic is ploted in Fig. 7. The data of B-mode transient characteristic as well as the C mode frequency vs. temperature dependence have been used to define the "static" frequency vs. time dependence, which is dotted in the same figure. Differense between these curves gives a "pure" temperature-dynamic shift, which is shown to has a maximum of $1{,}5 \cdot 10^{-8}$. So the $\bar{a}$-coefficient equals to $6 \cdot 10^{-9}$min/$^{\circ}C$. For the comparence the $\bar{a}$ of AT-cut resonators equals to $-1 \cdot 10^{-7}$ min/$^{\circ}C$.

The temperature-dynamic frequency shift under some other rate of temperature changing for the developed resonators can be defined as following:

$$\Delta F/F = 6 \cdot 10^{-9} \Delta T/t \quad (6)$$

Since the real rate of the ambient temperature changes usually doesn't exceed 5 $^{\circ}$/min the temperature-dynamic nonstability of the DTCXO is less than $3 \cdot 10^{-8}$.

Equivalent electrical parameters

The main electrical parameters of developed two-mode resonators are given in table.

As it follows from the data values of motional resistance on B and C-modes are quite acceptable for excitation in the oscillator. The capacitance relationship confidently allows to alter the frequency in the range of about 1 kHz that is sufficiently to compensate the resonator temperature-frequency nonstability in a wide temperature range.

Parameters	C-mode	B-mode
Frequency, MHz	10.0	10.9
Motional resistance, Ohm	8-15	20-30
Dynamic capacitance, pF	0.004	——
Q - factor, 10^3	300-400	——
Capacitance relationship C_m/C_o	0.001	——

Frequency hysteresis and resonators aging

Frequency hysteresis of resonator can be determined as difference in frequency measured at the same temperature, which appears under cyclic ambient temperature changes due to not returning alteration in the resonator properties. As result of precise compensation of the temperature nonstability the hysteresis is one of the factors that limits stability of DTCXO.

The frequency hysteresis of the developed resonators was measured by the following way. The resonators were placed into the thermal chamber, the temperature of which equals to their turnover temperature T_o, and their frequency was measured. Then the chamber temperature was changed in accordance with the diagram: T_o - (-60°C) - T_o - (+90°C)- T_o with holding at these points for an hour and measuring the frequency at temperature equaled to T_o. The difference between the frequency measurements defines the resonators frequency hysteresis.

The test of the developed resonators has shown that about half of 20 samples has hysteresis of $(0.5\div1.0)\cdot10^{-8}$ and the left part has $(1\div3)\cdot10^{-8}$.

There were no tests of B-mode hysteresis. However it doesn't influences noticeably on the oscillator frequency vs. temperature stability due to B-mode large frequencyvs. temperature slope.

Fig. 8 shows the typical aging curves of the two-mode resonators. As it can be seen from given curves after 8 months of test the frequency shift of most part of resonators doesn't exceed $3\cdot10^{-7}$, and the daily frequency drift after 30 days of initial operating time is $(1\div2)\cdot10^{-9}$.

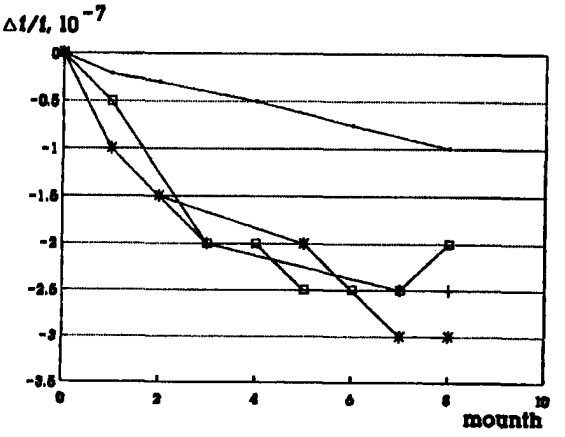


Fig. 8. Typical the dual-mode resonators aging.

Obtained results are much better then ones of the similar AT-cut resonators although they are somewhat worse than that of SC-cut resonators operating on the third mechanical harmonic [3].

Conclusion

The dual-mode yxbl/23°25'/34°-cut resonators have been developed, in which the quasi-shear B-mode having frequency versus temperature slope of about $-30\cdot10^{-6}$ 1/°C is used as a thermosenser.

The resonator design provided absence of motional resistance anomalies vs. temperature on the both C and B modes has been found.

The basic characteristics of the two-mode resonators (temperature-frequency, temperature-dynamic, electrical, aging, frequency hysteresis) have been studied, which are shown to be appropriate for DTCXO of $1\cdot10^{-7}$ total nonstability and $(2-3)\times10^{-7}$ per year aging.

Acknowledgement

The author thanks Nataly N.Kuchatovu for assistance at carring out the experiments.

Reference

1. Top Products 1990, Microwaves & RF, December, 1990, p.120.

2. J.A.Kusters, M.C.Fisher and J.G.Leach, "Dual-Mode Operation of Temperature and Stress Compensated Crystal", Proc. 32nd Ann. Symp. on Frequency Control, 1078, pp. 389-397.

3. R.l.Filler, J.R.Vig, "Resonators for the Microcomputer Compensated Crystal Oscilator", Proc. 43th Ann. Symp. on Frequency Control, 1989, pp. 8-15.

4. R.L.Filler, J.R.Vig,"SC-cut Quartz Resonators with Suppressed B-mode", Patent No 4,381,471 USA, 1983.

5. A.Ballato, "Doubly Rotated Thickness-mode Plate Vibrators", Physical Acoustics, Editted by W.Mason, 1979, pp. 115-181.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ANALYSIS OF START-UP CHARACTERISTICS OF CMOS CRYSTAL OSCILLATORS

Masahiro Toki and Yasuo Tsuzuki
Division of Electrical & Computer Engineering,
Yokohama National University, Yokohama 240,Japan

Abstract: In this paper, an analysis method of the start-up characteristics of CMOS crystal oscillators is described. First, the origin causing oscillation to start just after d.c. power supply is switched on has been described and confirmed by experiment. Then, this analysis method has been outlined which can calculate the overall start-up characteristics of the amplitude of oscillation. Finally, experimental results have been given to verify the usefulness of this analysis method.

Introduction

Crystal oscillators are in wide use in communication and measuring equipments as sources of standard frequency or of clock of the equipments. According to the development of mobile communication, switching on-off of the power of the equipment is made very much frequently for power saving. Accordingly, the analysis of the start-up behavior of the oscillators is desired. In one of the publications of IEC an approximation formula has been considered to calculate the start-up time[1].

Start-up behavior of the common *Colpitts* crystal oscillator using a bipolar transistor has been investigated[2]. In the present paper, the overall start-up behavior of CMOS crystal oscillators which are in wide use as clock sources is investigated.

In the first section, we discuss the start-up behavior, especially the origin causing oscillation to start, i.e., oscillation triggering. In the next section our analysis method of the overall start-up characteristics is outlined. In the last section, experimental results to verify the usefulness of the present analysis method are described.

Start-up Behavior of Oscillator

The oscillator circuit configuration considered in this paper is the common *Colpitts* crystal oscillator using a CMOS INVERTER, and is shown in Fig.1. When the input voltage is in the low level, the MOS-FET(FET1) on the side of the d.c. power supply conducts and the one(FET2) on the side of the ground is open circuited. Then, when the input voltage jumps to the high level, FET1 is made open circuited and FET2 conducts. A high resistance R_f is connected to produce the d.c. bias of the CMOS circuit at $V_G = V_D = V_{CC}/2$, and the CMOS circuit is used as an inverting amplifier of the crystal oscillator. The upper side of the terminals is the crystal resonator shown in the form of an equivalent circuit.

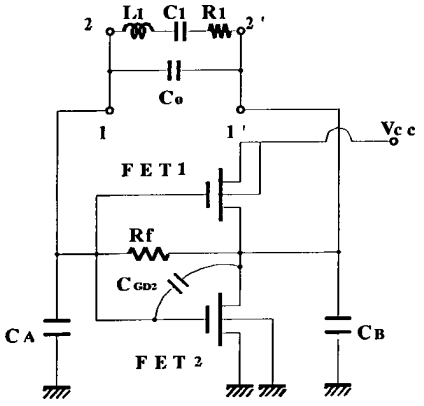


Fig.1. CMOS crystal oscillator circuit.

The overall start-up behavior of oscillator in this circuit has been measured to see how the behavior is. The measured result of the oscillator voltage V across the resonator is shown in Fig.2.

0-7803-0476-4/92 $3.00 © 1992 IEEE

As shown in this figure, the oscillation amplitude suddenly buildsup to a certain value much larger than a value in a noise level when d.c. supply is turned on. Then the amplitude of oscillation V increases in an exponential manner, in other words, it increases linearly in logarithmic scale. Finally the oscillation amplitude V gradually approaches the steady-state level.

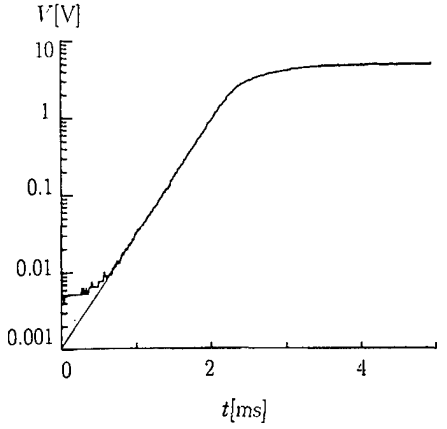


Fig.2. Start-up characteristics of oscillation current of motional arm of crystal resonator.

In order to clarify the origin causing oscillation to start, we consider how the d.c. bias of the CMOS device builds up. Immediately after the d.c. power supply is turned on in the circuit shown in Fig.1, FET1 is conducting while FET2 is open circuited. There are junction capacitances between the drain and the gate of FET1 and FET2 and a channel on-resistance of FET1, before the operating point of the CMOS circuit builds up. Therefore, the circuit shown in Fig.2 can be equivalently converted to a circuit shown in Fig.3 immediately after the application of d.c. supply. We include the capacitance C_{GS2} between the gate and the source of FET2 in the capacitance C_A.

The channel on-resistance is usually so small that the capacitances are charged fast, and hence, a step-like voltage appears across the resonator. A step-like voltage has been investigated for cases where the gate of one or both MOS-FETs are held on the d.c. power supply voltage V_{CC} prior to the switching-on[3]. However, the step voltage should be investigated for a more general case where V_{CC} is switched on to the whole CMOS circuit. The motional arm of the equivalent circuit of the resonator is triggered simultaneously by the step-like voltage across C_{GD1}. Both of the rise time and the value of the step-like voltage across the resonator can be estimated using this small signal equivalent circuit in Fig.3, and therefore, the initial amplitude of oscillation current I_0 can be calculated using the motional arm parameters, i.e., the $L_1C_1R_1$ equivalent circuit of the resonator with the applied step-like voltage across it, as shown in Fig.4. The rise time of the step-like voltage V_{X0} is so small as 10 to the minus 9th of second that it can be neglected.

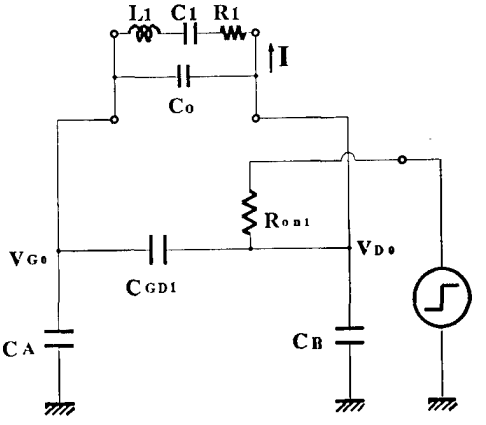


Fig.3. Equivalent circuit immediately after the application of d.c.power supply.

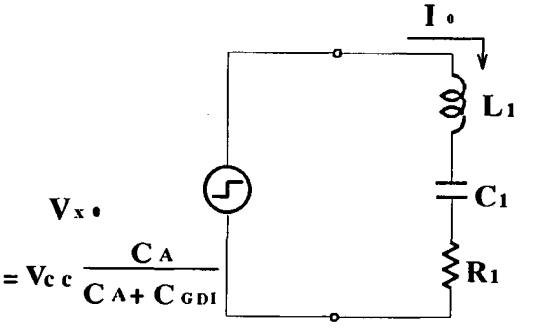


Fig.4. Initial excitation of resonator.

Based on these discussions, it has been confirmed clearly that d.c. supply switching is the most important oscillation trigger in this circuit configuration and a circuit stimulation can be performed without any random noise sources in the oscillator circuits.

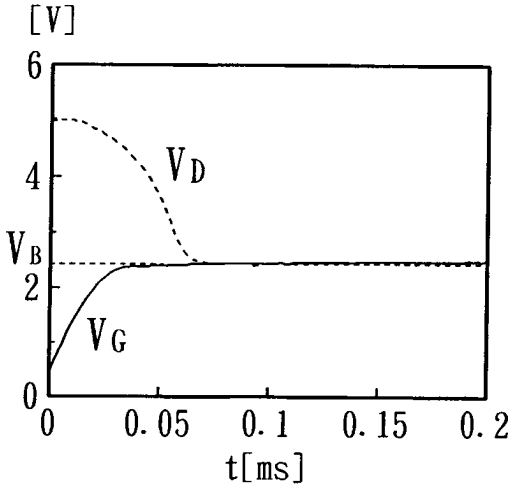


Fig.5. Transient characteristics of gate and drain bias voltages.

After the switching on, the d.c. voltage of the input and output terminals varies to reach the steady state bias voltage V_B at a certain time t_B as shown in Fig.5. Namely, switching on the supply, the drain voltage jumps up to V_{CC} while the gate voltage jumps up to V_{G0} where $V_{G0} = V_{CC}[C_{GD1}/(C_A + C_{GD1})]$. Then the gate voltage V_G grows in an exponential manner, with a time constant $\tau = C_A R_f$, while the drain voltage remains at V_{CC}. When the gate voltage reaches the threshold voltage of the FET2, FET2 begins to conduct and hence the drain voltage begins to fall off until both the drain and gate voltages reach the steady-state d.c. bias V_B. Then the CMOS device gets in a certain linear operation region to make the amplitude of oscillation grow. We consider how the amplitude grow in the next section.

Outline of Analysis Method

As shown in Fig.6, a crystal oscillator can be equivalently converted to a series LCR loop circuit where L is just the motional inductance L_1 of the resonator, C and R are determined by the parameters of the resonator, and the negative resistance $-R_L$ and the equivalent capacitance C_L of the active circuit, containing the shunt capacitance C_0 of the resonator in the active circuit.

From Fig.6, the start-up crystal current I is given as follows:

$$I = I_0 e^{\sigma t} \tag{1}$$

where

$$\sigma = [R_L(t) - R_1]/2L_1 \tag{2}$$

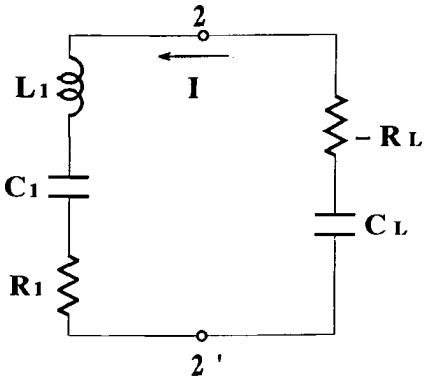


Fig.6. Equivalent circuit of crystal oscillator.

The negative resistance $-R_L(t)$ changes its value as the time goes by. We investigate the overall start-up behavior, dividing the start-up time into two time intervals i.e., an interval between the instance of switching on and the instance of the steady-state d.c. bias establishment and another interval thereafter up to the steady-state oscillation. First, according to the building up of the d.c. bias of the CMOS circuit, the negative resistance $-R_L$ varies, from a certain positive value, to a certain small signal negative resistance $-R_L(t_B) = -r_L$ at the steady-state d.c. bias. We assume that $r_L > R_1$ so that a sure oscillation occurs. In a short time interval Δt, however, the negative resistance $-R_L(t)$ can be treated as a constant. Then the increment ΔI of the oscillation current I can be calculated by using Equ.(1) in this short time interval Δt as shown below:

$$\Delta I = I_0 \sigma e^{\sigma t} \Delta t \tag{3}$$

where σ is given by Equ.(2).

By employing the transient characteristics of the negative resistance, the resonator current level I_B at the instance of the establishment of the d.c. bias can be obtained, by integrating the contributions ΔI in every short time intervals Δt. The current I_B becomes the initial current in the following time interval after the establishment of the d.c. bias. Characteristics of the negative resistance $-R_L(t)$ can be obtained by using the transient characteristics of the d.c. bias V_G as shown

in Fig.5, and the relationships between the d.c. bias voltage V_G and the negative resistance $-R_L$ as shown in Fig.7. The result is shown in Fig.8.

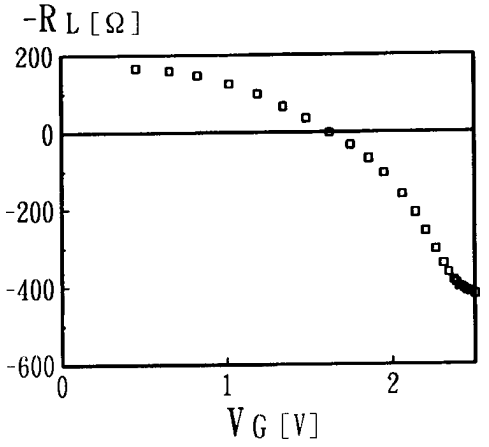


Fig.7. Relationships between gate voltage and negative resistance.

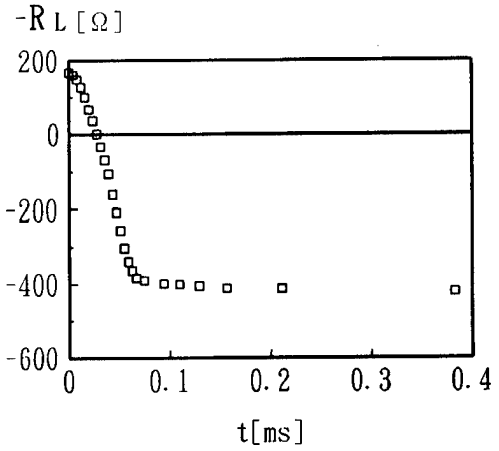


Fig.8. Transient characteristics of negative resistance.

Next, we investigate the start-up behavior in the time interval after the establishment of the steady-state d.c. bias. From Equ.(1), the start-up time measured from t_B, denoted as t, is given as follows:

$$t = (1/\sigma)\log(I_s/I_B) \qquad (4)$$

where σ is given in Equ.(2) and I_B is the amplitude of the oscillation current at the instance t_B as obtained above, and I_s is its amplitude at the steady-state. Due to the large signal nonlinearity of the CMOS device, both magnitudes of the equivalent capacitance C_L and that of the negative resistance $-R_L$ change with the oscillation signal level, i.e., the level of the oscillation current I as shown in Fig.9. Therefore, the start-up time

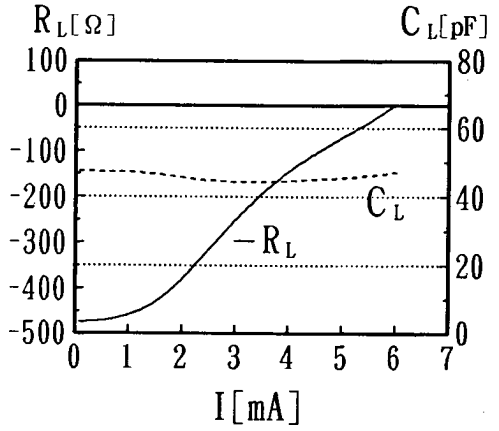


Fig.9. Relationships between signal level and negative resistance

calculated by Equ.(4) is valid only in the linearly increasing region as seen in Fig.2. In a relatively short time interval Δt, however, this LCR circuit can be treated as a linear circuit since Q is so high. In other words, both the equivalent capacitance C_L and the negative resistance $-R_L$ can be considered approximately as linear elements in such a short time interval. Therefore, the short time interval Δt can be calculated by using Equ.(4) in the small increment ΔI of the oscillation current I in this circuit as follows:

$$\Delta t = (1/\sigma)\log[(I + \Delta I)/I] \qquad (5)$$

where the negative resistance $-R_L(I)$ is a function of the current level I only and not of the increment ΔI. By employing the relation between the resonator current level and the negative resistance, the overall amplitude characteristics from the start-up to the steady-state can be calculated by integrating the contributions in every small increments ΔI of the oscillation current level I.

Experimental Results

The analysis method presented in the above section was applied to two circuit examples designed to have different values of the negative resistance $-r_L$ at small signal level. Their circuit parameters

are listed in Table 1. The equivalent circuit parameters of the crystal resonator which was used in the experiment are given in Table 2.

Table 1. Parameters of two circuit examples

#	C_A[pF]	C_B[pF]	$-r_L$[ohm]
1	75	75	-476
2	150	150	-167

#	V_{X0}[V]	I_0[μA]	t_B[ms]
1	4.83	2.85	0.36
2	4.58	2.71	0.13

Table 2. Equivalent circuit parameters of crystal resonator.

f_s[MHz]	L_1[mH]	C_1[fF]	R_1[ohm]	C_0[pF]
4.09	65.8	23.0	8.6	5.1

By using the present analysis method, the overall start-up characteristics of the oscillation current I in these two circuit examples were easily obtained. Comparison between experiment and calculation was made, and the results are shown in Fig.10. The curves show the amplitude of the voltage V across the terminal of the resonator by multiplying the amplitude of the current I by the input impedance of the terminals 2-2' shown in Fig.6. Good agreement between experiment and calculation demonstrated that the present analysis method is quite useful for practical use.

Conclusions

A simple analysis method of the overall start-up characteristics of CMOS crystal oscillators has been described. After clarifying the origin causing the oscillation current to start by means of a small signal circuit analysis and experiment, the analysis method, employing the relation between

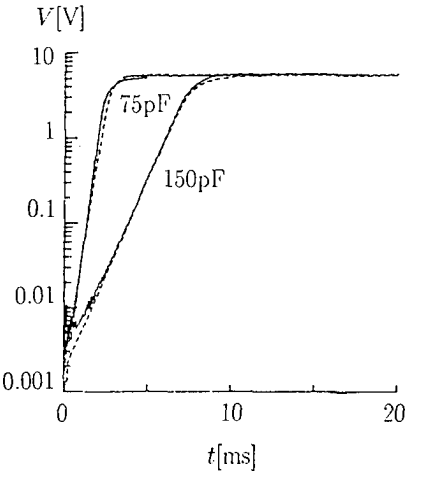


Fig.10. Start-up characteristics of voltage across terminal of resonator of two circuits.

the oscillation current and the negative resistance and the equivalent capacitance, has been developed. It has been demonstrated that the present analysis method can easily calculate the overall start-up characteristics of the amplitude of oscillation. The present method is useful for calculation of the rise time of CMOS crystal oscillators.

References

[1] International Electrotechnical Commission, "Quartz crystal units for frequency control and selection", IEC Standard Publication 122-2,Second Edition,1983.

[2] Jin-Qin Lu,Yasuo Tsuzuki, "Analysis of Start-up Characteristics of crystal oscillators", Proceedings of 39th Frequency Control Symposium,1991.

[3] Andreas Rusznyak, "Start-up Time of CMOS Oscillators", IEEE Trans. on Circuits and Systems,Vol.CAS-34,No.3,Mar.1987.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

UNIVERSAL, COMPUTER FACILITATED, STEADY STATE OSCILLATOR, CLOSED LOOP ANALYSIS THEORY

SOME APPLICATIONS TO PRECISION CRYSTAL OSCILLATORS

Benjamin Parzen
consulting engineer
3634 Seventh Avenue
San Diego, CA 92103
(619) 291 0567

ABSTRACT

This paper considers applications, to precision crystal oscillators, of the titled theory, presented in previous papers. The oscillators are familiar and unfamiliar types. The theory and applications are integrated by frequent in context reference to the appropriate sections of Ref 1 which last presented the theory.

1. INTRODUCTION

Ref 1 presented the complete theory of the title. It also included some applications in the microwave frequency region. Referenced portions of that paper are identified by an !.
This paper describes some applications to crystal precision oscillators and should be treated as a continuation of Ref 1.

The data was obtained with program BPT, also described in Ref 1, as directed by the user guided by the theory.
The circuit is entered into the computer as a NETLIST via a file or the keyboard. The computer translates the netlist into a PARTS LIST which is readily understood by any user. The user then interactively directs the computer to generate the desired data.

The program is basically an elaborate laboratory simulator with extensive stockroom, fabrication, instrument room, measurement, housekeeping, and recordkeeping facilities, unmatchable in any real laboratory. At present, the program is available for the IBM PC, AT etc. and compatible computers.
The user proceeds, controls and operates the program as if he or she were constructing, testing, and then modifying the "simulated breadboard circuit", as directed by the user and the program, in exactly the same manner as in a real laboratory, but much more expeditiously, accurately, thoroughly. and with much greater understanding. The important difference is that the simulated breadboard includes only the information as directed by the user but the real breadboard also includes intrinsic information, unknown to the user, such as stray and parasitic components and frequencies. This difference signifies that only about 90 % of the real laboratory testing can be eliminated by the computer simulation.

From this program description, is seen that the general analysis and design modification procedures consists of the following 5 steps all performed within the program environment.

1. Construction of the oscillator.
2. Trimming this oscillator to the desired frequency.
3. Analyzing and evaluating the oscillator performance with the aid of the extensive measurement facilities within the program.
4. Modifying the oscillator to improve the performance.
5. Repeating steps 3 and 4 until the desired performance is obtained.

6. The analysis is then confirmed by constructing and testing the real oscillator to check the correct entrance of the parts and layout data into the computer and to be alerted of important omissions in the data.

It will be noted that steps 1 to 5 are exactly those followed in the real laboratory but slightly modified for use with the theory. The effort and time required to perform these steps will be a small fraction of those for

0-7803-0476-4/92 $3.00 © 1992 IEEE

step 6.
The main difference between the this method of analysis and the customary present methods are:

1. The circuit of the device being analyzed is that of the full real oscillator and not a possibly poor approximation incapable of producing all the important and correct data.
2. The type of data obtained closely resembles that for the real oscillator and, in addition, types, practically unobtainable in the real laboratory, are available.

The difference is primarily due to the closed loop analysis and the noise source, amplifier and filter oscillator model, made possible by the computer and the theory, as contrasted with the customary open loop analysis . It should be remembered that the real oscillator operates closed, and not open, loop.

The applications are

1. A 10 MHz 1 resonator Colpitts oscillator.
2. A 10 MHz 1 resonator Pierce oscillator.
3. A 10 MHz 1 resonator DXO oscillator. (differential crystal oscillator)
4. A 10 MHz 2 resonator oscillator.

To facilitate comparison, all applications use the identical resonator, a 10 MHz SC cut, 3rd overtone crystal, XL, having R1 = 70 Ohms, C1 = 2.1E-16 Farad and Qx = 1.083E6. The value of Co is set at 1E-50 and may be modified as desired.

All the oscillators are provided with an ALC system to maintain the crystal current, Ix, at 1 mA. In the simulated oscillator, the ALC system is represented by dR and Ix is set by setting the magnitude of dR. (see Ref 1)

Application 1 has been and is being manufactured in very large quantities and it is difficult to appreciate the value of a detailed analysis at this stage in its design history. However, the analysis is still useful, at this time, in the following respects:
1. It provides a greater understanding of the oscillator operation.
2. It clearly demonstrates the validity of the complete design basis including the optimum noise performance.
3. It serves as a production control tool for quickly determining the effect of changes in part characteristics upon the total oscillator performance and thus providing information as to the permissibility of substituting parts with these changes. Such changes are very often found necessary during production.

Application 2 is the oscillator of application 1 but with identical components in a Pierce oscillator configuration.The purpose of application 2 is to permit the comparison of the performance of the Colpitts and Pierce oscillators with identical components.
Some writers maintain that there are substantial differences while others believe that the differences are minor and are due to the introduction of the grounded power supplies and to the effects of the parasitic elements, such as capacitances between terminals, which varies with the circuit and the location of the element with respect to ground. Application 2 can serve to help settle this controversy.

Application 3 is the balanced differential oscillator, DXO, described in Ref 7. It is configured in its most basic form and as similar to application 2 as possible.
Some writers claim substantial advantages for this circuit. The purpose of application 3 is to permit the comparison of the performance of the Pierce and DXO oscillators with identical major components.

Application 4 is an example of the use of the theory and program as a research and development tool. This oscillator has never been built and it is advisable to do a preliminary computer study to explore its behavior, including noise, and desirability prior to more intensive computer studies and expensive experimental efforts.

The data for these applications are presented in the form of simplified schematics, typical netlists, a typical parts list, and plots of the more important, and infrequently or not previously published, operating characteristics. Principal performance data and comments on the data are also included.

The oscillator plots are for 2 quantities versus the Fourier frequency, f.

1. The circuit transformation of residual noise at location m, CTRm (f).
2. The magnitude of the closed loop impedance, Zin (f), at the input terminals of the active device.

The CTRm function is described in Sect 3.12 !. If the noise performance of the oscillator, $\mathcal{L}_m$ (f), has been

experimentally determined and CTRm (f) has been calculated, then the residual noise can be calculated from Eq 25!.
The Zin quantity determines the contribution of the active device input noise current, In, to the oscillator noise as it produces a noise voltage, En = In * Zin, across the active device input terminals. It is therefore very important, when measuring the device residual noise, that the device be terminated to simulate the impedances present in the closed loop oscillator.

In this connection, the noise currents may be determined by measuring the residual noises at the calculated terminations and then calculating the corresponding noise currents (see Sect 3.12.3c!).

2. 10 MHz 1 RESONATOR COLPITTS OSCILLATOR

Its schematic is shown in Fig 1.

In addition to the crystal and the transistor, there are 5 components which are critical and therefore must be carefully controlled; CA, LA, CN, LN, and C'L.

CA and LA make up the resonator mode selector network, X1 (see Ref 3). CN and LN make up the resonator overtone selector network, X2. It is possible to combine the overtone and mode selector functions into 1 three element network, either in X1 or X2 (see Refs 3 and 6). However, in production, the control of the elements becomes more difficult.

C'L is the tuning element of the resonator. It may be a capacitance, inductance, or a network including a tuning diode.

While these 5 components have been classified as being critical, the stability of seemingly unimportant components should not be neglected. For example, a change of 1 % in bypass capacitor CB2 causes an oscillator frequency shift of 3E-10.

The use of elements consuming RF power has been minimized so that the calculated oscillator Qop, 9.95E5, is close to Qx. This is true only when RL is 1 Megohm. Table 1 gives the carrier performance data for decreasing values of RL. Qop quickly decreases with further reductions in RL.

The meanings of the symbols in Table 1, (the references are to items in Ref 1) are:

RL is the load resistance set to the values shown.
dR, described in Sect 3.5!, is set so that Ix = 1mA.
Calculated V_{RL} is the voltage across RL.
Calculated Vins is Vin, described in Sect 3.12!, = Vs described in Sect 3.5!.
Calculated RT, --> -RN, is described in Eqs 2, 13, and 17!.
Calculated Qop, is described in Sect 3.13!

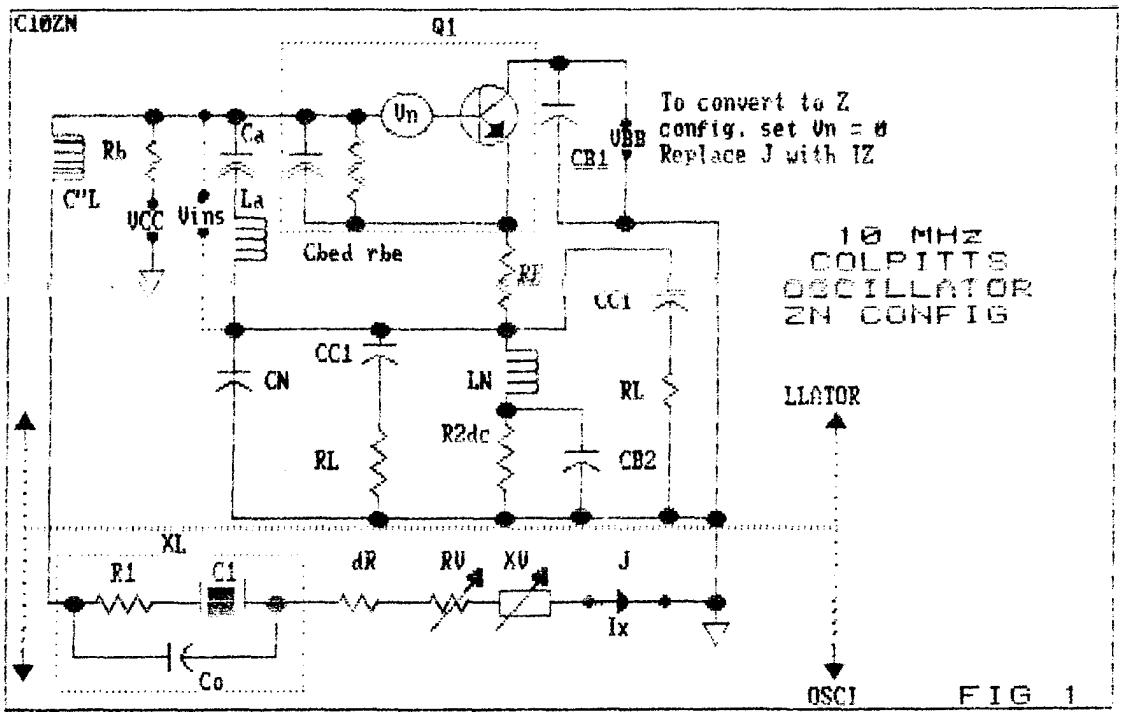
Fig 1 shows both Vin and Vs (see Sect 3.12!) defined as if RE were an integral part of the transistor Q1. This is done to ensure that the measurement of the residual noise, $\mathcal{L}_R$ (f), in Q1 includes the, well known, marked reduction in transistor flicker noise due to RE.

Fig 2 shows CTRm and Zin plotted versus f.

CTR data is presented for 2 locations, RL and Ix. RL is the normal output location. However, the curves indicate that the Ix noise performance is superior past f = 10 Hz and much superior at high values of f. Therefore consideration should be given to extracting the output from Ix. One method of doing this without significant deteriorating Qop and the low frequency noise performance is described in Ref 6. Ref 6 should also be studied as an example of good oscillator design. Worth particularly noting, is the low transistor input impedance and voltage which enables quasilinear transistor operation to minimize transistor multiplicative noise. Its only feature that can stand improvement, is the ALC system, which, because it is based upon the transistor gm variation with DC bias voltage, permits only a "small signal" transistor input level to maintain the quasilinear operation.

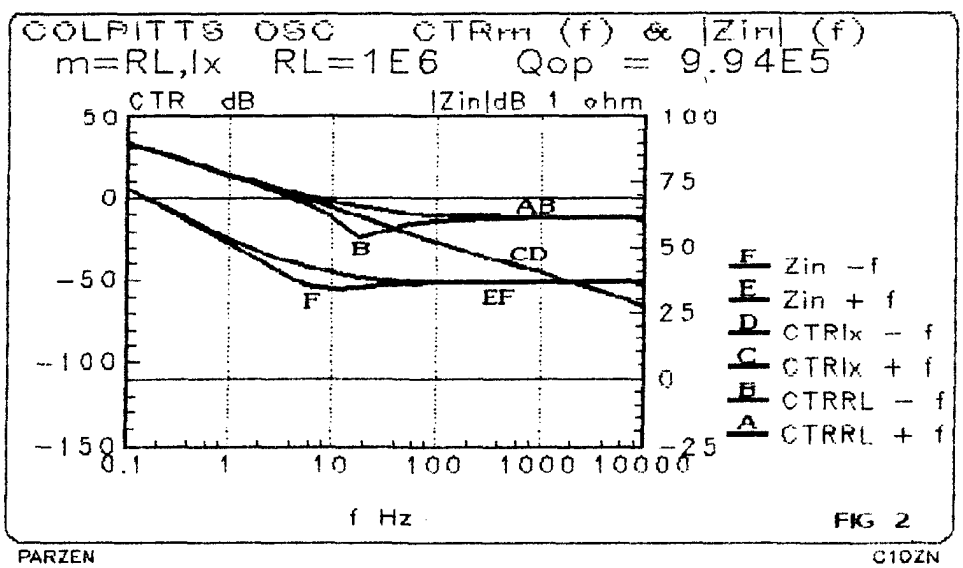
The curves include data for both the upper and lower sidebands, +f and -f, of the spectrum since they may not be symmetrical. The asymmetry is caused by the fact that the signal, at the location being observed, is the sum of at least 2 signals arriving via different paths. If there is only 1 major llator noise source then the signals are correlated and must be combined as phasors.

The relative phase varies with the frequency f, and at f = fa the signals will be in phase in one sideband and out of phase in the other sideband. The out of phase signals causes dips in the CTR function in the region of



RL	dR	V_{RL}	Vins	RT	Qop
1E6	3.49 E-6	0.298	.065	76.2	9.94 E5
1E4	3.88 E-6	0.300	.065	84.6	8.96 E5
1E3	7.44 E-6	0.319	.065	160.9	4.71 E5

Table 1 Operating Data for COLPITTS OSC1
Ix = 1 mA



fa. The value of fa has a strong dependence upon Qop, being closer to fo the greater the Qop, because the phase shifts more rapidly.

The magnitude of the dip is a function of the equality of the magnitudes of the 2 signals. Curve B of the plot shows a dip of about 20 dB at about 20 Hz below the carrier. There is no dip in the resonator current, Ix, noise because there is only 1 path.
This effect may be of great importance in systems which require an unusually low noise signal in a relatively narrow f region close to the carrier and unobtainable by any other means.

A strong dip also exists in curve G, the curve for Zin -f, at a somewhat higher fa, which is indirectly partly responsible for the dip in CTR.

The Zin plots show an increase of about X 100 in |Zin| as f varies from 100 to .1 Hz. The 100 decreases as the oscillator Qop increases.

3. 10 MHz 1 RESONATOR PIERCE OSCILLATOR

Fig 3 is the schematic diagram of this oscillator, called OSC2. The components are identical to those in OSC1 except for the values of the Xt trimming components, C'L (coarse) and XV(fine), and the Rt trimming components, gm(coarse) and RV(fine), which obviously must differ slightly because of the OSC1 larger losses.

The carrier performance data are given in Table 2.

Comparison of Table 2 and Table 1 discloses that the performances are very similar. The slightly superior Qop of OSC2 is due to the fact that the bias resistor R1 shunts only X1 in OSC2 and shunts X1+X2 in OSC1.

Since X2 is much larger than X1 the contribution of R1 to RT is much greater in OSC1 which causes a proportionate reduction in operating Q.
(See Chapters 5 and 7 of Ref 3.)

The X1, only, shunting action also produces greater isolation of the input and output networks which leads to the slightly better high frequency performance of OSC2.

Replacing R1 by an inductor will make the calculated performances identical but it is not recommended because an inductor will contribute more to the frequency instability and because the inductor provides a potential added resonance which would increase the possibility of spurious oscillations.

Fig 4 shows CTRm and Zin plotted versus f.
Comparison of Figs 2 and 4 discloses almost equal noise and Zin performance of OSC1 and OSC2. The only difference is the slightly better noise performance of OSC2, at f below 10 Hz, due to its better Qop.

An additional important characteristic of an oscillator circuit is the effect of stray elements on its frequency stability. Table 3 shows the frequency shift, due to adding 1 pF between ground and the base, b, emitter, e, and collector, c., for both OSC1 and OSC2.
Again, Table 3 shows that the performance of OSC2 is slightly better.

4. 10 MHz 1 RESONATOR DXO OSCILLATOR

Fig 5 is the schematic diagram of the oscillator, called OSC3.
The reader is cautioned that the circuit has been prepared to demonstrate only the unique features of the DXO type oscillator. The mode and harmonic selector networks are absent for simplicity.
The active device is an operational amplifier having an output impedance of 1E-6 ohms resistive and very high input impedances. Obviously A = 100 is too high to be realized in practice at most RF frequencies.

Figs 5 and 6 are the oscillator netlists configured for amplifier voltage gains, A, of 10 and 100 respectively. It will be noted that the NETLISTS (= the CIRCUIT FILES) are identical except for the magnitudes of A, dR, Rn, R1, R2, RV and XV. Note that, for the resistive networks shown, A = (Rn + Re)/ Re and Xe = 0 where Re and Xe are the effective values of the resonator network R and X.

The user sets the values of A, R1, and fo. The computer calculates the other values.
The load resistance, RL, has been set at 1E6 but, because of the very low amplifier output impedance, it can be any impedance, ZL, greater than 50 ohms without any performance change except for the power in ZL.

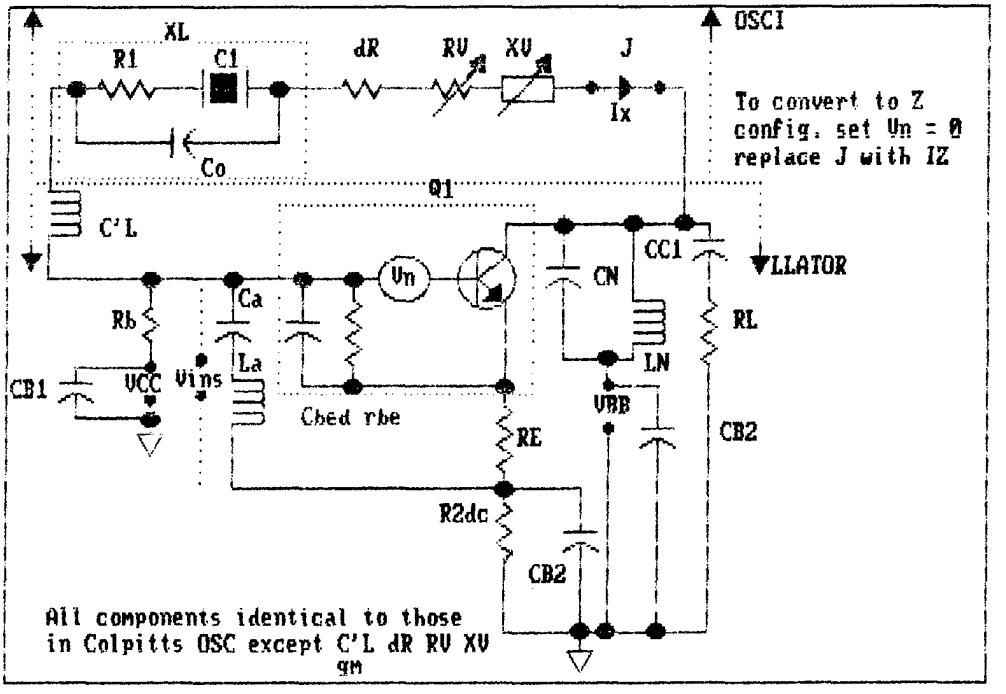
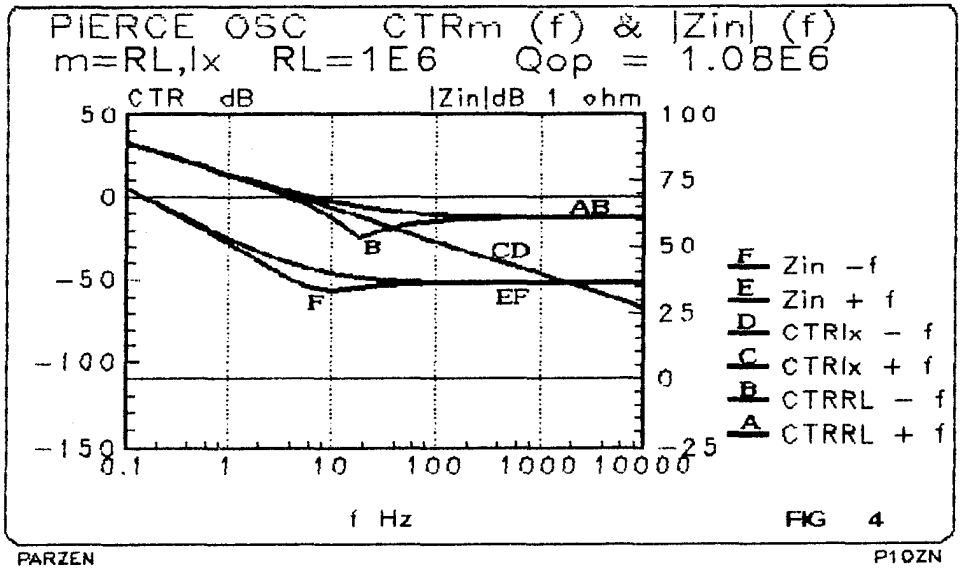


FIG 3 10 MHz PIERCE OSCILLATOR
ZN CONFIG

RL	dR	V_{RL}	Vins	RT	Qop
1E6	3.23 E-6	0.295	.065	70.3	1.08 E6
1E4	3.62 E-6	0.295	.065	78.7	9.62 E5
1E3	7.04 E-6	0.312	.065	152	4.99 E5

Table 2 Operating Data for PIERCE OSC2
Ix = 1 mA



OSC TYPE	b	e	c
Colpitts	-4.9 E-8	-3.1 E-8	0
Pierce	-1.8 E-9	-9.3 E-11	-3.3 E-8

Table 3 Fractional Frequency Shift Caused by Adding 1 pF to GRD at Locations in Heading

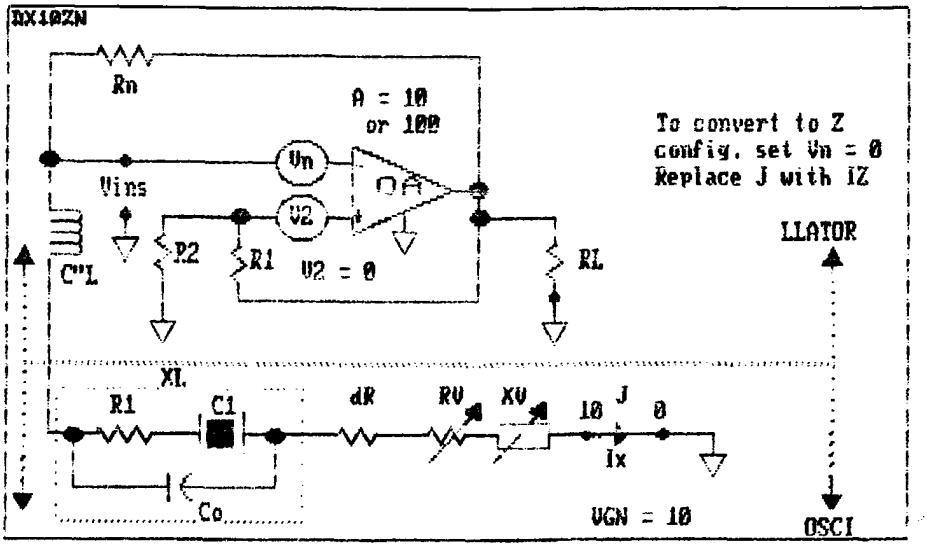


FIG 5 10 MHz 1 XTAL DXO OSCILLATOR ZN CONFIG

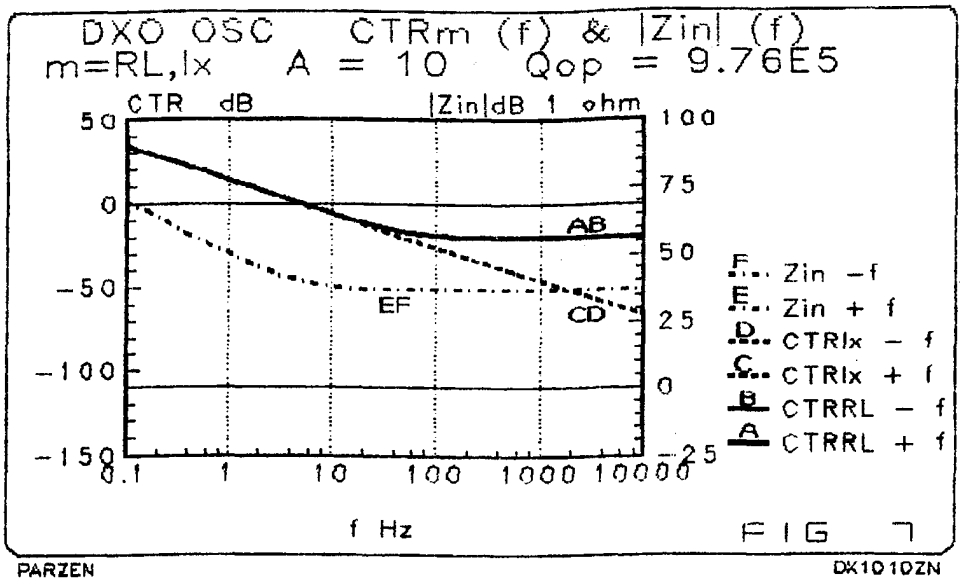
DXO 10 MHz OSC ZN CONFIG A = 10, 0
(Vn),VW, 5 , 1 , 0 , 0 , 0 , 0
v2,VW, 2 , 4 , 0 , 0 , 0 , 0
R1,R, 3 , 4 , 1000 , 0 , 0 , 0
R2,R, 4 , 0 , 249.9999968420093 , 0 , 0 , 0
RN,R, 3 , 5 , 630 , 0 , 0 , 0
(R1)XL,R, 5 , 6 , 70 , 0 , 9 , 5000000
(C1),C, 6 , 7 , 2.1E-016 , 0 , 9 , 10000000
(Co),C, 5 , 7 , 1E-050 , 0 , 9 , 0
{RV},R, 7 , 8 ,-1.745388608680036E-006 , 1 , 0 , 0
{XV},X, 8 , 9 , 7.578806813899777E-009 , 0 , 0 , 0
{dR},R, 9 , 10 , 3.22E-006 , 0 , 0 , 0
A1,A, 1 , 2 , 3 , 10 , 0 , 0
RL,R, 0 , 3 , 1000000 , 0 , 0 , 0
{IZ},I, 0 , 10 , 1 , 0 , 2 , 0
Vout,TP, 0 , 3 , 1E+020 , 0 , 0 , 0
Vins,TP, 0 , 5 , 1E+020 , 0 , 0 , 0

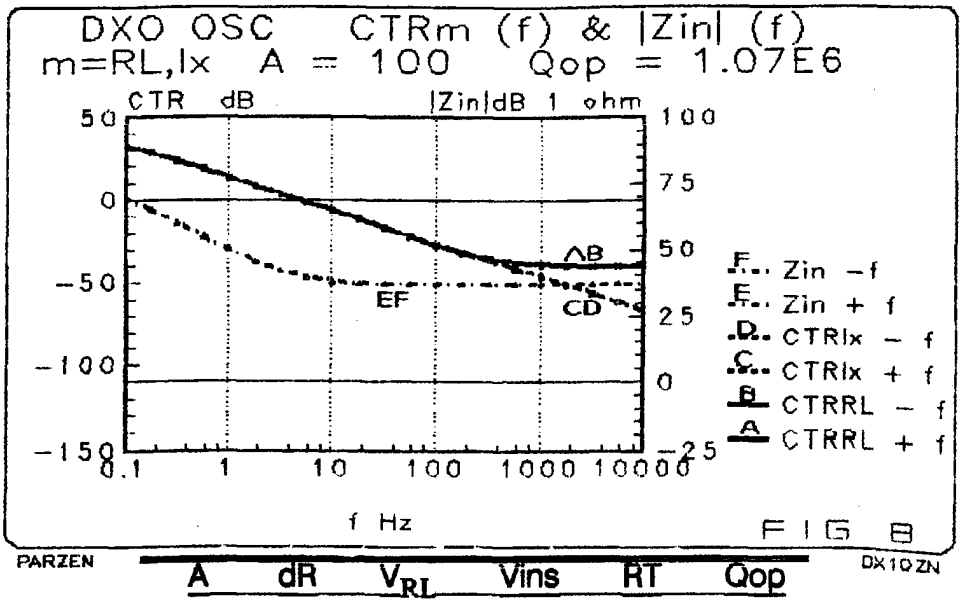
a. A = 10

DXO 10 MHZ ZN CONFIG A = 100, 0
(Vn),VW, 5 , 1 , 3E-009 , 0 , 0 , 0
v2,VW, 2 , 4 , 3E-009 , 0 , 0 , 0
R1,R, 3 , 4 , 490 , 0 , 0 , 0
R2,R, 4 , 0 , 10 , 0 , 0 , 0
RN,R, 3 , 5 , 6930 , 0 , 0 , 0
(R1)XL,R, 5 , 6 , 70 , 0 , 9 , 5000000
(C1),C, 6 , 7 , 2.1E-016 , 0 , 9 , 10000000
(Co),C, 5 , 7 , 1E-050 , 0 , 9 , 0
{RV},R, 7 , 8 ,-1.515858628263932E-007 , 1 , 0 , 0
{XV},X, 8 , 9 , 7.578806813898495E-009 , 0 , 0 , 0
{dR},R, 9 , 10 , 3.03E-006 , 0 , 0 , 0
A1,A, 1 , 2 , 3 , 100 , 0 , 0
RL,R, 0 , 3 , 1000000 , 0 , 0 , 0
{J},J, 0 , 10 , 9.999999999999999E-021 , 0 , 2 , 0
Vout,TP, 0 , 3 , 1E+020 , 0 , 0 , 0
Vins,TP, 0 , 5 , 1E+020 , 0 , 0 , 0

b. A = 100

Fig 6 Netlists for DXO Oscillators





A	dR	V_{RL}	Vins	RT	Qop
100	3.03 E-6	7.000	.070	70.7	1.07 E6
10	3.22 E-6	0.700	.070	75.2	9.76 E5

Table 4 Operating Data for DXO OSC3
Ix = 1 mA

GAIN	-	+	out
10	2 E-9	-1.3 E-8	0
100	2 E-9	2.0 E-9	0

Table 5 Fractional Frequency Shift Caused by Adding 1 pF to GRD at Locations in Heading for DXO OSC3

The carrier performance data are given in Table 4.

Note that the Qop is almost identical to that of the Pierce oscillator. The very high theoretical V_{RL}, 7 V, cannot be attained in practice because as pointed out above A = 100 cannot be attained in practice.

Table 5 displays the frequency shift, due to adding 1 pF between ground and the op amp terminals -, +, and out, for gains A = 10 and 100.
The performance for A = 100 is understandably much better.

Figs 8 and 9 show CTRm and Zin plotted versus f.
The data for Fig 8 is almost identical to that in Fig 4 of the Pierce oscillator. The important difference is that there are only single paths for all signals. Therefore the plots are symmetrical with respect to f and the dips of Fig 4 are absent in Figs 8 and 9.

Fig 9 shows that at high Fourier frequencies the RL noise is 20 dB less than that of Fig 8 corresponding to the 20 dB difference in the amplifier gain, A. This is due to a noise cancellation effect in the differential active device. Of course, this effect is limited by the gain that can be practically achieved.

5. 10 MHz 2 RESONATOR OSCILLATOR

The following reports on the result of a preliminary computer study to determine whether this oscillator merits additional computer and experimental studies.

Fig 10 is the working but unoptimized schematic diagram of the ac circuits of this oscillator, called OSC4.

This oscillator is the first result of an attempt to create a class of oscillators having noise properties superior to those of the best available oscillators.

A study discloses (see Sect 6 of this paper) that, for well designed oscillators, the low frequency noise is determined by the resonator frequency noise and the high frequency noise is the thermal noise of the output devices. This leaves only the intermediate frequency noise for possible improvement.

OSC4 is a modified Pierce type with 2 resonators, XL1 and XL2, identical to XL1 of Fig 3, capacitively coupled by Cc.
The oscillator PARTS LIST is shown in Fig 10.

For simplicity, the mode selector and overtone selector networks are not included but they can be similar to those of OSC2. It is interesting to observe that their omission is often tolerable in the computer oscillator but may be disastrous in the real oscillator.
C'L1 and C'L2 are the tuning adjustments for their respective resonators.
These adjustments also serve to set the oscillator frequency, fo, and to shape the oscillator phase noise curve at low Fourier frequencies.

A +1 Hz shift in the effective fs of XL1 corresponds to +.43 Hz shift in the oscillator fo.
A +1 Hz shift in the effective fs of XL2 corresponds to +.55 Hz shift in the oscillator fo.
This data shows that the resonators are almost equally important in determining the oscillator long term frequency stability.

The carrier performance data are given in Table 6.

Qop = 1.47E6 which is about 40 % higher than that of OSC2 because of the 2 resonators. Note that the usual Qop concept has little meaning in OSC4. A better concept would be Qopm (f).

The signal levels are rather low and should be increased.

Table 7 displays the frequency shift, due to adding 1 pF between ground and the transistor terminals. Comparison of this data with that of the Pierce oscillator (Table 3) shows that the 2 crystal oscillator performance is considerably better.

Fig 11 shows CTRm and Zin plotted versus f.

At f < 8 Hz the noise is identical at all locations and similar to those of OSC2 except for the 3 dB improvement due to the higher Qop.
At f > 20 Hz the noise performance may be much superior to that of OSC2.
At 10 KHZ, the best noise performance of Fig 11, that at location RL, is potentially better by about 60 dB than the best noise performance of Fig 4 at location Ix.

The Zin plots are appreciably better than those of OSC2.

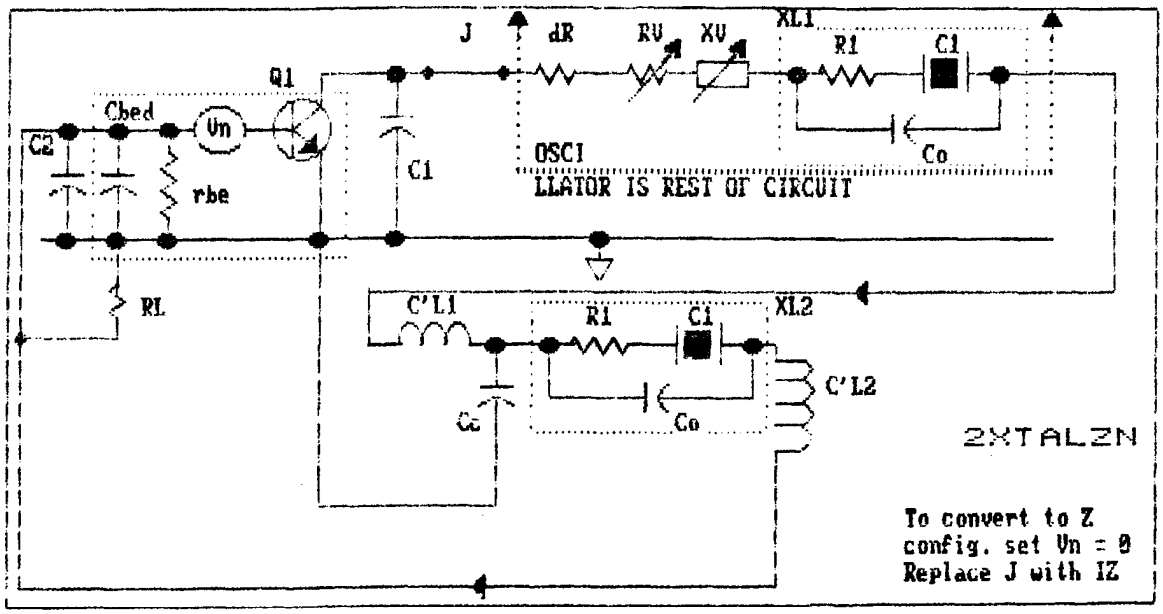


FIG 9 10 MHz 2 XTAL OSCILLATOR ZN CONFIG

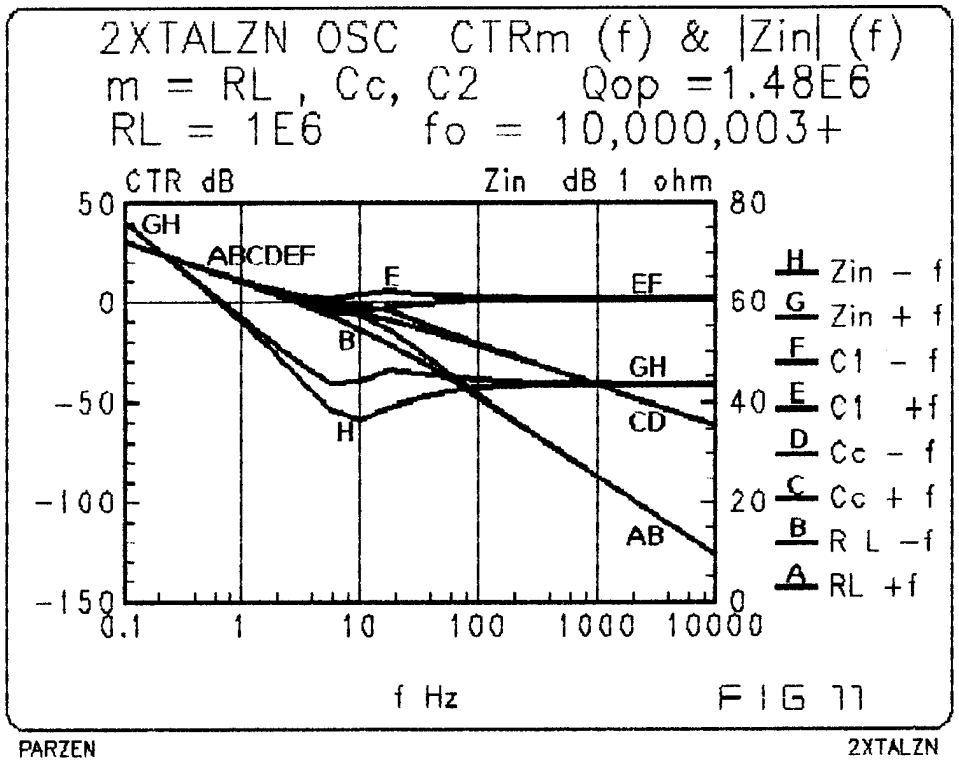
05-10-1992 18:49:57
OF COMPONENTS = 21 HIGHEST NODE # = 13 # OF VOLTAGE SOURCES = 1
2XTALZ FREQUENCY = 10000003.14691093
CIRCUIT NOTE : 10 MHz 2 XTAL OSC ZN CONFIG
COMPNT. CONNECTED TO NODE

#	SYB	N1-	+N2	TYPE	VALUE	PHASE ANGLE
1	c1	1	0	CAPACITOR	1E-010	
2	{J}	1	10	JUMPER, R =	9.999999999999999E-021	
3	{dR}	10	11	RESISTOR	1.366E-005	
4	{RV}	11	12	RESISTOR	-6.454289498532971E-005	
5	{XV}	12	13	X, REACTANCE	-2.3020853335073 91E-007	
6	(R1)XL1	13	2	RESISTOR	70	PO XL
7	(C1)	2	3	XTAL RSN'TOR	C1 = 2.1E-016 fs = 10000000	
8	(Co)	1	3	CAPACITOR	1E-050	PO XL
9	C'L1	3	4	INDUCTOR	5.388510659948518E-007	
10	Cc	4	0	CAPACITOR	1E-009	
11	(R1)XL2	4	5	RESISTOR	70	PO XL
12	(C1)	5	6	XTAL RSN'TOR	C1 = 2E-016 fs = 10000000	
13	(Co)	4	6	CAPACITOR	1E-050	PO XL
14	C'L2	6	7	INDUCTOR	1E-006	
15	C2	7	0	CAPACITOR	1E-010	
16	RL	0	7	RESISTOR	1000000	
17	mm	B = 9 E = 0 C = 1		N, NPN BIP TRANSISTOR		
		gm(o) = 2.861425368072061E-002 BETA = 100 FT (MHz) = 1000				
18	(rbe)	0	7	RESISTOR	3494.761775575397	PO BIP
19	(Cbed)	0	7	CAPACITOR	4.554099916172146E-012	PO BIP
20	Vn	7	9	VW, WH NS V	3E-009	PO BIP
21	Vins	0	7	TESTPOINT SET, R =	1E+020	

Fig 10
2 XTAL OSCILLATOR
PARTS LIST

RL	Ix2	dR	V_{RL}	V_{Cc}	V_{C1}	Vins	RT	Qop
1E6	1.69 E-4	1.37 E-5	.026	.015	.099	.026	117	1.48 E6
1E3	1.44 E-4	1.68 E-5	.021	.015	.096	.021	120	1.31 E6

Table 6 Operating Data for 2 CRYSTAL OSC4
Ix1 = 1 mA



b	e	c
-2.6 E-9	0	- 1.9 E-9

Table 7 Fractional Frequency Shift Caused by Adding 1 pF
to GRD at Locations in Heading for 2CRYSTAL OSC4

6. ADDITIONAL NOISE SOURCES

A large part of the just reported very good noise performance of OSC1 to OSC3 and the even better performance of OSC4 may be nullified by the following important additional noise sources:

Resonator noise (See Section 3.3!)- Resonator noise, which is mainly flicker frequency noise, produces f^{-3} phase noise which, in good circuit designs, swamps the circuit flicker noise and thus effectively determines the total oscillator phase noise, at low f.

Additive noise- Noises, produced by passive component thermal and other noise sources and noises generated in active devices such as buffer and output amplifiers, set effective limits to the total oscillator noise floors.

Those readers, not used to the CTRm and residual noise concepts but are familiar with the customary $\mathcal{L}$ (f) noise data, are reminded that, since $\mathcal{L}_R$ (1E4) of a good active device is better than -140 dBc, the 10 KHz point on curve AB of Fig 11 corresponds to a highly improbable $\mathcal{L}_{RL}$ (1E4) of (-140 -130) = -270 dBc.

7. CONCLUSIONS

The plots demonstrate that the oscillator noise sidebands may not be symmetrical.

Zin is a sharp function of the Fourier frequency and may be very large at close to carrier frequencies.

Application 1- This Colpitts oscillator has been analyzed and its design justified.

Application 2- The differences in performance of similar Colpitts and Pierce oscillators are minor. The choice of circuit is strongly dependent upon practical considerations, such as crystal switching or frequency muultiplication in the collector circuit, and weakly dependent upon frequency, the Pierce circuit being superior at the higher frequencies because of the then slightly better performance and the greater effect of stray elements in the Colpitts oscillator.

Application 3- The Qop of the Pierce oscillator is 1.08E6 and that of the DXO is 1.07E6. It is seen that they are almost identical. The theoretical advantvanges, discussed in Sect 4, appear to be nullified by practical considerations.

Application 4- The CTR plot for the 2 crystal oscillator shows a f^2 response for f up to 10 Hz. For larger f, the response becomes f^4 which disagrees with Leeson's noise model. In view of this potentially extraordinarily low noise performance above 10 Hz, further computer, and possibly experimental, including the possibility of also using the 2 resonators as part of a vibration noise cancellation system, studies are in order.

Much additional effort is desirable to decrease the effect of the noise sources described in Sect 6.

8. REFERENCES

[1] B. Parzen, "Universal, Computer Facilitated, Steady State Oscillator Analysis Theory and Some UHF and Microwave Applications", Proc. 45th Annual Frequency Control Symposium, pp. 368-383. May 1991.

[2] D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proc. I.E.E.E, vol 54, pp. 329-330, Feb. 1966.

[3] B. Parzen, Design of Crystal and Other Harmonic Oscillators. New York:Wiley, 1983.

[4] W.P. Robins, Phase Noise in Signal Sources. London:Peter Peregrinus Ltd, 1982.

[5] G. Montrose et al., "Residual Noise Measurements of VHF, UHF, and Microwave Components", Proc. 43rd Annual Frequency Control Symposium, pp. 349-359, May 1989.

[6] R. Burgoon and H. l. Wilson, "Design Aspects of an Oscillator Using the SC Cut Crystal", Proc. 23rd Annual Frequency Control Symposium, pp. 411-416, 1979.

[7] A. Benjaminson, "Balanced Feedback Oscillators", Proc. 38th annual Frequency Control Symposium, pp. 327-333, May 1984.

[8] B. Parzen, Oscillator and Stability Analysis in the Immittance Domain. In preparation.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LONG-TERM PERFORMANCE OF PRECISION CRYSTAL OSCILLATORS IN A NEAR-EARTH ORBITAL ENVIRONMENT

Lauren J. Rueger

Bendix Field Engineering Corporation, Columbia, Maryland 21045-1897

Jerry R. Norton

Johns Hopkins University, Applied Physics Laboratory
Laurel, Maryland 20723-6099

Paul T. Lasewicz

Naval Satellite Operations Center, Pt. Mugu, California 93042-5013

ABSTRACT

The Navy Navigation Satellite System (NNSS) uses precision quartz-crystal oscillators to provide time and frequency in the orbiting spacecraft. The frequency changes for multiple oscillators during many years of operational service in the orbital environment are discussed in this paper. The primary frequency changes are believed to be caused by mass transfer to and from the resonator, stress relief in the resonator mounting structure and electrodes, and ionizing radiation of the quartz resonator [1].

SUMMARY

The drift of precision quartz oscillators has been observed for 28 years aboard the spacecraft of the Navy Navigation Satellite System (NNSS). Observations to a resolution of 10^{-13} have been made from 1963 to 1991 on 20 operational satellites in near-Earth orbit. Before 1967, satellite service life was limited to one year or less owing to short-lived spacecraft power system solar cell arrays. Since 1967, satellite service life has primarily been limited by the longevity of spacecraft power system rechargeable batteries. No oscillator failures have occurred during the entire program life of nearly 30 years. One oscillator provided continuous operational service for over 21 years, and several have served more than 15 years. No oscillator changed frequency more than two parts in 10^7 while in operational service. Development of a new oscillator was initiated in the early 1970's. Both the electrical and mechanical design concepts were different from previous APL oscillators. The quartz resonators used in the new oscillator were made from cultured quartz that exhibits better frequency stability in the presence of ionizing radiation. One of the best performing oscillators, which had a predictable drift rate of $-9 \times 10^{-13} \pm 1 \times 10^{-13}$ per day after three years of service, was one of these later designs. The frequency measurements were made relative to Cesium standards located in the ground tracking stations of the NNSS. Long-term drift in these Cesium standards was corrected by periodic comparisons with the United States Naval Observatory master clock.

The oscillators that were observed for long periods of time (Figure 1) had similar aging characteristics. The output frequency for most of the oscillators changed at relatively high rates in the first six months to a year after the spacecraft was launched. A reversal of the aging sign or direction then occurred and was followed by an exponential decay for the next three years because of ionizing radiation. Finally, a long-term linear drift dominated. The time to complete this pattern of frequency change varied widely as illustrated in Figures 2, 3, and 4.

BACKGROUND

The NNSS was conceived in 1958 and became operational in 1964; it is the current Department of Defense system for global navigation and will remain so until the Global Positioning System becomes operational [2]. The NNSS system consists of spacecraft orbiting the Earth at an altitude of 600 nautical miles (nmi) transmitting 150 MHz and 400 MHz RF carriers derived from quartz oscillators. Tracking stations receive the RF signals and precisely measure the carrier Doppler frequency shift as the spacecraft moves past each station; a computing center uses the Doppler shift data to determine the spacecraft orbit ephemeris and to predict

0-7803-0476-4/92 $3.00 © 1992 IEEE

what it will be in the future; the predicted orbit is transmitted from the ground and stored in the spacecraft. The predicted ephemeris data are used to modulate the RF carriers. A ship or submarine near the surface can receive the RF carriers and the modulation. The geometry between the spacecraft and ship is determined from the measured Doppler shift, and the position of the spacecraft in Earth coordinates is derived from the modulation. Combining these two quantities enables the ship's position in Earth coordinates to be ascertained. System accuracy is directly related to the frequency stability of the quartz oscillator in the spacecraft. A 1×10^{-9} error in the ship's measurement of the spacecraft frequency results in a navigation fix uncertainty of about 0.1 nmi. The position error is proportional to the frequency error.

OSCILLATOR DESIGN

The oscillator performance presented in this paper is based on three oscillator designs. Two of the oscillator designs incorporated one kind of quartz resonator, whereas a different resonator was used in the third oscillator type. When the oscillators described in this report were designed (1963 through 1980), the best quartz resonators and technology available at the time of manufacture were incorporated. Each was designed to survive the Scout rocket launch, to have high performance, and to maintain this performance for a minimum of five years in the space environment of a 600 nmi orbit. Other critical design parameters included power consumption, weight, and size. Both the electrical and the mechanical designs for the early oscillators were quite different from the current oscillators. The three design types share one other common trait, since the quartz resonators for all the oscillators were made by Bliley Electric.

The design concept of the first two oscillator types used in the Oscar series of spacecraft was similar. The major difference between the two is the method of precise temperature control for the quartz resonator and oscillator circuits. The first type used an active, thermostatically controlled oven, whereas the second type, starting with the Oscar-12 launch, had an active, proportionally controlled oven. The oven material for both types was Monel, which has a very high thermal mass and acts as a filter to reduce the thermal fluctuations generated in the heater control circuits. The oscillator circuit is a modified Pierce circuit. A Bliley BG61AH-5, 5 MHz, 5th overtone, AT cut resonator was used in both oscillators.

The third oscillator type was designed for the NOVA spacecraft. The oscillator has dual ovens (one inside the other) controlled by a proportional circuit in a different configuration; it also has higher gain. Mechanically, the oven is smaller than previous ovens and is made from copper to reduce thermal gradients and provide radiation shielding for the quartz resonator. The oscillator circuit is a modified Colpitts type with both ac and dc negative feedback to reduce noise and stabilize gain. A Bliley BG61AH-5S, 5 MHz, 5th overtone, AT cut resonator is the frequency control element in this oscillator. The resonators were fabricated from Electronic Grade, swept, cultured quartz to reduce the frequency changes induced by ionizing radiation. A report on the radiation characteristics of quartz resonators used in this and other oscillators may be found in [3]. Design details and test results on this third type of oscillator performance were reported in [4].

PERFORMANCE DATA COLLECTION

The data presented in this paper were obtained by ground stations observing spacecraft passing within radio range. The Doppler shift of the spacecraft carriers, as received on the ground, was measured against the station atomic-controlled frequency and time standard as a time function of frequency. The ground stations collecting the data are within radio range of each spacecraft four to six times each day. Twice each day for Oscars and once each day for NOVAs, the NNSS system computer complex determines an updated orbit for each spacecraft that includes the latest evaluation of the spacecraft oscillator frequency. From late 1963 to late 1976, these frequencies were hand plotted in archival records at JHU/APL. Since late 1979, these frequencies have been a part of archival records of orbit updates for each satellite kept by the Naval Satellite Operations Center (NAVSOC).

PERFORMANCE DATA REDUCTION

For the periods over which data were available, more data exist than can usefully be plotted to describe the drift of a given oscillator. The frequency data in the orbital archives contain about 60 bits of data in an orbital record of more than 50,000 bits. For the Oscar spacecraft, two sets of orbital data are generated each day, and a frequency–time program is sent to each spacecraft twice a day. For the NOVA spacecraft, the frequency measurements are made once each day, but since these spacecraft employ a programmed frequency compensation synthesizer following the output of the quartz oscillator, a frequency correction is only needed about once a month. It is this frequency correction event that provides the data for the NOVA drift rate plots. NOVA 1 experienced a failure in the synthesizer, which required a "work around" to operate the spacecraft time and frequency system. Thus, no NOVA 1 oscillator drift data were available for this paper.

To reduce the data to a manageable size, each year was limited to multiples of 30 days for the Oscar spacecraft and to the monthly measurements of the NOVA spacecraft.

In 1988, several Oscar-type spacecraft were stored in orbit. The oscillator in each spacecraft is on continuously, but the oscillator frequency is only measured at two-year intervals or whenever the spacecraft is transferred to active service. The Scout launch rocket typically launches the

spacecraft slightly off the desired polar orbit. The Oscar spacecraft must operate with this slight orbit error, and, as a result, their orbit precesses slowly in longitude in inertial space. For service lives of a year or so, this orbit precession is inconsequential, but for service lives in excess of five years, the orbit precession creates conflicts with the orbits of other spacecraft in the system. When these conflicts arise, the spacecraft selected for service may change to one stored in orbit. The NOVA spacecraft do not have to accept the Scout rocket launch errors because they have thrust-controllable rockets attached to the spacecraft that can correct the launch error after separation from the last stage of the Scout.

DRIFT DATA

Oscillator drift data are presented in Figures 2 through 4. Table 1 contains data for oscillators whose drift rates have stabilized and are clearly established both in magnitude and direction. The service history of the NNSS spacecraft is presented in Figure 1. Note that none of the spacecraft provided more than a year of service before 1967. The 21-year service life of Oscar 13 is the best achieved in this program to date. Oscillator performance for the NNSS Oscar spacecraft is shown in Figure 2. NOVA 3 has a drift rate of -9×10^{-13}/day—one of the best achieved in this program to date. The NNSS NOVA spacecraft oscillator performance is illustrated in Figure 3. During thermal vacuum evaluation of the NOVA oscillators before launch, NOVA 2 had a drift rate of -4.4×10^{-11}/day, but after stabilization in orbit the drift rate was $+5.0 \times 10^{-12}$/day. The drift rate before launch for the NOVA 3 oscillator was -4.7×10^{-11}/day, and the in-orbit drift rate is -9×10^{-13}/day. The in-orbit drift rates are much better; however, the oscillators were evaluated on the ground for less than three months. The best drift rate for an Oscar oscillator was -8×10^{-13}/day. Figures 2 and 3 show that the direction of oscillator drift is both positive and negative. The spread in the value of oscillator drift rate is large with no clear link between the value of the drift and drift direction. Figure 4 is an isolated plot of Oscar 13 and 20 oscillator performance. These plots clearly show the initial, the exponential, and the linear phases of oscillator drift in orbit.

Two types of quartz resonators were used in the oscillators, a BG61AH-5 and a BG61AH-5S. The basic design of the quartz resonator disk in both types is very similar if not identical. The enclosures for the resonator disk and the way the resonators were processed during manufacture are very different. The drift rate and other operating characteristics of quartz resonators are very closely related to the quality of the vacuum inside the resonator enclosure and the degree of cleanliness maintained for the quartz disk throughout its manufacturing cycle. The BG61AH-5S resonator enclosure is much smaller than the BG61AH-5; therefore, the enclosed volume of the resonator is much smaller. Since there is less volume to evacuate and less surface area to contaminate, the quality of the vacuum is superior for the BG61AH-5S. The manufacturing process for the BG61AH-5S is also cleaner. The net result for the BG61AH-5S is a resonator whose drift becomes stable in less operating time, and the drift rates are lower. Comparison of Figures 2 and 3 illustrates this performance. The superior drift rate for the NOVA oscillators that used the BG61AH-5S resonator implies that a major cause of resonator aging is mass transfer to and from the quartz resonator disk.

Table 1 Oscillator Service Life and Final Drift Rate

Spacecraft I.D.	Service (years)	Final Drift Rate (10^{-12}/Day)
O-11	8.3	−30.4
O-12	11.0	−0.8
O-13	21.4	15.4
O-14	16.5	−10.0
O-18	8.7	−31.8
O-19	5.2	3.3
O-20	17.3	6.0
NOVA 1	10.5	work around
NOVA 2	3.6	5.0
NOVA 3	7.4	−0.9

ACKNOWLEDGMENT

The authors wish to acknowledge Mr. B. J. Condon for his contribution in extracting the frequency performance data from the voluminous NNSS archival files. This work was supported by the U. S. Navy Strategic Systems Programs, for which we are grateful.

REFERENCES

[1] J. R. Vig and T. R. Meeker, "The Aging of Bulk Acoustical Wave Resonators, Filters, and Oscillators," in the *45th Annual Symposium on Frequency Control Proceedings*, 1991, pp. 77–101.

[2] L.L. Pryor, "The Transit System: A Status Report," presented at the AIAA Space Programs and Technologies Conference, Huntsville, AL, March 24–27, 1992.

[3] J. R. Norton, J. M. Cloeren, and J. J. Suter, "Results From Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators," in the *38th Annual Symposium on Frequency Control Proceedings*, 1984, pp. 63–72.

[4] J. R. Norton, "An Ultrastable Low Power 5 MHz Quartz Oscillator Qualified for Space Usage," in *30th Annual Symposium on Frequency Control Proceedings*, 1976, pp. 275–278.

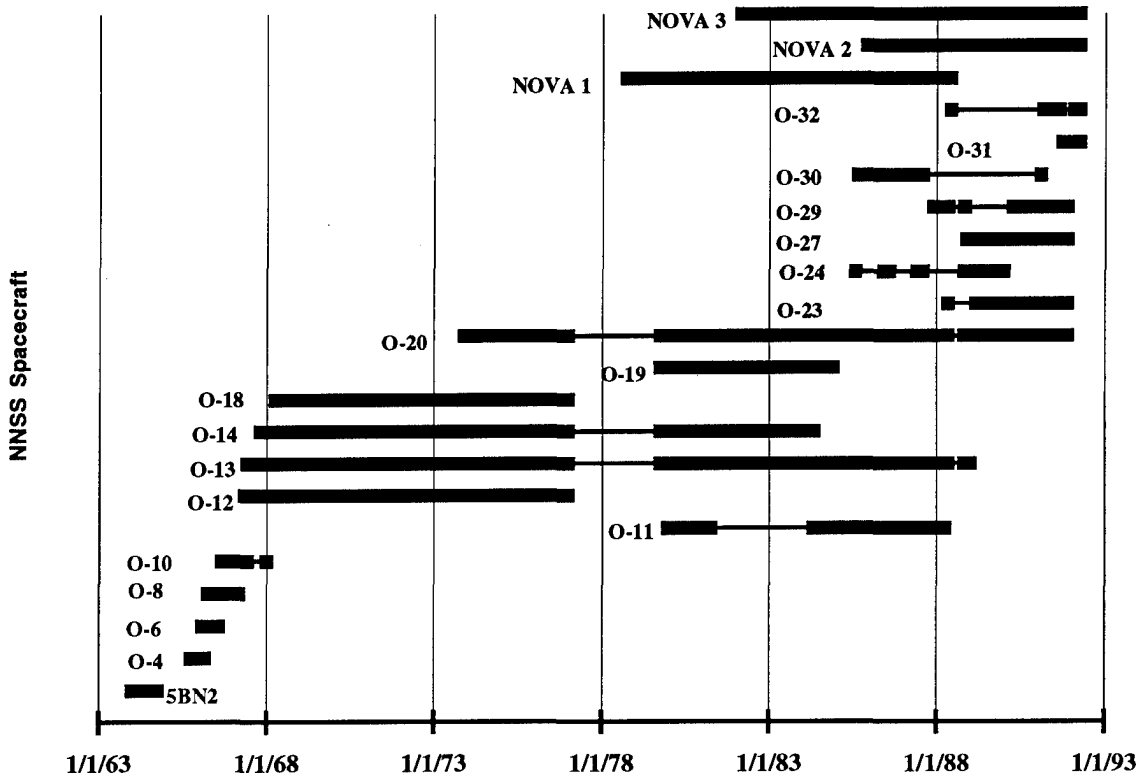


Figure 1. Service history of NNSS spacecraft.

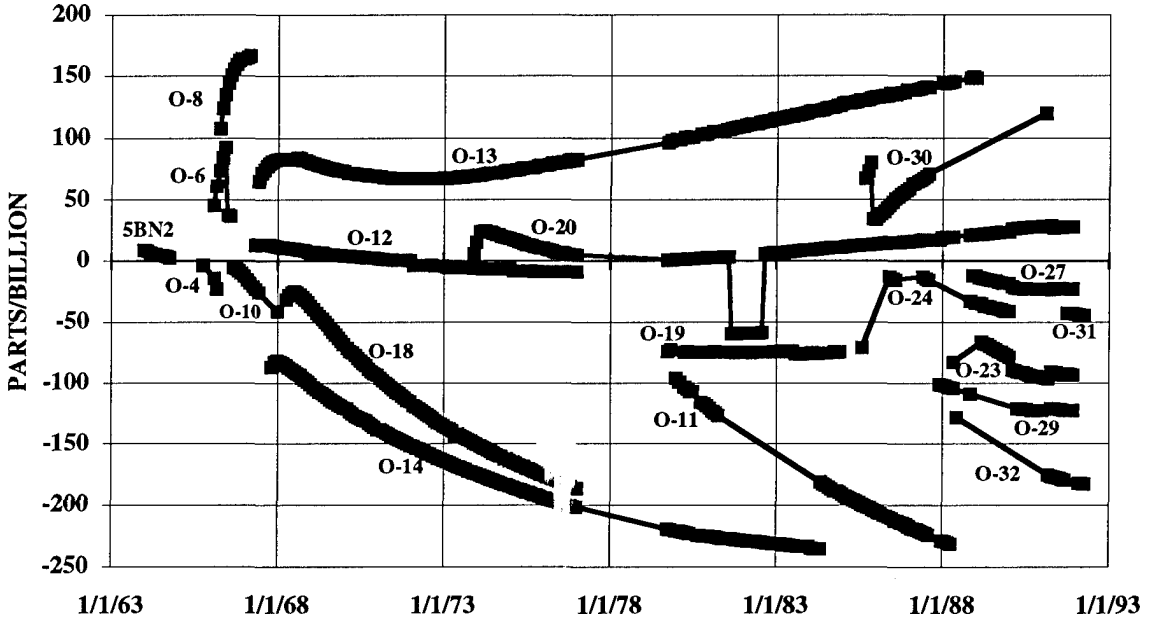


Figure 2. Oscillator performance for NNSS Oscar spacecraft.

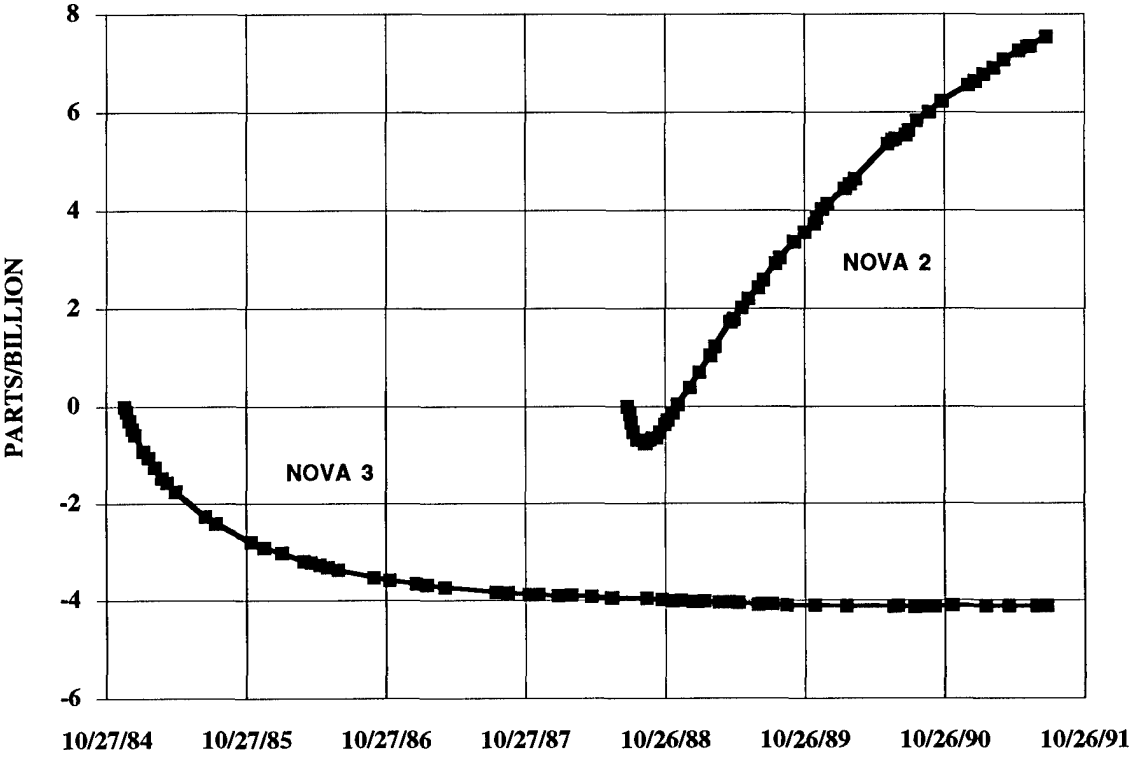


Figure 3. Oscillator performance for NNSS NOVA spacecraft.

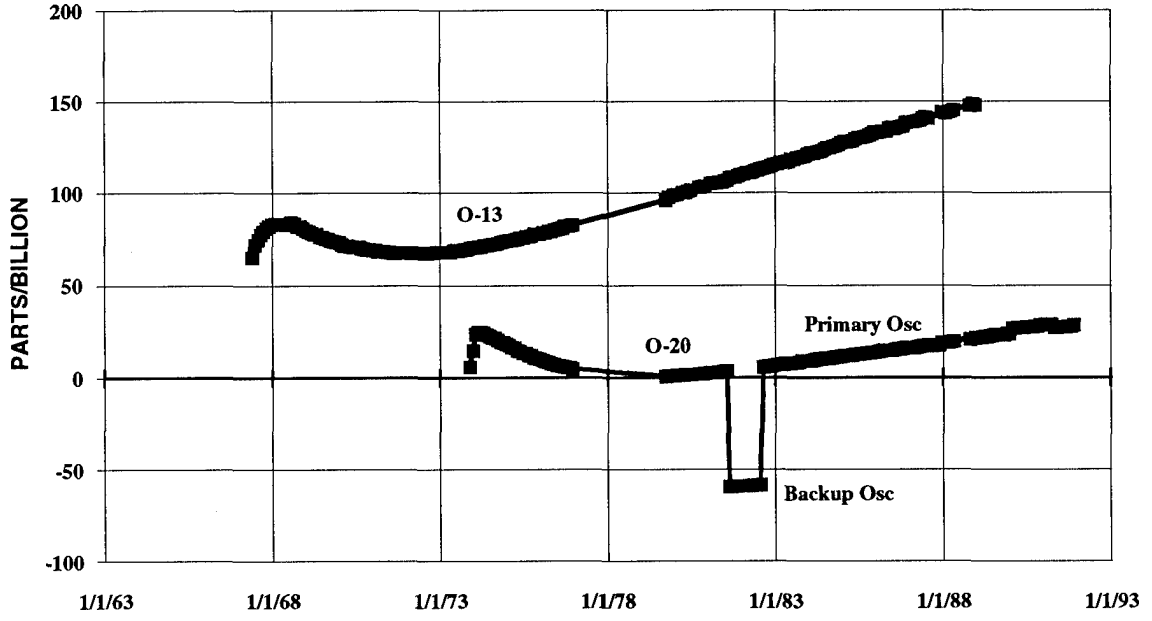


Figure 4. Oscillator performance for Oscar 13 and 20 spacecraft.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LONG TERM AGING of OSCILLATORS

Raymond L. Filler and John R. Vig
US Army Electronics Technology & Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5601

ABSTRACT

A search performed in connection with a recent review of the literature on aging has revealed very few reports on the long term (e.g. for periods greater than 1 year) aging of oscillators. The purpose of this paper is to report aging results for more than 40 oscillators, from a variety of sources, for periods ranging from 1 year to more than 10 years.

The aging data were accumulated with an automated aging facility. The oscillators that have been on test include TCXO's and OCXOs. The TCXO's were maintained in a controlled temperature environment. Several of the TCXO's were built for a gun-launched sensor application and have been shown to be capable of surviving more than 30,000 g shock levels of 12 ms duration. The aging of these ruggedized TCXOs are surprisingly good ($<2 \times 10^{-10}$ per day). The better OCXOs exhibit long term aging of a few parts in 10^{12} per day.

INTRODUCTION

Aging is the systematic variation of frequency with time when all environmental parameters are held constant.[1] A search performed in connection with a recent review of the literature on aging[2] has revealed very few reports on the long term (e.g. for periods greater than 1 year) aging of oscillators. The purpose of this paper is to report representative aging results from tests on more than 40 oscillators, from a variety of sources, for periods ranging from 1 year to more than 10 years. (Since many of the commercial oscillators were purchased 10 or more years ago, the aging results reported below may not have much relevance to current capabilities.) Some of the oscillators were not well behaved, i.e., exhibited short-term instabilities much greater than the aging per day. Oscillators that were not well behaved initially did not improve upon extended aging. Only the aging of well behaved oscillators are shown in Figures 1 - 33.

The oscillators that have been on test include commercially available and prototype temperature compensated crystal oscillators (TCXO), commercially available ovenized crystal oscillators (OCXO) and prototype bulk-wave crystal resonators in specially built ovenized test oscillators.[3]

Data was collected while the TCXOs were maintained in a controlled temperature environment at +60° or at -40°C. The ovenized oscillators were in laboratory ambient where the internal thermal control circuit maintained a nominally constant temperature. In several instances, oscillators were subjected to extreme temperature changes, either a return to room temperature or an exposure to temperatures below -40°C. After the -40°C storage, the OCXOs were returned to room ambient before restarting.

The aging data were accumulated with an automated aging facility established in 1980.[4]

AGING GRAPHS

The aging graphs show aging behaviors and the effects of interruption and temperature change. On all of the aging graphs the ordinate is the reduced frequency, y, in units of parts per billion (ppb), where

$$y_k = \frac{\Delta f_k}{f_o} = \frac{f_k - f_o}{f_o},$$

f_o is the first frequency, measured (at time t_o), and the abscissa is the elapsed time from t_o.

On each graph there is a straight line included for reference. The slope and intercept of this straight line are determined from a least squares fit to all of the data between the days indicated. This straight line is not intended to indicate that the aging rate is constant but is included to facilitate comparison of aging behaviors.

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

LOGARITHMIC FIT

The function $y = A \ln(Bt + 1)$ has been proposed as a candidate for extrapolation of the initial 30 days of aging data to periods in excess of 1 year.[2,5] For many oscillators, extrapolation of the log function obtained form fitting the first 30 days of data is a poor predictor of the subsequent aging. In general, for well behaved oscillators that are kept on continuously, curve fitting the data for a longer period results in a log function that may provide a better approximation to the actual long term aging. For example, as is shown in Figure 34 and in Table I, the log function obtained from the first 30 days of aging data is a poor predictor of the long term aging. The log function obtained from the initial 60 and 300 days provide better and better ability to predict the actual aging. Figure 35 shows an anomaly when the first 60 days is used although the 300 day fit is very good.

A review of the aging graphs will show that "non-logarithmic" behavior seems to be the norm not the exception. If the temperature changes or the power is interrupted, the situation is even worse. The parameters of the logarithmic aging model obtained from the first 30 days of data may be a good indicator of process control but are not very useful as a performance indicator during field use.

CONCLUSIONS

In conclusion, there were no major surprises in the data. The oscillators which started out good stayed good and the poor performers stayed poor. It is important to note that aging direction can reverse in time. This usually occurs early, but in one case it happened after 4.5 years. A logarithmic fit to the first 30 days is a poor indicator of long term performance especially with large environmental disturbances such as temperature changes. It is true, however, that the aging rate generally decreases with time, although for at least one case, it has been increasing for over eight years.

ACKNOWLEDGEMENTS

During the years of this project's existence, many of our colleagues contributed at one time or another. We thank P. Thompson for collecting the data on the high-shock TCXOs and for his work on the aging test oscillators, V. Rosati and J. Messina for TCXO aging measurements, J. Kosinski and R. Lindenmuth for OCXO and resonator aging measurements and software maintenance.

REFERENCES

1. CCIR Recommendation No. 686, "Glossary," CCIR 17th Plenary Assembly, Vol. VII, "Standard Frequency and Time Signals (Study Group 7)," (1990). Consultative Committee on International Radio (CCIR); copies available from: International Telecommunications Union, General Secretariat - Sales Section, Place des Nations, CH-1211 Geneva, Switzerland.

2. J. R. Vig and T. R. Meeker, "The Aging of Bulk Acoustic Wave Resonators, Filters and Oscillators," Proc. 45th Annual Symposium on Frequency Control, pp. 77-101, 1991.

3. H. W. Jackson, "Update on the Tactical Miniature Crystal Oscillator Program," Proc. 36th Annual Symposium on Frequency Control, pp. 492-498, 1982.

4. R. L. Filler, et al., "Aging Studies on Quartz Crystal Resonators and Oscillators," Proc. 38th Annual Symposium on Frequency Control, pp. 225-231, 1984.

5. "MIL-C-49468, Military Specification, Crystal Units, Quartz, General Specification for," the latest revision is available from Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

Table I
Difference between predicted frequency and actual frequency for different curve fitting intervals

	1 - 30 days (ppb)	1 - 60 days (ppb)	1 - 300 days (ppb)
Oscillator A	-29	-22	+3
Oscillator B	-9	+16	-1

Figure 1. Brightline OCXO

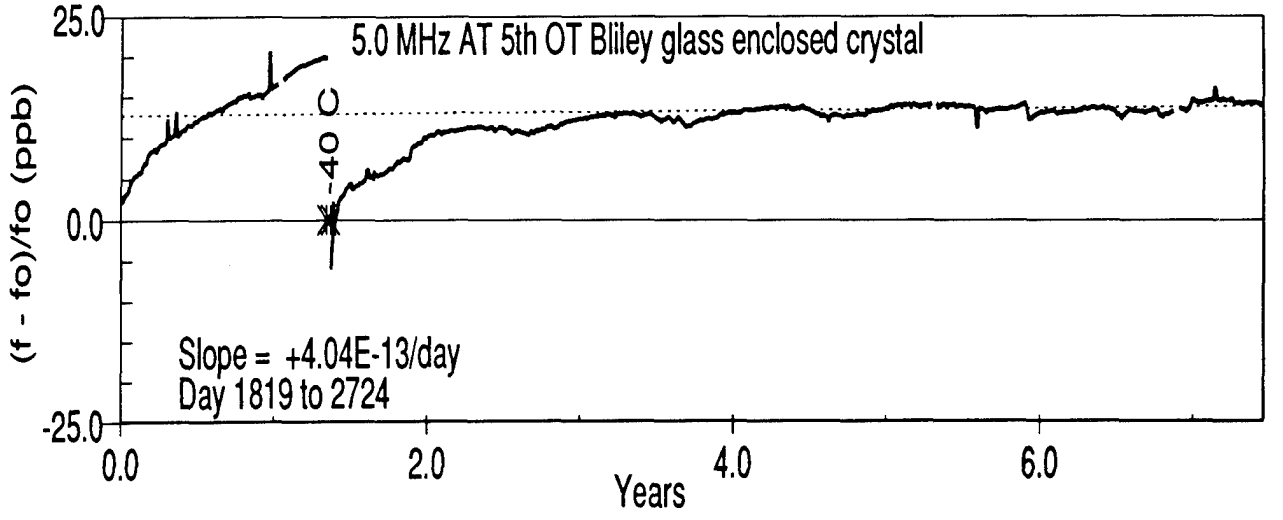


Figure 2. FTS OCXO

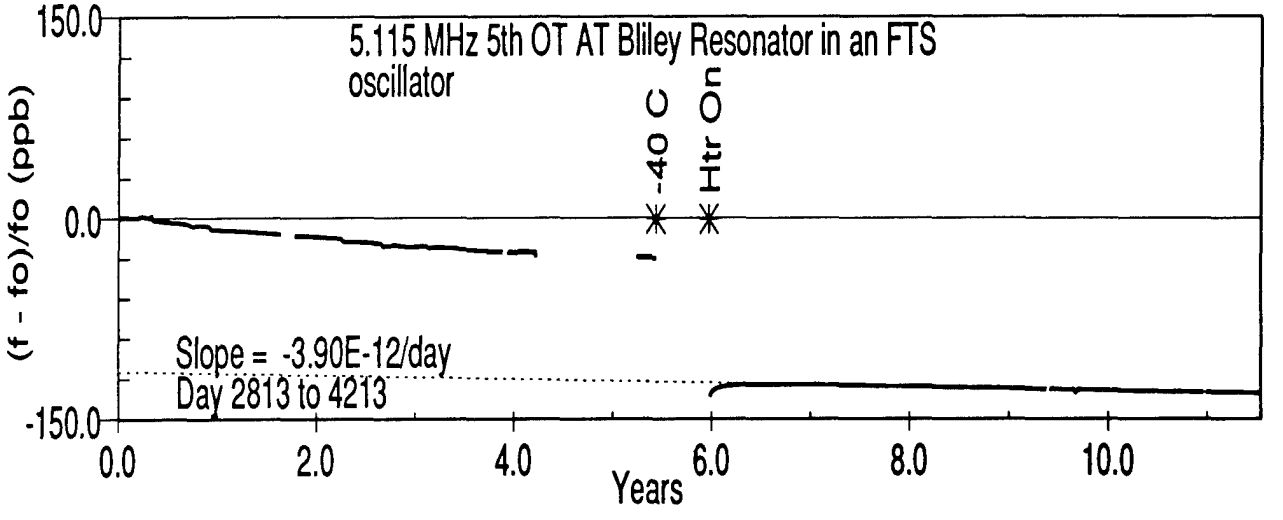


Figure 3. FTS OCXO

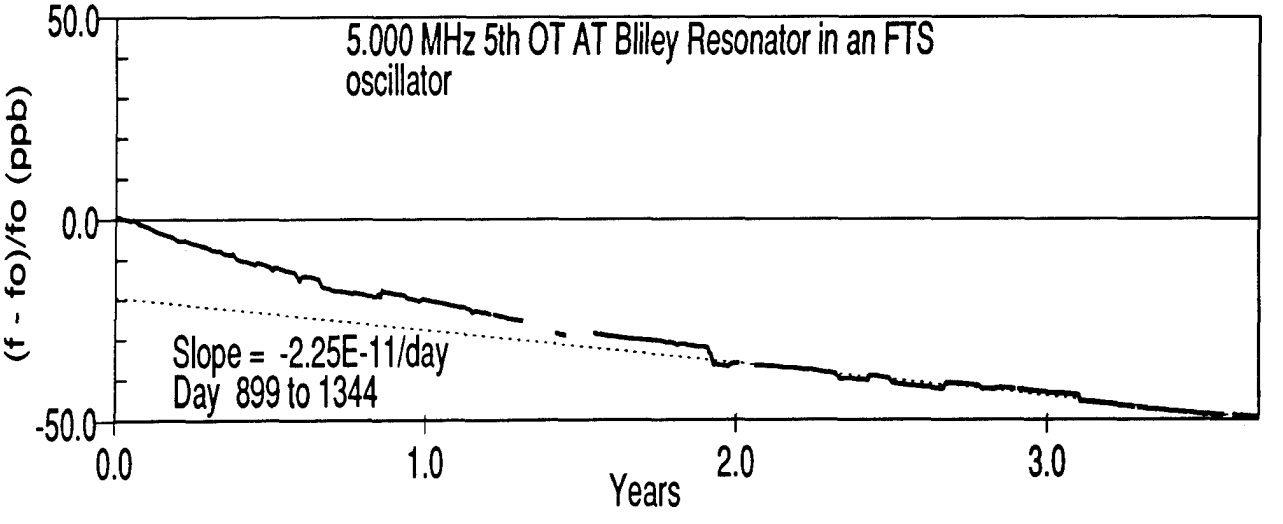


Figure 4. FEI 'MAXI' Resonator in a test OCXO

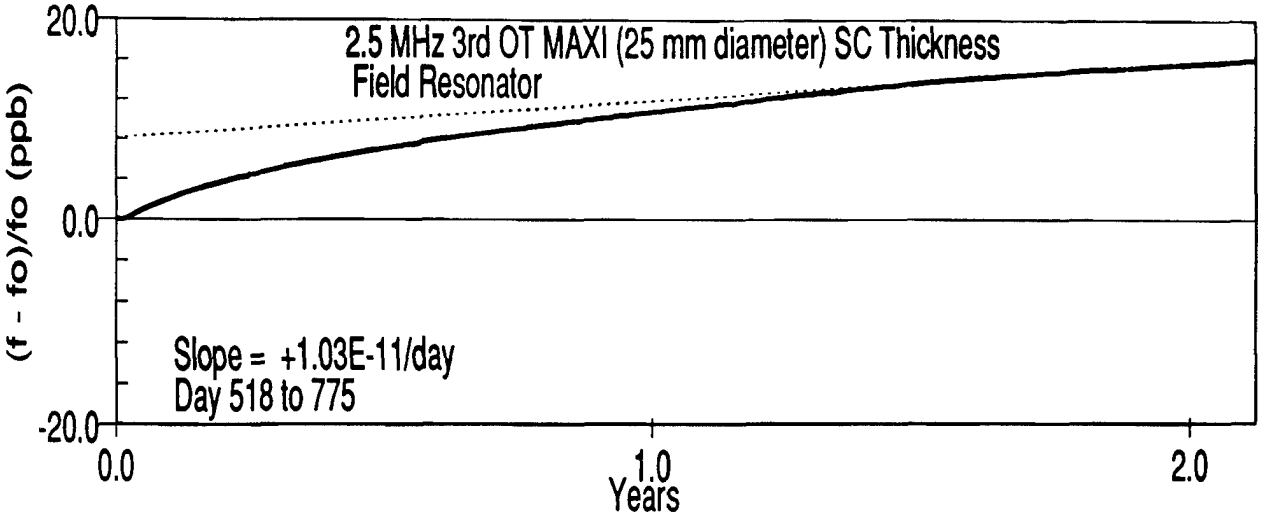


Figure 5. FEI 'MAXI' Resonator in a test OCXO

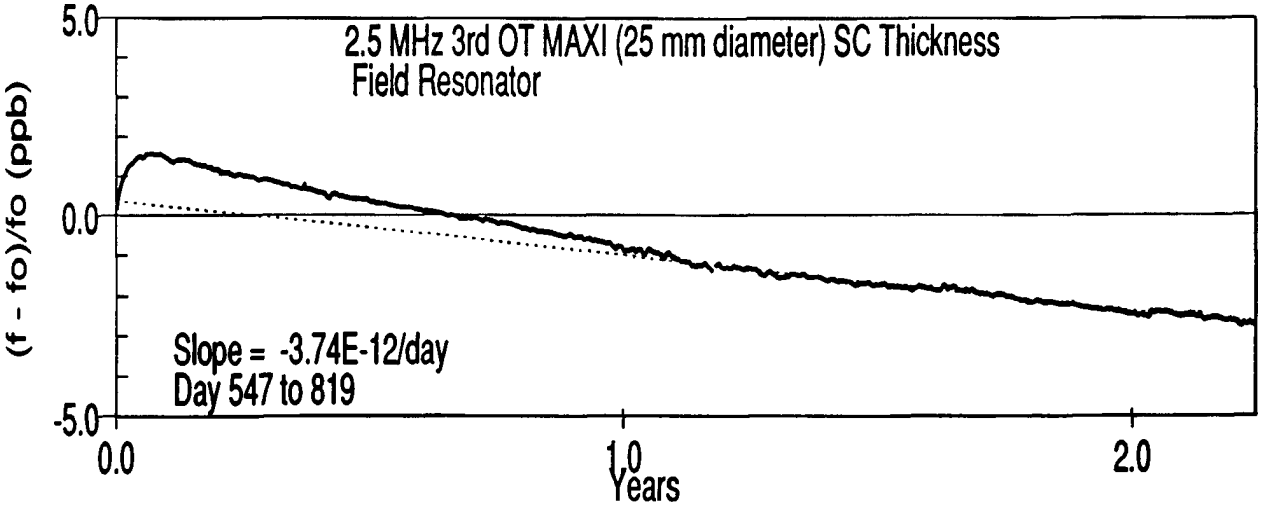


Figure 6. FEI Ceramic Flatpack with sapphire covers

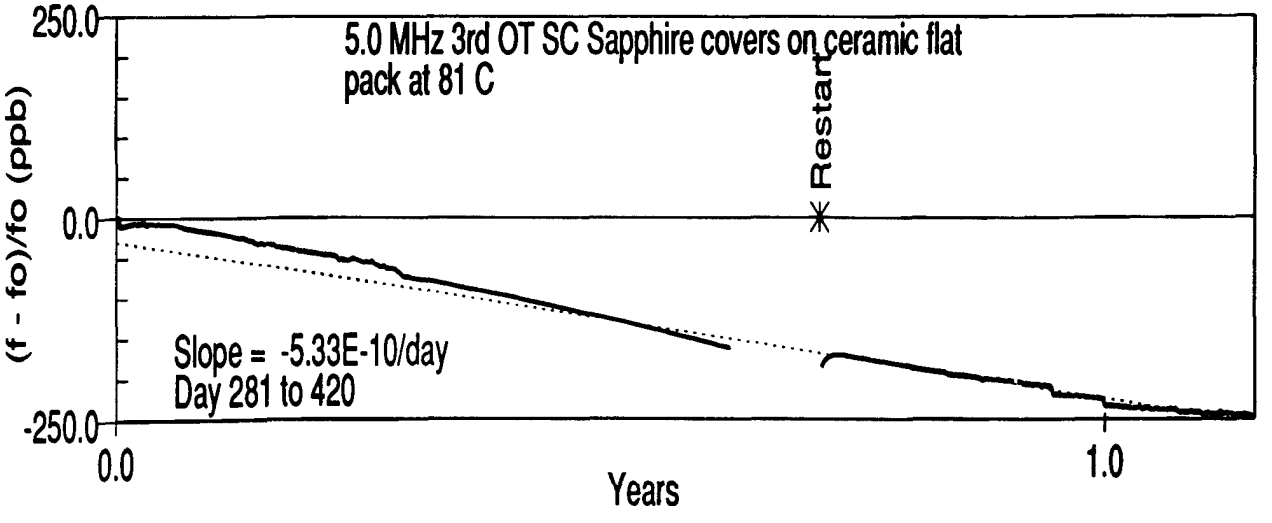


Figure 7. HP10811A OCXO

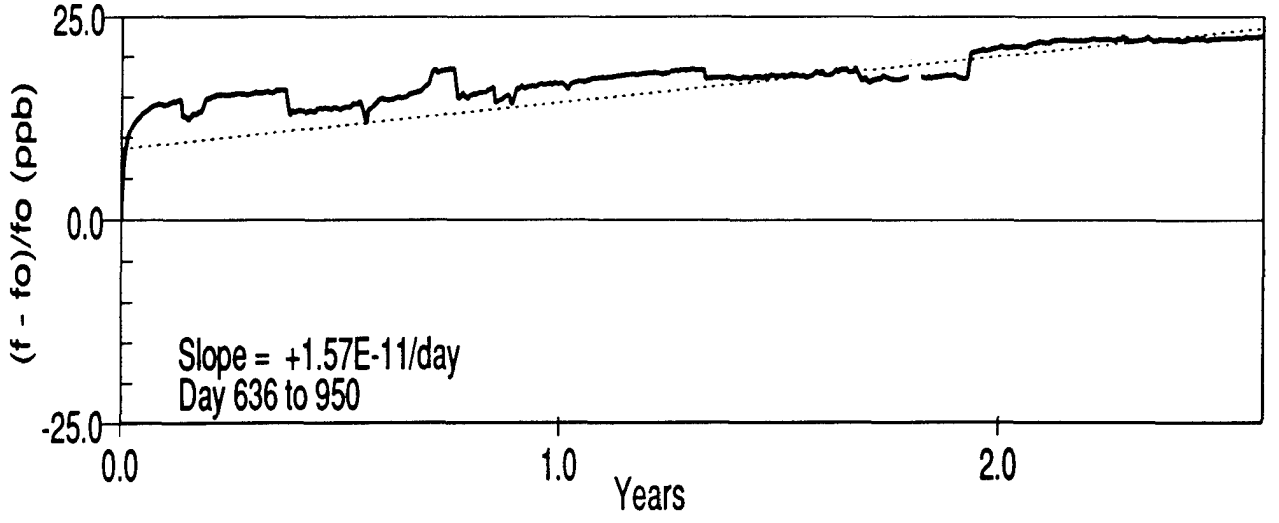


Figure 8. HP10811A OCXO

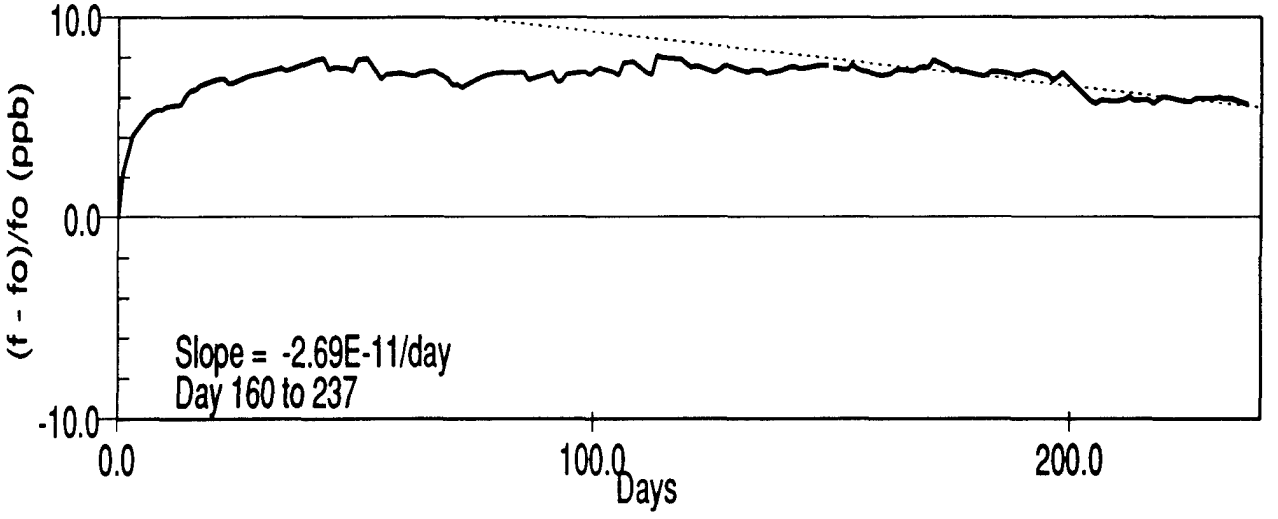


Figure 9. HP10811A OCXO

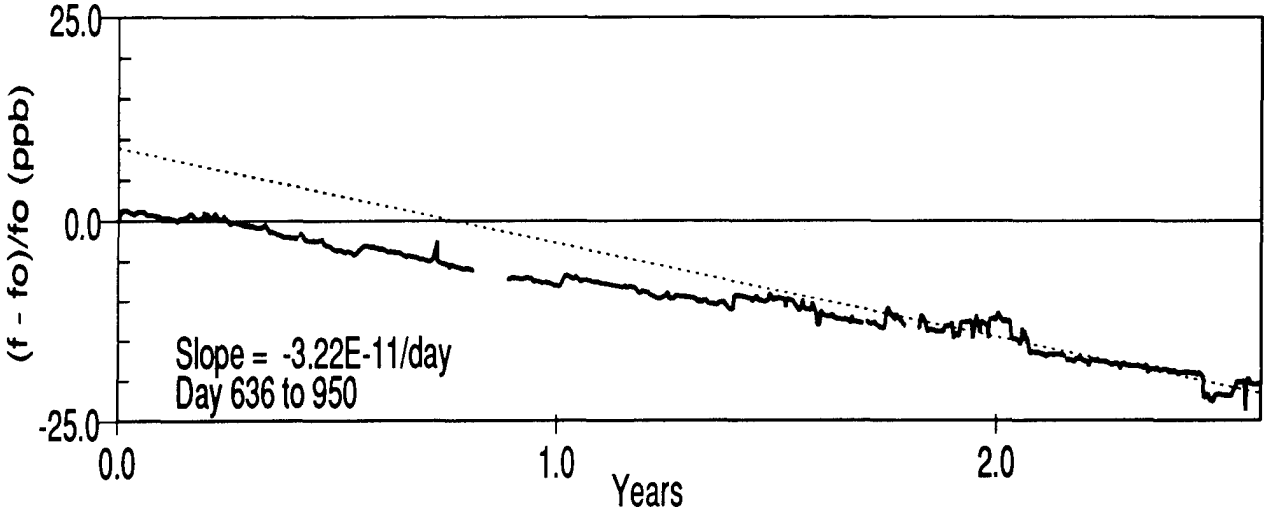


Figure 10. Piezo Crystal OCXO

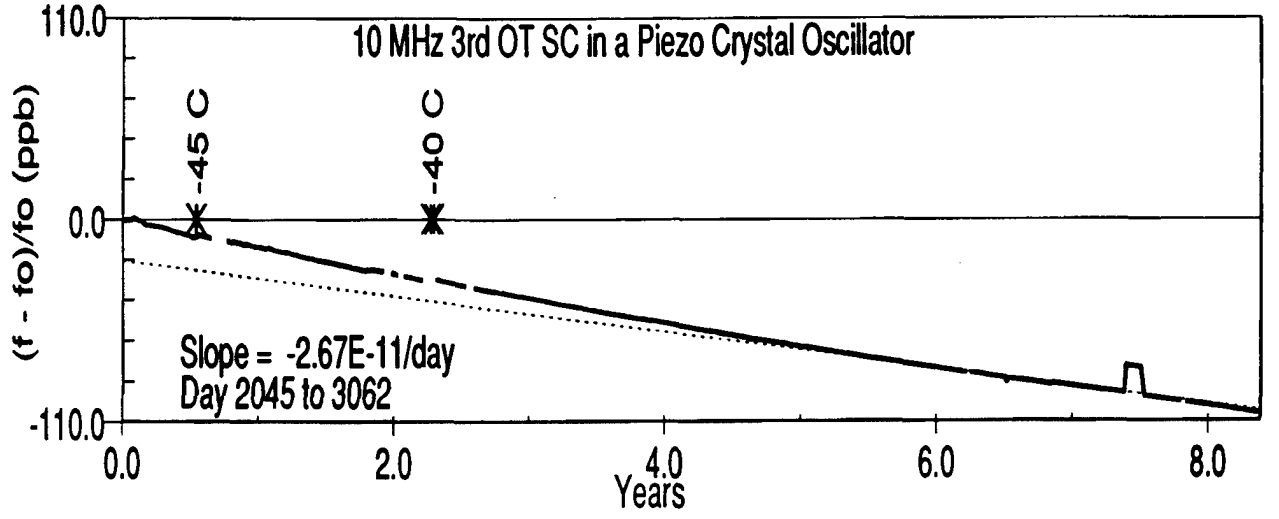


Figure 11. Piezo Crystal OCXO

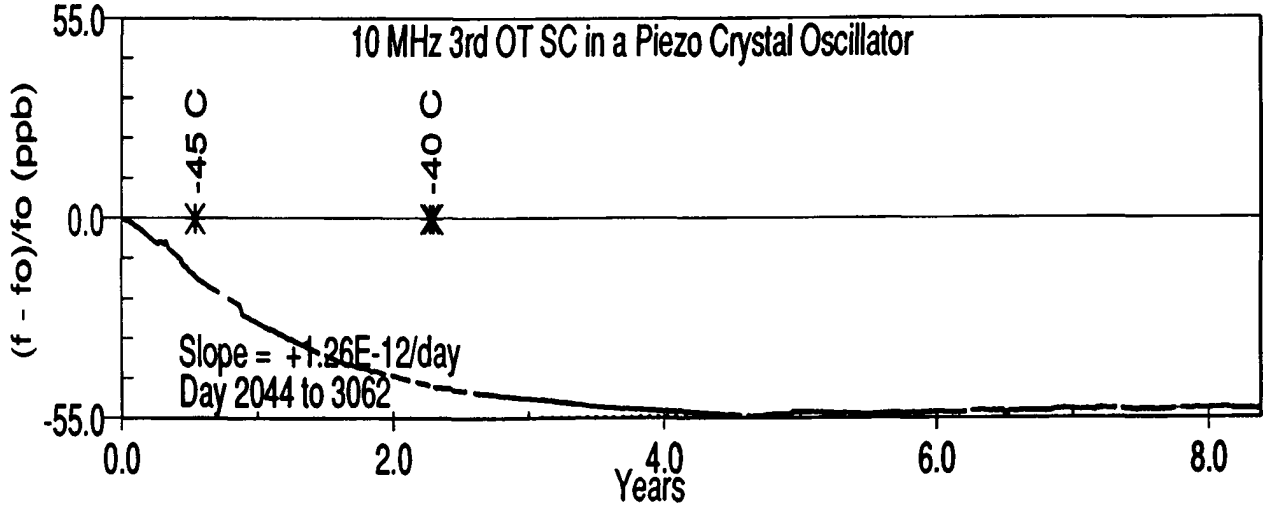


Figure 12. Piezo Crystal OCXO

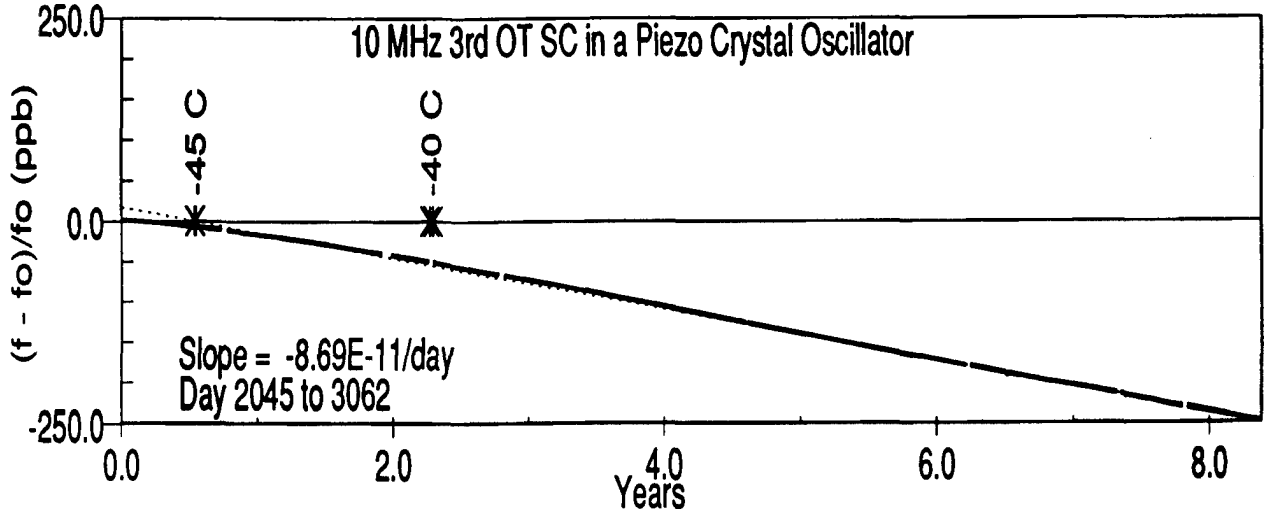


Figure 13. TMXO with Ceramic Flatpack Resonator

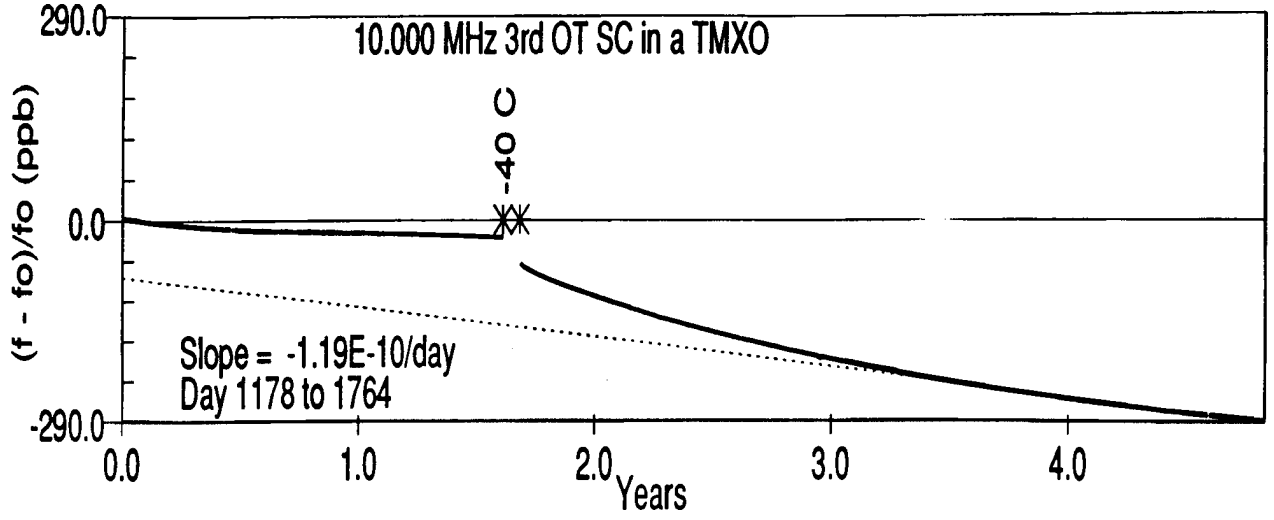


Figure 14. TMXO with Ceramic Flatpack Resonator

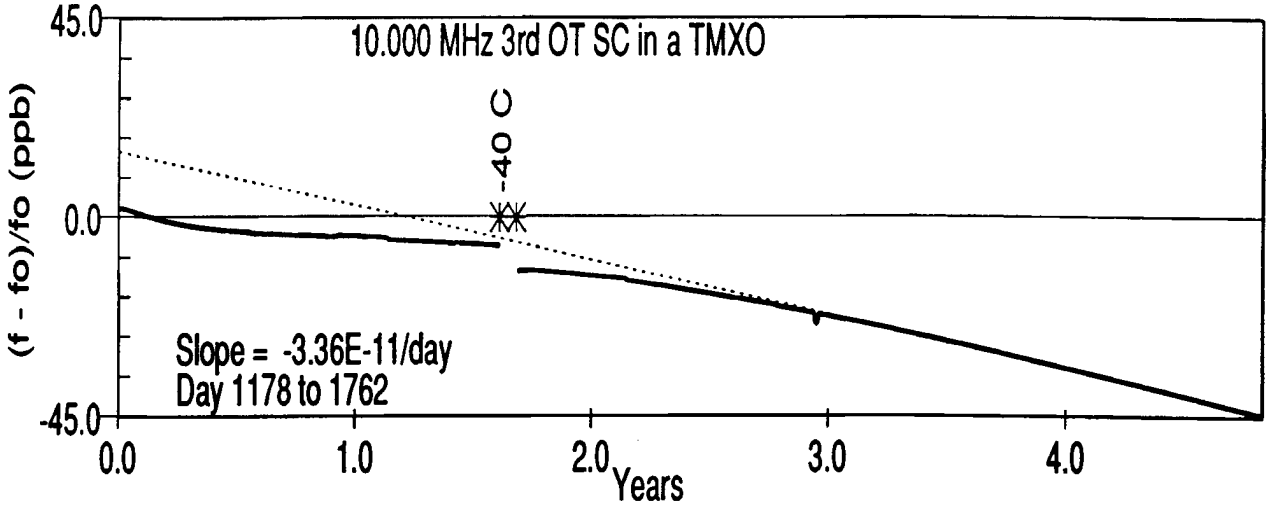


Figure 15. TMXO with Ceramic Flatpack Resonator

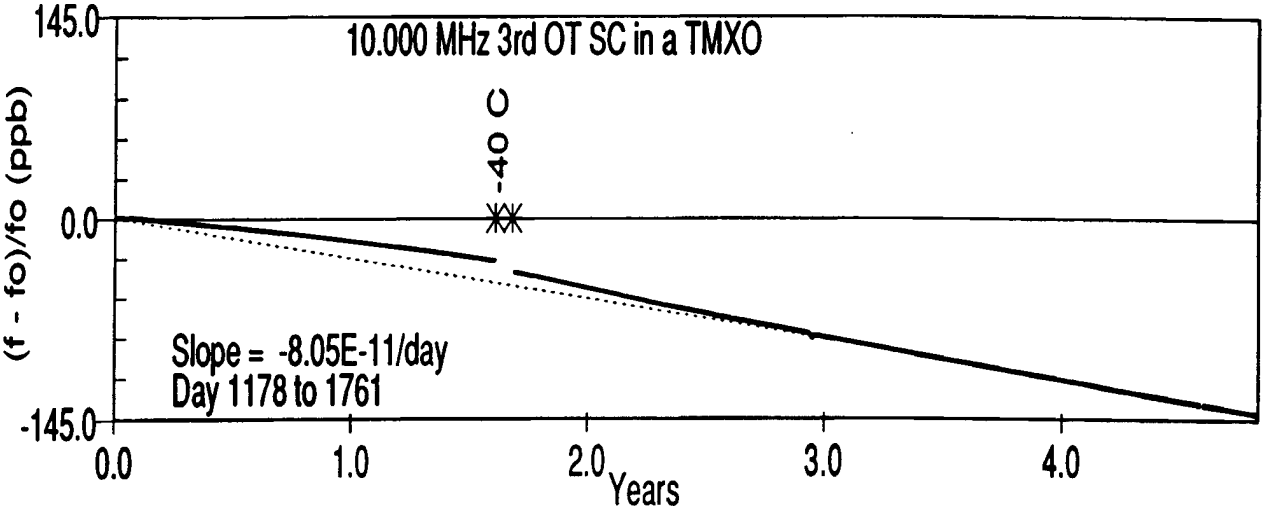


Figure 16. Shock Resistant Resonator in a TCXO

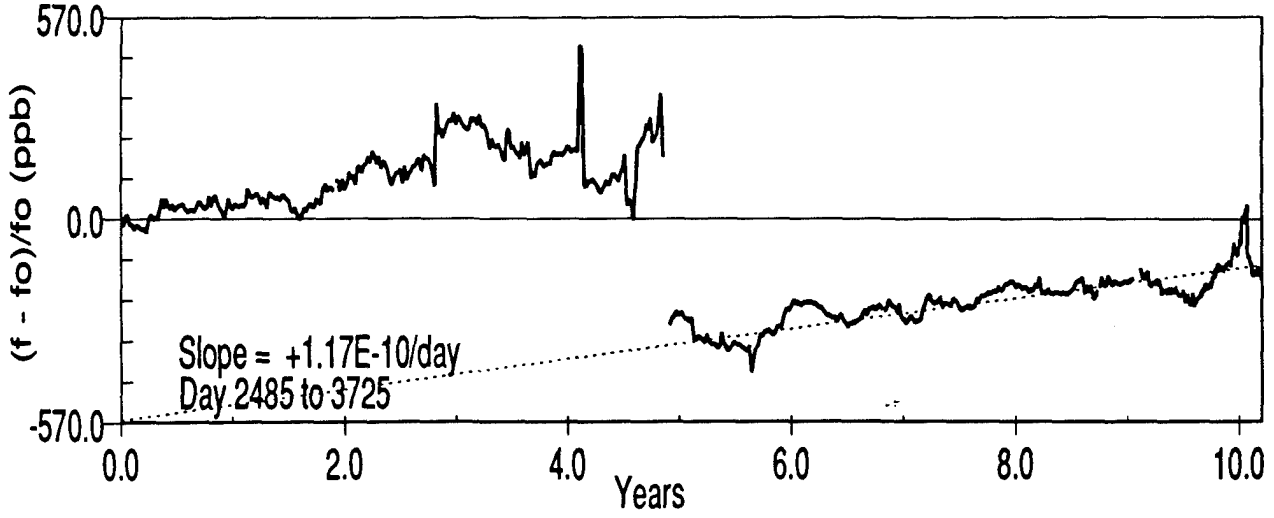


Figure 17. Shock Resistant Resonator in a TCXO

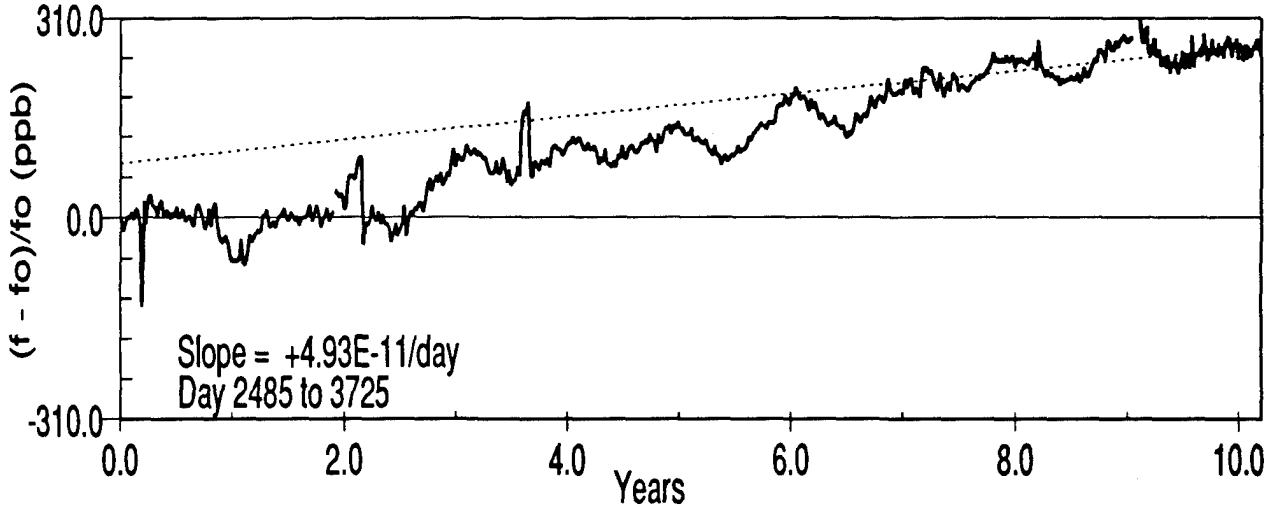


Figure 18. Shock Resistant Resonator in a TCXO

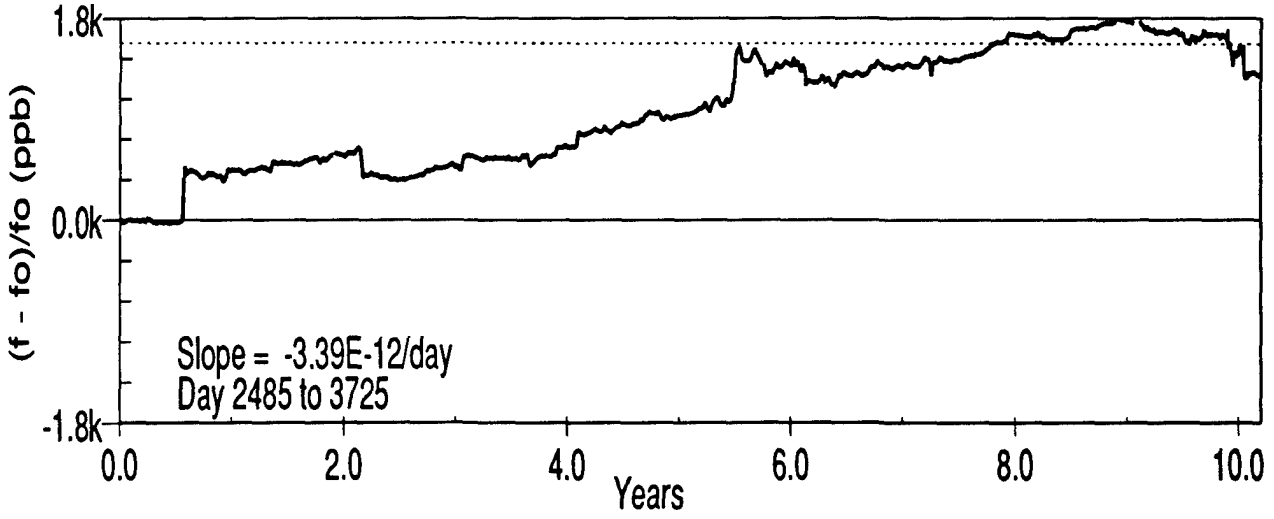


Figure 19. Shock Resistant Resonator in a TCXO

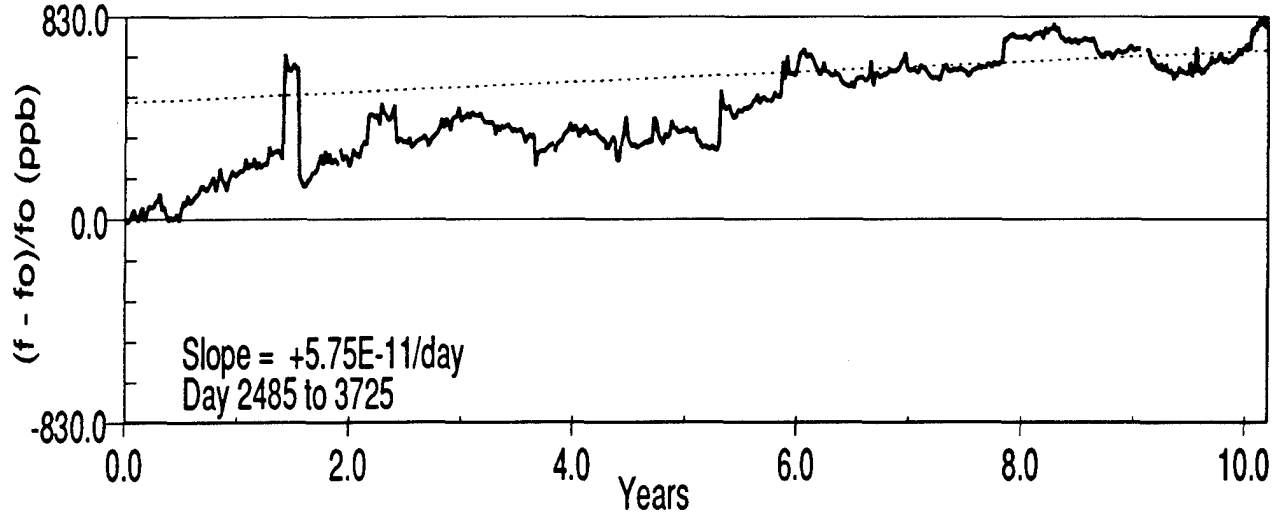


Figure 20. Shock Resistant Resonator in a TCXO

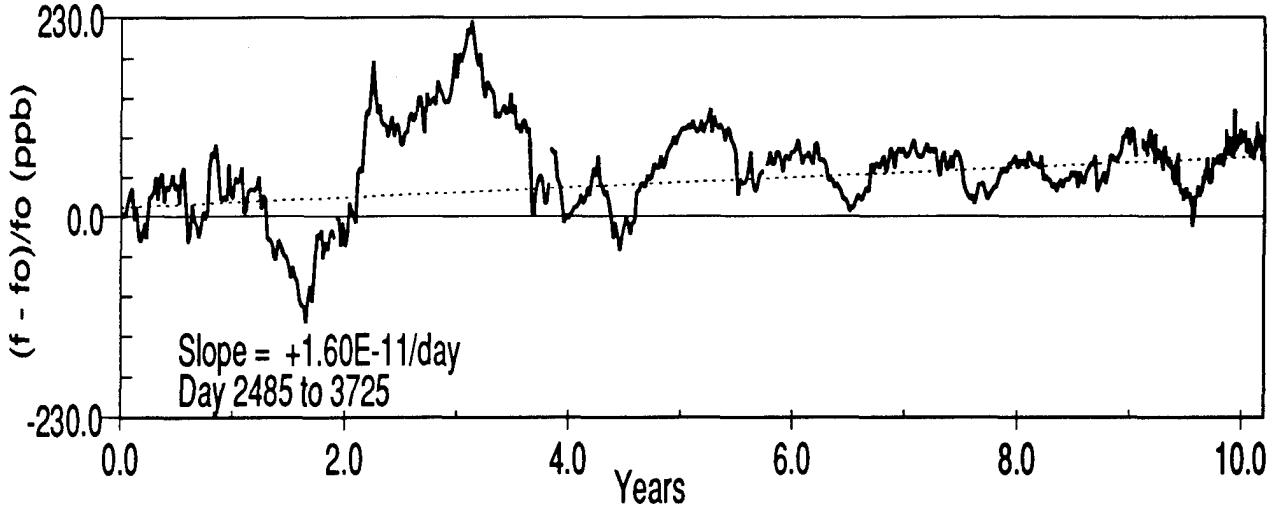


Figure 21. Shock Resistant Resonator in a TCXO

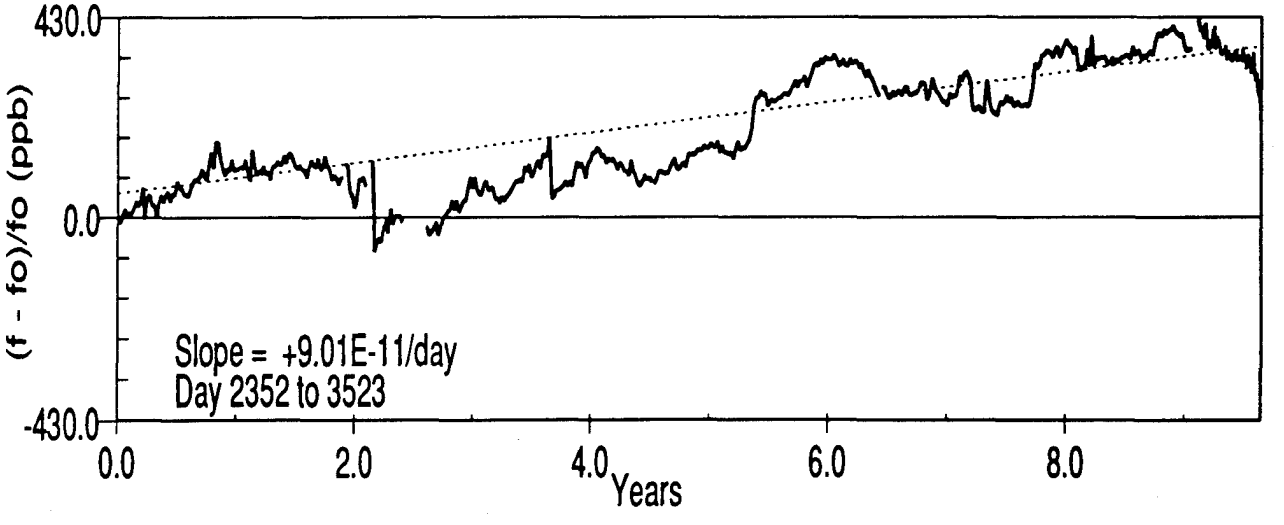


Figure 22. CTS TCXO

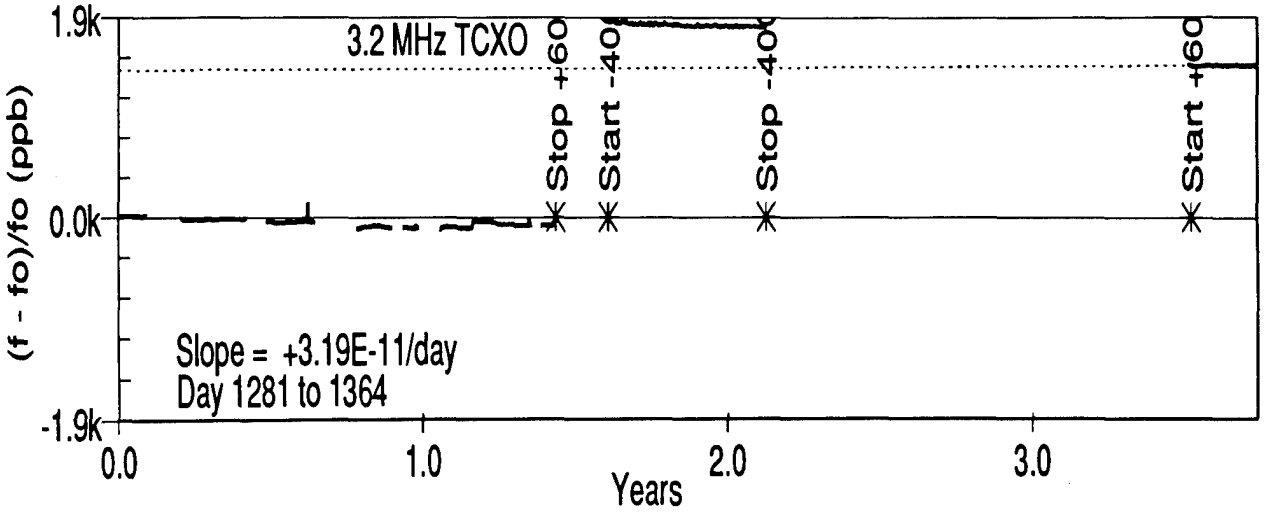


Figure 23. CTS TCXO

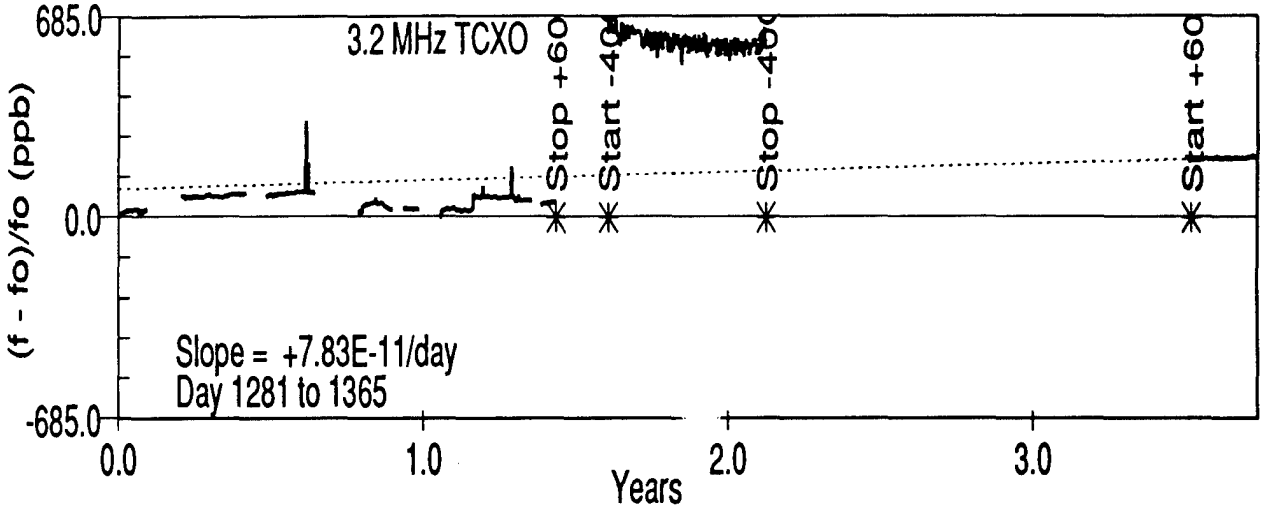


Figure 24. CTS TCXO

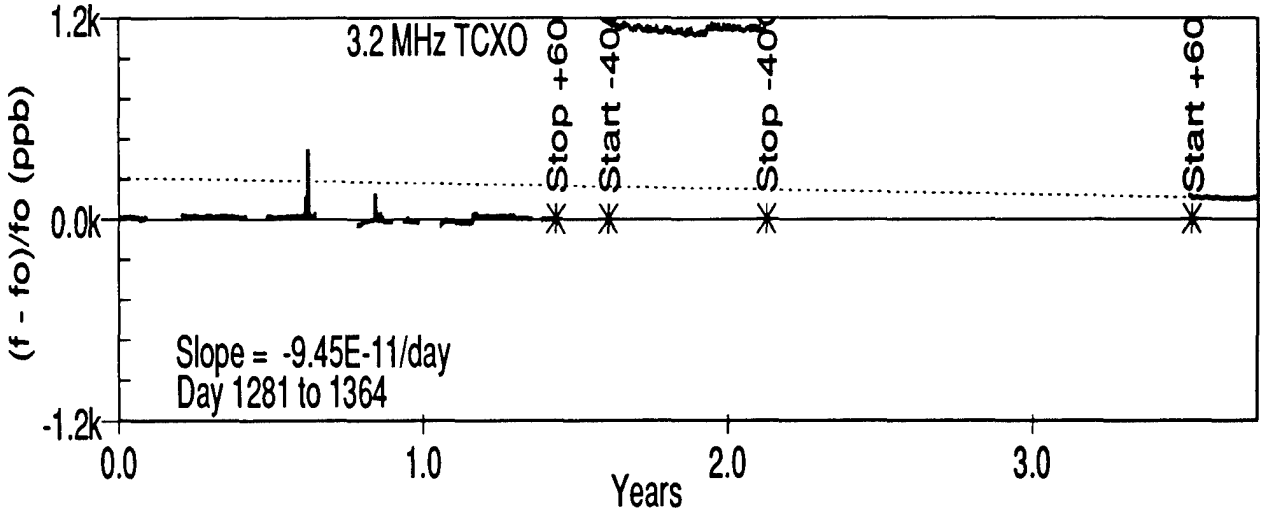


Figure 25. CINOX TCXO

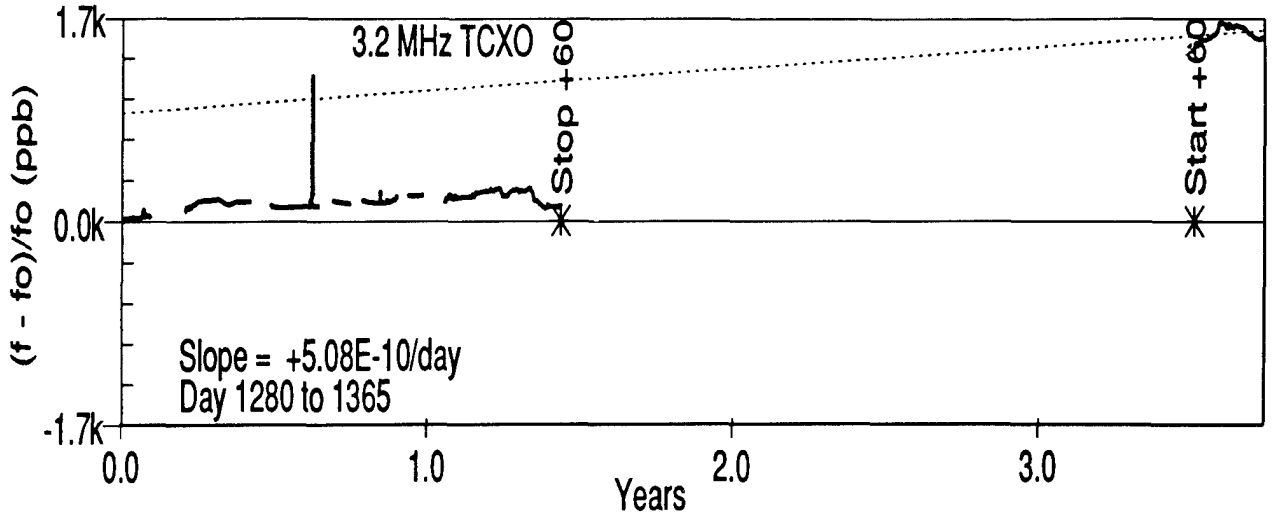


Figure 26. CINOX TCXO

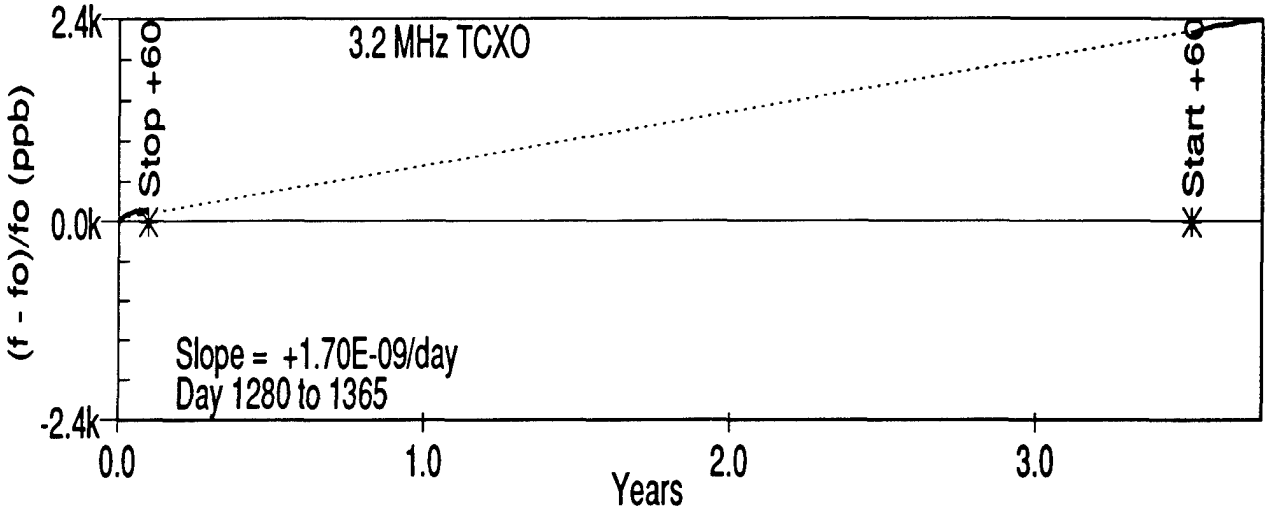


Figure 27. CINOX TCXO

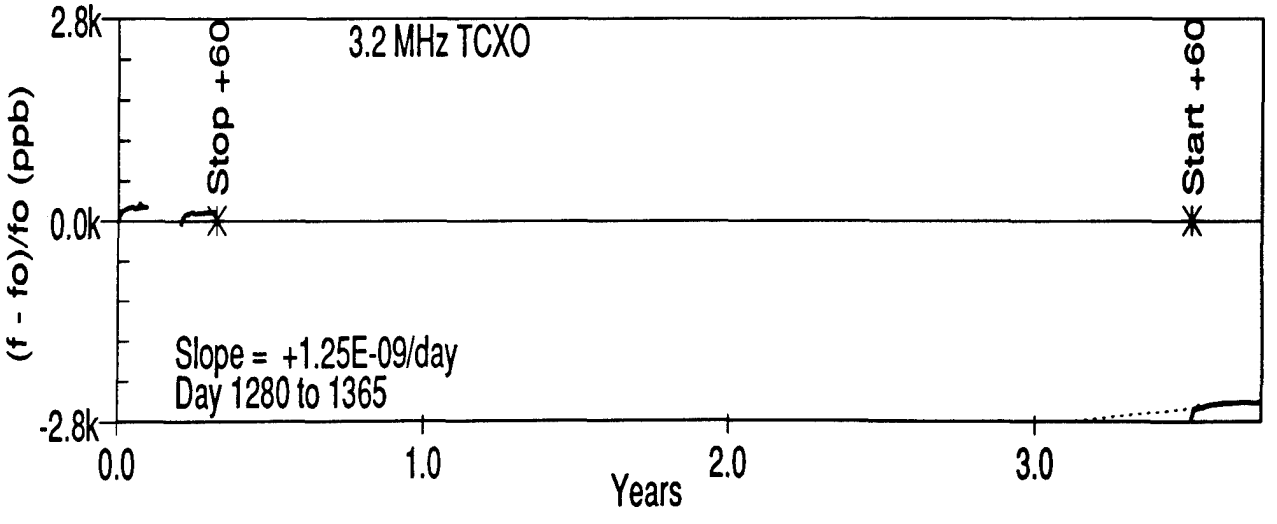


Figure 28. STC TCXO

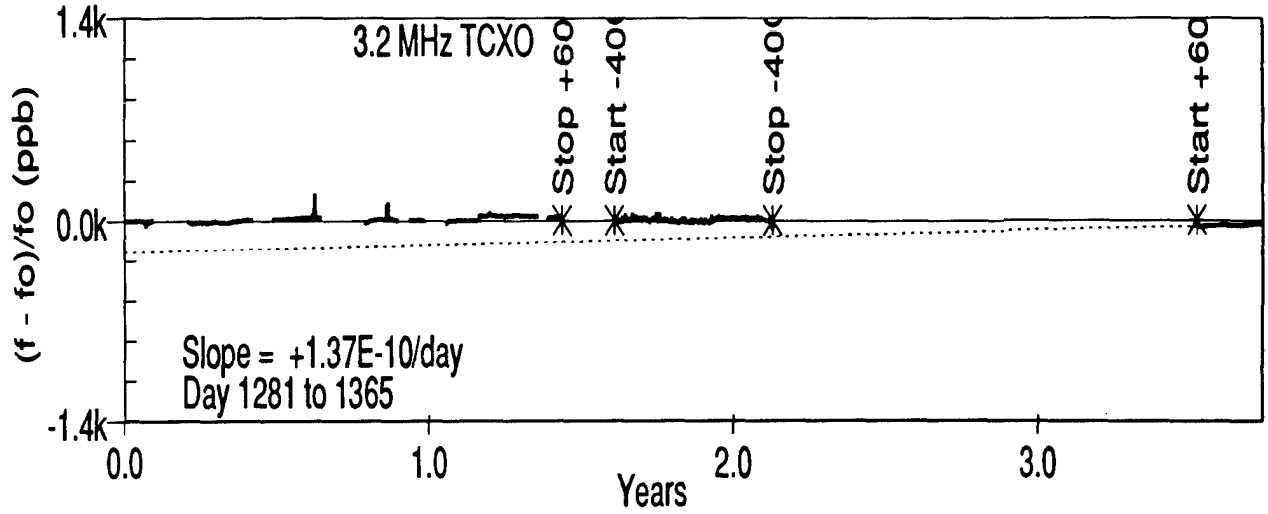


Figure 29. STC TCXO

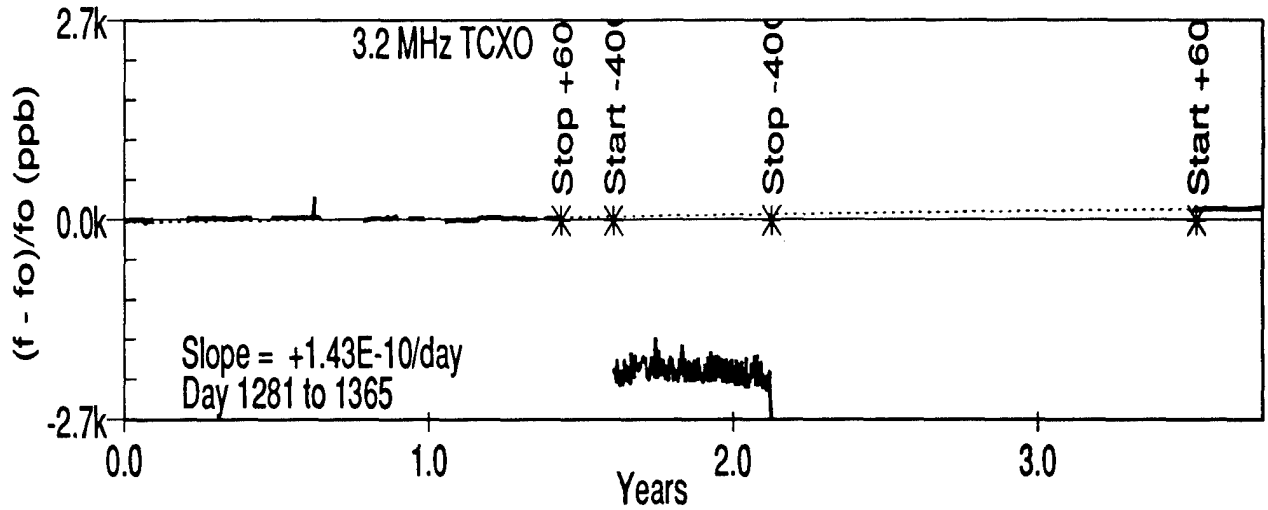


Figure 30. STC TCXO

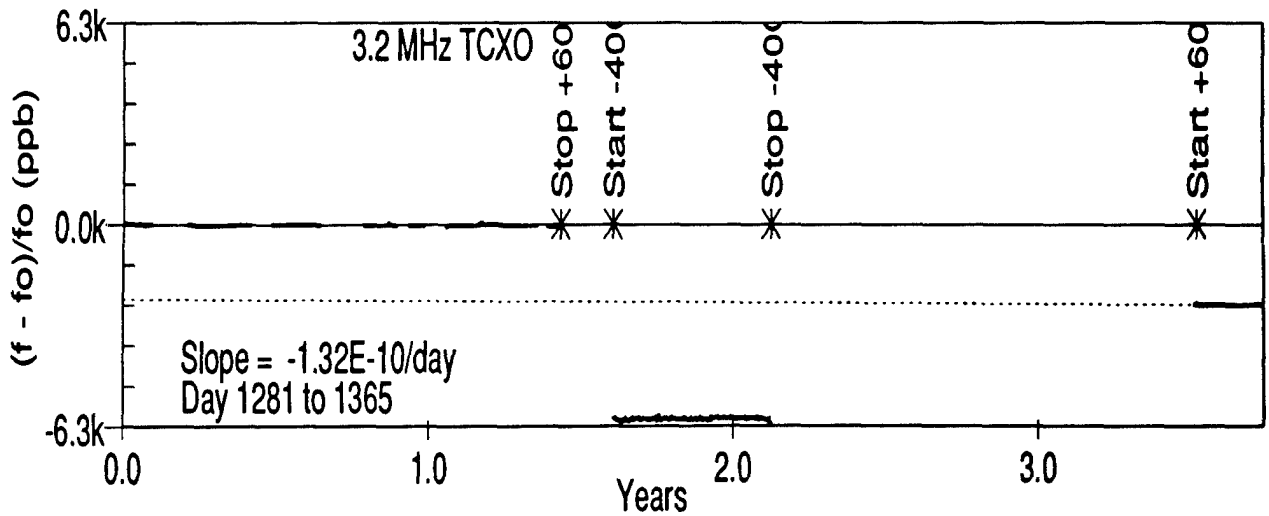


Figure 31. TFL TCXO

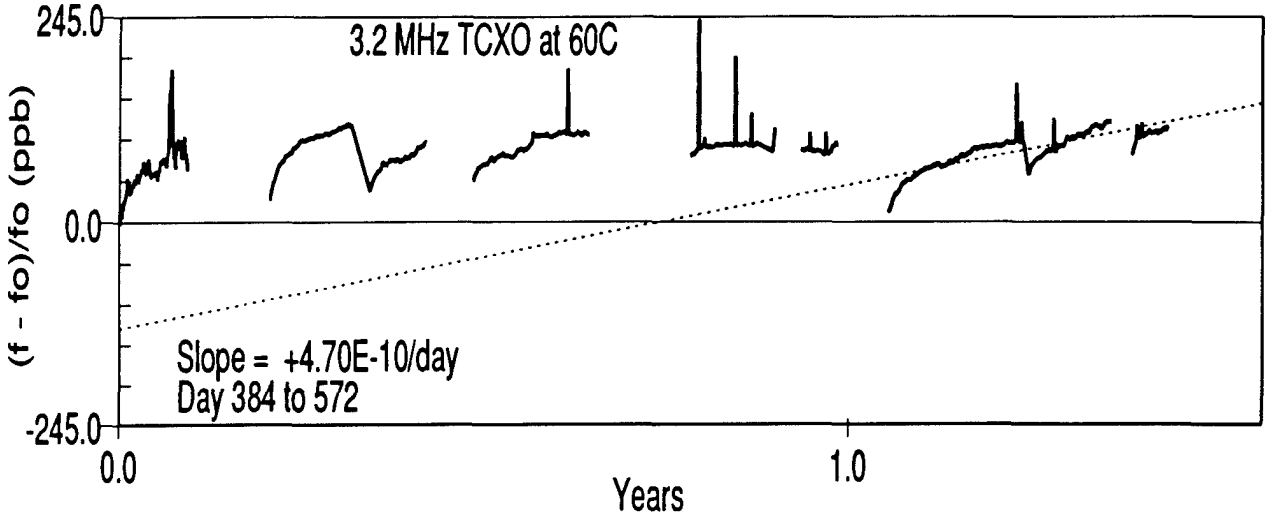


Figure 32. TFL TCXO

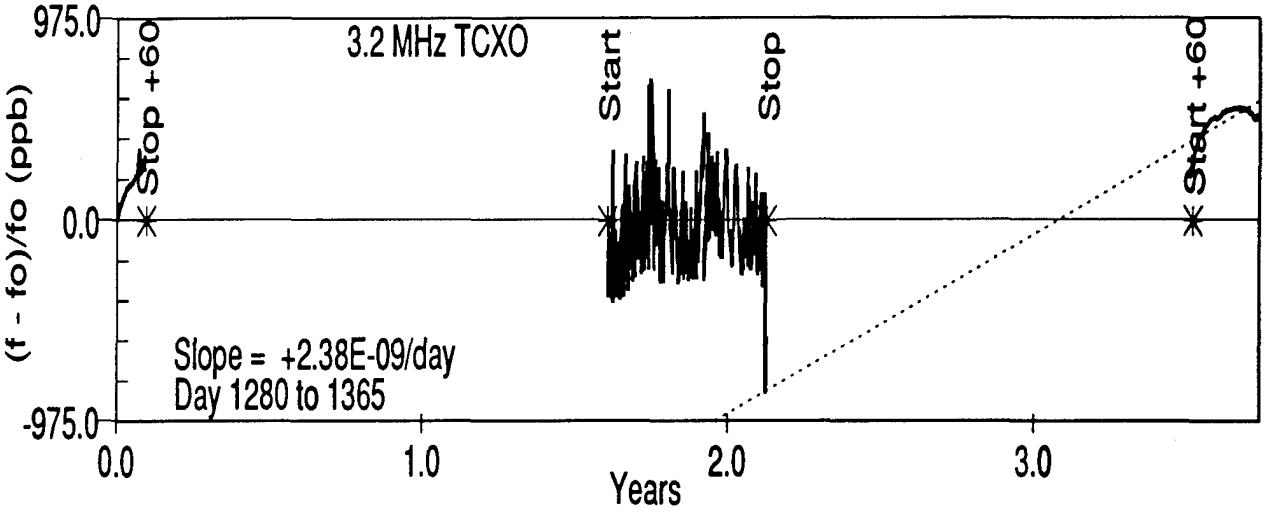
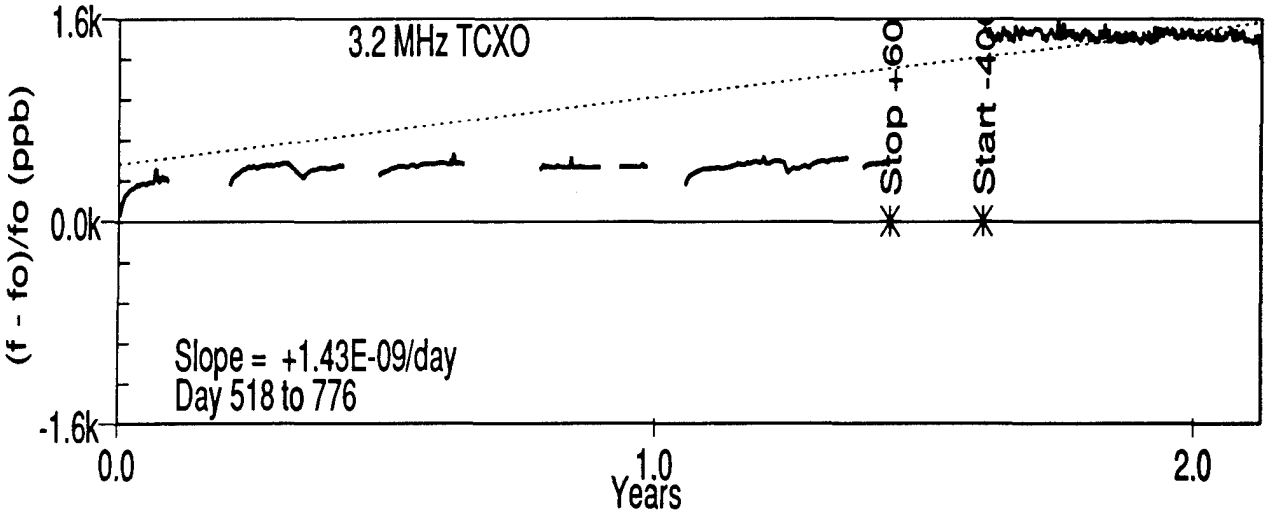


Figure 33. TFL TCXO



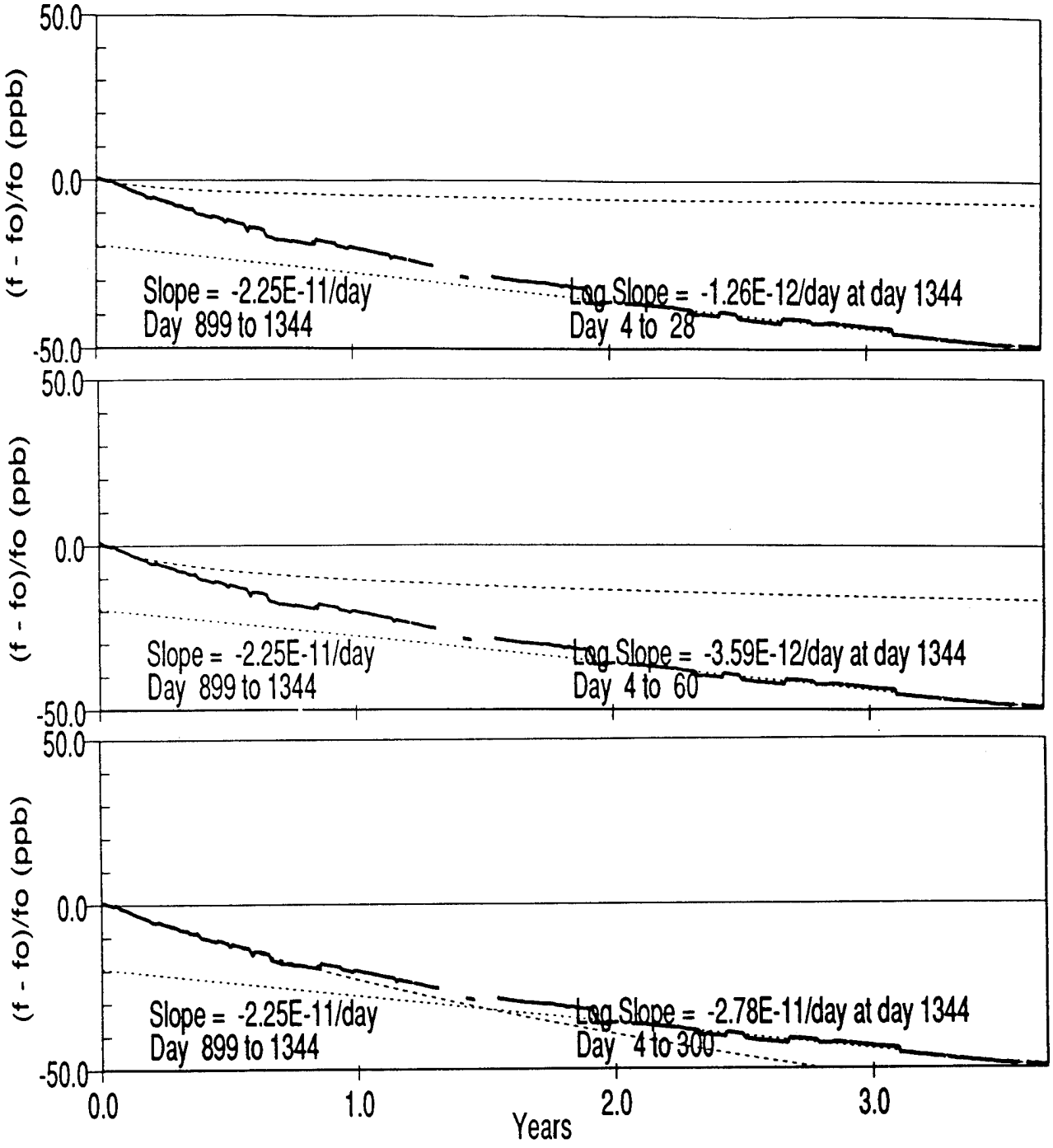


Figure 34. Oscillator A - Log fit using 30, 50 and 300 days of data .

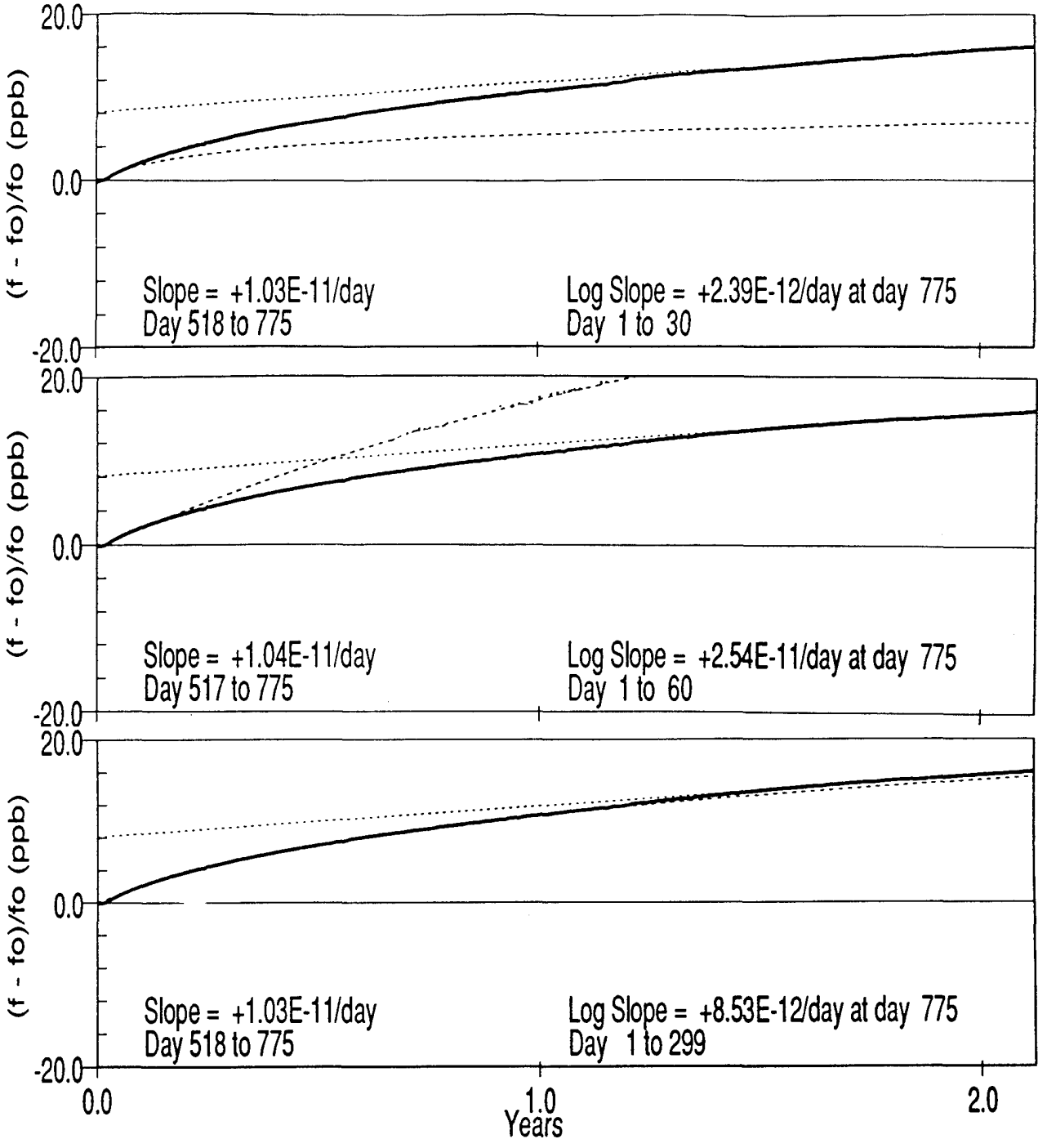


Figure 35. Oscillator B - Log fit using 30, 50 and 300 days of data.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ANALYSIS AND RESULTS OF COMPENSATED QUARTZ CRYSTAL OSCILLATOR OVENS

R. Brendel*, G. Marianneau*, F. Djian* and E. Robert**

* Laboratoire de Physique et Métrologie des Oscillateurs du CNRS
associé à l'Université de Franche-Comté-Besançon - 32,avenue de l'Observatoire - 25044 Besançon Cedex - France
** Centre National d'Etudes Spatiales - Dept AE/TT/TF - 18 avenue E. Belin - 31055 Toulouse Cedex - France

Abstract

This paper shows how it is possible to improve the performance of thermal enclosures by using a compensating system the principle of which has been described by Fred Walls a few years ago (41st AFCS, 1987). It is shown that because of the thermal network between the outside temperature, the temperature sensor and the device to be regulated, the latter may undergo residual temperature variations which reduce the overall thermal efficiency of the oven.

The paper shows how thermal transfer functions can be measured by using an experimental setup in which the node temperatures are measured by thermal sensors. By identifying the thermal response of the nodes with the theoretical transfer function under external temperature or heater excitation, the components of the equivalent R-C network can be determined. By knowing these thermal transfer functions, it is then possible to make use of a compensating system which can eliminate the parasitic static as well as dynamic thermal effects.

Validating measurements and experimental results are presented which show the strong improvement achieved by this compensating system with respect to the conventional approach.

Introduction

Over the years, it has been shown that the frequency stability of a quartz crystal oscillator is, to a large extent, dependent on the way that variations in the external temperature are handled. It is well known that the best results are obtained with ovens in which the temperature is kept constant by using a thermal servo loop. In such a system, the temperature of the oscillator is measured by a thermal sensor placed as close as possible to the device to be regulated. The temperature setpoint is compared with the actual value measured by the sensor and the error signal is amplified and converted into a current controlling the heating elements.

Nevertheless, because of the parasitic thermal impedance between the device to be regulated, the sensor, and the external medium, only the temperature of the sensor is kept constant, so that external temperature fluctuations may in fact be transmitted to the device. Fred Walls presented at the 41st AFCS [1] an interesting idea intended to reduce these parasitic effects: by using a second sensor measuring the external temperature, the temperature setpoint is modified so as to compensate the residual temperature variations of the device.

Electrical equivalent of the thermal enclosure

It is possible to describe the thermal behavior of a device by using an electrical analogy in which temperatures are equivalent to voltages, heating power to current sources, thermal resistance to electrical resistance and heat capacity to capacitance. Different methods can be used to determine the thermal impedances : (a) by calculation from the geometric and thermal properties of the materials used [2] or (b) by measurement of the thermal transfer functions.

The complexity of the various thermal transfer mechanisms (conduction, convection and radiation) and the lack of knowledge of the thermal properties of the materials used makes the calculations somewhat difficult and inaccurate. Although the measurement method can, to some extent, modify slightly the thermal behavior of the device because of the thermal sensors required to get the temperature at various nodes, it proves to be easier and much more accurate than the calculations.

Because of the perturbations induced by the thermal sensors, their number has to be judiciously chosen. If they are too few the thermal transfer functions can be inaccurate, and if they are too many the perturbations can be unacceptable.

Figure 1 shows the accepted thermal model of the single-layer oven used for the experiments.

0-7803-0476-4/92 $3.00 © 1992 IEEE

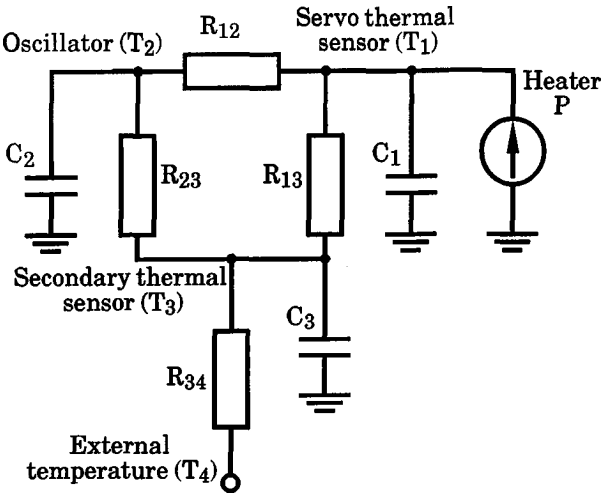


Fig. 1 : *Electrical equivalent network of the thermal model of a single-layer oven*

The relevant thermal nodes and temperatures are :

(a) T_1, the temperature of the thermal sensor used for the servo-loop. It is closely tied to the heating power transistor.

(b) T_2, the oscillator node. In some cases the resonator has been replaced by a thermal sensor. In other cases the resonator itself is used as a thermal probe.

(c) T_3, the secondary thermal sensor strongly coupled with the inner side of the outer shell.

(d) T_4, the external temperature.

Because of their small size and strong coupling, the sensor thermal impedance has been shown to be negligible.

C_1, C_2 and C_3 represent the thermal capacities of the corresponding nodes and R_{ij} are the thermal resistances between nodes. The heater is represented by a current source of heating power P.

Thermal impedance measurements

By using the electrical analogy of the thermal network it is possible to write the heat transfer equations in the following form

$$C_1 \frac{dT_1}{dt} = \frac{1}{R_{12}}(T_2 - T_1) + \frac{1}{R_{13}}(T_3 - T_1) + P \quad (1)$$

$$C_2 \frac{dT_2}{dt} = \frac{1}{R_{12}}(T_1 - T_2) + \frac{1}{R_{23}}(T_3 - T_2) \quad (2)$$

$$C_3 \frac{dT_3}{dt} = \frac{1}{R_{13}}(T_1 - T_3) + \frac{1}{R_{23}}(T_2 - T_3) + \frac{1}{R_{34}}(T_4 - T_3) \quad (3)$$

When the thermal steady state is reached, the derivatives vanish so that the remaining terms permit to obtain the thermal resistances. Measurements are done in a temperature-programmable climatic chamber which permits the variation of the external temperature T_4. Also, the experimental setup permits variation of the heating power P so as to obtain enough equations to lead to a solution. Table I shows the results thus obtained. In the experiments, steady state is usually reached after one hour of stabilizing time, the temperature T_1 is about 60°C.

Table I
Measured thermal resistances of the oven (in K/W)

R_{12}	R_{13}	R_{23}	R_{24}
35	83	83	5

The records of temperature transients are needed to measure the heat capacities. They can be obtained by giving the heater a power step function. The temperature-vs-time curves thus obtained are then used to measure the heat capacity. Equation 1 shows that the slope at the origin of the curve $T_1(t)$ gives $C_1 = P/(dT_1/dt)_0$. C_2 (resp. C_3) can be determined by observing that the right hand side of Eq. (2) (resp. Eq. (3)) shows a maximum when dT_2/dt (resp. dT_3/dt) is maximum. This is an inflexion point of the curve $T_2(t)$ (resp. $T_3(t)$) so that the heat capacity C_2 (resp. C_3) is given by [3]

$$C_i = \frac{(R.H.S.)_{max}}{(dT_i/dt)_{inflexion\ point}}$$

It should be noted that the right hand side of equations (2) and (3) can be obtained from the measured $T_i(t)$ and the thermal resistances previously measured. The power step amplitude is 0.2 W.

Table II summarizes the heat capacitances thus obtained.

Table II
Measured heat capacities of the oven (in J/K)

C_1	C_2	C_3
2	1.9	70

Verifications

It is possible to check for the validity of the model as well as for the measured thermal impedances by using an electrical simulator. Figures 2 and 3 show the experimental and simulated step responses of nodes 1 and 2. Given the measurement uncertainties, the results look fairly good. The slight observed differences can be partly attributed to the fact that the thermal resistances are not always constant but usually depend on the temperature.

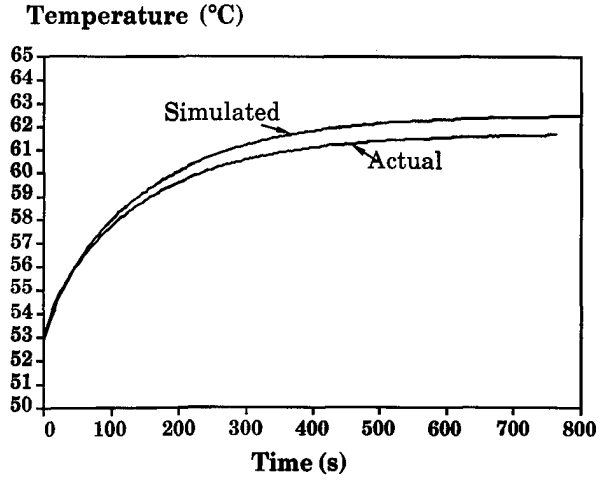


Fig. 2 : *0.2 W step response of node 1*

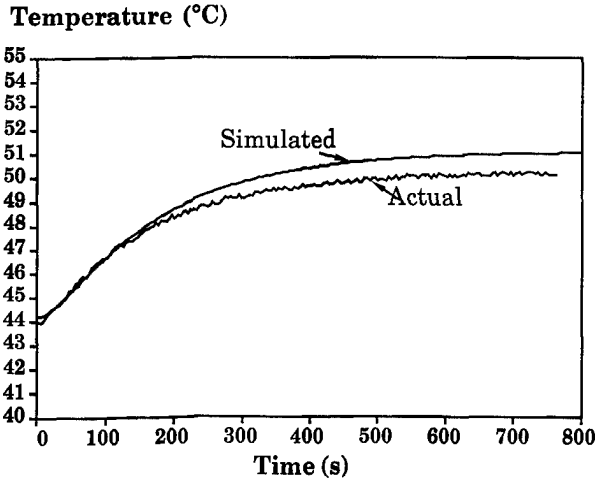


Fig. 3 : *0.2 W step response of node 2*

Compensation principle

Figure 4 shows the schematic representation of a conventional thermal enclosure. Here we see that only the servo-sensor temperature T_s is in fact kept constant. the resonator itself is only a part of the remaining thermal structure which does not belong to the servo-loop so that its temperature T_q can be different from the temperature setpoint T_c. T_q may even vary under the temperature perturbations T_a.

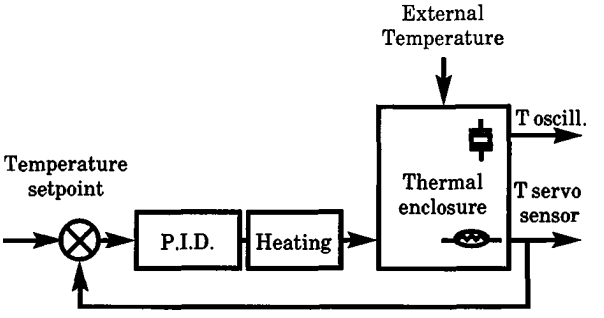


Fig. 4 : *Conventional thermal enclosure*

This fact can be expressed as

$$T_q = a \,.\, T_c + b \,.\, T_a \qquad (4)$$

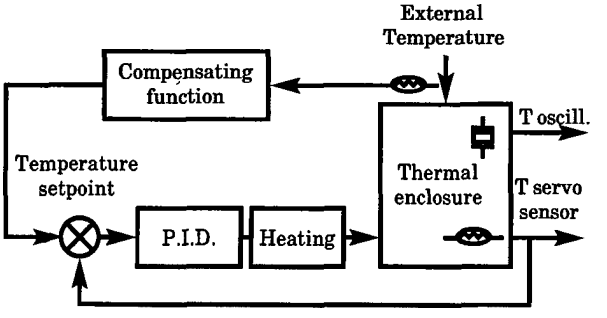


Fig. 5 : *Compensated thermal enclosure*

The compensation principle consists of changing the temperature setpoint T_c of the servo-loop as a function of the temperature perturbation T_a so as to compensate the residual temperature variations of the resonator [1]. Such a system is shown in Fig. 5. The temperature variations T_a are measured by means of a secondary thermal sensor, which controls the temperature setpoint T_c through the compensating function circuit.

This compensating function can be easily determined when the thermal impedances of the oven are known.

As a first approximation, let us consider only the quasistatic perturbations. In such a case the electrical equivalent circuit is composed only of thermal resistances (see Fig. 6) and the actual resonator temperature is a linear function of both setpoint and perturbation temperatures T_c and T_a. That means that the coefficients a and b of Eq. (4) are constant. Here we see that the resonator temperature T_q can be kept constant if

$$\Delta T_q = a.\Delta T_c + b.\Delta T_a = 0 \qquad (5)$$

This equation gives the compensating function

$$\Delta T_c = -\frac{b}{a}.\Delta T_a \qquad (6)$$

The coefficients a and b are functions of the thermal resistances of the circuit shown in Fig. 6 which is quite similar to Fig. 1 where heat capacities have been removed.

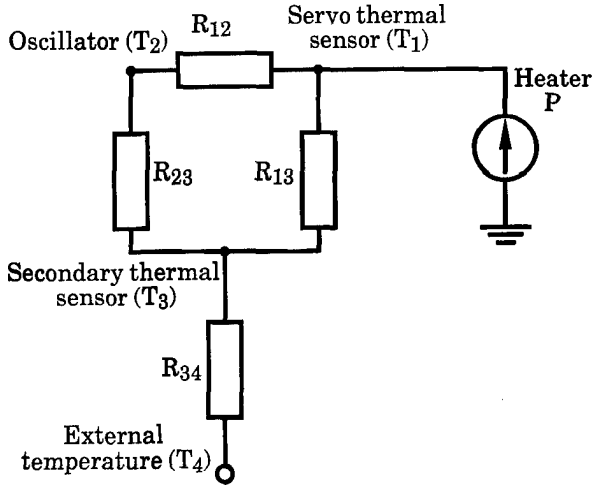


Fig. 6 : *Steady state or quasistatic thermal model of the oven*

By identifying the nodes we see that node 2 is the resonator ($T_q = T_2$). Assuming an ideal servo-loop, the servo-sensor temperature is identical to the setpoint temperature $T_c = T_s = T_1$. In addition, Table I shows that the thermal resistance R_{34} is much smaller than the others so that for quasistatic perturbations, $T_3 \approx T_4 = T_a$. Equation (2) then reduces to

$$\frac{1}{R_{12}}(T_c - T_q) + \frac{1}{R_{23}}(T_a - T_q) = 0 \qquad (7)$$

which can be written as

$$T_q = \frac{R_{23}\,T_c + R_{12}\,T_a}{R_{12} + R_{23}} \qquad (8)$$

Comparing (4) and (8) we can see that the compensating function (6) becomes

$$\Delta T_c = -\frac{R_{12}}{R_{23}}\Delta T_a \qquad (9)$$

By using numerical values listed in Table I we obtain $\Delta T_c = -0.42\,\Delta T_a$.

Experimental verification of the quasistatic compensation

Figure 7 shows the external temperature perturbation used in the verification experiments. This perturbation, whose maximum slope is about 10°C/hour, can be considered as quasistatic.

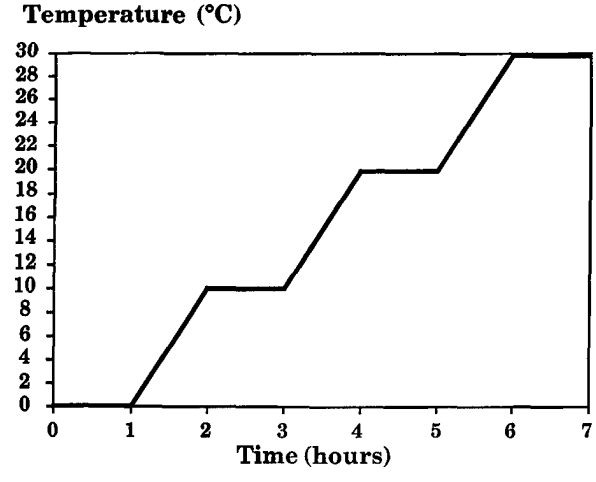


Fig. 7 : *External temperature perturbation*

The response of the resonator node temperature is plotted in Fig. 8 with and without using a compensating circuit to demonstrate the improvement thus obtained. Figure 9 shows the response of the servo thermal sensor when the system is compensated.

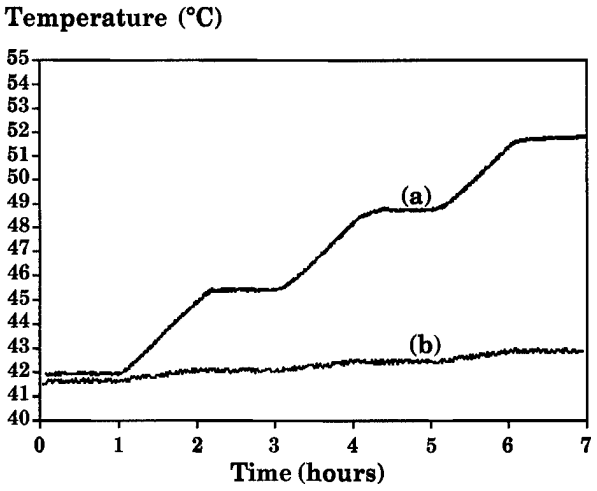


Fig. 8 : *Resonator node temperature response (a) without compensation (b) with compensation*

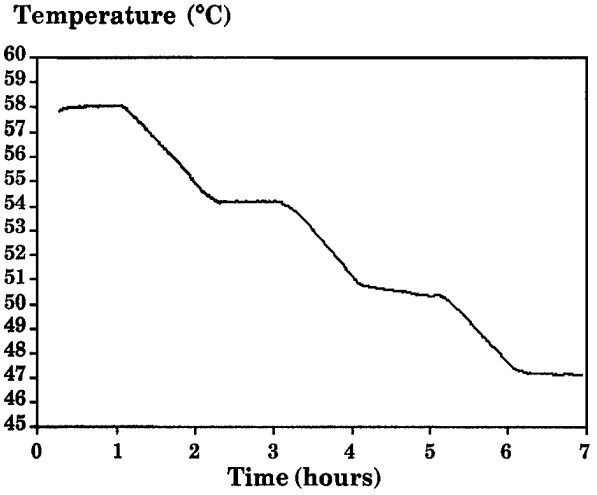


Fig. 9 : *Servo thermal sensor temperature response when using the compensating circuit*

Dynamic compensation

The quasistatic compensating system previously presented does not let us take into account dynamic perturbations. In this case heat capacities can no longer be neglected because of the time delays they induce in the system. Nevertheless the resonator temperature T_q is still a function of T_c and T_a as in Eq. (4). However the coefficients a and b are no longer constant but are functions of time so that Eq. (4) can be rewritten as

$$T_q = a(p) \, . \, T_c + b(p) \, . \, T_a \qquad (10)$$

where p is Laplace's variable. The compensating function Eq. (6) now takes the form

$$\Delta T_c = - \frac{b(p)}{a(p)} . \Delta T_a \qquad (11)$$

In the case of the thermal circuit of Fig. 1 it can be shown that Eq. (11) takes the following form

$$\Delta T_c = - \frac{R_{12}}{R_{23}} \frac{1 + \tau_1 p}{1 + \tau_2 p} . \Delta T_a \qquad (12)$$

where τ_1 and τ_2 are functions of the thermal impedances of the circuit. Here it should be noted that when p becomes very small (quasistatic perturbation) Eq. (12) reduces to Eq. (9).

Experimental verification of the dynamic compensation

The compensation function Eq. (12) has been implemented on an experimental oven. The oven has been submitted to external temperature ramps of higher and higher rates of change as shown in Fig. 10. The temperature response of the resonator node is plotted in Fig. 11 in three cases (a) without compensation (b) with quasistatic compensation only and (c) with dynamic compensation. It can be seen that the single quasistatic compensation does not work very well for fast temperature changes whereas the dynamic compensation does.

Results obtained with ovenized crystal oscillators

In order to demonstrate the value of the thermal compensating principle, this method has been implemented on a commercially-available ovenized crystal oscillator. The frequency-vs-temperature curve of the resonator used shows that the turnover point is located at about 75°C (see Fig. 12). Nevertheless, the working temperature setpoint has been shifted down to about 63°C so that the resonator acts as a thermal sensor with a sensitivity of about 3 Hz/°C. The thermal behavior of the enclosure is thus magnified. Fig. 13 represents the frequency-vs-temperature curve of the non-compensated ovenized oscillator submitted to a temperature ramp having a rate of 0.2°C/min within the range [-40°C, +80°C]. This curve shows that the servoloop does not keep the resonator temperature constant (and no longer acts when the external temperature rises above the temperature setpoint).

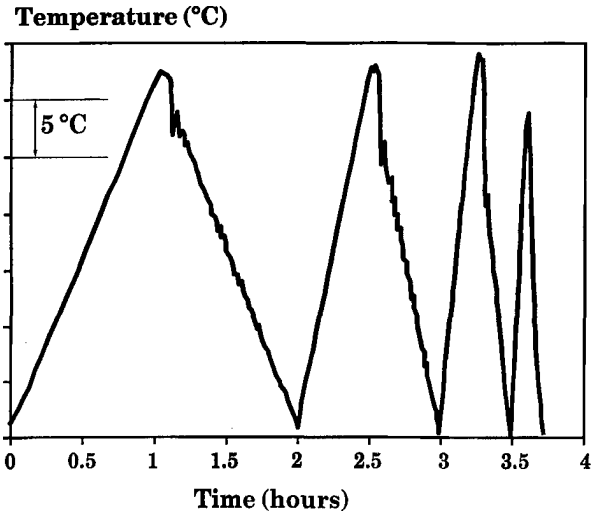


Fig. 10 : *External temperature perturbation*

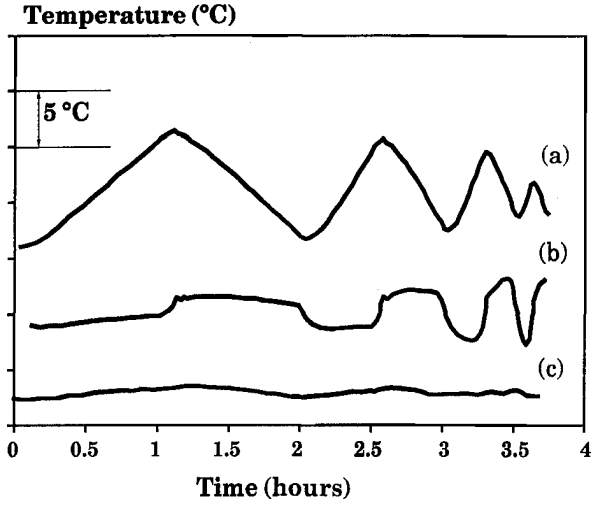


Fig. 11 : *Resonator node temperature response (a) without compensation, (b) with static compensation only, (c) with dynamic compensation*

Figure 14 shows the curves obtained for different values of the compensating coefficient (Eqs. 9 or 11). We can see here that the system can be (a) under-compensated, (b) overcompensated or (c) properly adjusted. The performance improvement of the thermal enclosure can be estimated by using a parameter which could be called "Thermal efficiency" and defined as the ratio between the total temperature variation "seen" by the device to be regulated and the corresponding external temperature variation for a given temperature range. By using this definition it can be shown that in the range [-40°C, +60°C] the thermal efficiency goes from 10 for an uncompensated OCXO (Fig. 13) up to 300 for the compensated unit (Fig. 14c).

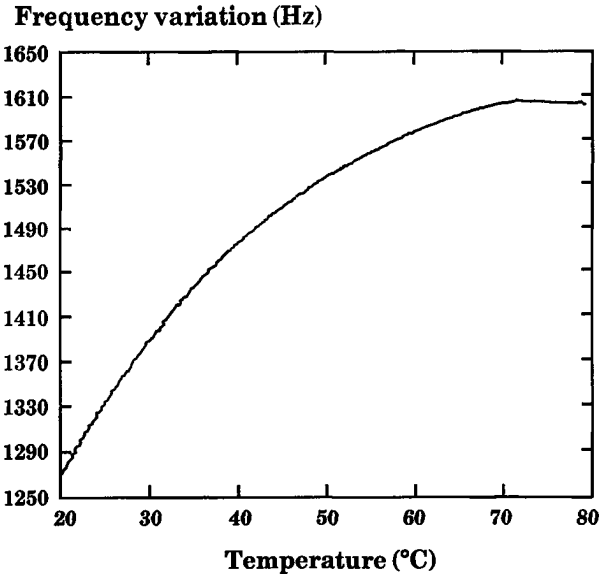


Fig. 12 : *Free running frequency-temperature characteristic of the oscillator under test*

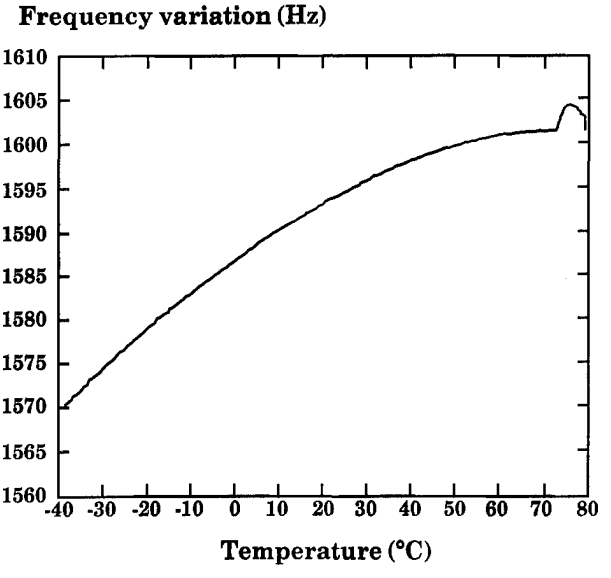


Fig. 13 : *Frequency-temperature characteristic of the uncompensated ovenized oscillator*

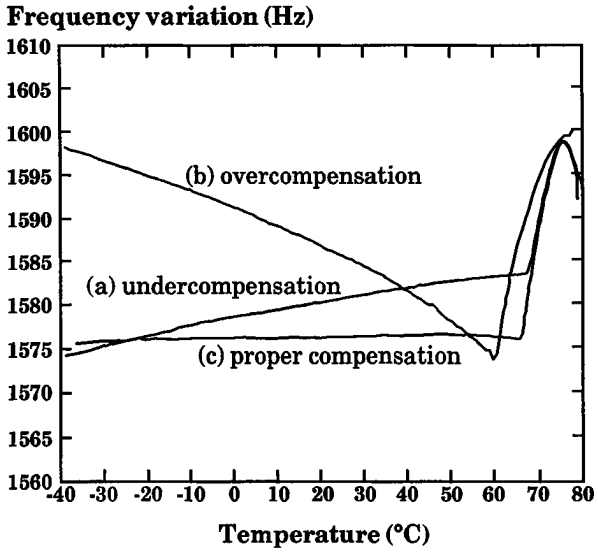


Fig. 14 : *Frequency-temperature characteristic of the compensated ovenized oscillator*
(a) undercompensated
(b) overcompensated
(c) properly adjusted

Conclusion

This paper has shown that thermally-compensated OCXO's achieve much better performance than their conventional counterparts. Although the work presented in this paper deals with an analog servoloop, the compensating principle can also be used in digital systems. In either cases, the electronics involved can be very simple as in the case of static compensation, or a bit more sophisticated as in the case of dynamic compensation.

The compensating system serves as an active thermal insulation. That means that for a given amount of insulating material better performance is obtained. Or, from another viewpoint, a compensated oven achieves the same performance which might be expected from a much more bulky thermal enclosure. This latter point is especially useful wherever volume or weight considerations are important.

Until now, the compensation of high-performance thermal enclosures such as Dewar flasks has not yet been investigated. Nevertheless, for these ovens, it seems that even if some improvements can be expected, they will not be as dramatic as those presented here.

Acknowledgements

This work was supported by CNES and CNRS under contracts # 932 88 53 474 00 and # 832 90 0815 00. The authors wish to thank Nick Yanonni for kindly revising this paper.

References

[1] F.L. Walls, "Analysis of high performance compensated thermal enclosures", Proc. 41st Annual Symposium on Frequency Control, pp. 439-443 (1987).

[2] 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).

[3] F. Djian, "Modélisation thermique des thermostats pour oscillateurs à quartz et applications", Thèse de Doctorat de l'Université de Franche-Comté (1991).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A PULSED CARRIER RECOVERY USING A DISCRETE CONCEPTION OF THE PLL

A. Aguasca, J.M. Plana, A. Broquetas

A.M.R Group, dept. of T.S.C, Polytechnic University of Catalonia. Spain
PO. BOX 30002, Barcelona 08080, Spain.

ABSTRACT

A bistatic Radar system needs an accurate synchronization between the transmitter and a the remote receiver, in order to achieve a correct coherent target detection and positioning [1]. This paper presents an unlinked receiver carrier phase/frequency synchronization method, from the radiated pulsed signal of a remote radar transmitter when its antenna beam illuminates the receiver. A modified discrete conception of the PLL technique is used, based on the special properties of the signal to be recovered.

INTRODUCTION

Coherent target detection and processing is used in modern radar systems to reject static clutter echoes and estimate target velocities from the doppler frequency shift of the scattered signals. Coherent detectors require precise phase and frequency references of the transmitted pulses [2]. In conventional monostatic systems both emitter and receiver share the same reference oscillators. This is not the case for bistatic or multistatic systems in which a remote receiver must use dedicated links to achieve coherent operation. An alternative is to synchronize a high stability oscillator from the radiated signals by the transmitter when the antenna is in the receiver direction. A typical radar signal consists of a basic rectangular RF pulse of 1 μs width which is repeated with a period of 1ms. The transmitter antenna rotation speed and beamwidth limits the availability of the signal from the receiver to windows of around 30 ms every 4.5 s (figure 1). Thus the synchronism technique proposed must acquire the phase of the signal in the short visibility time and maintain the coherence until the next antenna scan. Similar synchronization problems with limited access to reference are encountered in other important areas like TDMA and mobile communication systems [3].

Phase and frequency recovery are usually based on PLL techniques [4][5] (figure 2), where the synchronized signal is obtained with a VCO controlled with the filtered phase error between the reference signal and its output. PLL systems have been extensively studied for incoming continuous wave (CW) signals, this is not the case when the signal has a pulsed characteristic like the transmitted radar signal, which has an extremely low duty cycle, on the order of 10^{-3}. A discrete-time PLL technique is proposed to achieve the synchronization of pulsed signals considering the pulse modulation as a temporal sampling.

THE PLL AS A PULSED CARRIER SYNCHRONIZER

A second order PLL with a perfect integrator loop filter (figure 2) is able to maintain the synchronism when the reference signal has an amplitude modulation, whenever the Acquisition process has been done, and the system is under Tracking mode. The best way to use the property of immunity to amplitude modulation is to perform the acquisition process during just one received pulse, and ensure it has been completed when signal vanishes. However, the non ideal performance of the loop filter and the VCO, in the PLL, reduce considerably this characteristic. It can be seen a reduction of the loop bandwidth, that increases the acquisition time, and the appearance of a peak at the cut-off frequency, that implies a potential source of unstabilities. Thus a PLL designed for large bandwidth, is not a well solution to synchronize pulsed signals.

At the closed-loop, For low duty cycle rectangular amplitude modulation, the output signal from the phase detector is the sampled version of the phase error when the incoming signal has CW characteristic, with a sampling interval T equal to the modulating signal. From this consideration, it is possible to reach the synchronization making use of the whole burst of pulses in the received signal window. We must use theories usually employed in discrete and digital system. In the

0-7803-0476-4/92 $3.00 © 1992 IEEE

loop filter (figure 2), y(t) can be expressed as:

$$y(t) = y[(n-1)T] + \frac{\tau_2}{\tau_1}x(t) + \frac{1}{\tau_1}\int_{(n-1)T}^{t} x(t)dt \;;\; for\; t \geq (n-1)T \quad \textbf{(a)}$$

And at the VCO, we have:

$$\theta_0(nT) = \theta_0((n-1)T) + K_0\int_{(n-1)T}^{nT} y(t)dt \quad \textbf{(b)}$$

using (a) and (b) the output phase of the VCO, at the instant of sampling, can be written as:

$$\theta_0(nT) = \theta_0((n-1)T) + K_0\left[Ty((n-1)T) + \left[\frac{\tau\tau_2}{\tau_1} + \frac{T\tau}{\tau_1} - \frac{\tau^2}{2\tau_1}\right]x((n-1)T)\right]$$

where τ is the received pulse duration. Last expression represents the evolution of the VCO phase at the instants of sampling, as a function of the signals at previous instants. We can use the Z-transform in order to obtain the relationship between signals:

$$\theta_0(z) = \theta_0(z)z^{-1} + K_0[Ty(z)z^{-1} + Kx(z)z^{-1}]$$

$$where\; K = \frac{\tau\tau_2}{\tau_1} + \frac{T\tau}{\tau_1} - \frac{\tau^2}{2\tau}$$

We can express y(nT) with its Z-transform as:

$$y(z) = \frac{\tau}{\tau_1}\frac{z^{-1}}{1-z^{-1}}x(z)$$

Then, the VCO phase is:

$$\theta_0(z)(1-z^{-1})^2 = K_0\left[K + \left(\frac{T\tau}{\tau_1} - K\right)z^{-1}\right]x(z)z^{-1}$$

Where $x(z)=K_d[\phi(z)]$, and $\phi(z)=\theta_i(z) - \theta_0(z)$. After some operations we get the phase error evolution as:

$$\phi(z) = \frac{(1-z^{-1})^2}{1-[2-KK_0K_d]z^{-1} + \left[1+K_0K_d\left(\frac{T\tau}{\tau_1} - K\right)\right]z^{-2}}\theta_i(z) \quad \textbf{(c)}$$

Using Z-transform theory, the system is stable if the denominator has its roots with a module less than 1. And using the last value theorem [6], it is seen that phase error tends to zero for step and ramp variation of the input phase θ_i. The loop transfer function is the eq.(d).

Loop filters parameters must be modified respect to values obtained for analog PLL in order to ensure loop stability and acquisition velocity, a well method to calculate them is to obtain the equivalent analog system

$$H(z) = \frac{\theta_0(z)}{\theta_i(z)} = \frac{K_0K_d\left[K + \left(\frac{T\tau}{\tau_1} - K\right)z^{-1}\right]z^{-1}}{1-[2-KK_0K_d]z^{-1} + \left[1+K_0K_d\left(\frac{T\tau}{\tau_1} - K\right)\right]z^{-2}} \quad \textbf{(d)}$$

by replacing z by $(1+j\omega T)$ [7], if the equivalent analog bandwidth B_L is $B_L << 1/T$, and calculate them for a fixed natural frequency and damping factor of the loop transfer function.

However the system is still under technological limitations and the system cannot maintain the coherence between two consecutive signal windows, because the real O.A reduces the ideal integrator performance of the loop filter.

FIRST APPROXIMATION TO SOLUTION

Next step is to seek some solutions to avoid these limitations, that degrades the system performance. The best way to do this is to replace analog loop filter by a digital conception filter, because it permits an easy implementation of integration performance better than with an A.O. Using Z-transform, the PLL can be represented as in figure 2, replacing F(s) and V(s) by F(z) and V(z), which are the loop filter and VCO transfer function in Z domain. The Closed-Loop Transfer Function and Phase Error, in linear regime, are:

$$H(z) = \frac{K_dF(z)V(z)}{1+K_dF(z)V(z)}$$

$$\phi(z) = \frac{1}{1+K_dF(z)V(z)}\theta_i(z)$$

If we design a loop filter with a relation in time domain as:

$$y(t) = y((n-1)T) + K_f[x(nT) - ax((n-1)T)] \quad for \quad nT \leq t < (n+1)T$$

that implies in the Z-domain:

$$F(z) = \frac{y(z)}{x(z)} = \frac{K_f(1-az^{-1})}{1-z^{-1}}$$

In the VCO we have:

$$\theta_0(nT) = \theta_0((n-1)T) + K_0 T y((n-1)T)$$

$$V(z) = \frac{\theta_0(z)}{y(z)} = \frac{K_0 T z^{-1}}{1 - z^{-1}}$$

Obtaining:

$$H(z) = \frac{K_0 K_d K_f T(1 - a z^{-1}) z^{-1}}{1 - (2 - K_0 K_d K_f T) z^{-1} + (1 + a K_0 K_d K_f T) z^{-2}}$$

$$\phi(z) = \frac{(1 - z^{-1})^2}{1 - (2 - K_0 K_d K_f T) z^{-1} + (1 + a K_0 K_d K_f T) z^{-2}} \theta_i(z)$$

These expressions for Loop Transfer Function and Phase Error evolution, are similar to (c) and (d), which tell us that this configuration is able to synchronize the VCO to the phase and frequency of the pulsed carrier. And it is similar to the general baseband model of a Digital PLL (DPLL)[7].

Systems which are described by the Z-transform can be easily simulated numerically. Using the properties of discrete time signals, the general expressions relating the PLL signals can be written as:

$$\begin{aligned} x(n) &= K_d \sin[\theta_i(n) - \theta_0(n)] \\ y(n) &= y(n-1) + K_f(x(n) - a x(n-1)) \\ \theta_0(n) &= \theta_0(n-1) + K_0 T y(n-1) \end{aligned} \qquad \text{(e)}$$

System simulation using these expressions allows to calculate loop filter parameters in order to reach stable performance, and optimize the phase and frequency acquisition time.

frequency acquisition uncertainty

It can be seen that the system possibly acquires the reference phase and frequency with an error always multiple of the sampling frequency. It can be explained observing the signal from the phase detector in two different situations, when the VCO phase close to the input phase, and VCO phase with a frequency shift multiple to the sampling frequency, the input signal to the loop filter cannot be distinguished under these two situations, and the loop acts at the same manner. This behavior is encountered in all discrete systems where there is a periodic spectral performance. However this frequency discrete ambiguity is not a handicap in the specific case where we use it, since the radar processor compensates internally this frequency shifts.

THE PULSED CARRIER SYNCHRONIZER

Figure 3 shows a hardware solution that implements the proposed loop filter, based on Sample & Hold circuits acting as delay tapped lines, that maintain constant the necessary signals between two consecutive pulses. The sampling is synchronized to the instants of pulse reception.

Several measured results of the loop performance are presented in figures 4 and 5, that can be compared with the simulation results using the expressions (e), for the same VCO, phase detector and loop filter parameters, and similar initial conditions. It is shown that the measurements are very close to simulation. In these figures are shown that, for small frequency errors, the process can be considered as linear (Sinus operand of the phase detector is approximated by its angle), and a pull-in process is observed for large frequency errors.

This configuration theoretically solves the coherency maintenance when no direct illumination is done. However the drift of the Sample & Hold circuit as a controller of the VCO produces a progressive frequency drift, unacceptable for this application.

The discrete consideration of the system allows to use A/D-D/A converters and digital registers as delay lines, to implement the loop filter transfer function. A loop filter using digital-analog converters has been developed, based on a D/A input-output PC card (figure 6), and it has seen that it is able to be used (figure 7). Using this loop filter, the frequency drift due to the non-ideal integrator performance is solved, but a quantization error effect appears as a limitation of performance, because the VCO control voltage is restricted to the finite range of digitized levels. It can be reduced using large number of bits for a defined discretization window.

CONCLUSIONS

A PLL technique suitable to periodic pulsed signals has been presented, based on a discrete-time treatment of the involved signals. This consideration allows an easy and realistic simulation of the PLL acquisition and tracking processes. At the same time, the loop filter may be implemented by delay tapped lines. Two experimental loop filters has been implemented, one based on S&H circuits, and another on D/A and digital registers, that show a close agreement between simulated and actual performance. The proposed approach can be applied to pulsed signal synchronization for radar and communication coherent systems.

ACKNOWLEDGEMENT

This work has been supported by the Spanish Commission for Research CICYT, TIC 481/89.

REFERENCES

[1] N.J. Willis, Bistatic Radar, Artech House, 1991.

[2] M.I. Skolnik, Radar Handbook, Second Edition, McGraw-Hill, 1990.

[3] A.B. Carlson, Communication Systems, third Edition, McGraw-Hill, 1986.

[4] F.M. Gardner, Phaselock Techniques, John Wiley & Sons, 1979.

[5] A. Blanchard, Phase-Locked Loops, John Wiley & Sons, 1976.

[6] D.J. DeFatta, J.G. Lucas, W.S. Hodgkiss, Digital Signal Processing: a System Design Approach, John Wiley & Sons, 1988.

[7] W. C. Lindsey, C. M. Chie, "A survey of Digital Phase-Locked Loops", Proceedings of the IEEE, vol 69. April 1981.

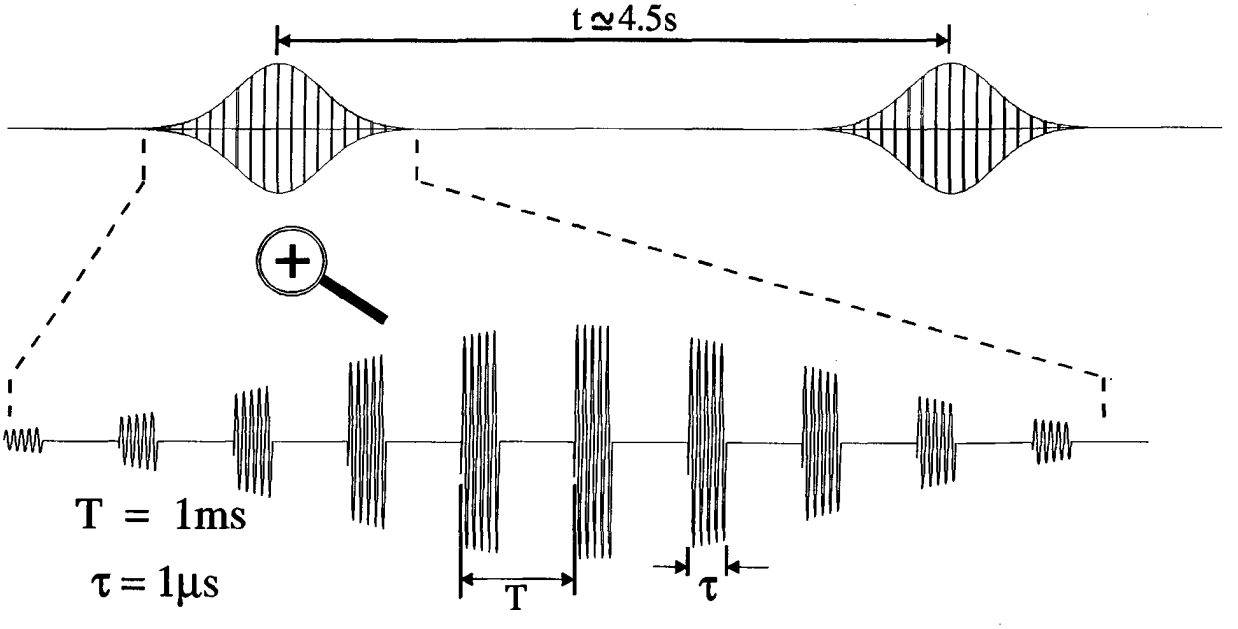


figure 1. Received signal aspect.

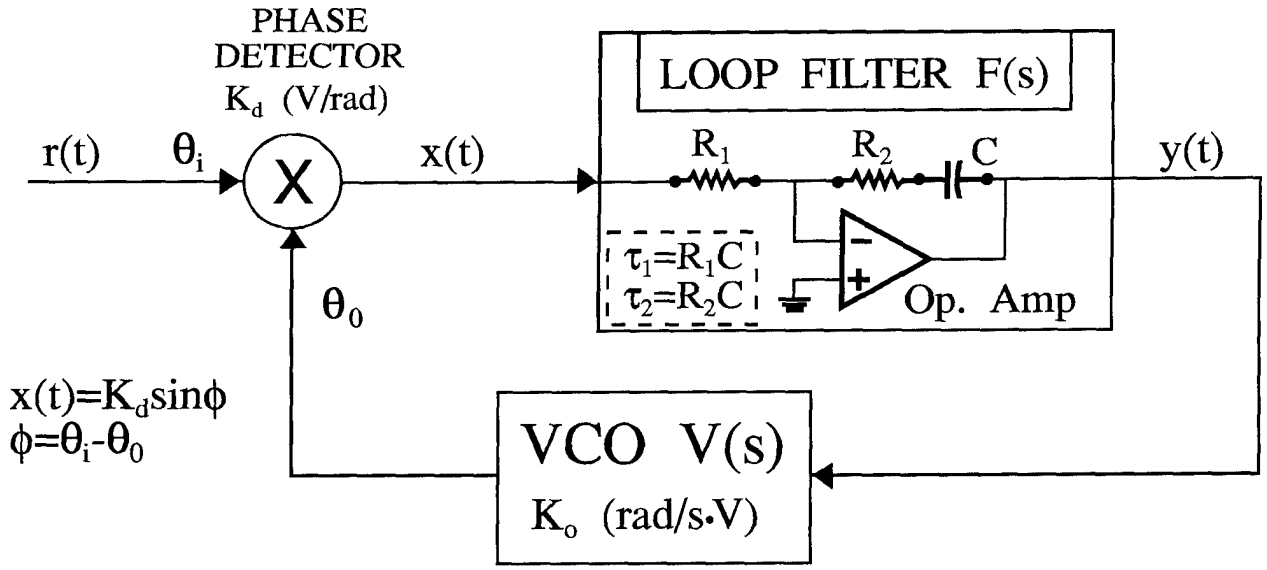


figure 2. General diagram of the Phase-Locked Loop.

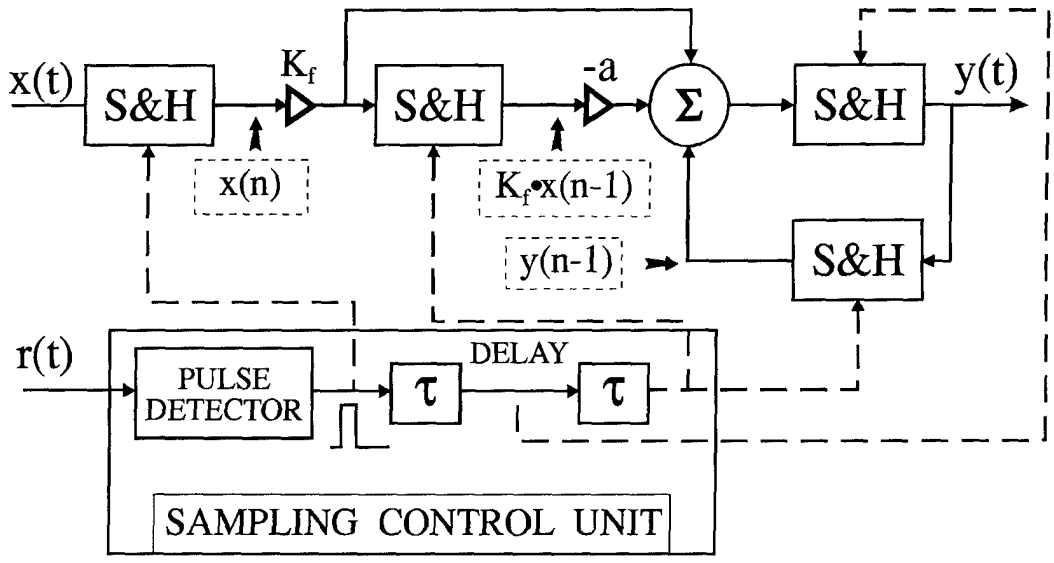


figure 3. Proposed new Loop Filter, based on Sample and Hold circuits.

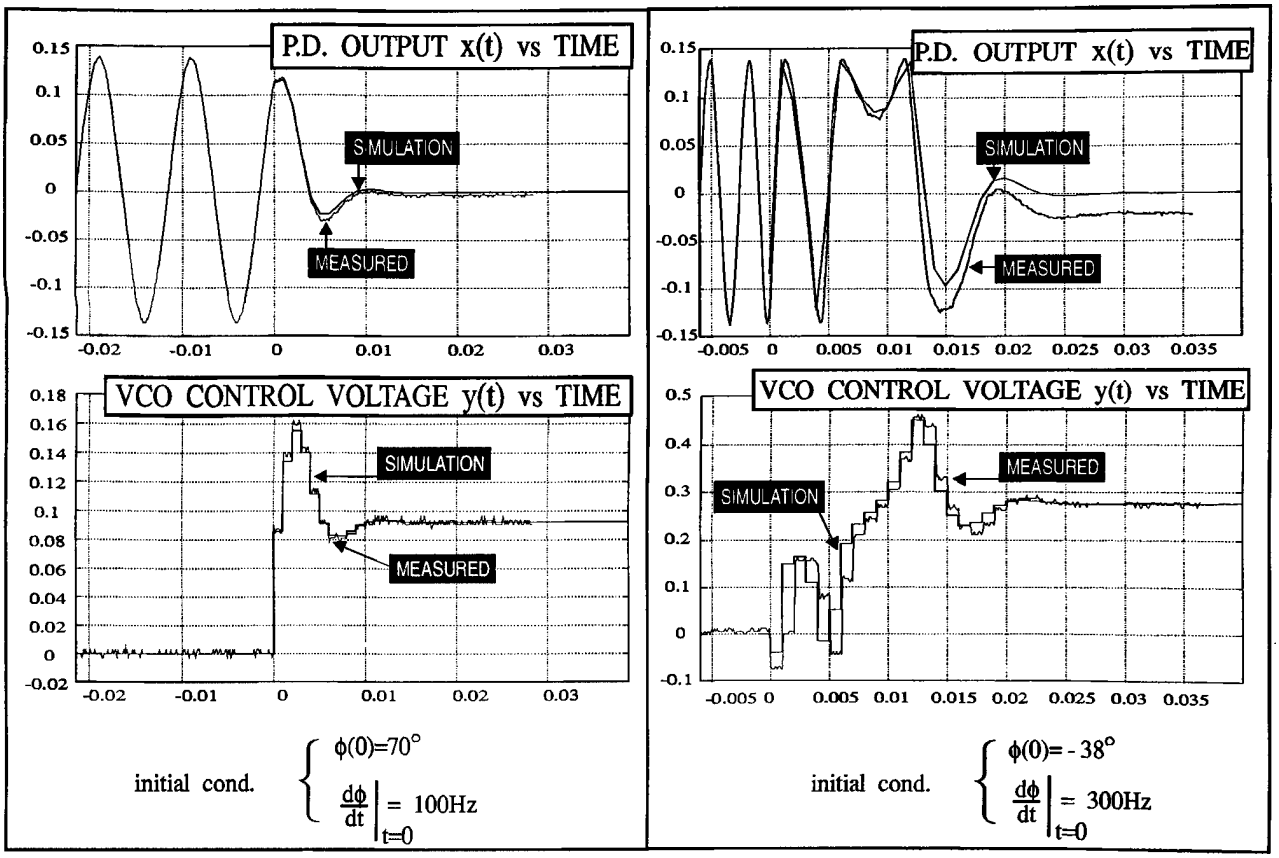


figure 4 and 5. Measurement and simulation results. K_0=6800 rad/s·V, K_d=0.1375 V/rad, K_f=1, a= 0.6, f=30MHz.

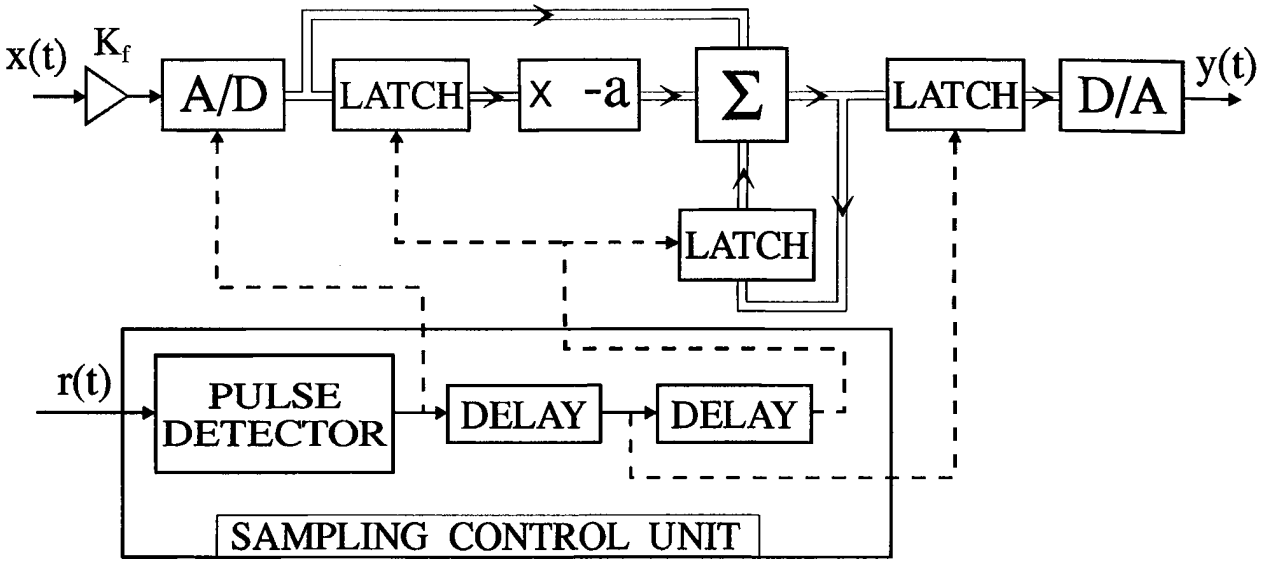


figure 6. Proposed All-Digital Loop Filter.

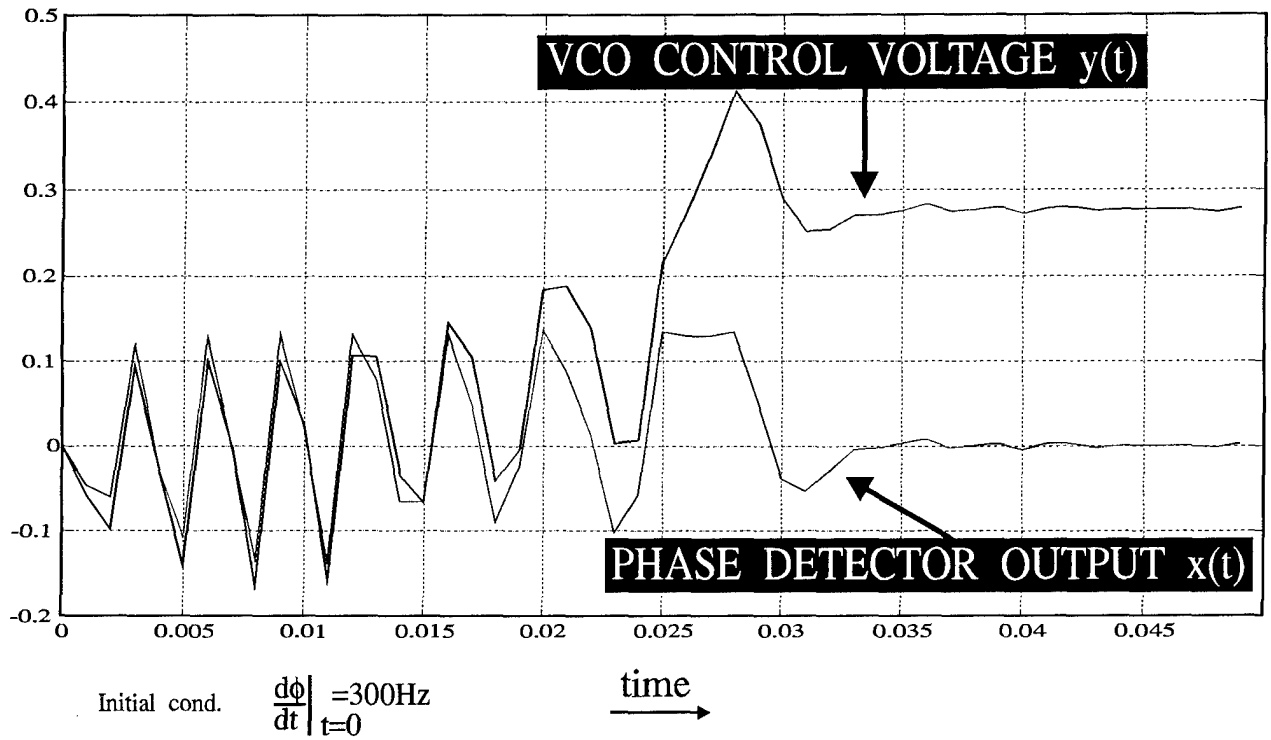


figure 7. System response with the Loop Filter of fig. 6. Same loop parameters.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

IMPROVEMENT OF CHARACTERISTICS OF QUARTZ RESONATOR-THERMOSTATE WITH DIRECT HEATING PIEZOELEMENT

Igor V. Abramson and Anatoly N. Dikidzhi

Omsk Research Institute of Device Making
231 Maslenikov St.
Omsk 644063, Russia

Abstract: The goal of the present work was to improve temperature- frequency stability, power consumption and warm up time of direct heated resonators-thermostate.

A noval design of the resonator on the basis of 10 MHz SC-cut plate combaining constructive featuries of the resonators with direct and indirect heating of the plate has been developed. Mathematical model describing transient processes in the resonator has been obtained. The plate form and the heaters configuration providing frequency vs. transient temperature well behavior have been found.

The noval resonator is shown to has warm-up time less than 15 s, frequency vs. temperature stability better then $3 \cdot 10^{-8}$ and heating current about 5 mA.

Introduction

The quartz resonator-thermostate with the direct heating of piezoelement is a device in which the piezoplate placed in vacuum volume is temperature controlled by means of film heaters and the thermosensor deposited on it's surfaces.

Practically full superposition of the heaters with the thermostatting object - the plate active area - ensures fast warming-up of the resonator and the device small sizes. However, the same constructive feature causes the main disadvantage of the resonator - significant temperature gradients in the plate, caused by the intensive thermal flows from the heaters into environment. Inhomogeneous thermal field provides in the plate the field of temperature stresses, influencing the resonator frequency, configuration of which changes during the resonator warming-up as well as ambient temperature changing. As a result the temperature frequency stability of the resonator is comparatively low $(1 \div 2) \cdot 10^{-7}$ even in a narrow temperature range.

Another disadvantage of such resonator is thermal losses on the power transistor of the temperature control system being placed outside the resonator, thereat the lower the film heaters resistance the greater the losses. So, the contradiction appears between the resonator consumption and the warm-up time. Indeed, in order to reduce the power consumption it is necessary to increase the film heaters resistance but this results in decreasing the start power and hence, raising the warming-up time. For example, to limit the power, consumed by the oscillator at stationary operation by 200-250 mW, the heaters resistance should not be less than 400-500 Ohms that limits the warm up time by 45-60 s.

The object of this paper is significant improving the direct heated resonator parameters, first of all, the temperature-frequency stability, time of warming-up and the consumption using new constructive solutions and optimizing the piezoplate design.

Construction and operation of the noval resonator

The resonator construction

The construction of the new resonator is outlined in Fig 1. It contains the

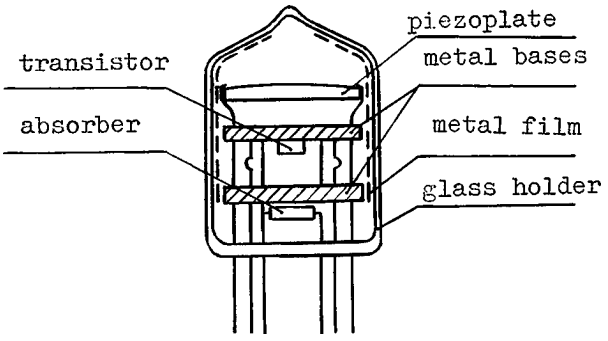


Fig. 1. The construction of the noval resonator-thermostate.

crystal element with exciting electrodes,

heaters and thermosensor deposited on it's surfaces. The plate is placed on the metal base, on the back side of which the power transistor of the temperature control system is attached. The base is mounted in vacuum glass holder, the inner surface of which is coated by thin aluminium film in order to reduce the radiation thermal losses. In order to obtain high vacuum in the volume the gas absorber is placed in it.

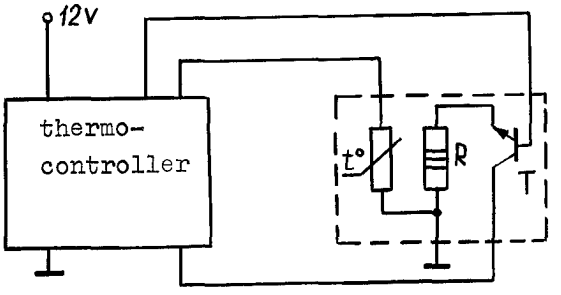


Fig. 2. The thermocontrol system of the oscillator.

The resonator operates in the following way. Just after switching on, the temperature control system (Fig. 2) holds the power transistor in on state, and almost whole thermal power dissipates in the film heaters that provides fast heating of the plate. When the plate obtains the given temperature the temperature-control system cuts the transistor off. As a results of the power redistribution the most part of the power dissipats on the transistor, causing the base intense heating and the less part does in film heaters. So far as the thermal coupling of the thermostatting object - active area of the plate - with the heaters is very strong, then the operating temperature is set very fast - in few seconds . At the same time under the stationary conditions the thermal power dissipated on the plate is very low due to well thermal coupling with the metal base, temperature of which is close to the plate temperature.

To calculate the transient process in the resonator the mathematical model connecting heating current and internal temperature with ambient temperature and constructive featuries of the resonator has been developed.

Mathematical model

As it follows from previous consideration the thermal trancient process in the resonator consists of two stages,

1.) At the first stage the plate is heating by the start power to temperature being equal to the resonator lower turnover temperature T_e.

The equation describing the thermal balance on the plate has the form:

$$\frac{dT_p}{dt} = \frac{U^2}{R_h c_p m_p}\cdot\frac{1}{} - \frac{T_p - T_b}{\rho_p} + \frac{(T_b - T_o)}{\rho_b} \qquad (1)$$

where T_p - temperature of the plate; T_b - temperature of the base; T_o - ambient temperature: c_p, m_p - thermal capacity and mass of the plate; ρ_p - thermal resistance between the plate and the base; ρ_b - thermal resistance between the base and environment, U - input voltage on the resonator (Fig. 2); R_h - resistance of the heaters; t - time.

Neglecting the third item of Eq. 1 due to quite small value of $T_b - T_o$ during the plate warming up and solving the task with the electrical-thermal analogy method we obtaine:

$$T_p = \frac{U^2}{(c_p m_p + c_b m_b) R_h}\left[t + z\frac{c_b m_b}{c_p m_p}(1 - e^{-\frac{t}{z}})\right]; \qquad (2)$$

$$T_b = \frac{U^2}{(c_p m_p + c_b m_b) R_h}\left[t - z\cdot(1 - e^{-\frac{t}{z}})\right]; \qquad (3)$$

where

$$z = \frac{\rho_p c_p m_p c_b m_b}{c_b m_b + c_b m_b}, \qquad (4)$$

c_b, m_b - thermal capacity and mass of the base.

Using Eq. 2, time (t_o) of the plate warming up to T_e can be found. Then from Eq. 3 the base temperature (T_b^o) at the time being equal to t_o can be defined. The recieved values of t_o and T_b^o are used at the second stage as the initial conditions.

2.) At the second stage the plate temperature remains constant, and metal base temperature T_b is rising by following way:

$$\frac{dT_b}{dt} = \frac{p}{c_b m_b} + \frac{T_e - T_b}{\rho_p} - \frac{T_b - T_o}{\rho_b}, \qquad (5)$$

$$P = IU - I^2 R_h, \qquad (6)$$

where I - heating current through the

film heaters and the transistor; P - power dissipated on the plate.

Taking into account that in stationary condition

$$I^2R_h = \frac{T_e - T_b}{\rho_p}, \quad I = \sqrt{\frac{T_e - T_b}{R_h \cdot \rho_p}} . \qquad (7)$$

expression (6) takes the form:

$$P = U \sqrt{\frac{T_e - T_b}{\rho_p \cdot R_h}} - \frac{T_e - T_b}{\rho_p} . \qquad (8)$$

Using Eq. (8) in (5), we have

$$\frac{dT_b}{dt} = \frac{U}{c_p m_p} \cdot \sqrt{\frac{T_e - T_b}{\rho_p \cdot R_h}} - \frac{T_b - T_o}{c_p m_p \rho_t} . \qquad (9)$$

Taking $T_e - T_b = \theta$, we recive the final equation:

$$-c_p m_p \frac{d\theta}{dt} = \frac{U\,\theta^{1/2}}{\rho_p R_h)^{1/2}} + \frac{\theta}{\rho_b} - \frac{T}{\rho_b} , \qquad (10)$$

where $\Delta T = T_e - T_o$.

Solving the differential equition we have:

$$-\frac{t}{c_b m_b} + C = \frac{1}{b} \ln\left|b\theta + \sqrt{\theta} - d\right| - \frac{a}{b\sqrt{a^2+4bd}} \times$$

$$\ln \frac{2b\sqrt{\theta} + a - \sqrt{a^2+4bd}}{2b\sqrt{\theta} + a + \sqrt{a^2+4bd}} , \qquad (11)$$

where $a = \frac{U}{(\rho_p \cdot R_h)^{1/2}}$; $b = \frac{1}{\rho_b}$; $d = \frac{\Delta T}{\rho_b}$.

Constant C we find substituting in Eg. (11) t=0; $\theta_o = T_e - T_b^o$:

$$C = \frac{1}{b} \left\{ \ln\left|b\theta_o + a\sqrt{\theta_o} - d\right| - \frac{a}{\sqrt{a^2+4bd}} \times \right.$$

$$\left. \times \ln\left| \frac{2b\sqrt{\theta_o} + a - \sqrt{a^2+4bd}}{b\sqrt{\theta_o} + a + \sqrt{a^2+4bd}} \right| \right\} . \qquad (12)$$

Then Eg.12 takes the form:

$$t = \frac{c_p m_p}{b} \left\{ C + \frac{a}{\sqrt{a^2+4bd}} \cdot \ln\left|b\theta + a\sqrt{\theta} - d\right| - \right.$$

$$\left. - \frac{a}{\sqrt{a^2+4bd}} \cdot \ln\left| \frac{2b\sqrt{\theta} + a - \sqrt{a^2+4bd}}{2b\sqrt{\theta} + a + \sqrt{a^2+4bd}} \right| \right\} \qquad (13)$$

Equitions (2), (3), (13) with (7) allow to determine the transient processes in the resonator in dependance on ambient temperature, input voltage and constructive parametors of the resonator.

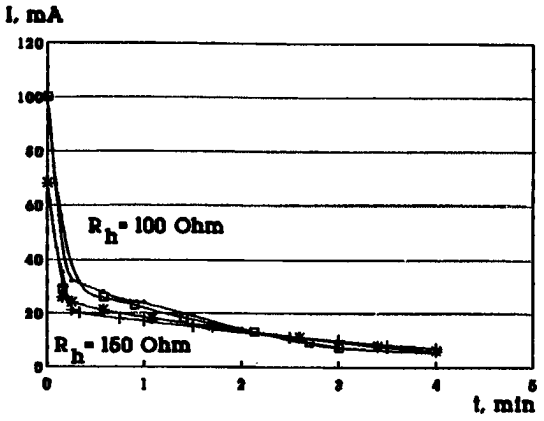


Fig. 3. The heating current vs. time calculeted and empirical curves.

In Fig. 3 the calculated base temperature vs. time at different resistance of the heaters are described with the measured characteristics. As it's seen the theoretical and empirical curves almost coinside, that proves the correctness of the model.

Thus, the recieved model can be used sucessfuly to predict resonator transient and stationary behavior at different ambient conditions and also to determine optimal resonator construction to improve characteristics of the ocsilator.

The resonator characteristics

As it can be seen from Fig.3, when the heaters resistance reduces the initial power proportionally increases but the current at stationary operation remainsalmost permanent. This allows to use low resistance heaters, significantly raising the start heating power.

When the heaters resistance is 100 Ohms, for example, the power dissipated on the plate at the stationary operation is about 5 mW. This is one order lower

than in direct heated resonators of conventional construction, that provides significant increasing the device temperature stability. Fig. 4 shows the frequency vs. ambient temperature. It is seen, that in a wide temperature range the frequency nonstability doesn't exceed 3 10^{-8} that approaches the parameters of indirect heated resonators-thermostate [1].

The power dissipated on the resonator at the stationary operation depends on the ambient temperature (fig. 5) and at - 60°C is 150 mW. It should be noted that given value includesalmost whole the power consumed by the temperature control system that provides low consumption of the oscillator.

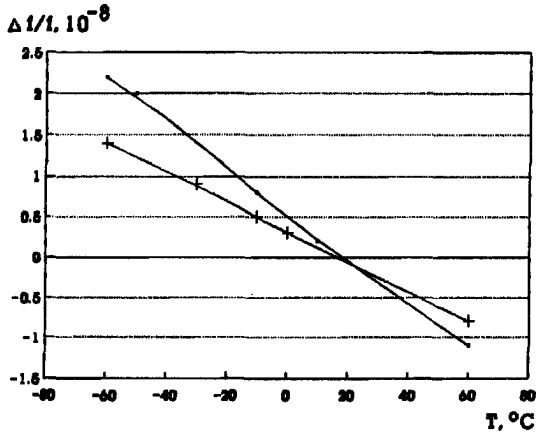


Fig. 4. The resonator frequency vs. ambient temperature.

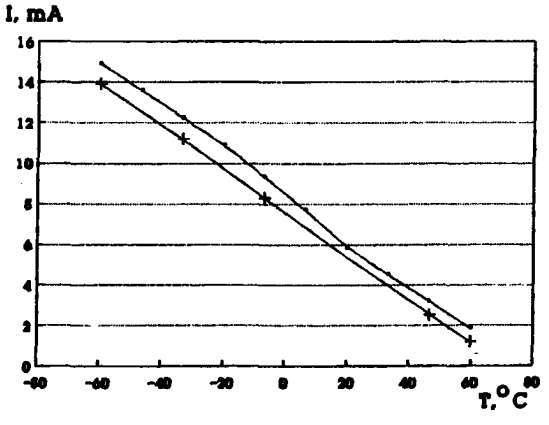


Fig. 5. The heating current vs. ambient temperature.

Thus, in the noval resonator disagreements between the start power, the consumption and frequency vs. temperature stability, limiting improvements in parameters of the direct heated resonators, dissapear. That allows to improve significantly it's basis characteristics.

However, to obtain the best warm up characteristic the new piezoplate design having optimal temperature-frequency and temperature-dynamic characteristics must be developed.

3. Optimization of piezoplate design

One of the main requirements, which is to be fulfilled in order to obtain the resonator high frequency stability, is smooth frequency vs. temperature in the temperature controlled point. The plates of double-rotated SC-cut meet this requirement most fully if the work temperature of the plate is equal to the lower turnover temperature.

The resonator temperature-frequency dependence is known to describes correctly the frequency changing when theambient temperature changes quite slowly so that the thermal equilibrium has time to settle on the whole plate. However, with the increasing the warm up rate the dependence form is distorted.This distortion is explained by appearing in the plate the time-varying temperature gradients, causing the mechanical stresses and resulting the resonator frequency changing. These temperature- dynamic effects occur especially strong in the direct heated reronators where the plate temperature after the switching on rises at a rate of about 10 °/s.

The SC-cut is known as the one having the minimun sensitivity to temperature-dynamic influence, which is 10-100 times less than one of AT-cut [2]. However, to obtain the high temperature-dynamic stability in the resonator- thermostate the optimal plate form and heater configuration must be found. It is connected with that what the SC-cut plates exellent characteristics take place only when the temperature mechanical stresses field in the plate is of axial symmetric. But in direct heated resonators the field depends on the heaters configuration and the plate form.

In order to define the optimal piezoplate construction the local heating characteristic (LHC) of SC-cut plano-convex plates operating at a frequency of 10 MHz on the 3^{d} overton were studied. The LHC were studied in the following way [3]. The piezoelectric plate was excited at an operating resonance and then different points of its surface were impulse heated by the infrared ray heater, measuring there at the induced frequency

shifts. Fig. 6 shows the obtained characteristic.

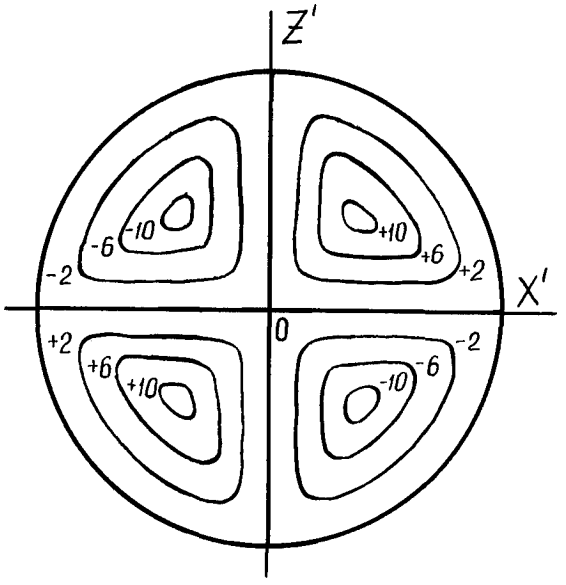


Fig. 6. SC-cut plate characteristic of local heating.

As it can be seen from Fig.6 the LHC has the positive and negotive areasand may be described by symmetry of group $\bar{4}$. The positive and negative fields are separated by two zero sensitivity lines being perpendicular each other and located close to the X', Z' crystal axis. It should be noted that the values of frequency shifts decrease with approaching the plate edges.

Evidently the heaters configuration may be considered as optimal if they provide minimum integral frequency shift:

$$\Delta F/F = \iint_S \Delta F/F(x,z)\cdot dx\cdot dy \ , \qquad (14)$$

where $\Delta F/F(x,z)$ is the LHC value in the point with x,z coordinates; s - the area of the heaters location.

Equition (14) may be fulfilled when the heaters are located along the zero sensitivity lines or, so that positive and negative frequency shifts are mutually compensated. The first condition is not practically fulfulled due to the technological complexity of the heaters relization. The realization of the second condition is theoretically provided by an infinite number of heaters location. However, only simple configurations are acceptable. sinse they can be easily fabricated and are more stable to the plate's form manufacture deviations.

Taking into account all cosidered above, the following SC-cut piezoplate designs have been chosen: the circular plate with the ring heaters coated on its edge; the rectangular plate with the sides parallel to the zero sensitivity lines and heaters, located symmetrically at the edges of the long sides (Fig. 7).

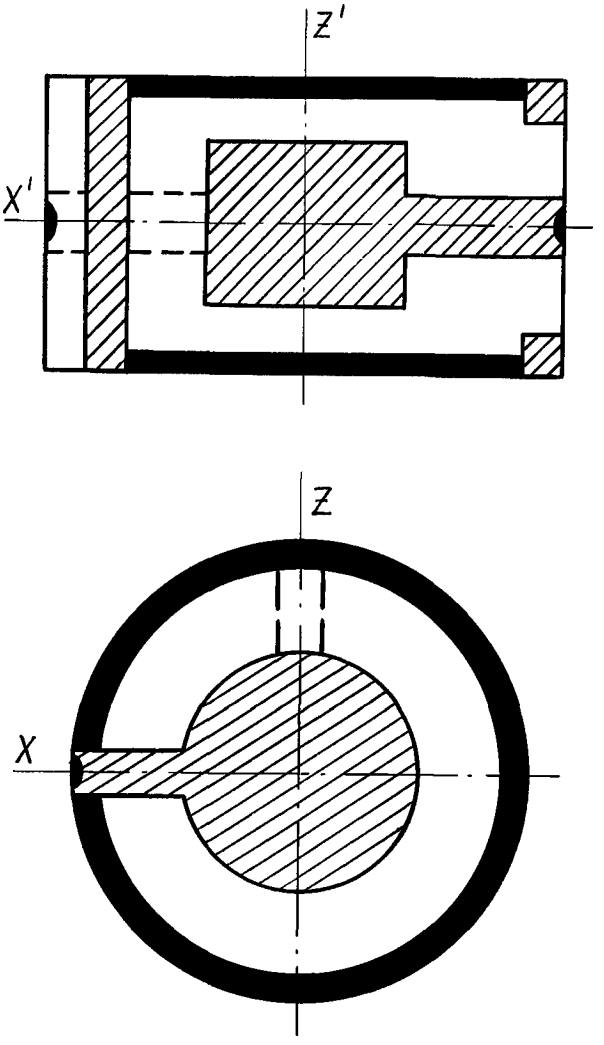


Fig. 7. The optimal piezoplate designs.

In order to examine the correctness of the adopted decisions, the new plates have been tested as a part of the noval resonator. Fig. 8 shows the curves of the resonator frequency setting after switching on when the start power is about 1W (the heatersresistance is 120 Ohms).As it follows from the curves the frequency settling time with accuracy of $1\ 10^{-7}$ doesn't exceed 12 s.

The important condition of receiving the minimum warm up time of the new resonator is the choice of the optimal plate fixture on the heated base. In fact,as can be seen from Fig. 3 the base temperature after the plate fast warming up rises during 4-5 min. This causes slow temperature increasing at the plate fix-

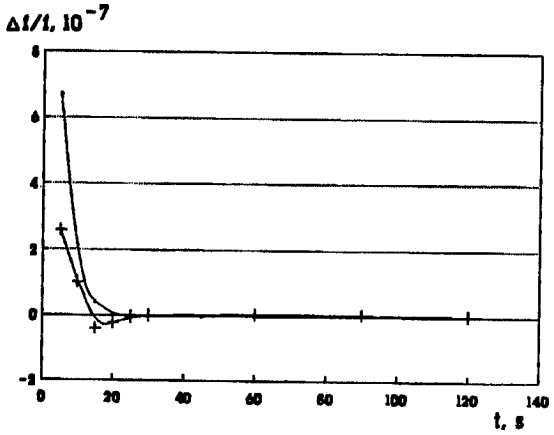


Fig. 8. The resonanor frequency vs. time.

ture places that results in the resonator frequency drift. It is clear that the best fixture places are on the zero sensitivity lines where the crystal has the minimum sensitivity to the heating. In the developed resonators the X'-axis two--point mounting system has been used that provide the warm-up time with accuracy $2 \cdot 10^{-8}$ less than 30 s.

Thus, as a result of performed studies the piezoplate constructions having well frequency vs. trancient plate temperature behavior have been developed, that provided significant improvement of the new resonators warm up time.

Conclusion

The new construction of the resonator-thermostate combining the advantages of the direct heated and indirect heated resonators - high temperature-frequency stability, low power consumption and minimum warm-up time has been developed.

The temperature-dynamic characteristics of the direct heated SC-cut plates were studied. On the basis of the studies the optimal piezoplate and film heaters configurations providing the best warm up characteristics has been found.

On the basis of the noval resonator there has been developed new high stability crystal oscillators having following characteristis: the operating frequency - 10MHz, the temperature range $(-30 \div +60)^{\circ}C$, the total frequency instability - $5 \cdot 10^{-8}$, the warm up time - 15 s, consumed power - 170 mW, aging - $1 \cdot 10^{-7}$ per year, volume - 30 cm^3.

Reference

1. Y.Vorokhovsky, I.Petrosian, B.Krackhlis, A.Volkov "The Futher Development of Internally Heated Quartz Resonators and the Oscillators on Their Basis For Use in Up-to-date Search-and-Rescue and Communication Systems", elsewhere in this Proceedings.

2. J.Kusters, "Transient Thermal Compensation for Quartz Resonators", IEEE Trans. on Sonics and Ultrasonics, 1976, n.4, pp.273-276.

3. V.Terenko, V.Bagaev, "The Local Heating Influence on the Frequency of the Circular Plates with thickness-shear vibrations". Electron. Teckhnika, 1978 Seria 5, Issue 1(26), pp. 34-37.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

THE FURTHER DEVELOPMENT OF INTERNALLY HEATED QUARTZ RESONATORS AND THE OSCILLATORS ON THEIR BASIS FOR USE IN UP-TO-DATE SEARCH-AND-RESCUE AND COMMUNICATION SYSTEMS

Ya. L. Vorohkovsky, Inna G. Petrosian, Boris G. Drakhlis, Alexander A. Volkov

Morion Mfg. Co.
13a Kima Avenue, St.Petersburg 199155, Russia

Abstract: Internally Heated Quartz Resonators (IHQR) open wide opportunities for design of high-stability quartz oscillators with low power consumption, fast warm-up, small size, resistance to severe operating conditions [1,2].

This paper describes further development of IHQRs and oscillators on their basis to meet the requirements of new electronic systems, such as SARSAT/COSPAS search-and-rescue system and a number of communication systems.

During this work performance of IHQRs and quartz oscillators on their basis was significantly improved: long-term frequency stability up to $2x10^{-10}$ per day and $3x10^{-8}$ per year, power consumption of IHQR to 150 mW at -60°C and 65 mW at +25°C, IHQR overall dimensions to 10 cm^3. Transition to "binary code" frequencies 16.384 MHz and 6.4 MHz was made, transfer function of VCXO on IHQR was linearized. This shows that the opportunities for further developments in this field are far from being exhausted.

1.INTRODUCTION

Internally heated quartz resonators (IHQR) represent a new type of piezoelectronic devices that in one design combine functionally and technologically a quartz resonator proper and an oven [1,2]. Here the heater, the temperature sensor, and heat-distributing elements of the inner oven are mounted on the holder together with the piezoelement, resonator vacuum being a natural and most efficient heat insulation of the inner oven, and the resonator enclosure being at the same time the enclosure of the entire device. Accurate temperature stabilization necessary for ensuring frequency stability of the order of $1x10^{-7}$ and better is provided by means of the oven control circuit [3] which is usually placed outside the IHQR case. In the basic IHQR the temperature frequency stability of the order of $1x10^{-6}$ to $3x10^{-7}$ is provided by a self-regulating positive TC thermoresistor heater without oven control circuit [4].

In the case of combination of severe requirements to the quartz oscillator, both temperature compensation and conventional ovenizing by placing the quartz resonator and the most temperature-sensitive parts of oscillator circuitry in a special oven often turn out to be inept. Temperature compensation - due to comparatively low frequency stability (temperature, long-term and short term), conventional ovenizing - due to sharp rise in power consumption, warm-up time, overall dimensions and weight, worsening of resistance to mechanical effects and reliability. Besides, both temperature compensation and traditional ovenizing make manufacture considerably more difficult and quartz oscillators costly.

An effective way of implementation of combination of severe requirements typical of modern radio electronic equipment consists in developing IHQRs and functionally complete quartz oscillators on their basis. IHQRs and quartz oscillators on their basis developed by the authors are already widely used in radio electronic equipment for communication, control, navigation and instrumentation. The experience of their use is very positive and research, development and manufacture are in full swing.

This paper presents some general principles of IHQR design connected with the specific way of providing the required heat conditions and mainly the development of IHQR and quartz oscillators on their basis during the period after the report on these devices on the 45th Annual Symposium on Frequency Control [2]. Research and development of this period were determined by the needs of newly developing radio electronic systems: SARSAT/COSPAS search-and-rescue system and a number of communication systems. As a result, a range of highly efficient industrial types of devices have been developed. In this process some essential problems of general importance for this field have also been solved:

- long-term frequency stability was noticeable improved (up to $2x10^{-10}$ per day and $3x10^{-8}$ per year);
- power consumption of IHQR was reduced (to 150 mW at -60°C and 65 mW at +25°C);
- IHQR overall dimensions were reduced (to 10 cm^3);
- transition to "binary code" frequencies 16.384 MHz and 6.4 MHz was realized;
- transfer function of VCXO on IHQR was linearized.

2.THE MAIN SPECIFICS AND PRINCIPLES OF ENSURING IHQR HEAT CONDITIONS. TEMPERATURE FREQUENCY STABILITY

In their time, in spite of seeming simplicity

0-7803-0476-4/92 $3.00 © 1992 IEEE

of the task, the attempts of practical realization of IHQR came across considerable difficulties and not without reason [5 - 8]. First of all, many technical solutions, even working with precision oven control circuits, revealed low frequency stability. Surprisingly, low power consumption regarded as one of "doubtless" and fundamental advantages of this devices became another serious problem of IHQR construction.

The conducted studies have shown that providing of the necessary thermal condition for IHQR significantly differs from the case of resonator placed in external oven [1]. First of all, the main differences are due to the fact that in IHQR in the steady state the heat flux from the heater runs directly through the holder and the inner oven. As a result, significant temperature differences in IHQR construction are inevitable. The specific nature of destabilizing influence of this temperature differences on the frequency consists in their sharp change with the ambient temperature of IHQR in accordance with the change of IHQR heat losses compensated by the heater.

There are other specifics in providing IHQR thermal regime. For instance, in case of external ovenizing of the resonator the thermal insulation of the piezoelement from the enclosure, i.e. the holder design and the resonator vacuum, does not affect power consumption, while improvement of thermal insulation of the piezoelement from the enclosure delays warm-up. In the case of IHQR the opposite is true.

The following general heat physics principles of design of precision IHQR with low power consumption have been developed [1]:

1. Thermal resistance of the part of the crystal holder on the way of stationary heat flux from the heater to the place of piezoelement installation must be minimal.

Full thermal resistance between the heater and the piezoelement may be comparatively high (50-100°C/W) due to thin posts in which the piezoelement is mounted. Here the design improvement is achieved not by elimination of this thermal resistance but by removal of the stationary heat flux through it.

2. In the steady state the thermal fluxes through the piezoelement itself or at least in its active zone must be reduced to minimum.

3. The thermal resistance to the heat transfer from the inner oven to the IHQR environment must be maximal because the values of the temperature gradients arising in the inner oven are proportional to the dissipated power; besides, reduction of power consumption is one of the main requirement to IHQR.

The first principle is satisfied by mounting the piezoelement and the heater on the same highly heat-conductive base of the crystal holder provided that the contact between the heater and the crystal holder has minimal thermal resistance.

The second principle leads to mounting of the piezoelement inside the highly heat-conductive case (cap) with low degree of outer surface blackness. The IHQR internal chamber consisting of the cap joined to the base with mounted piezoelement and heater is a thermal screen creating a zone with uniform thermal field around the piezoelement.

The first two principles determine the basis for IHQR design and can be generalized: ideally the piezoelement has to be removed from the zone of stationary heat flux from the heater to the IHQR environment, which is schematically shown in Fig. 1.

The third principle stipulates the necessity of limiting to the minimum of all components of heat transfer from the inner oven to the environment (IHQR enclosure): through gas, by radiation, by conduction through mounting supports of the inner oven and current leads. For this purpose the inner oven must be mounted in thoroughly evacuated enclosure with a thermoreflective coating on its surface. The holder and the current leads must have as high thermal resistance as possible.

On the basis of above principles the basic IHQR design presented in [1,2] was developed. The temperature frequency stability of IHQR of this design with AT-cut piezoelements and the static error of oven control circuit 0.10-0.18°C in the ambient temperature range from -60°C to +70°C is $(1\text{-}2)\times10^{-8}$ for 5 MHz at the 5th overtone, about 5×10^{-8} for 10 MHz at the 3rd overtone and 1.5×10^{-7} for 10 MHz at the fundamental frequency.

SC-cut and its modifications allow to reduce the temperature frequency instability of 5 MHz 3rd overtone IHQR to $(0.5\text{-}1)\times10^{-8}$.

3.HEAT LOSSES AND POWER CONSUMPTION OF IHQR IN THE STEADY STATE

The heat flux from the inner oven to the IHQR environment is transferred by residual gas (Q_{μ}), radiation (Q_{ε}) and conduction through mounting supports and current leads (Q_{λ}). The IHQR total heat losses (Q) are equal to its power consumption in the steady state (P) and can be expressed as

$$Q = P = Q_{\mu} + Q_{\varepsilon} + Q_{\lambda}$$

It can be shown [1] that for the residual pressure less than 1×10^{-4} mm Hg the heat transfer trough the residual gas is negligible (about 2 mW for the ambient temperature -60°C).It is important that from the point of view of thermal insulation the requirements to the vacuum inside IHQR are no more severe than for common precision resonators.

The practical opportunities for sharp reduction of IHQR radiating heat losses lie mainly in reduction of the degree of blackness of IHQR enclosure by covering it with the thermoreflective coating, Fig. 2 [1]. The absolute values of the radiation heat losses for the basic design at -60°C are: 73 mW without the thermoreflective

coating, 13 mW for the aluminum coating and 4 mW for the silver coating.

Heat transfer by conduction is the main and the most difficult to reduce component of the heat losses [1]. In general the ways of reduction of heat transfer through the mounting supports (posts of the holder) and leads are clear: 1) making them of the materials with as low heat conductance as possible; 2) reduction of their cross-section and total number; 3) increase of their length. The second and the third ways are contrary to mechanical strength of the construction which limits in principle the possibilities of reduction of IHQR conductive heat losses. However, here also certain opportunities are opened by the choice of materials, improvement of design (including shape) of the heat-removing parts of the holder, optimization of holder design in accordance with required level of mechanical effects. The obtained results can be seen in Fig.3 by comparison of curves 1 and 3. In this case it turned out that the improved versions of the basic IHQR design corresponding to the curves 2 and 3 have rather high mechanical strength successively withstanding the shocks with acceleration of 1000 g.

As it is seen in Fig. 3, the total heat losses (the IHQR power consumption) were almost halved as compared with the initial version of the basic design bringing them to the level of approximately 155 mW at -60°C and 65 mW at +25°C. Such IHQRs - with the power consumption corresponding to curve 3 - are intended mainly for use where the power sources are very much limited. In particular, they are used in oscillators for emergency radio beacons of the SARSAT/COSPAS project (see para.7.3 below).

4.LONG-TERM FREQUENCY STABILITY

As a result of research and advancements of technology carried out recently the long-term frequency stability of IHQR has been significantly improved. In this respect the oscillators on IHQR have closely approached conventional precision ovenized quartz oscillators. By discarding silver-containing contact paste burnt into the piezoelement and changing over to vapor deposited metal bonding pads as well as by improving cleaning and degassing of the structure the long-term frequency stability of precision AT-cut IHQRs at frequencies about 5 MHz at the 5th overtone in the continuous mode of operation has been brought to the level of about 2×10^{-10} per day and 3×10^{-8} for the first year of operation. Typical characteristics of the long-term frequency stability are presented in Fig. 4 and Fig. 5. Before turn on IHQRs were turned off at least for two days. As it is seen in Fig. 4, in the initial period the frequency stability per day is sharply improving with time and is about 5×10^{-9} one hour after turn on, 1×10^{-9} 24 hours after turn on, 5×10^{-10} 5 days after turn on and 2×10^{-10} 30 days after turn on. The long-term frequency stability per month of operation (Fig. 5) is about 1×10^{-8} one day after turn on, 3×10^{-9} 30 days after turn on and about 1×10^{-9} 90 days after turn on.

5.WARM-UP TIME

Advancements in the IHQR technology introduced to raise the long-term frequency stability together with the increase of the structure thermal resistance to heat transfer to the environment have also led to acceleration of the warm-up process. As a result, at -60°C the warm-up time for AT-cut IHQRs is: with accuracy 1×10^{-7}, no more than 7 min; with accuracy 5×10^{-8}, no more than 8 min; under normal conditions 1.5 times less. Fast warm-up of IHQR ensured successful realization of a quartz oscillator for radio beacons of the SARSAT/ COSPAS project (see para. 7.3, the characteristics of the frequency drift, Fig. 9).

6.SIMPLE SELF-REGULATING IHQR

At the level of frequency stability of the order of 1×10^{-6} simple IHQRs with a self-regulating positive TC thermoresistor heater [1] often successfully compete with temperature compensated quartz oscillators. This is due to low cost and very simple design of quartz oscillators on such IHQRs.

Further extension of use of simple IHQR with temperature frequency stability about 1×10^{-6} necessitated transition to "binary code" frequency 6.4 MHz accompanied by significant improvement of the design of the holder, inner oven and piezoelement so that power consumption under normal conditions has been reduced to about 85 mW (against former 150 mW), warm-up time to 1.5 min (against former 3 min) and maximum diameter of the enclosure to 19 mm (against former 22,5 mm).

7.QUARTZ OSCILLATORS ON IHQRs

7.1. Design

Circuit design deliberations as well as economical considerations do not allow to design precision quartz oscillators on IHQRs in the form of a single or several crystals of semiconductor IC. The developed design of the quartz oscillator (Fig. 6) is simple at sight but exhibits high stiffness which in the transverse and vertical directions is provided by frames 4 and in the longitudinal direction (along the IHQR axis) - by printed circuit boards 6; dimensions of such a quartz oscillator are 63*53*29.5 mm (98.5 cm^3). The significant reduction of volume of quartz oscillator on IHQR is achieved by the use of a sandwich construction, arranged along the IHQR axis, with printed circuit boards placed orthogonally to the IHQR axis (Fig. 7); here the dimensions of the oscillator are 81.5*26.5*26.5 mm. Free inner space of the quartz oscillator is filled with plastic foam providing reliable fixing of elements of the construction, effective protection of the quartz oscillator from mechanical and environmental effects, high reliability.

The presented designs (Fig. 6 and 7) have became basic; using them a range of industrial

types of precision quartz oscillators have been developed.

7.2.Temperature Frequency Stability

In developed quartz oscillators inside IHQR, in addition to the piezoelement, the varicap diode of the electronic frequency adjuster can also be ovenized [1,2]. The remaining elements of the circuit are not ovenized. Practically realizable temperature frequency stability of such quartz oscillators over the temperature range from -60°C to +70°C taking into account the IHQR stability discussed earlier (Section 2) is better than $2x10^{-8}$ for SC-cut and its modifications and better than $3x10^{-8}$ for AT-cut.

7.3.Quartz Oscillators for Emergency Radio Beacons of SARSAT/COSPAS Project

High frequency stability (including short-term stability) under significant and fast changes of working temperature inherent in quartz oscillators on IHQRs combined with low power consumption, fast warm-up and compact sizes has allowed to develop the specialized quartz oscillator for radio beacons of the satellite search-and-rescue system SARSAT/COSPAS. A structural diagram of this quartz oscillator is presented in Fig. 8, the mechanical design corresponds to that discussed earlier (para. 1.7, Fig. 7). These oscillators employ 5.075312 or 5.012654 MHz, 5th overtone AT-cut IHQRs, output signal - standard LSTTL type.

Here the main quartz oscillator parameters are mid-term frequency instability (drift) for 15 min and residual noise (root mean square error of the operating frequency relative to the least square straight line fit to the measurement points). The typical performance data for this parameters versus time after turn on for the most severe working conditions (at -40°C) are shown in Fig.9. As it should be expected the frequency drift rapidly decreases with time. After 10 min of quartz oscillator operation the residual noise is practically not dependent on time. Short-term frequency stability (Allan variance) for 100 ms 15 min after turn on does not exceed $5x10^{-10}$.

The necessity of ensuring the required level of quartz oscillator performance in conditions of fast changes of the working temperature in a wide range imposes rather strict limitations on the temperature frequency stability which for the quartz oscillator under consideration is better than $\pm 5x10^{-8}$ over the temperature range from -40°C to +60°C. Due to effective system of oven control and thermal insulation of IHQR, the quartz oscillator practically maintains the output signal parameters in conditions of working temperature gradient of 9°C/hour and also 15 min after the direct thermal shock of 30°C made at the moment of turning on the oscillator.

Because of the very limited power resources of the beacon battery supply the reduction of the quartz oscillator's power consumption was one of the important problems. The characteristics of the IHQR achieved in this respect were presented in Fig. 3 (curve 3). Typical values of current consumed by the quartz oscillator as a whole are: about 30 mA at -40°C, 25 mA at 0°C and 20 mA at +25°C. Taking into account different variants of the beacon battery supply, the supply voltage may change from 9 to 18 V, and the oscillator ensures the required parameters at 3 V intermittent drops ("dips") of the supply voltage without transients or frequency drift. The latter is ensured mainly by the IHQR oven control circuit design in accordance with that proposed in [3].

It should be noted that in addition to low power consumption the developed oscillators are capable of withstanding severe mechanical effects, in particular, the shocks with acceleration of 1000 g, which is essential for a number of applications.

The results obtained during development of oscillators for beacons of the SARSAT/COSPAS project can be effectively and quickly used for creation of a family of specialized quartz oscillators for radio beacons of satellite navigational systems mounted on various mobile objects where, like in the SARSAT/COSPAS system, the desired coordinates of an object are determined by results of satellite measurements of the Doppler frequency shift of the radio signal radiated by the object.

7.4.New Quartz Oscillators for Communication Systems

Here the main efforts have been concentrated in two areas:

A) precision voltage controlled quartz oscillator at "binary code" frequency of 16.384 MHz with the linear transfer function of the frequency adjuster;

B) high-stability compact oscillator at "binary code" frequency of 6.4 MHz with fast warm-up.

7.4.A. The structural diagram of the precision voltage controlled quartz oscillator at 16.384 MHz is presented in Fig.10, the design corresponding to the one shown in Fig.6. The output signal is ALS TTL type.

Here the long-term frequency stability is significantly improved - to the level of about $2x10^{-10}$/day and $3x10^{-8}$/year which was ensured by the advancements of IHQR technology and discussed earlier (Section 4). Typical characteristics of long-term stability correspond to those presented in Fig.4 and Fig.5.

The difficult problem in development of a voltage controlled quartz oscillator on IHQR was to ensure the high temperature frequency stability and the linearity of the transfer function of the frequency adjuster within the range of operating temperatures. The reason is that the linearization of the transfer function of the frequency adjuster requires an additional circuit of fine frequency adjuster which is not ovenized. In this case, as before [1,2], in IHQR besides the piezoelement

only one varicap diode is ovenized - that of the coarse frequency adjuster.

By optimization of the circuit design and careful selection of the fine adjuster parts, using, in particular, the mutual compensation of the introduced components of temperature frequency instability, the mentioned problem was solved to a considerable degree. The achieved temperature frequency stability of the oscillator is better than $\pm 1 \times 10^{-8}$ in the temperature range from +5°C to +45°C and $\pm 2 \times 10^{-8}$ in temperature range from -10°C to +60°C and the fine frequency adjuster transfer function slope is within $(7 \pm 1.5) \times 10^{-8}$/V. The achieved linearity of transfer function of fine frequency adjuster is at least not worse than that of the precision voltage controlled quartz oscillators with conventional ovenizing where the fine frequency adjuster is inside the oven.

The quartz oscillator frequency instability due to supply voltage changes 12 V ± 10% is no more than $\pm 1 \times 10^{-9}$. The quartz oscillator power consumption is about 0.4 W at -10°C and 0.3 W at +25°C. The warm-up time is no more than 5 min with accuracy 1×10^{-7}, no more than 10 min with accuracy 2×10^{-8}.

7.4.B. The structural diagram of high-stability compact quartz oscillator at 6.4 MHz with fast warm-up is presented in Fig.11, the design corresponding to that shown in Fig.7. The output signal is standard LSTTL type. The quartz oscillator is based on IHQR with a double rotated cut representing a modification of the SC-cut. The operating frequency of crystal plate is 6.4 MHz at the 3rd overtone.

The design of this quartz oscillator is mostly based on the results obtained during development of quartz oscillator presented in [2], last column of Table 1. A number of measures, however, were taken to decrease power consumption and improve frequency stability, including corresponding improvements of the IHQR. The power consumption of the quartz oscillator from 12 V supply at - 40°C is about 350 mW, from 5 V supply - about 25 mW (independent of the operating temperature). The temperature frequency stability in the temperature range from -40°C to +60°C is better than $\pm 2.5 \times 10^{-8}$. The frequency instability due to ±10% changes of the supply voltage is no more than $\pm 1 \times 10^{-9}$. Warm-up time at -40°C with accuracy 5×10^{-8} is no more than 3 min.

8.INDUSTRIAL TYPES OF DEVELOPED DEVICES

On the basis of the results presented above the following new devices have been developed:

1) The quartz oscillators of ГК40-ТС type intended for emergency radio beacons of the SARSAT/COSPAS system (see para. 7.3.). With a sufficient margin they satisfy all conventional requirements of the SARSAT/COSPAS project. Beacons with ГК40-ТС quartz oscillators were granted the international certificate of the SARSAT/COSPAS Committee in London.

2) Precision voltage controlled quartz oscillators of ГК41-УИТС type at 16.384 MHz (see para. 7.4.A).

3) High-stability compact quartz oscillator of ГК50-ТС type at 6.4 MHz with fast warm-up (see para.7.4.B).

4) Simple self-regulating IHQR of РК357СВ type with stability about 1x10-6 (see Section 6). Characteristics of a number of other, earlier, IHQRs and quartz oscillators on them are presented in [1,2].

9.CONCLUSION

The developed IHQRs and quartz oscillators on their basis are already widely used in different important radio electronic systems.They are characterized by combination of high frequency stability with low power consumption,fast warm-up, small size and resistance to severe operating conditions together with a relatively simple design and low cost.

The paper, representing mainly the devices developed after the previous, 45th, Symposium on Frequency Control, shows that the opportunities for further developments in this field are far from being exhausted.

References

[1] Я.Л. Вороховский. Резонаторы-термостаты и *пьезоэлектронные устройства на их основе.* - В кн.: Пьезоэлектрические резонаторы. Справочник. Москва: Радио и связь, 1992. Стр. 318 - 386.

[2] Ya.L. Vorokhovsky and B.G. Drakhlis, "High-stability Quartz Oscillators on Internally Heated Quartz Resonators with AT and SC Cuts," in Proc. of the 45th Annual Symposium on Frequency Control, USA, 1991, pp. 447-451.

[3] Я.Л. Вороховский, Б.Г. Драхлис "Выбор и построение терморегулятора для резонатора-термостата прецизионного кварцевого генератора," Электронная техника. Сер. Радиодетали и радиокомпоненты, 1987, вып.2 (67), с.67 - 70.

[4] Я.Л. Вороховский, В.Б. Грузиненко, И.Г. Петросян "Кварцевый резонатор-термостат с саморегулирующимся позисторным нагревателем," Электронная техника. Сер. Радиодетали и радиокомпоненты, 1977, вып.3 (22), с. 18- 29.

[5] F.G. Tinta, A.S. Matistic, and G.A. Lagasse, "The Direct Temperature Control of Quartz Crystals in Evacuated Enclosures," in Proc. of the 24th Annual Symposium on Frequency Control, USA, 1970, pp. 157-167.

[6] H.M. Greenhouse, R.L. McGill, and D.P. Clark, "A Fast Warm-up Quartz Crystal Oscillator," in Proc. of the 27th Annual Symposium on Frequency Control, USA, 1973, pp.199-217.

[7] H.W. Jackson, "Tactical Miniature Crystal Oscillator," in Proc. of the 34th Annual Symposium on Frequency Control, USA, 1980, pp.449-456.

[8] A. Stali, M. Brunet, " An Ultra Stable Crystal Oscillator for Beacons Program 'SARSAT', " in Proc. of the 37th Annual Symposium on Frequency Control, USA, 1983, pp.454-458.

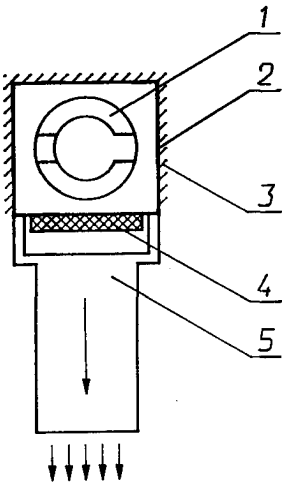


Fig.1. Principle of IHQR design - removal of piezo - element from the zone of passage of the heat flux from the heater to the environment:
1 - piezoelement
2 - isotermic chamber - heat screen
3 - heater and temperature sensor
4 - heat insulation
5 - holder of oven. Arrows show heat fluxes to the environment.

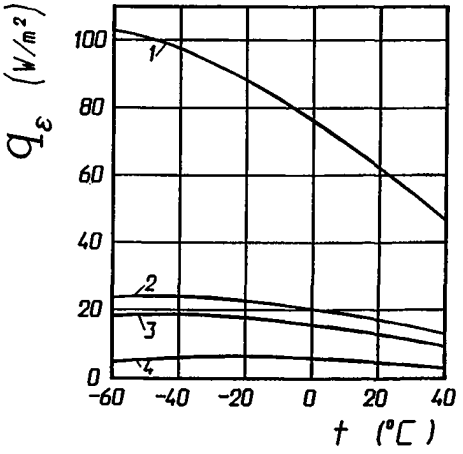


Fig.2. Density of radiant heat flux $/q_\varepsilon/$ from the surface of the IHQR inner oven in the range of the ambient temperatures $/t/$:
1 - glass tube without coat
2 - coat of nickel
3 - coat of aluminum
4 - coat of silver.

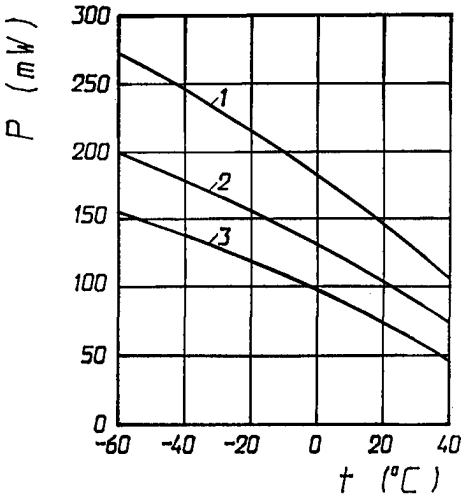


Fig.3. Typical IHQR power consumption $/P/$ in the range of ambient temperatures $/t/$:
1 - initial version of the basic design
2,3 - consequently improved versions with reduced conduction through the holder /and leads/.

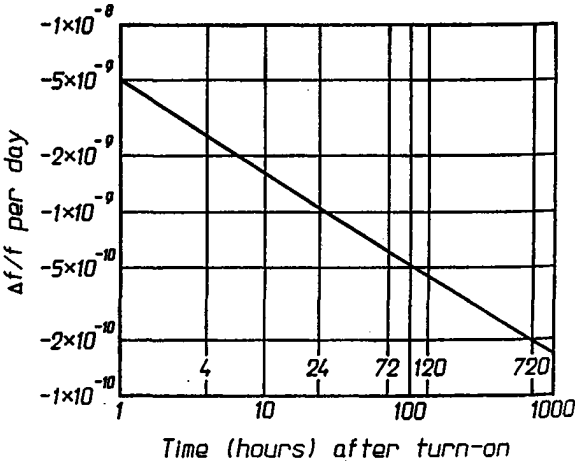


Fig.4. Typical characteristic of frequency stability per day as function of the time after turn-on /improved IHQR with AT-cut, 5.46 MHz, 5th overtone/.

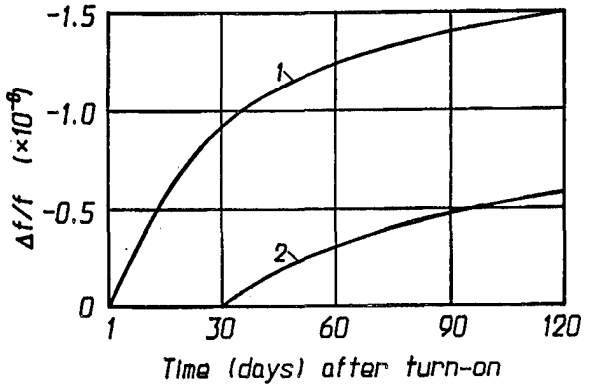


Fig.5. Typical characteristic of long-term frequncy stability /improved IHQR with AT-cut, 5.46 MHz, 5th overtone/:
1 - after 24 h of continuous operation
2 - after 30 days of continuous operation.

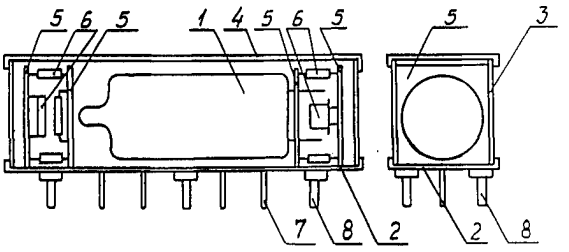


Fig.7. Miniature version of mechanical design of quartz oscillator on IHQR:
1 - IHQR
2 - base
3 - case
4 - cover
5 - printed circuit board
6 - radio elements
7 - lead
8 - stud; vacant inner space is filled with plastic foam.

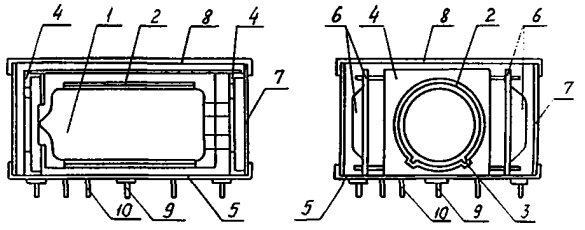


Fig.6. Mechanical design of quartz oscillator on IHQR:
1 - IHQR
2 - metal screen
3 - temperature stress compensator
4 - frame
5 - base
6 - printed circuit boards with radio elements
7 - shell
8 - cover
9 - fastening screw
10 - lead; vacant inner space is filled with plastic foam.

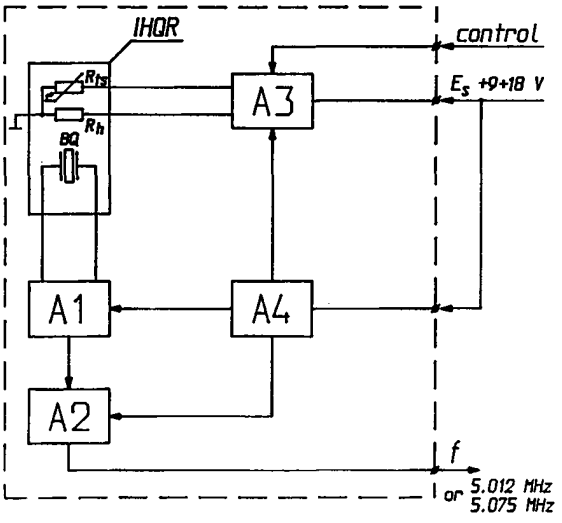


Fig.8. Structural diagram of quartz oscillator for radio beacons of SARSAT/COSPAS project:
A1 - active part of oscillator
A2 - former of TTL-signal
A3 - proportional oven control circuit
A4 - reference voltage source.

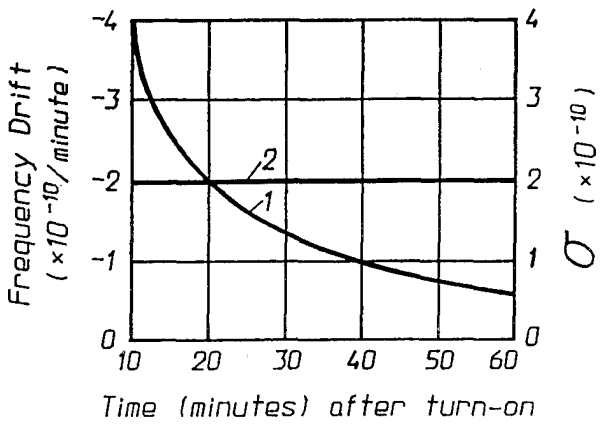


Fig.9. Typical characteristics of mid-term frequency instability /drift/ in 15 min and residual noise /root mean square error of the operating frequency relative to the least square straight line fit to the measurement points/:
1 - frequency drift
2 - residual noise /σ/.

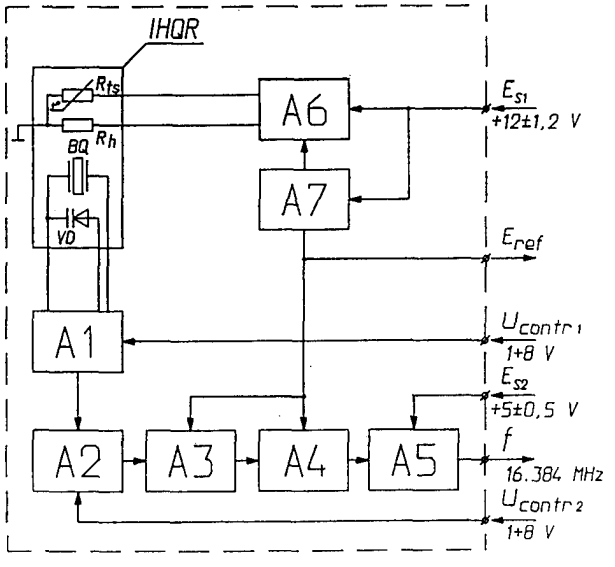


Fig.10. Structural diagram of precision VCXO at 16.384 MHz:
A1 - coarse electronic frequency adjuster
A2 - fine /linear/ electronic frequency adjuster
A3 - active part of oscillator
A4 - frequency multiplier by 3
A5 - former of TTL-signal
A6 - proportional oven control circuit
A7 - reference voltage source.

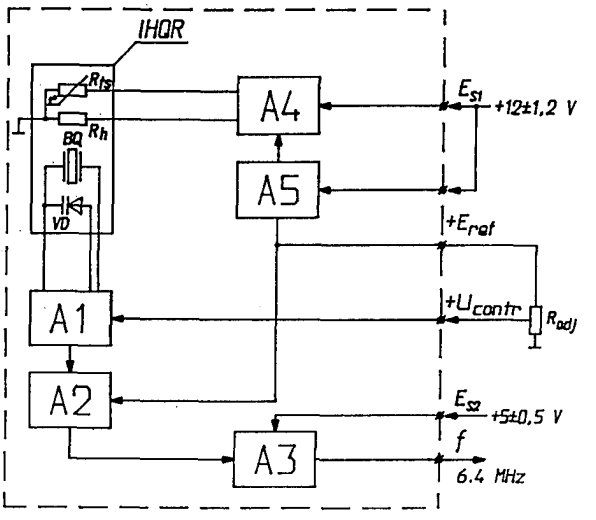


Fig.11. Structural diagram of high-stability compact oscillator at 6.4 MHz with fast warm-up:
A1 - electronic frequency adjuster
A2 - active part of oscillator
A3 - former of TTL-signal
A4 - proportional oven control circuit
A5 - reference voltage source.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PIEZOELECTRIC SENSORS AND ACTUATORS: SMART MATERIALS

Robert E. Newnham

Materials Research Laboratory
The Pennsylvania State University
University Park, PA 16802

Abstract: "Smart" materials have the ability to perform both sensing and actuating functions. Passively smart materials respond to external change in a useful manner without assistance, while actively smart materials have a feedback loop which allows them to both recognize the change and initiate an appropriate response through an actuator circuit.

One of the techniques used to impart intelligence into materials is "Biomimetics," the imitation of biological functions in engineering materials. Composite ferroelectrics fashioned after the lateral line and swim bladders of fish are used to illustrate the idea. "Very smart" materials, in addition to sensing and actuating, have the ability to "learn" by altering their property coefficients in response to the environment. Field-induced changes in the nonlinear properties of relaxor ferroelectrics and soft rubber are utilized to construct tunable transducers. Integration of these different technologies into compact, multifunction packages is the ultimate goal of research in the area of smart materials.

Introduction

It has been said that life itself is motion, from the single cell to the most complex organism: man. This motion, in the form of mobility, change, and adaptation, is what elevates living beings above the lifeless forms [1]. This concept of creating a higher form of materials and structures by providing the necessary life functions of sensing, actuating, control, and intelligence to those materials is the motivation for studying smart materials.

Smart materials are part of smart systems - functional materials for a variety of engineering applications. Smart medical systems for the treatment of diabetes with blood sugar sensors and insulin delivery pumps. Smart airplane wings that achieve greater fuel efficiency by altering their shape in response to air pressure and flying speed. Smart toilets that analyze urine as an early warning system for health problems. Smart structures in outer space incorporating vibration cancellation systems that compensate for the absence of gravity and prevent metal fatigue. Smart toys like "Altered Beast" where one is awakened from the dead and must learn to survive in the hostile environment of a different age. Smart houses with electrochromic windows that control the flow of heat and light in response to weather changes and human activity. Smart tennis rackets with rapid internal adjustments for overhead smashes and delicate drop shots. Smart muscle implants made from rubbery gels that respond to electric fields, and smart dental braces made from shape memory alloys. Smart hulls and propulsion systems for navy ships and submarines that detect flow noise, remove turbulence, and prevent detection. Smart water purification systems that sense and remove noxious pollutants. A number of smart systems have already been developed for automobiles, but there are many more to come. In a recent newspaper cartoon, Blondie and Dagwood encountered a smart automobile that drives itself back to the finance company when the owner misses a payment!

In this article the idea of "smartness" in a material is discussed, along with a number of examples involving electroceramic components. Some of these smart ceramics are in production, while others have great potential but are thus far limited to laboratory investigations.

To begin our discussion, we first define "smart" to set the limits for classifying smart materials.

0-7803-0476-4/92 $3.00 © 1992 IEEE

How Smart is Smart?

The short answer is "not very." Webster's dictionary gives several definitions for the word SMART, including "alert, clever, capable," "stylish," and "to feel mental distress or irritation." All three definitions are appropriate for the currently fashionable subject, "smart materials." They are "stylish," they are - in some cases - "clever," and it does cause some of us "mental distress" to think that a ceramic might somehow possess intelligence, even in rudimentary form.

There are many words in the English language denoting various degrees of intelligence. Beginning at the bottom, an intelligence scale might look like this: stupid - dumb - foolish - trivial - sensible - smart - clever - intelligent - wise. Many modern day materials have been cleverly designed to carry out useful functions and, it seems to us anyway, that we are justified in calling them "smart." They are decidedly better than "sensible" materials, but calling them "intelligent" seems rather presumptuous and self-serving. Perhaps in the future - when we are able to integrate information - processing and feedback circuitry into our sensor and actuator materials - perhaps then we will be justified in calling our materials "intelligent." As pointed out later, such a time is not far off.

To clarify the concept of smart materials, we describe a few examples of passive and active smartness.

Passive Smartness

A passively smart material has the ability to respond to environmental conditions in a useful manner. A passively smart material differs from an actively smart material in that there are no external fields or forces or feedback systems used to enhance its behavior. The "S" words in Table I summarize some of the meanings of passive smartness.

Table I. Some Attributes of Passive Smartness

Selectivity
Self Diagnosis
Self Tuning
Sensitivity
Shapeability
Self Recovery
Simplicity
Self Repair
Stability and Multistability
Stand-by Phenomena
Survivability
Switchability

Many passively smart materials incorporate self-repair mechanisms or stand-by phenomena which enable the material to withstand sudden changes in the surroundings. The crack-arresting mechanisms in partially stabilized zirconia are a good example. Here the tetragonal-monoclinic phase change accompanied by ferroelastic twin wall motion are the stand-by phenomena capable of generating compressive stresses at the crack tip. In a similar way, toughness can be improved by fiber pull-out or by multiple crack-branching as in the structural composites used in aircraft, or in machinable glass-ceramics.

Ceramic varistors and PTC thermistors are also passively smart materials. When struck by high-voltage lightning, a zinc oxide varistor loses most of its electrical resistance and the current is by-passed to ground. The resistance change is reversible and acts as a stand-by protection phenomenon. Varistors also have a self-repair mechanism in which its highly nonlinear I-V relationship can be restored by repeated application of voltage pulses. Barium titanate PTC thermistors show a very large increase in electrical resistance at the ferroelectric phase transformation near 130°C. The jump in resistance enables the thermistor to arrest current surges, again acting as a protection element. The R(V) behavior of the varistor and the R(T) behavior of the PTC thermistor are both highly nonlinear effects which act as standby protection phenomena, and make the ceramics smart in a passive mode.

Active Smartness

A smart ceramic can also be defined with reference to sensing and actuating functions, in

automatically senses the amount of rain falling and adjusts the windshield wipers to the optimum speed. [4]

Controlled compliance with piezoelectric ceramics is utilized in Toyota's piezoTEMS (Toyota Electronic Modulated Suspension), a system which has been developed to improve the drivability and stability of the automobile, and at the same time enhance passenger comfort. [5] The TEMS is basically a road stability sensor and shock adjustor, which detects bumps, dips, rough pavement, and sudden lurches by the vehicle, then rapidly adjusts the shock absorbers to apply a softer or firmer damping force, depending on what is necessary to minimize discomfort while maintaining control of the vehicle. The shock absorbers are continuously readjusted as the road conditions change so that rocking or wobbling on soft shocks is eliminated.

The TEMS road surface sensor consists of a five-layer piezoelectric sensor mounted on the piston rod of the shock absorber. When a bump in the road is encountered, the resulting stress applied to the sensor produces a voltage which is fed into an electronic control unit that amplifies the signal and supplies a high voltage to the piezoelectric actuator. The 88-layer PZT actuator produces a 50μm displacement on the oil system which is hydraulically enlarged to two millimeters, enough to change the damping force from firm to soft; the entire process takes only about twenty milliseconds (not even enough time to slam on the brakes!) Also figured into the actuator output are the vehicle speed and the driver's preference for a generally softer (American) or firmer (European) ride.

Alternatively, it is possible to damp stresses and vibrations without the need for a sensor-actuator feedback loop; materials which can perform this function are called passive damping materials. In a piezoelectric passive damper, a piezoelectric ceramic is connected in parallel with a properly matched resistor. The external stress creates a polarization in the piezoelectric, which induces a current in the resistor, leading to energy dissipation. A high piezoelectric coupling coefficient is required to induce the maximum voltage and energy dissipation.[6]

Actuator Materials

There are many approaches to controlling vibration and structural deformation. Actuation strain can be controlled by piezoelectric materials, [7] electrostrictive materials, [8] magnetostrictive materials, [9] shape-memory metal alloys, [10] and thermally-controllable materials. [11] Utilizing a system with distributed actuators, it is possible to design structures with intrinsic vibration and shape control capabilities. Among the most important actuator materials are shape memory metals and ceramics. The shape memory effect is exhibited by alloys which undergo thermoelastic martensite transformations. This is a first order displacive transformation in which a body centered cubic metal transforms by shear on cooling to a martensitic phase. When deformed in the martensitic low temperature phase, shape memory alloys will recover this deformation and return to the original shape when heated to a temperature where the martensite reverts back to the parent body-centered cubic structure. Unlike most ferroelectric and ferromagnetic transitions, the shape memory transformation has a large hysteresis which can be troublesome in practice.

Alloys exhibiting the shape memory effect fall into two general classes: non-ferrous and ferrous. Non-ferrous alloys currently in commercial use are Ni-Ti, Cu-Zn-Al, and Cu-Ni-Al. Ferrous shape memory alloys under development include Fe-Pt, Fe-Ni-C and Fe-Ni-Co-Ti.

Non-ferrous shape memory alloys of nickel-titanium alloy (Nitinol) have been developed by Goodyear Aerospace Corporation for spacecraft antennae. [10] A wire hemisphere of the material is crumpled into a tight ball, less than five centimeters across. When heated above 77°C, the ball opens up into its original shape--a fully formed antenna. Although it has seldom been used in service, this antenna demonstrates the magnitude of deformation and reformation possible in shape memory alloys.

While shape memory alloys are more like a solution looking for a problem, it has been suggested that transient and steady state vibration control can be accomplished with hybrid structures in which the shape memory alloy is embedded inside the material [12].

Some ceramic materials also possess a sizeable shape-memory effect; of particular interest are materials which are simultaneously ferroelectric and ferroelastic. Their ferroelasticity ensures that recoverable spontaneous strain is available for contributing to the shape memory effect, and the ferroelectricity implies that their spontaneous strain can be manipulated not only by mechanical forces but also by electric fields.

Shape-memory has been demonstrated in PLZT ceramics, an important ferroelectric-ferroelastic because of the tremendous potential for applications due to the formation of microdomains smaller than the wavelength of light. In one experiment, a 6.5/65/35 PLZT helix was heated to 200°C, well above the transition temperature for recovery, T_F (= T_C), then mechanically loaded and cooled to 38°C (well below T_F) - the "brittle" PLZT helix was deformed by 30% after the load was removed. Upon heating to 180°C, above T_F, the helix transformed back to its original shape, dramatically demonstrating the shape memory effect in brittle ceramics. [13]

Researchers at Sophia University in Tokyo have created a multilayer shape memory actuator with a (Pb, Nb)(Zr, Sn, Ti)O_3 ceramic. Twenty rectangular plates of the ceramic were stacked in an MLCC-type structure; the magnitude of the strains that were induced were small by comparison to most shape memory alloys (3-4 microns) but were three times larger than the strains produced using conventional piezoelectric actuators. In this case, a ferroelectric to antiferroelectric phase change is responsible for the shape change. [14]

Electrorheological Fluids

One of the criteria which separates "smart" materials from "very smart" or "intelligent" materials is the ability of the material to not only sense a change and actuate a response, but to automatically modify one or more of its property coefficients during the sensing/actuating process. In effect, this type of material not only warns the user of a change in its environmental conditions and responds to it, but can in addition adjust itself to compensate for future change.

Electrorheological (ER) fluids [15] and their magnetic analog, ferrofluids, [16] are an example of materials that have great potential for use in smart materials and systems. ER fluids are typically suspensions of fine particles in a liquid medium; the viscosity of the suspension can be changed dramatically by applying an electrical field. The electric field causes alignment of the particles in fibril-like branches in the direction of the applied field. The alignment disappears when the electric field is removed, thus creating the desired property of complete cyclic reproducibility.

ER fluids represent an advanced class of composite materials with self-tuning properties, that will find considerable use in vibration control applications. In addition, the compatibility of this technology with modern solid state electronics makes it an attractive component for integration into multifunction, self-contained smart material packages.

Biomimetics - Fish Ears

The word "biomimetic" is not found in most dictionaries so it needs to be defined. It comes from the greek words "bios," meaning "life," and "mimetikos," meaning "to imitate." Biomimetic means to imitate life, or to use the biological world as a source of ideas for device concepts.

Fish and the other inhabitants of the underwater world have some interesting ways of talking and listening which have been copied using piezoelectric ceramics. For most fish, the principal sensors are the lateral line and the inner ear coupled to the swim bladder. The pulsating

swim bladder also acts as a voice, as do chattering teeth in certain fish species.

The lateral line runs fom the head to the tail of the fish and resembles a towed array with sensing organs (stitches) spaced at intervals along the nerve fiber. Each stitch contains several neuromasts made up of gelatinous cupulae resembling pimples in shape within each cupula are a number of fibers which vibrate as the fish swims through water and acts as sensors for flow noise. The hair-like fibers are extremely thin in diameter ranging from 0.5 to 10 μm. When stimulated by turbulence, the motion of the hairs produces changes in the synapses which are in turn connected to the nerve fiber. The electric signal originates from impedance changes in cell walls which modulate the flow of K^+ ions. The lateral line is especially sensitive to low frequency fluid motion parallel to the length of the fish. In the 50 Hz range, threshold signals are observed for displacements as small as 30 nm! [17,18]. The 1-3 composite hydrophones described later are patterned after the hair-filled cupulae of the lateral line. Thin PZT fibers embedded in polymer provide excellent electromechanical coupling to a liquid medium and can be used as both sensors and actuators.

Dr. Dolittle And Fish Talk

Among the most popular children's books of all time are the Doctor Dolittle books written by Hugh Lofting. Doctor Dolittle could talk to the animals. He started with "Pig" and went on to "Duck" and "Cow", and eventually mastered 498 languages. There is a marvelous scene in the movie where young Tommy Stebbins meets Dr. Dolittle in his laboratory and watches him talk to goldfish by blowing bubbles in the fish bowl through a rubber tube. Between bursts of bubbles, Dr. Doolittle listened intently to through a stethoscope pressed against the side of the bowl.

In recent years, great advances have been made in recording and understanding fish talk, largely because of the development of improved hydrophone arrays and high speed spectrum analyses. Much of the talk is in the form of low frequency grunts below 200Hz. Sound functions in a variety of ways for fish, both in offense and in defense, for warning and intimidation. Many fish speak differently during breeding season, and appear to use coded repetition rates to communicate. Our ability to "farm the oceans" could be greatly enhanced by learning how to talk to fish and control their movements and feeding habits.

Although they do not possess a larynx, many species of fish produce high-pitched sound by grinding their teeth, but the vibration of the swim bladder wall provides the greatest repertoire of noises or calls. The croakers of Chesapeake Bay make tapping noises like a woodpecker, by contracting their drumming muscles attached to the swim bladder, and the twilight choruses of sea robins caused great confusion among the operators of antisubmarine echo-location devices during World War II. [19]

Inner Ears And Swim Bladders

The nature of sound transmission in water has had a great influence on the evolution of hearing in fish. Sound, especially low frequency sound, travels faster and farther than in air. "Near-field" sound consists of small fluid motions or vibrations and are characterized by a displacement direction. They are detected by the inner ear or by the lateral line. The hydrostatic component or "far-field" sound is detected best through the swim bladder.

The inner ear is made up of inertia-sensing chambers resembling accelerometers. Within each chamber is a dense ear stone (otolith) which vibrates in a near-field sound wave. The inertia of the ear stone causes it to lag behind the motion of the fish, and to push against hair cells which line the chamber (sacculus). On bending, the hair cellular membranes deform, stimulating neural transmissions to the brain. Connections to the swim bladder improve the sensitivity to far-field sound.

The primary function of the gas-filled swim bladder is to provide buoyancy, but it is also

used for sound and pressure reception and in some species is equipped with drumming muscles for sound production. The flexible swim bladder responds to hydrostatic pressure waves by changing volume. Fish with swim bladders can perceive relative pressure changes equivalent to less than 0.5% of the ambient hydrostatic pressure.

Direct or indirect linkages from the swim bladder to the inner ear promote the hearing sensation. Fish with no connections perceive low frequency sound (less than 500 Hz), while those with good connections have an upper frequency response of 5000 Hz. As might be expected, the swim bladder is reduced in size with depth, and loses much of its sensitivity as a sensor.

Hydrophone Materials

The knowledge which comes from the understanding of "fish talk" can be directly applied to research in materials destined to someday "sleep with the fishes". Hydrophones are underwater listening devices made from piezoelectric materials which respond to hydrostatic pressure waves. Among the applications for hydrophones are sonar systems for submarines, off-shore oil platforms, geophysical prospecting equipment, fish finders, and earthquake monitors.

As the earth's population continues to increase, mankind must continue to search for new and efficient sources of food and nutrition. The world's oceans may provide a solution to this problem, not only through fish farming but through the use of new and varied salt water vegetation that could provide an abundant source of food, especially for third world countries in which poor soil and harsh climates prohibit conventional farming. Smart hydrophone transceivers will receive and transmit fish talk and monitor the growth of underwater vegetation.

The figure of merit for hydrophone materials is the product of the hydrostatic piezoelectric charge coefficient (d_h) and the piezoelectric voltage coefficient (g_h). While good piezoelectric materials such as PZT have high d_{33} and d_{31} piezoelectric coupling coefficients, the d_h value is only about 45 pC/N because d_{33} and d_{31} are opposite in sign, and $d_h = d_{33} + 2d_{31}$. $d_h g_h$ is also inversely related to the dielectric permittivity, ε_{33}, so that low dielectric constants are desirable as well.

Rather than abandon PZT in search of the ultimate hydrophone material, avoiding this problem is often a matter of clever (=smart?) engineering of existing materials. Too often in the field of materials research we put too much emphasis on the synthesis of new materials and too little emphasis on new and unique designs for old materials.

A composite design with 1-3 connectivity is similar in design to the hair-filled gelatinous cupula with thin PZT rods embedded in a polymer matrix. The 1-3 piezocomposites have excellent sensitivity to pressure waves in water (20, 21). The large d_{33} is maintained because the parallel connection results in stress transfer from polymer to piezoceramic, while the d_{31} is destroyed because of series connection in the lateral dimension where the mechanical load is absorbed by the polymer and not transferred to the PZT rods. Finally, ε_{33} is minimized due to the large volume of low ε_{33} polymer present. The $d_h g_h$ values are improved by more than an order of magnitude when very thin rods are used.

Another piezoelectric hydrophone composite maximizes d_h by simply re-directing the applied stresses using specially shaped electrodes [25]. These are flextensional transducers which mimic the motions of the swim bladder. Shallow air spaces are positioned under the metal electrodes while the PZT ceramic plays the role of the muscle lining the swim bladder. When subjected to a hydrostatic stress because of waves, the thick metallic electrodes convert a portion of the z-direction stress into large radial and tangential stresses of opposite signs. The result is that d_{31} changes from negative to positive, so that its contribution now adds to d_{33} rather than subtracting from it. The $d_h g_h$ of these

composites is approximately 250 times that of pure PZT.

Very Smart Composites: The Tunable Transducer

By building in a learning function, the definition of a smart material can be extended to a higher level of intelligence: A very smart material senses a change in the environment and responds by changing one or more of its property coefficients. Such a material can tune its sensor and actuator functions in time and space to optimize future behavior. With the help a feedback system, a very smart material becomes smarter with age, something even human beings strive for. The distinction between smart and very smart materials is essentially one between linear and nonlinear properties. The physical properties of nonlinear materials can be adjusted by bias fields or forces to control response.

To illustrate the concept of a very smart material, we describe the tunable transducer recently developed in our laboratory. Electromechanical transducers are used as fish finders, gas igniters, ink jets, micropositioners, biomedical scanners, piezoelectric transformers and filters, accelerometers, and motors.

Five important properties of a transducer are the resonant frequency f_r, the acoustic impedance Z_A, the mechanical damping coefficient Q, the electromechanical coupling factor k, and the electrical impedance Z_E. The resonant frequency and acoustic impedance are controlled by the elastic constants and density, as discussed in the next section. The mechanical Q is governed by the damping coefficient (α) and is important because it controls "ringing" in the transducer. Definitions of the coefficients are given in Table II.

Table II. Important characteristics of an electromechanical transducer.

Fundamental resonant frequency, f, of the thickness mode

$$f = \frac{1}{2t}\sqrt{c/\rho}$$

t = thickness dimension
c = elastic stiffness
r = density

Acoustic Impedance Z_A

$$|Z_A| = \sqrt{\rho c}$$

Mechanical Q

$$Q = \pi/\lambda_a \alpha$$

λ_a = acoustic wavelength
α = damping coefficient

Electromechanical coupling coefficient k

$$k = d \sqrt{c/\varepsilon}$$

d = piezoelectric charge coefficient
ε = electric permittivity

Electrical impedance Z_e

$$|Z_e| = t / \omega \varepsilon A$$

ω = angular frequency
A = electrode area

Electromechanical coupling coefficients are controlled by the piezoelectric coefficient which, in turn, can be controlled and fine-tuned using relaxor ferroelectrics with large electrostrictive effects. The dielectric "constant" of relaxor ferroelectrics depends markedly on DC bias fields, allowing the electrical impedance to be tuned over a wide range as well. In the following sections we describe the nature of nonlinearity and how it controls the properties of a tunable transducer.

Elastic Nonlinearity: Tuning The Resonant Frequency

Information is transmitted on electromagnetic waves in two ways: amplitude modulation (AM) and frequency modulation (FM). There are a number of advantages to FM signal processing, especially where lower noise levels are important. Atmospheric static is considerably lower in FM radio than in AM.

Signal-to-noise ratios are also important in the ultrasonic systems used in biomedical and nondestructive testing systems, but FM-modulation is difficult because resonant frequencies are controlled by stiffness (c) and (transducer) dimensions (t). Neither c, t, nor the density (ρ) can be tuned significantly in ceramics and most other materials, but rubber is an exception. To tune the resonant frequency of a piezoelectric transducer, we designed and built a composite transducer incorporating thin rubber layers exhibiting nonlinear elasticity. [23]

Rubber is a highly nonlinear elastic medium. In the unstressed compliant state, the molecules are coiled and tangled, but under stress the molecules align and the material stiffens noticeably. Experiments carried out on rubber-metal laminates demonstrate the size of the nonlinearity. Young's modulus ($E = 1/s_{1111}$) was measured for a multilayer laminate consisting of alternating steel shim and soft rubber layers each 0.1 mm thick. Under compressive stresses of 200 MPa, the stiffness is quadrupled from about 600 to 2400 MPa. [24] The resonant frequency f is therefore doubled, and can be modulated by applied stress.

Rubber, like most elastomers, is not piezoelectric. To take advantage of the elastic nonlinearity, it is therefore necessary to construct a composite transducer consisting of a piezoelectric ceramic (PZT) transducer, thin rubber layers, and metal head and tail masses, all held together by a stress bolt.

The resonant frequency and mechanical Q of such a sandwich structure was measured as a function of stress bias (Fig. 13). Stresses ranged from 20 to 100 MPa in the experiments. Under these conditions the radial resonant frequency changed from 19 to 37 kHz, approximately doubling in frequency as predicted from the elastic nonlinearity. At the same time the mechanical Q increases from about 11 to 34 as the rubber stiffens under stress.

The changes in resonance and Q can be modeled with an equivalent circuit in which the compliance of the thin, rubber layers are represented as capacitors coupling together the larger masses (represented as inductors) of the PZT transducer and the metal head and tail masses. Under low stress bias, the rubber is very compliant and effectively isolates the PZT transducer from the head and tail masses. At very high stress, the rubber stiffens and tightly couples the metal end pieces to the resonating PZT ceramic. For intermediate stresses the rubber acts as an impedance transformer giving parallel resonance of the PZT - rubber - metal - radiation load.

Continuing the biomimetic theme, it is interesting to compare the change in frequency of the tunable transducer with the transceiver systems used in the biological world. The biosonar system of the flying bat is similar in frequency and tunability to our tunable transducer. The bat emits chirps at 30 kHz and listens for the return signal to locate flying insects. To help it differentiate the return signal from the outgoing chirp, and to help in timing the echo, the bat puts an FM signature on the pulse. This causes the resonant frequency to decrease from 30 to 20 kHz near the end of each chirp. Return signals from the insect target are detected in the ears of the bat where neural cavities tuned to this frequency range measure the time delay and flutter needed to locate and identify its prey. Extension of the bat biosonar principle to automotive, industrial, medical and entertainment systems is underway.

Piezoelectric Nonlinearity: Tuning The Electromechanical Coupling Coefficient

The difference between a smart and a very smart material can be illustrated with piezoelectric and electrostrictive ceramics. PZT (lead zirconate titanate) is a piezoelectric ceramic in which the ferroelectric domains have been aligned in a very large poling field. Strain is linearly proportional to electric field in a fully poled piezoelectric material which means that the piezoelectric coefficient is a constant and cannot be electrically tuned with a bias field. Nevertheless it is a smart material because it can be used both as a sensor and an actuator.

PMN (lead magnesium niobate) is not piezoelectric at room temperature because its Curie temperature lies near 0°C. Because of the proximity of the ferroelectric phase transformation, however, and because of its diffuse nature, PMN ceramics exhibit very large electrostrictive effects. The nature of this large nonlinear relationship between strain and electric field, and of its underlying atomistic origin, will be described later.

Electromechanical strains comparable to PZT can be obtained with electrostrictive ceramics like PMN, and without the troubling hysteretic behavior shown by PZT under high fields. The nonlinear relation between strain and electric field in electrostrictive transducers can be used to tune the piezoelectric coefficient and the dielectric constant.

The piezoelectric d_{33} coefficient is the slope of the strain-electric field curve when strain is measured in the same direction as the applied field. Its value for $Pb(Mg_{0.3}Nb_{0.6}Ti_{0.1})\ O_3$ ceramics is zero at zero field and increases to a maximum value of 1300 pC/N (about three times larger than PZT) under a bias field of 3.7 kV/cm.

This means that the electromechanical coupling coefficient can be tuned over a very wide range, changing the transducer from inactive to extremely active. The dielectric constant also depends on DC bias. The polarization saturates under high fields causing decreases of 100% or more in the capacitance. In this way the electrical impedance can be controlled as well.

Electrostrictive transducers have already been used in a number of applications including adaptive optic systems, scanning tunneling microscopes, and precision micropositioners [8].

To summarize, two types of nonlinearity are utilized in the fully tunable transducer: elastic nonlinearity and piezoelectric nonlinearity. By incorporating thin rubber layers in an electrostrictive transducer several important properties can be optimized with bias fields and bias stresses. Electromechanical coupling coefficients and electric impedance are tuned with electric field, and mechanical damping, resonant frequency, and acoustic impedance with stress bias.

Smart Electroceramic Packages

Up to this point, our discussion has focused primarily on piezoelectric transducers in which the sensing and actuating functions are electromechanical in nature. But the idea of a smart material is much more general than that. There are many types of sensors and many types of actuators, and many different feedback circuits.

Many of these sensors and actuators can be fabricated in the form of multilayer ceramic packages. Until recently multilayer packages consisted of low permittivity dielectric layers with metal circuitry printed on each layer and interconnected through metallized via holes between layers. Buried capacitors and resistors have now been added to the three-dimensional packages, and other components will follow shortly. Smart sensors, adaptive actuators, and display panels, together with thermistors and varistors to guard against current and voltage overloads, are next in line for development. [25]

Integration and miniaturization of electroceramic sensors and actuators is an ongoing process in the automotive and consumer electronics areas. Multilayer packages

containing signal processing layers made up of low-permittivity dielectrics and printed metal interconnections are in widespread production. Further integration with embedded resistors and capacitors are under development, and its seems likely that intelligent systems will make use of the same processing technology. Tape casting and screen printing are used most often. Varistors, chemical sensors, thermistors, and piezoelectric transducers can all be fabricated in this way, opening up the possibility of multicomponent, multifunction ceramics with both sensor and actuator capabilities. Silicon chips can be mounted on these multifunctional packages to provide all or part of the control network. Processing is a major challenge because of the high firing temperatures of most ceramics, typically in the range 800°C to 1500°C. Differences in densification shrinkage and thermal contraction, together with adverse chemical reactions between the electroceramic phases, create formidable problems. Nevertheless, the rewards for such an achievement are substantial. An all-ceramic multifunction package would be small, robust, inexpensive, and sufficiently refractory to withstand elevated temperatures.

Electrodes are both a problem and a challenge. At present, precious metals such as palladium and platinum are used in multilayer ceramic components, greatly adding to the cost of the device. Copper and nickel electrodes require that the ceramic be fired in a reducing atmosphere, which may reduce the electroceramic layers and adversely affect the electrical properties. Copper and silver have high electrical conductivity but the melting points (~1000°C) require lower firing temperatures and make it necessary to alter the ceramic compositions and fabrication procedures. Some headway has been made on this problem, but further work is needed. One interesting approach to the problem is ceramic electrodes. There are a number of ceramic phases with excellent conductivity which could be used, including the copper-oxide superconductors. In actuator devices, there are some special advantages in having electrodes and piezoceramics with matched elastic properties.

Composites are another approach to making sensor-actuator combinations. These can be formed at lower temperatures using low-firing ceramics and high temperature polymers such as polyimides. Sol-gel and chemical precipitation methods are helpful in preparing ceramic powders with low calcining temperatures, but further work on composite fabrication is required to obtain reliable and reproducible electrical behavior.

To miniaturize the sensors and actuators, and to obtain complex shapes, we recommend the use of photolithography and other processing methods employed in the semiconductor industry. Ultraviolet curable polymers incorporated into the tape-casting process make photolithographic processing comparatively easy and should find wide use in preparing ceramic or composite packages for intelligent systems.

The next logical step is to combine the sensor and actuator functions with the control system. This can be done by depositing electroceramic coatings on integrated circuit silicon chips.

Reliability is a major requirement in all complex systems. Further research on electrical and mechanical breakdown of sensor-actuator materials is needed to elucidate the mechanisms responsible for failure. Fail-safe and self-repair phenomena would be very helpful.

Intelligent systems for hostile environments is another topic area of interest. Sensor and actuator systems that can operate at high temperatures inside engines or furnaces are required to monitor combustion and to provide pollution controls. Radioactive and chemical waste sensors are another problem. Intelligent systems for oceanographic studies and oil exploration must withstand high pressures and salinity conditions.

Electroceramics have a vital role to play in intelligent systems, and many new developments will take place in the coming decade and the next century.

References

[1] C. A. Rogers, "From the Editor," *Journal of Intelligent Material Systems and Structures* **1**, 3 (1990).

[2] T. Takagi, "The Concept of Intelligent Materials and the Guidelines on R&D Promotion", Japan Science and Technology Agency Report, (1989).

[3] R.E. Newnham, Q.C.Xu, S.Kumar, and L.E.Cross, "Smart Ceramics," *J.Wave-Matl. Int.* **4**, 3-10 (1989).

[4] M.Taguchi, "Applications of High-Technology Ceramics in Japanese Automobiles," *Adv.Ceram.Mat*. **2**, 754-762 (1987).

[5] H.Tsuka, J. Nakomo, and Y.Yokoya, "A New Electronic Controlled Suspension Using Piezoelectric Ceramics," IEEE Workshop on Electronic Applications in Transportation, (1990).

[6] A.R. Ramachandran, Q.C.Xu, L.E.Cross, and R.E.Newnham, "Passive Damping with Piezoelectric Ceramics," Presented at the 7th International Symposium on Applications for Ferroelectrics, June 6-8, Urbana, Illinois, (1990).

[7] S. Burke, and J. E. Hubbard, "Active Vibration Control of a Simply-Supported Beam Using a Spatially-Distributed Actuator," *IEEE Control Systems Magazine* **7**, 25-30 (1987).

[8] K. Uchino, "Piezoelectric/Electrostrictive Actuators," Morikita Publishers, Tokyo, (1986).

[9] J. L. Butler, "Application Manual for Design of Magnetostrictive Transducers," Edge Technologies, Extrema Division, N.Marshfield, MA (1988).

[10] L. Schetky, "Shape Memory Alloys," *Scientific American* **241**, 74-82 (1979).

[11] D. L. Edberg, "Control of Flexible Structures by Applied Thermal Gradients," *AIAA Journal* **25**, 877-883 (1987).

[12] C. A. Rogers, , C. Liang, and D. K. Barker, in "Smart Materials, Structures, and Mathemetical Issues," ed.by C.A.Rogers, Technomic Publishers, Lancaster, PA, 39-62 (1989).

[13] V. K. Wadhawan, M. C. Kernion, T. Kimura, and R. E. Newnham, "The Shape Memory Effect in PLZT Ceramics," *Ferroelectrics* **37**, 575-578 (1981).

[14] A. Furuta, K. Oh, and K. Uchino, "Mechanical Clamper Using Shape Memory Ceramics,"

[15] M. V. Ghandi, and B. S. Thompson, in "Smart Materials, Structures, and Mathemetical Issues," ed. by C. A. Rogers, Technomic Publishers, Lancaster, PA, 63-68 (1989).

[16] R. E. Rosenweig, "Ferrohydrodynamics," Cambridge University Press, New York, (1985).

[17] C. E. Bond, "Biology of Fishes," Saunders College Publishing Co., Philadelphia, (1979).

[18] P. B. Moyle, and J. J.Cech, Jr., "Fishes: An Introduction to Ichthyology," Prentice Hall, Englewood Cliffs, N.J. (1988).

[19] R. J. Urick, "Principles of Underwater Sound," McGraw Hill, New York, (1975).

[20] K. A. Klicker, , J. V. Biggers, and R. E. Newnham, Composites of PZT and Epoxy for Hydrostatic Transducer Applications," *J.Amer.Cer.Soc.* **64**, 5-9 (1981)

[21] T. R. Gururaja, W. A. Schulze, L. E. Cross, R. E. Newnham, B. A. Auld, and J.Wang, "Resonant Modes in Piezoelectric PZT Rod-Polymer Composite Materials," Proc. of IEEE Ultrasonic Symposium, 523-527 (1984).

[22] Q. C. Xu, J. Belsick, S. Yoshikawa, T. T. Srinivasan, and R. E. Newnham, "Piezoelectric Composites with High Sensitivity and High Capacitance for Use at High Pressures,"*IEEE Proceedings*, in press.

[23] Q. C. Xu, R. E.Newnham, M. Blaskiewicz, T. T. Fang, T. T. Srinivasan, and S. Yoshikawa, "Nonlinear Multilayer Composite Transducers," Presented at the 7th International Symposium on Applications for Ferroelectrics, June 6-8, Urbana, Illinois, (1990).

[24] E. I. Rivin, "Properties and Prospective Applications of Ultra Thin Layered Rubber-Metal Laminates for Limited Travel Bearings," *Tribology Int.* **16**, 17-25, (1983).

[25] R. E. Newnham, "The Golden Age of Electroceramics," *Adv.Ceram.Mat.* **3**, 12-16 (1988).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

New Cuts for Width-Extensional Mode Quartz Crystal Resonators

Hirofumi KAWASHIMA

Seiko Electronic Components Ltd.
Tochigi-shi, Japan 328

Abstract

This paper describes various characteristics for a width-extensional mode quartz crystal resonator. The object of this paper is to propose a new cut for a width-extensional mode quartz crystal resonator, which is designated the "KT-Cut" and to clarify its frequency characteristics, frequency temperature behavior and electrical characteristics.

First, a frequency equation is derived under the conditions that width x_0 is much larger than thickness z_0 but much smaller than length y_0 by solving partial differential equations for a new shape GT cut quartz crystal resonator already-obtained with variational principle, because the modified shape for the present resonator is refered to as that of the new shape GT resonator.

Next, from this equation, relationships of resonant frequency versus width x_0 and the first- and second-order temperature coefficients α, β versus a cut angle θ are analyzed theoretically. As a result, a cut angle θ which gives $\alpha=0$ approximately is $\theta=24.5^\circ$ at the reference temperature of 20℃ and the β has a value of $\beta=-0.60\times10^{-8}/℃^2$. This value is then compared with the measured values of $\beta=-1.25\times10^{-8}/℃^2$ and the absolute value in the experiments is found to be about twice of the calculated value. In addition, α and β are very dependent on a thickness-to-width ratio R_{zx} (thickness z_0/width x_0), and have a parabolic curve versus the ratio R_{zx}. Thus, α and β are also found to vary with R_{zx}.

Finally, series resistance R_1 and a quality factor Q are principally examined in the frequency range of approximately 8.7 to 32MHz. Consequently, e.g., when f=16.3MHz, a width-extensional mode quartz crystal resonator is found to be successfully obtained with satisfactory values of $R_1=89.7\Omega$ and Q=180,000.

§1 Introduction

Quartz is a very stable material physically and chemically. Therefore, in resonators of various cuts that are manufactured from this quartz, an AT cut quartz crystal resonator[1], [2] is widely used as source of a standard signal of communication equipment, consumer products, measurement apparatus and so on, because it has advantages such as ease of a supporting means of the resonator, frequency stability of short and long terms and excellent frequency temperature behavior. In addition, since resonant frequency for the resonator which vibrates in thickness shear mode is approximately determined by thickness, it is particularly suitable over about 4 MHz, when the resonator's size is considered. At present, as an innovative shape face shear mode quartz crystal resonator[3] formed by an etching process, is put to practical use until approximately 6 MHz, actually, the thickness shear mode was the only vibration mode wherein a zero temperature coefficient exists and resonant frequency is realized over 6 MHz. Recently, according to miniaturization, low consumption power and high precision of these various equipment and products, requirements of miniaturization and so forth are also being performed for an AT cut quartz crystal resonator.

However, the miniaturization may cause further difficulty of a supporting means of the resonator, increase of series resistance R_1 by vibration energy losses and more spurious responses. In order to overcome the subjects, processing precision, an improvement of a supporting means and so forth are tightly required in addition to an optimum design of resonator's shape and dimension. Thus, some subjects to be solved are left so as to realize the miniaturization. From the situation, a new type quartz crystal resonator over about 6 MHz that is formed by an etching process is, therefore, desired and expected, because a means of the process makes it possible to form an arbitrary resonator shape and facilitates a supporting means of the resonator.

In this paper, a study is performed with a view to clarifying theoretically and experimentally frequency characteristics and frequency temperature behavior of a new cut for a quartz crystal resonator which vibrates in a width-extensional mode (hereafter call width-

0-7803-0476-4/92 $3.00 © 1992 IEEE

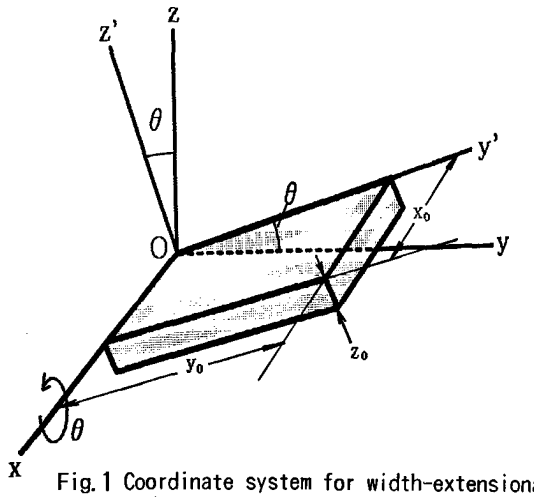


Fig. 1 Coordinate system for width-extensional mode quartz crystal resonator.

extensional mode quartz crystal resonator). Specifically, an examination is performed in the vicinity of Z-plate that is easily formed by an etching process. First, let us describe derivative procedure of a frequency equation.

A partial differential equation is derived under the conditions that resonator's thickness is much smaller than the width and the length. Furthermore, a frequency equation is obtained by solving the partial differential equation under the condition that the length is much larger than the width. From this equation, resonant frequency and frequency temperature coefficients are calculated. As a result, a new cut for a width-extensional mode quartz crystal resonator is theoretically found to exist with a zero temperature coefficient and the existence is confirmed in the experiments. This cut is designated the "KT-cut". Simultaneously, frequency characteristics, frequency temperature behavior and electrical equivalent circuit parameters are clarified. A detail description will be given below.

§ 2 Frequency Equation

Figure 1 shows a width-extensional mode quartz crystal resonator and its coordinate system. The resonator has a dimension of width x_0, length y_0 and thickness z_0 and also is formed from rotated Z-plate with a rotation angle θ around the x-axis. In addition, this resonator is expressed in IEEE expression of (ZYw) θ.

Figure 2 shows a mode shape for a rectangular width-extensional mode quartz crystal resonator with a cut angle θ=0 analyzed by the finite element method. The broken line illustrates a steady-state of a resonator shape, while the solid line illustrates a mode shape of the resonator in vibration. As is obvious from

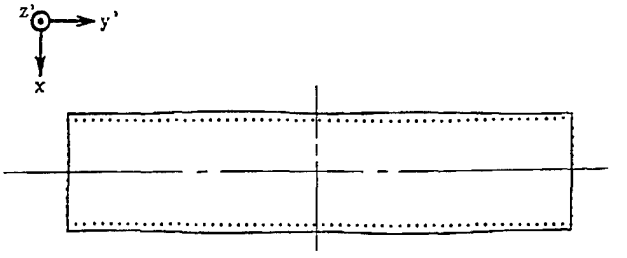


Fig. 2 A mode shape for width-extensional mode quartz crystal resonator.

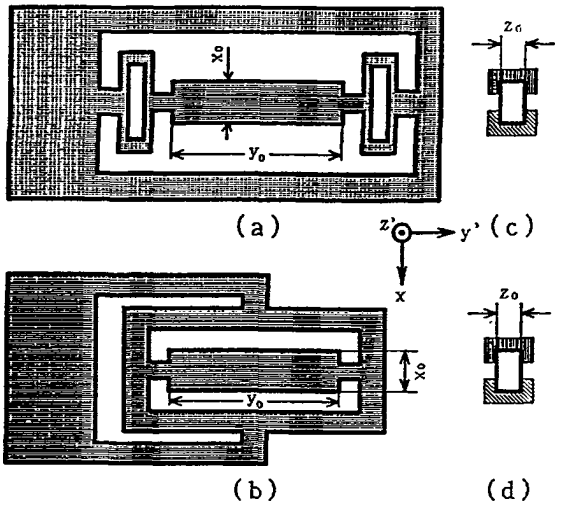


Fig. 3 Actual resonator's shapes (a), (b) of width-extensional mode resonator and their electrode constructions (c), (d) of section for vibrational portion.

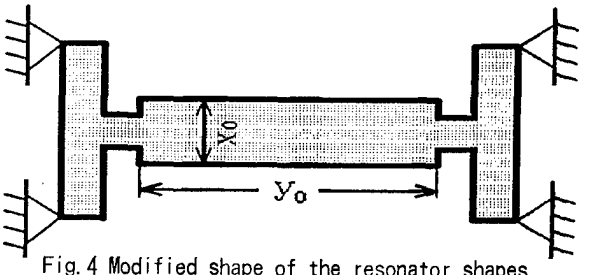


Fig. 4 Modified shape of the resonator shapes (a), (b).

Fig. 2, the width-extensional mode resonator extends and contracts largely in the width x_0 direction. According to a motion of the extension and contraction, a motion in the length y_0 direction is also caused somewhat. Therefore, a supporting means of resonators which does not suppress the motion of the length y_0 direction, is needed in order to obtain a small series resistance R_1. Good supporting means is shown in Fig. 3.

Figure 3 shows two kinds of shapes (a), (b) for a width-extensional mode quartz crystal resonator to be treated in this paper and electrode constructions (c), (d) of section for

the vibrational portions. The resonator consists of a vibrational portion and supporting portions, a dimension of the vibrational portion is width x_0, length y_0 and thickness z_0, which correspond to rotated crystal axes x, y' and z' of quartz, respectively. Further, the excitation electrodes are, as shown in Fig. 3(c), (d), disposed on obverse, reverse and sides of the vibrational portion dividedly at the center of the obverse and the reverse. The shape (b) is smaller in size of length y_0 than the shape (a). Where a width-extensional mode is a vibration mode to extend and contract largely in the width x_0 direction by an electric field of the x-axis direction, as illustrated in Fig. 2.

Figure 4 shows a modified shape of the resonator shapes (a), (b). Similar to a new shape GT cut quartz crystal resonator[4]-[6] already-introduced by the authors, it is conceivable that the resonator has the supporting portions vibrating in a flexural mode with each mass m at both ends of the length y_0 and also with the boundary conditions of "both hinged ends". That is, the supporting portions each have mass m and resistance to vibration of the vibrational portion (work done by external force). As derivative procedure of a partial differential equation for the resonator under the condition of $x_0, y_0 \gg z_0$ is described in detail in the reference[4], here, its procedure is explained briefly.

Now, when width x_0 and length y_0 are much larger than thickness z_0 ($x_0, y_0 \gg z_0$), stresses T_3, T_4 and T_5 are given as

$$T_3 = T_4 = T_5 = 0 \tag{1}$$

Therefore, when forced terms are neglected, stress T_i and strain S_i (i=1, 2, 6) have the following relation:

$$\begin{aligned} T_1 &= \gamma_{11}S_1 + \gamma_{12}S_2 + \gamma_{16}S_6 \\ T_2 &= \gamma_{21}S_1 + \gamma_{22}S_2 + \gamma_{26}S_6 \\ T_6 &= \gamma_{61}S_1 + \gamma_{62}S_2 + \gamma_{66}S_6 \end{aligned} \tag{2}$$

where γ_{ij} is related to the elastic stiffness constants c_{ij} and is generally given by the following relation:

$$\begin{aligned} \gamma_{ij} &= \hat{c}_{ij} - \hat{c}_{5i}\hat{c}_{j5}/\hat{c}_{55} \quad i, j = 1, 2, 6 \\ \hat{c}_{ij} &= \bar{c}_{ij} - \bar{c}_{4i}\bar{c}_{j4}/\bar{c}_{44} \quad i, j = 1, 2, 4, 6 \\ \bar{c}_{ij} &= c_{ij} - c_{3i}c_{j3}/c_{33} \quad i, j = 1, 2, 4, 5, 6 \end{aligned} \tag{3}$$

Therefore, when taking displacements U, V in x- and y-axes direction of the vibrational portion neglecting a time term, the displacement $\bar{V}$ of the supporting portion, density ρ of quartz, angular frequency ω, and force per unit volume $\bar{F}_0$ operated on the vibrational portion by the supporting portions, which is so designed that it gets very weak, in order to obtain a small series resistance R_1, partial differential equations from variational principle is, taking into account that stress T_6 is entirely separated from stresses T_1 and T_2, given as

$$\begin{aligned} -\rho\omega^2 U &= \gamma_{11}\frac{\partial^2 U}{\partial x^2} + \gamma_{12}\frac{\partial^2 V}{\partial x \partial y} \\ -\rho\omega^2 (V+\bar{V}) + \bar{F}_0 &= \gamma_{21}\frac{\partial^2 U}{\partial y \partial x} + \gamma_{22}\frac{\partial^2 V}{\partial y^2} \end{aligned} \tag{4}$$

where γ_{16}, γ_{61}, γ_{26} and γ_{62} reach zero when the resonator is singly rotated around the x axis. In addition, since length y_0 is much larger than width x_0 ($y_0 \gg x_0$), equation (4) is simplified to

$$\rho\omega^2 U + \gamma_{11}\frac{\partial^2 U}{\partial x^2} = 0 \tag{5}$$

Accordingly, resonant frequency f ($=\omega/2\pi$) for a width-extensional mode quartz crystal resonator is given as

$$f = \frac{n}{2x_0}\sqrt{\frac{\gamma_{11}}{\rho}} \tag{6}$$

where n is a positive number, n is odd for symmetry mode and even for asymmetry mode.

§ 3 Results and Discussion

According to the following procedure the results are discussed. First, it is shown theoretically that for a width-extensional mode, a cut angle θ which gives a zero temperature coefficient exists and the results are compared with the measured values. An absolute value of the second order temperature coefficient β with $\alpha=0$ is found to be smaller than that of DT cut at room temperature. In addition, a relationship of α, β versus a thickness-to-width ratio R_{zx}(thickness z_0/width x_0), a relationship between a cut angle θ and R_{zx} giving $\alpha=0$ and the β are shown. Furthermore, α and β are found to reach approximately zero at high temperature of about 65℃. Therefore, the present width-extensional mode quartz crystal resonator has a cubic curve as well as an AT cut quartz crystal resonator at higher temperature, simultaneously, frequency deviation is found to be smaller in wider temperature range because the third order temperature coefficient γ is smaller than that of an AT cut quartz crystal resonator. Next, relationships of frequency constant ($f \cdot x_0$) versus a cut angle θ and resonant frequency versus width x_0 are shown. Finally, typical values of electrical equivalent circuit parameters are shown. In the calcula-

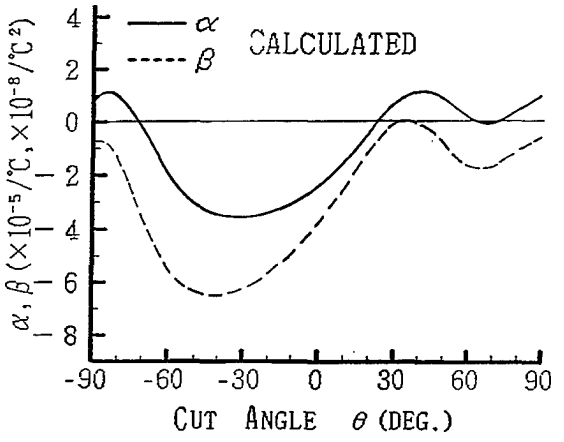


Fig. 5 Relationship between cut angle θ and the first and second order temperature coefficients α, β. (Calculated values)

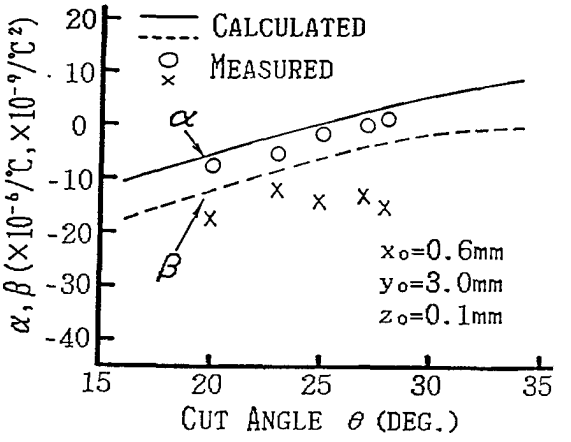


Fig. 6 Relationship between cut angle θ and the first and second order temperature coefficients α, β.

tion of frequency temperature coefficients, the references[6], [7] are used. Taylor's series is performed at 20℃. Also, various characteristics are measured at fundamental vibration.

3.1 Frequency Temperature Coefficients

Figure 5 shows a theoretical relationship of the first- and second- order temperature coefficient α, β versus a cut angle θ for a width-extensional mode quartz crystal resonator. α reaches zero at the cut angle θ of -72.4°, 24.5°, 66.3° and 72.7° in the calculation. Also, an absolute value of the β has a minimum value at the cut angle of θ=24.5°, simultaneously, which cut angle is very easy to form resonators by an etching process, but another cut angles are impossible by the etching process. Thus, as it was predicted that a width-extensional mode quartz crystal resonator with a zero temperature coefficient exists, fortunately with a small absolute value of the β, the existence is, in detail, confirmed in the experiments.

Figure 6 shows a relationship of the first- and second-order temperature coefficients α, β versus a cut angle θ for a width-extensional mode quartz crystal resonator (Resonator shape(a)) with width x_0=0.6mm ($f \fallingdotseq 4.5$MHz), length y_0=3.0mm and thickness z_0=100 μm of the vibrational portion. The solid and broken lines indicate the calculated values, while ○ and × are the measured ones. A cut angle θ where α =0 is θ=24.5° in the calculation and θ=27° in the measured values, an error of 2.5° takes place between them, but it was confirmed experimentally that a cut angle θ where α=0 exists. In addition, the β has a value of -0.60×10^{-8}/℃² and -1.25×10^{-8}/℃², respectively, the measured value has about twice value of the calculated one in the absolute value. This may be because the piezoelectric term was neglected in derivatives

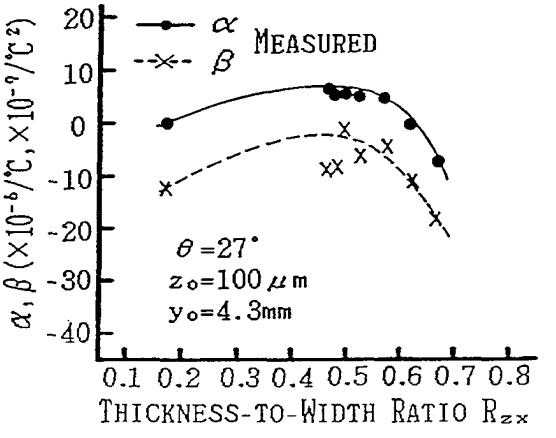


Fig. 7 Relationship between thickness-to-width ratio R_{zx} ($=z_0/x_0$) and the first and second order temperature coefficients α, β.

of the frequency equation. But, an absolute value of β measured is smaller than that of DT cut with -1.8×10^{-8}/℃², the present width-extensional mode quartz crystal resonator, therefore, shows good frequency temperature behavior at room temperature. Further, this result suggests that $\alpha=\beta=0$ can be expected in a higher temperature range. This will be described in detail in a term of frequency temperature behavior.

Figure 7 shows a relationship of α and β versus a thickness-to-width ratio R_{zx}(thickness z_0/width x_0) for a width-extensional mode quartz crystal resonator with a cut angle θ=27°, length y_0=4.3mm and thickness z_0=100 μm. α and β versus R_{zx} represent a convex-parabolic curve and the maximum in the vicinity of R_{zx}=0.5. In addition, when θ=27°, a thickness-to-width ratio R_{zx} where α=0 is R_{zx}=0.629 and 0.167, but, a larger R_{zx} achieves reduction of series resistance R_1 because of larger thickness z_0. Thus, α and β are dependent on R_{zx}. In other words, as will be described later, since

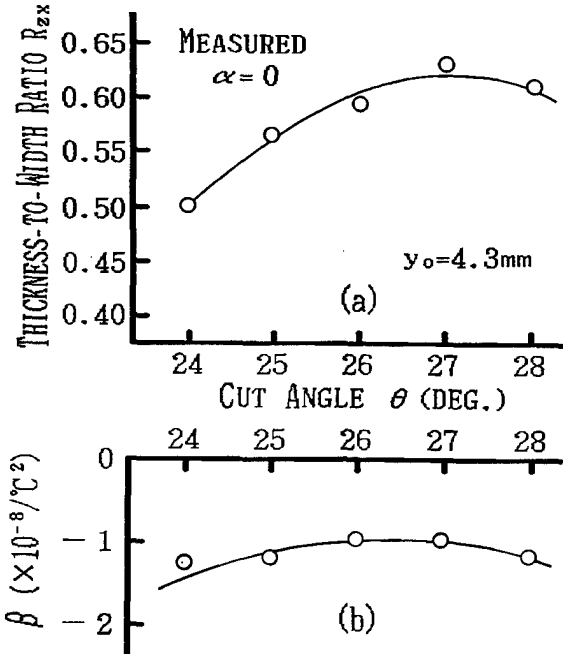


Fig. 8 Relationship between cut angle θ and thickness-to-width ratio R_{zx} when α=0 (a) and the β (b).

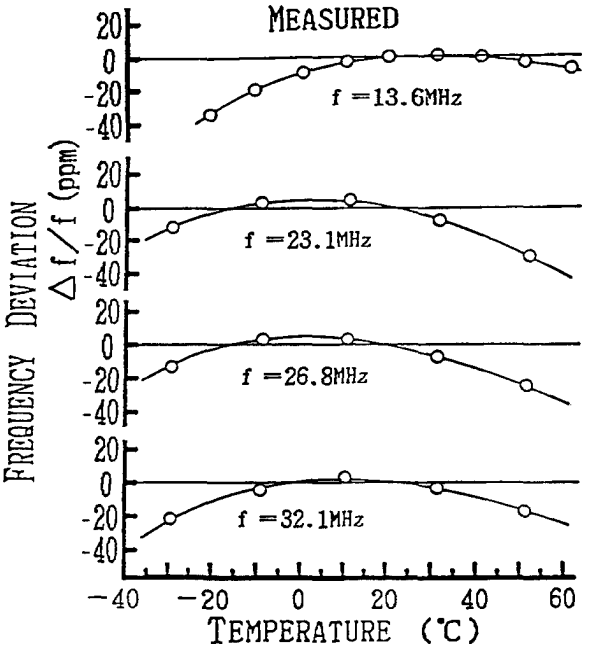


Fig. 9 Exampls of frequency temperature behavior.

resonant frequency is determined by width x_0 and a cut angle θ, the optimum value of the first order temperature coefficient α is, therefore, determined by thickness z_0 for a given frequency.

Figure 8 shows a relationship between a cut angle θ and R_{zx} for a width-extensional mode quartz crystal resonator(Resonator shape (b)) with length y_0=4.3mm of the vibrational portion when α=0, and the β. As shown in Fig. 7, a cut angle θ to give α=0 at a thickness-to-width ratio smaller than R_{zx}=0.5 exists, but, the ratio is not plotted because R_1 becomes larger and practical use is poor. As is illustrated in Fig. 8(a), according to increase of a cut angle θ, a thickness-to-width ratio R_{zx} to give α=0 also gets large, but represents the maximum in the vicinity of θ=27°, and the ratio gets rather small at θ=28°. Further, the β is, as plotted in Fig. 8(b), shown by a parabolic curve, an absolute value of β represents the minimum at θ=26°~27° and its value is approximately β=$-1.0\times10^{-8}/°C^2$. Thus, α=0 is obtained by combination of a cut angle θ and a thickness-to-width ratio R_{zx}. Simultaneously, an absolute value of the β is found to be very small.

3.2 Frequency Temperature Behavior

Figure 9 shows examples of frequency temperature behavior for width-extensional mode quartz crystal resonators of the resonator shape(a) with frequency f=13.6MHz and the resonator shape(b) with frequencies fs=23.1, 26.8, 32.1MHz. The temperature behavior is shown by an approximate parabolic curve. As already-explained in Figs 5~8, the present resonator has α=0 in the vicinity of room temperature and because an absolute value of the β becomes comparatively small, e.g., when f=13.6MHz, frequency deviation has a value of approximately +1.0/-20ppm in a temperature range of -10 ℃ to +60 ℃, in addition, when fs=26.8MHz~32.1MHz, frequency deviation is small at somewhat low temperature and e.g., has a value of about +5.0/-30ppm in a temperature range of -20℃ to +50℃, because a turnover temperature point T.p is set up to temperature lower than room one. Thus, width-extensional mode quartz crystal resonators excellent in frequency temperature behavior are successfully obtained with the frequencies of 13.6MHZ to 32.1MHz. Needless to say, it goes without saying that a turnover temperature point T.p can be set up to an arbitrary temperature by selection of α.

Figure 10 shows an example of frequency temperature behavior in so designing a resonator of the resonator shape(b) that α has a slightly positive value in the vicinity of room temperature. This is of frequency f=16.3MHz. α, β and γ have a value of $\alpha=4.21\times10^{-7}/°C$, $\beta=-1.12\times10^{-8}/°C^2$ and $\gamma=7.02\times10^{-11}/°C^3$ at the reference temperature of 20℃ in Taylor's series, moreover, as this resonator has an inflection point in the vicinity of 65℃, α, β and γ have a value of $\alpha=-1.56\times10^{-7}/°C$, $\beta=-1.21\times10^{-9}/°C^2$ and $\gamma=7.5\times10^{-11}/°C^3$ at this reference temperature. Therefore, the present resonator

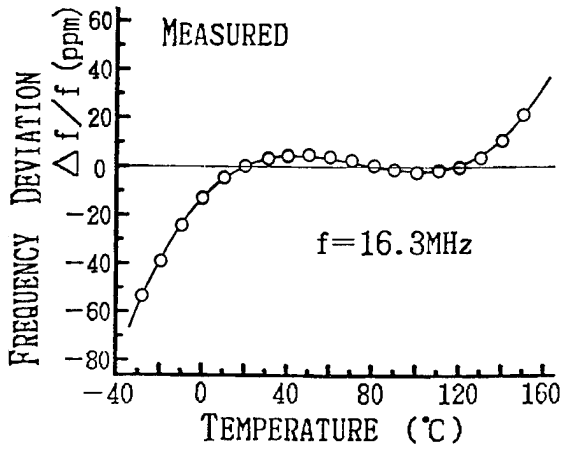


Fig. 10 Example of frequency temperature behavior.

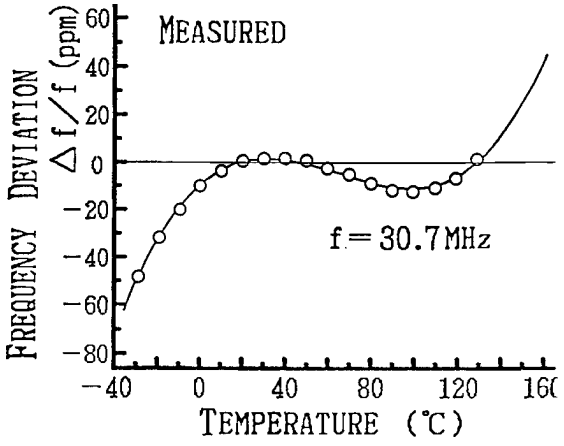


Fig. 11 Another example of frequency temperature behavior.

shows a cubic curve in a temperature range of room temperature to high temperature, as is illustrated in Fig. 10. In addition, since the third order temperature coefficient γ for the present resonator is somewhat smaller than that of an AT cut quartz crystal resonator, frequency deviation $\Delta f/f$ is small in a temperature range wider than that of an AT cut quartz crystal resonator, namely, a width-extensional mode quartz crystal resonator excellent in frequency temperature behavior is successfully obtained with frequency deviation of +4.8/-5.3ppm in a temperature range of +10℃ to +130℃. Thus, the present width-extensional mode quartz crystal resonator shows excellent frequency temperature behavior, particularly, in a range of room temperature to high one.

Figure 11 shows another example of frequency temperature behavior for a width-extensional mode resonator of f=30.7MHz with the same design concept as the resonator in Fig. 10. As well as that in Fig. 10, the present resonator also shows a cubic curve.

3.3 Frequency Constant

Figure 12 shows a relationship between a cut angle θ and frequency constant ($f \cdot x_0$) for a width-extensional mode resonator with the same dimension and shape as the resonator in Fig. 6. Frequency constant ($f \cdot x_0$) has a value of 271.5 to 271.3 kHz·cm at θ=20° ~28° in the calculation, while it has a value of 269.5 to 273kHz·cm in the experiments. The experiments cause a slight distribution, but, the calculated and measured data agree well within an error less than 1%.

Figure 13 shows a relationship between width x_0 and resonant frequency f. The experimental data of the signs ○ and × are those for a width-extensional mode quartz crystal resonator of the resonator shape(b) with length y_0=4.3mm, thickness z_0=100 μm for the sign ○, thickness z_0=50 μm for the sign × and a cut angle θ=27°

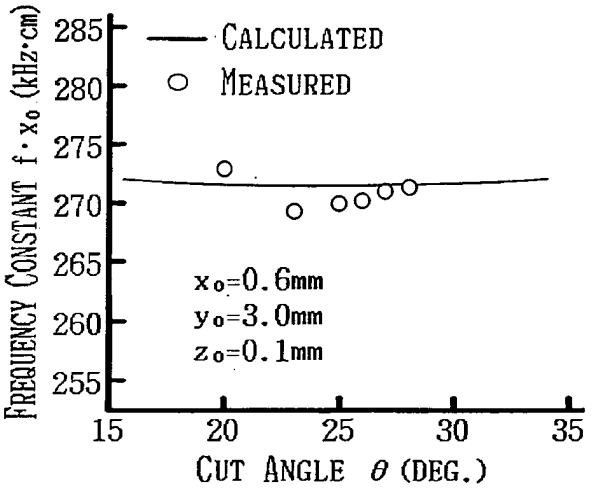


Fig. 12 Relationship between cut angle θ and frequency constant ($f \cdot x_0$). (Resonator shape (a))

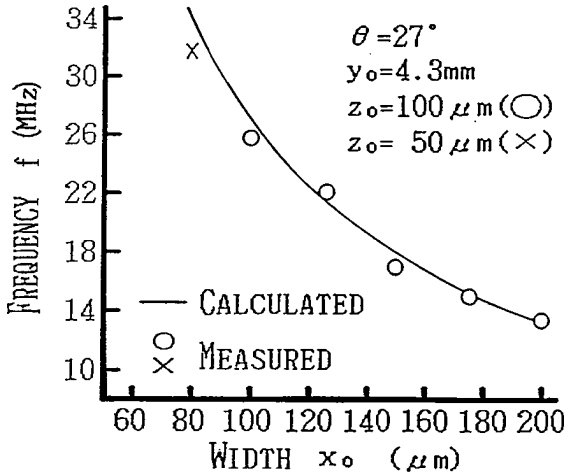


Fig. 13 Relationship between width x_0 and resonant frequency f.

As is apparent from Fig. 13, a variation of x_0=200 to 80 μm gives frequencies of about f=13 to 33MHz.

3.4 Electrical Equivalent Circuit Parameters

Table 1 shows typical values of electrical equivalent circuit parameters for width-extensional mode quartz crystal resonators (Shapes(a) θ=25° /(b) θ=27°) with frequencies of 8.67 to 32.1MHz. Series resistance R_1 and a quality factor Q have values of R_1=89.7 to 281Ω and Q=45,000 to 180,000, respectively. R_1 has a slight large value for some frequency, this may be because incompatibility of the vibrational and supporting portions incorporated is a principal cause. However, sufficiently satisfactory characteristics are obtained as a resonator such as R_1=89.7Ω and Q=180,000, when good compatibility of the vibrational and supporting portions is accomplished like f=16.3 MHz. The influence upon various characteristics due to the supporting portions for each frequency will be a coming subject.

§4 Conclusions

In this paper, a new cut for a width-extensional mode quartz crystal resonator designated the "KT-cut" was proposed and its frequency temperature behavior, frequency characteristics and electrical various characteristics were clarified.

First, partial differential equations are derived under the condition that the resonator's thickness is extremely small as compared with its width and length, and a frequency equation was derived by solving the partial differential equations under the condition that the length is then much larger than the width.

Next, from the obtained frequency equation, it was shown theoretically and confirmed experimentally that a cut angle θ where the first order temperature coefficient α reaches zero exists. An absolute value of β is smaller than that of DT cut. It was also shown that α and β vary with a thickness-to-width ratio R_{zx}. In addition, there is many combination of a cut angle θ and R_{zx} to give α=0, an absolute value of the β is comparatively small. Furthermore, frequency temperature behavior was shown experimentally for frequencies of 16 to 32MHz and width-extensional mode quartz crystal resonators with excellent frequency temperature behavior were found to be obtained successfully. By so designing a resonator that α has a slight positive value in the vicinity of room temperature, a width-extensional mode quartz crystal resonator is obtained with small frequency deviation in a range of low temperature to high one.

Then, a relationship between a cut angle θ and frequency constant was shown and frequency constant was found to have an almost constant value at the cut angle of θ=20° to 28° . Therefore, the choice of width x_0 gives a width-extensional mode resonator with a frequency of 8.67 to 32.1MHz at fundamental vibration.

Table 1 Typical values of electrical equivalent circuit parameters for width-extensional mode quartz crystal resonators.

CUT ANGLE θ(DEG.)	FREQUENCY f (MHz)	R_1 (Ω)	L_1 (mH)	C_1 (fF)	Q (x10^3)	r
25	8.67	205	467	0.722	124	380
	13.5	281	265	0.527	80	390
27	16.3	89.7	157	0.608	180	340
	23.1	116	72.3	0.651	90	290
	26.8	212	57.5	0.612	45	400
	32.1	136	48.2	0.510	71	370

Finally, electrical equivalent circuit parameters for the present resonators were examined. The present resonators has are frequencies of 8.67 to 32.1MHz, comparatively small series resistance R_1 and a high quality factor Q. Therefore, the present resonator is especially available for use in handy-type equipment such as cordless telephone and so on.

Acknowledgements

The author wishes to express his thanks to Mr. M. Matsuyama and Mr. M. Nakazato for their help in the experiments and also to Mr. K. Sunaga for his help in the calculation by computer.

References

[1] I. Koga: "Piezoelectricity and High Frequency", Ohm-sha, Japan(1932).

[2] R. A. Sykes; "Quartz Crystal for Electrical Circuits", ed. R. A. Heising, pp. 10-230, D. Van Nostrand New York(1946).

[3] H. Kawashima, M. Matsuyama and M. Nakazato: "A Variational Analysis of a New Shape Face Shear Mode Quartz Crystal Resonator Formed by an Etching Method", Proc. the 42nd Annual Symp. on Frequency Control, pp. 65-72(1988).

[4] H. Kawashima: "Variational Analysis of GT Cut Quartz Crystal Resonators with the Supporting Portions at the Ends", Proc. the 40th Annual Symp. on Frequency Control, pp. 193-200 (1986).

[5] H. Kawashima: "An Analysis of Frequency Temperature Characteristics and Electrical Equivalent Circuit Parameters of a New Shape GT Cut Quartz Crystal Resonator Formed by an Etching Method", IEEE Ultrasonics Symp. Proceedings pp. 465-470(1989).

[6] H. Kawashima: "Vibration Analysis of GT Cut quartz Crystal Resonator Taking Account of Dissipation Term", Trans. IEICE(A), Japan, J72-A, 4, pp. 633-639(1989).

[7] M. Aruga: "On Elastic Stiffnesses and their Temperature Characteristics of Quartz Crystal Resonator", Journal of Tokyo Institute of Technology, Japan, A-2, pp. 88-182(1956).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

FORCED VIBRATIONS OF THICKNESS-FLEXURE, FACE-SHEAR AND FACE-FLEXURE IN RECTANGULAR AT-CUT QUARTZ PLATES

Hitoshi SEKIMOTO and Yasuaki WATANABE
Tokyo Metropolitan University, Tokyo, 192-03 Japan
Mitsuo NAKAZAWA
Shinsyu University, Nagano, 380 Japan

Summary

In this paper, the two-dimensional spurious vibrations of thickness-flexure, face-shear and face-flexure in a rectangular AT-cut quartz plate are theoretically studied. The input capacitance characteristics and the equivalent inductance are calculated, and compared with the experimental data.

I. Introduction

Strip AT-cut quartz resonators have been recently developed and widely used in frequency applications because of their small size. Many spurious vibrations can be supported in the resonators employing rectangular plates. Some vibrations are electrically strong modes that Mindlin and Gazis[1] have investigated. The others belong to the modes with phase reversals along both the X and Z' axes, and therefore have weak electrical responses. Such a weak mode, however, may affect the fundamental thickness-shear mode when the frequencies of two modes approach each other, as well as the strong mode. This motivates the theoretical study of the two-dimensional spurious vibrations. Based on the Mindlin's two-dimensional plate equations [1], we have already developed a one-dimensional finite element program for calculating a mode chart of the thickness-flexure, face-shear and face-flexure vibrations in a rectangular AT-cut plate [2]. We have also clarified that most of spurious resonances belong to these families with phase reversals along both the X and Z' axes. However, the program was applicable only to the analysis of resonant frequencies and vibrational patterns. The analysis of the electrical characteristics is necessary to evaluate the strength of spurious modes.

In this paper, we analyze the forced vibrations of spurious modes to estimate their electrical behavior. The input capacitance of a rectangular plate is evaluated by solving the Mindlin's two-dimensional plate equations under the assumption of weak piezoelectric coupling [3]. Then the equivalent inductance of each resonance can be determined from the capacitance characteristics. In the present analysis we consider two models based on the plate equations:

1) a model with the elastic coupling of thickness-flexure to face-shear and face-flexure taken into account, which consists of five coupled equations.

2) a model with the coupling omitted, which is represented by the three coupled equations and/or the plane-stress equations.

Using a new version of the one-dimensional finite element program, the input capacitance and equivalent inductance are calculated. The results show that the former model gives better agreement with experiment.

II Input Capacitance (C_{in})

Fig. 1 shows a rectangular AT-cut quartz plate of thickness 2b, length(x_1) 2a and width(x_3) 2c that is fully plated on top and bottom surface by electrode platings of infinitesimal thickness. The input capacitance C_{in} can be derived from the relation between the surface charge q and the driving voltage V.

Under the approximations of weak piezoelectric coupling [3] and Mindlin's plate equations [1], the relation may be written as

$$q = C_{in} \cdot V = -\int D_2 dA_e$$
$$= C_d \cdot V - \int (k_1 e_{26} S_6 + e_{25} S_5) dA_e, \qquad (1)$$

where $\int \cdot dA_e$ means the integration over the plated area, D_2 is the electric displacement

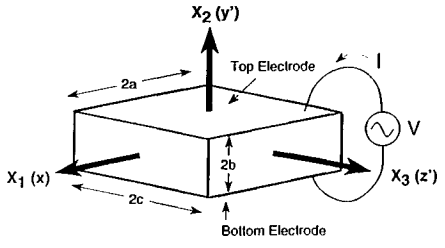


Fig.1 A rectangular AT-cut quartz plate

0-7803-0476-4/92 $3.00 © 1992 IEEE

along the x_2 direction, C_d (= $2\varepsilon_{22}ac/b$) is the clamped capacitance, e_{25},e_{26} are the piezoelectric constants, ε_{22} is the dielectric constant and $k_1^2(=\pi^2/12)$ is the correction factor. The thickness-shear strain S_6 and face-shear strain S_5 are represented by $u_{2,1}+\phi_1$ and $u_{1,3}+u_{3,1}$, respectively, according to Mindlin's notation. These two shear strains can be determined from the solutions of the following coupled equations

$$\begin{matrix} N_{1,1}+N_{5,3}+2b\rho\omega^2u_1 \\ N_{5,1}+N_{3,3}+2b\rho\omega^2u_3 \end{matrix} \quad (2)$$

$$\begin{matrix} Q_{1,1}+Q_{3,3}+2b\rho\omega^2u_2=0 \\ M_{1,1}+M_{5,3}-Q_1+(2/3)b^3\rho\omega^2\phi_1=0 \\ M_{5,1}+M_{3,3}-Q_3+(2/3)b^3\rho\omega^2\phi_3=0 \end{matrix} \quad (3)$$

Since we treat the forced vibrations, the expressions for two shear stress resultants N_5 and Q_1 in the Ref.[1] must be modified to

$$\begin{matrix} N_5=2b[c_{55}(u_{1,3}+u_{3,1})+k_1c_{56}(u_{2,1}+\phi_1)]+e_{25}V \\ Q_1=2b[k_1c_{56}(u_{1,3}+u_{3,1})+k_1^2c_{66}(u_{2,1}+\phi_1)]+k_1e_{26}V, \end{matrix} \quad (4)$$

where c_{ij} are the elastic constants.

In the present analysis, at first the one-dimensional forced solution ($V \neq 0$) and a prescribed number of two-dimensional free solutions ($V=0$), both of which satisfy the boundary conditions on x_1 edges ($N_1=N_5=Q_1=M_1=M_5=0$), are numerically determined by utilizing a one-dimensional finite element approach [2]. Then the amplitudes of two-dimensional free solutions are adjusted according to the method in Ref.[2], so that the boundary conditions on x_3 edges ($N_3=N_5=Q_3=M_3=M_5=0$) are approximately fulfilled with a sum of aforementioned solutions. As the result of this procedure, the two-dimensional forced solutions are obtained and thus C_{ln} in Eq.(1) is evaluated at a frequency ω. Also, the equivalent inductance can be obtained from the capacitance values near resonance.

The capacitance and inductance obtained here include the effect of the coupling of thickness-flexure with face-shear and face-flexure. The omission of the coupling may greatly simplify the calculations. In this case, Eq.(3) with $u_1=u_3=0$ and Eq.(2) under the plane stress approximation ($Q_1=Q_3=0$) can be separately solved.

III. Numerical Results

Fig.2 shows a mode chart of a rectangular AT-cut quartz plate without electrode plating. The experimental data denoted by dots and crosses were obtained by one of the authors, Nakazawa [2]. The plate thickness 2b was 1.737 mm, making the fundamental thickness-shear frequency about 0.95 MHz, and X-length 13.964 mm. The plate Z'-width 2c was carefully reduced from 15.3 mm to 3.2 mm in steps of approximately 0.1 mm. The curves of (a) and (b) in the figure were computed from the five coupled equations in Sec.II, and from Eq.(3) with $u_1=u_3=0$ and/or Eq.(2) under the plane stress approximation, respectively. The difference between the curves of (a) and (b) is small and both agree well, all in all, to the measured frequency spectrum. This result means that almost all of spurious resonances in a rectangular plate belong to the families of thickness-flexure, face-shear and face-flexure with phase reversals along both the X and Z' axes.

To make the difference of two approximating models clear, the input capacitance characteristics and equivalent inductances were

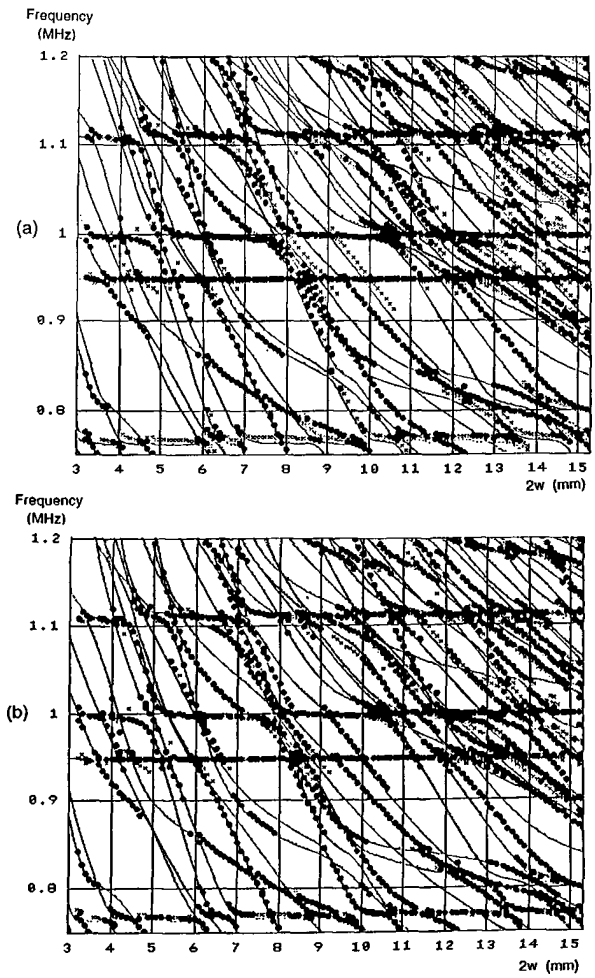


Fig.2 Comparison of calculated and measured resonant frequencies in rectangular AT-cut quartz plates (2a=13.964, 2b=1.737), as a function of Z'-width 2c. The curves of (a) and (b) in the figure were computed from the five coupled equations, and from Eq.(3) with $u_1=u_3=0$ and/or Eq.(2) under the plane stress approximation, respectively.

calculated and compared with the measured ones. The measurements were performed on three rectangular AT-cut quartz plates with full electrodes. Their Z' dimensions were 7, 10 and 14 mm, while the plate thickness and X-length were set to the same as before.

Figs. 3, 4 and 5 show the frequency characteristics of C_{in} in the vicinity of the fundamental thickness-shear resonance. The C_{in} in the figures is normalized to the clamped capacitance C_d. The curves of (a) represents the measured characteristics. The (b) and (c) were computed from the model of five-coupled equations and the model of three-coupled and/or plane-stress equations, respectively. The two strong responses can be observed in both the measured and calculated characteristics. One is the fundamental thickness-shear at about 0.95 MHz and the other is the thickness-flexure at about 1 MHz, whose displacements are almost uniform along the Z' axis. The main profiles of the calculated curves (b) and (c) appear to be the same and to fit well to those of the measured ones, since the main profiles are dominated by the two strong responses. However, in details (b) and (c) are different from each other. As seen from Fig.6, the curves (b) agree with the measured characteristics better than the curves (c).

The similar comparison for the equivalent inductances is shown in Tables I, II and III. In these tables, each inductance L_x is normalized to L_o(= $\rho b^3/4ace_{26}^2$), which means that of the pure thickness-shear mode in an infinite plate. The calculated values for the weak responses with $L_x/L_o>10$ generally disagree to the measured ones except for two strong responses. The discrepancy between the calculated and measured inductances is not entirely resolved but may be due to the limitations of plate equations. Figure 7 shows the distribution diagram of the calculated values versus the measured ones. For the model of five-coupled equations an 80 percent part of the calculated data lies within the range from 0.4 to 3.4 times as large as the measured values, while for the model of three-coupled and/or plane-stress equations is within the range from 0.2 to 8.2. This result indicates that although both models cannot give the good estimation of equivalent inductances, the former model may predict the strength of weak spurious responses better than the latter.

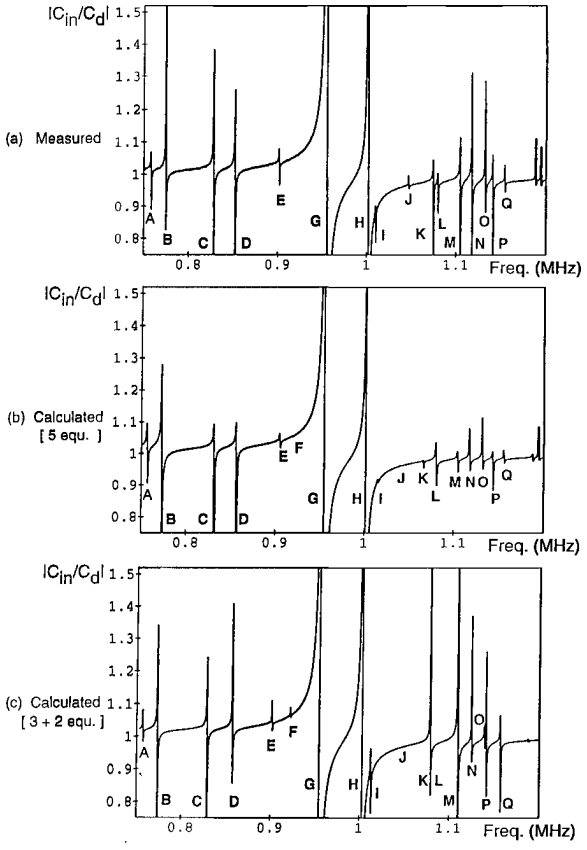


Fig.3 Input capacitance characteristics
(2a=13.964, 2b=1.737, 2c=7[mm])

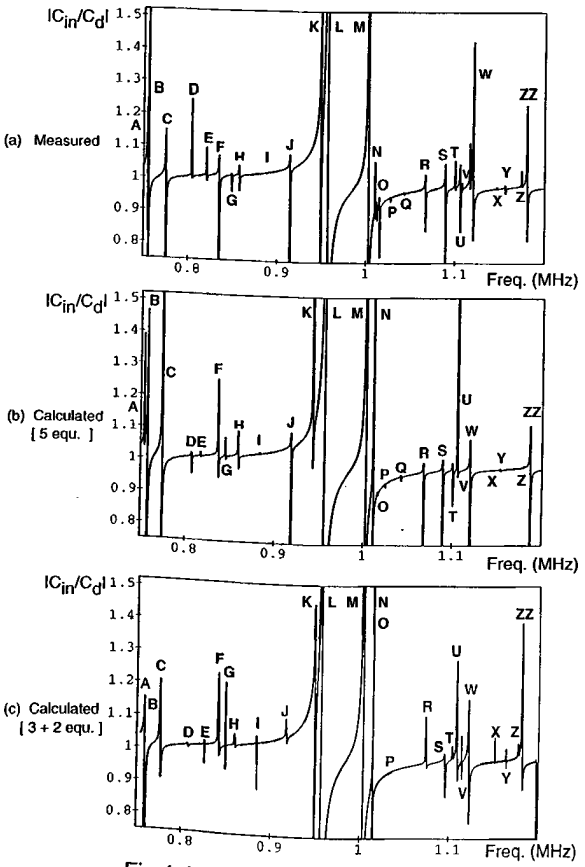


Fig.4 Input capacitance characteristics
(2a=13.964, 2b=1.737, 2c=10[mm])

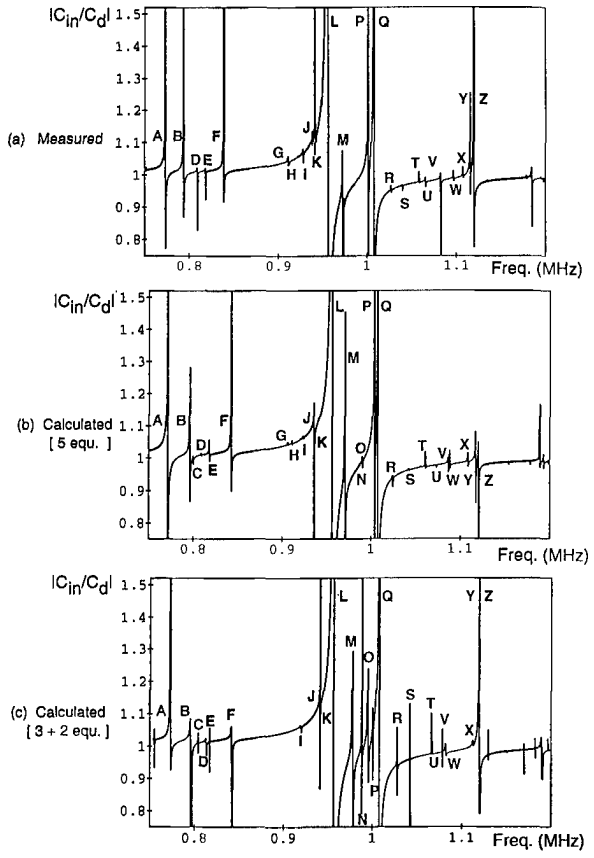


Fig.5 Input capacitance characteristics
(2a=13.964, 2b=1.737, 2c=14[mm])

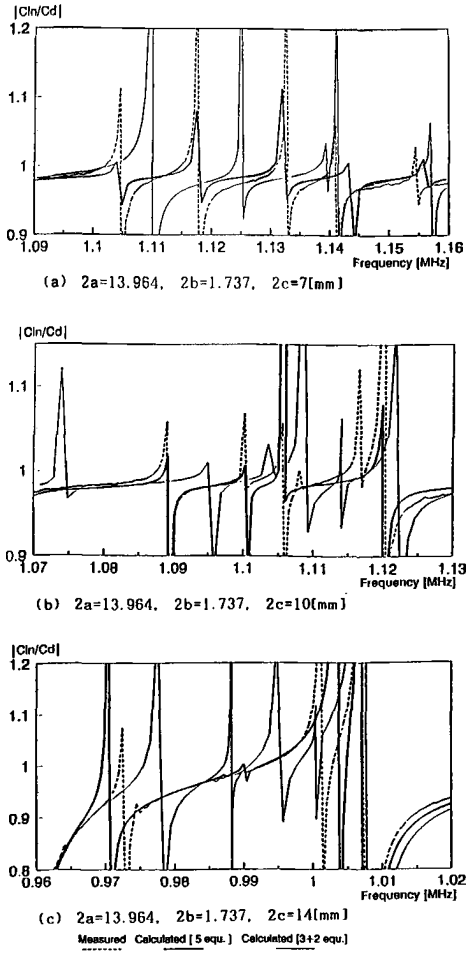


Fig.6 Input capacitance characteristics in regions where thickness-flexure is strongly coupled with face-shear and face-flexure.

Table I. Normalized Inductance L_x/L_0, L_0=1.97(H), 2b=1.737, 2a=13.964, 2c=7(mm)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
(a)	186	52	42	44	240	-	1.7	4.0	350	740	86	183	54	37	68	73	257
(b)	91	24	54	51	317	3440	1.8	3.6	880	6630	947	86	191	100	104	140	282
(c)	348	53	55	61	388	1330	1.8	3.2	125	$3\cdot10^5$	$3\cdot10^4$	44	21	106	293	52	88
	TF	TF	FS	FS	FS	TF	TS1	TF	FS	TF	TF	FS	TF	TF	TF	FS	FS

(a)Measured
(b)Calculated by using five coupled equations
(c)Calculated by using three and/or two equations
TS1:Thickness-Shear(Main), TF:Thickness-Flexure,
FS:Face-Shear and Face-Flexure

Table II. Normalized Inductance L_x/L_0, L_0=1.38(H), 2b=1.737, 2a=13.964, 2c=10(mm)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	ZZ
(a)	98	41	68	810	802	123	1300	363	6500	160	18	2.1	5.3	287	218	2140	2350	158	89	422	111	235	31	5900	899	2200	49
(b)	326	24	28	725	2020	138	845	263	4980	158	48	1.9	4.5	32	1490	3760	1600	209	162	291	288	-	62	$7\cdot10^5$	2910	7420	54
(c)	37	214	58	1270	3890	125	$8\cdot10^4$	635	$9\cdot10^4$	250	17	2.0	3.4	65	1774	$5\cdot10^4$	-	226	152	950	54	849	29	8180	1910	679	398
	FS	TF	TF	FS	TF	FS	TF	FS	TF	FS	TF	TS1	TF	TF	FS	FS		TF	FS	TF	TF	TF	TF	TF	TF	FS	FS

Table III. Normalized Inductance L_x/L_0, L_0=0.98(H), 2b=1.737, 2a=13.964, 2c=14(mm)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
(a)	44	66	-	656	925	76	957	5740	733	575	120	1.9	53	-	-	5.4	6.5	1180	2680	1730	799	281	815	820	90	26
(b)	18	82	1390	3180	531	75	8650	2020	3110	100	4060	1.9	34	4820	688	15.1	6.1	947	4930	1100	9730	513	563	578	104	94
(c)	48	60	2470	523	10^5	103	-	-	774	192	6310	1.8	24	161	37	610	4.0	560	2990	10^5	1950	10^5	320	638	$5\cdot10^4$	19
	TF	FS	TF	FS	TF	FS			FS	TF	FS	TS1	TF	FS	TF	FS	TF	TF	TF	TF	FS	TF	FS	TF	FS	TF

IV. Conclusions

In this paper, the Mindlin's two-dimensional plate equations were solved and the forced vibrations of a rectangular AT-cut plate were analyzed to investigate the electrical behavior of spurious modes. The comparison between calculation and experiment showed that although the values of equivalent inductances for weak vibrations could not be accurately evaluated from the plate equations, the model of five-coupled equations would predict the frequency and strength of weak spurious modes better than the model of three-coupled and/or plane-stress equations. In practice we need only a rough estimation of spurious strength to design a resonator. Therefore, the present analysis of the input capacitance and equivalent inductance, based on the Mindlin's plate equations, still serves as a guide for designing the rectangular AT-cut resonators.

References

[1] R.D.Mindlin and D.C.Gazis:" Strong resonances of rectangular AT-cut quartz plate", Proc. U.S. Natl. Congr. Appl. Mech., 4th, pp.305-310(1962)

[2] H.Sekimoto, Y.Watanabe and M.Nakazawa: "Two-dimensional analysis of thickness-shear and flexural vibrations in rectangular AT-cut quartz plates using a one-dimensional finite element method ", Proc. of 44th AFCS, pp.358-362(1990)

[3] H.F.Tiersten:" Linear Piezoelectric plate vibrations",Plenum(1969)

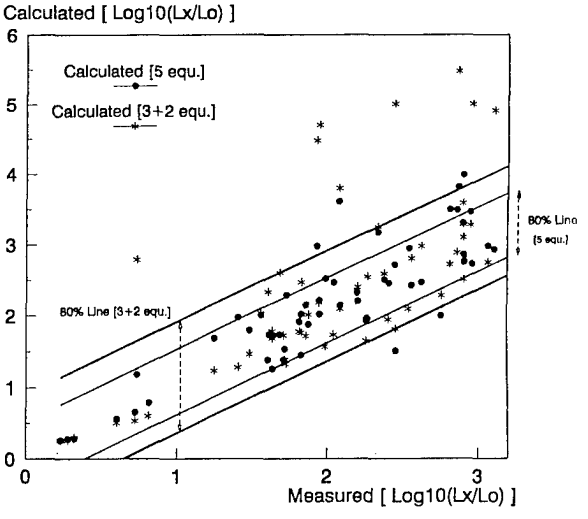


Fig.7 Equivalent inductance. (calculated vs measured)

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

VIBRATIONS OF AT-CUT QUARTZ STRIPS OF NARROW WIDTH AND FINITE LENGTH

P. C. Y. Lee and Ji Wang

Department of Civil Engineering and Operations Research
Princeton University
Princeton, NJ 08544

Abstract

A system of one-dimensional equations of motion for AT-cut quartz strip resonators with narrow width and for frequencies upto and including the fundamental thickness-shear is deduced from the two-dimensional, first-order equations for piezoelectric crystal plates(by Lee, Syngellakis, and Hou, 1987) by expanding the mechanical displacements and electric potentials in series of trigonometric functions of the width coordinate.

By neglecting the piezoelectric coupling and the weak mechanical coupling through c_{56}, four groups of coupled equations of motion are obtained. For the equations of each group, closed form solutions are obtained and the traction-free conditions at four edges are accommodated.

Dispersion curves, the resonance frequency as a function of the length-to-thickness ratio or as a function of the width-to-thickness ratio of the quartz strips are computed. Predicted results are compared with experimental data by Yamashita *et al.* with good agreement.

I. Introduction

Straight-crested wave solutions for crystal strips with a pair of parallel free edges were obtained from Mindlin's two-dimensional first-order equations for crystal plates and from Lee, Syngellakis, and Hou equations for piezoelectric crystal plates[1,2]. These solutions have been shown to predict with great accuracy the frequencies of modes of vibration that are associated with reflections from one pair of parallel edges as demonstrated by Mindlin and Gazis[1].

In the case of X-length strip resonators, for which the width is usually much smaller than the length but greater than the thickness, the wave reflections from the additional pair of free edges, which are perpendicular to the width coordinate, must be taken into account. However, two-dimensional solutions which satisfy the conditions for all four free edges are, in general, very complex and very few in existence.

In the present paper, one-dimensional equations for the modes of vibration in strip resonators with narrow width and for frequencies upto and including the fundamental thickness-shear are deduced from the two-dimensional, first-order equations for piezoelectric crystal plates, by Lee, Syngellakis, and Hou[3], by expanding the mechanical displacements and electric potentials in series of trigonometric functions of the width coordinate.

These equations are first reduced to those for crystal plate by setting piezoelectric constants e_{1p}=0. Then, since the values of the elastic stiffness c_{56} is small for AT-cut quartz, the couplings due to this elastic constant are neglected.

In order to understand the effect of length-to-thickness(a/b) and width-to-thickness(c/b) ratios of the strip on the thickness-shear resonance, four groups of the modes are selected for the detailed study. They are the thickness-shear and thickness-length flexure vibrations(TSh-tlF) and their first twist-overtone, the length-extension, width-stretch, and symmetric width-shear vibrations(E-WSh-s. WSt) and the width-shear, width-length flexure and anti-symmetric width-stretch vibrations (WSh-wlF-a. s. WSt).

For each of these four groups of coupled equations, closed form solutions are obtained and the traction-free conditions at four edges are accommodated. Dispersion curves, the resonance frequency as function of a/b ratio with fixed c/b, and the rosonance frequency as function of c/b ratio with fixed a/b are

0-7803-0476-4/92 $3.00 © 1992 IEEE

computed and compared with measured data by Yamashita *et al.*[4] with good agreement.

II. Two-Dimensional Equations of Motion

Two dimensional equations of motion of successively higher order approximation for piezoelectric crystal plates were derived by Lee, Syngellakis, and Hou [3].

A system of first-order equations may be extracted from the infinite system by letting

$$u_1^{(n)} = u_3^{(n)} = 0,\ \phi^{(n)} = 0, \text{ for } n>1,$$

$$u_2^{(n)} = 0, \qquad \text{for } n>2,$$

$$u_2^{(2)} = 0,\ T_2^{(1)} = 0,$$

$$T_{ij}^{(n)},\ \bar{T}_{ij}^{(n)},\ D_i^{(n)},\ \bar{D}_i^{(n)} = 0, \text{ for } n>1. \tag{1}$$

Hence, we have the field equations

$$T_{1,1}^{(0)} + T_{5,3}^{(0)} + \frac{1}{b}F_1^{(0)} = 2\rho(1+R)\ddot{u}_1^{(0)},$$

$$T_{6,1}^{(0)} + T_{4,3}^{(0)} + \frac{1}{b}F_2^{(0)} = \frac{2\rho}{\alpha_2^2}(1+R)\ddot{u}_2^{(0)},$$

$$T_{5,1}^{(0)} + T_{3,3}^{(0)} + \frac{1}{b}F_3^{(0)} = 2\rho(1+R)\ddot{u}_3^{(0)},$$

$$D_{1,1}^{(0)} + D_{3,3}^{(0)} + \frac{1}{b}D^{(0)} = 0,$$

$$T_{1,1}^{(1)} + T_{5,3}^{(1)} - \frac{\alpha_1}{b}T_6^{(0)} + \frac{1}{b}F_1^{(1)} = \rho(1+2R)\ddot{u}_1^{(1)},$$

$$T_{6,1}^{(1)} + T_{4,3}^{(1)} - \frac{\alpha_1}{b}T_2^{(0)} + \frac{1}{b}F_2^{(1)} = \rho(1+2R)\ddot{u}_2^{(1)},$$

$$T_{5,1}^{(1)} + T_{3,3}^{(1)} - \frac{\alpha_1}{b}T_4^{(0)} + \frac{1}{b}F_3^{(1)} = \rho(1+2R)\ddot{u}_3^{(1)},$$

$$D_{1,1}^{(1)} + D_{3,3}^{(1)} - \frac{\alpha_1}{b}D_2^{(0)} + \frac{1}{b}D^{(1)} = 0, \tag{2}$$

and the constitutive relations for rotation Y-cut of quartz

$$\frac{1}{2}T_1^{(0)} = c_{11}u_{1,1}^{(0)} + \frac{\alpha_1}{b}c_{12}u_2^{(1)} + c_{13}u_{3,3}^{(0)} + c_{14}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + e_{11}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_2^{(0)} = c_{21}u_{1,1}^{(0)} + \frac{\alpha_1}{b}c_{22}u_2^{(1)} + c_{23}u_{3,3}^{(0)} + c_{24}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + e_{12}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_3^{(0)} = c_{31}u_{1,1}^{(0)} + \frac{\alpha_1}{b}c_{32}u_2^{(1)} + c_{33}u_{3,3}^{(0)} + c_{34}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + e_{13}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_4^{(0)} = c_{41}u_{1,1}^{(0)} + \frac{\alpha_1}{b}c_{42}u_2^{(1)} + c_{43}u_{3,3}^{(0)} + c_{44}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + e_{14}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_5^{(0)} = c_{55}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{56}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{35}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{25}\phi^{(1)},$$

$$\frac{1}{2}T_6^{(0)} = c_{65}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{66}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{36}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{26}\phi^{(1)},$$

$$\frac{1}{2}D_1^{(0)} = e_{11}u_{1,1}^{(0)} + \frac{\alpha_1}{b}e_{12}u_2^{(1)} + e_{13}u_{3,3}^{(0)} + e_{14}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) - \varepsilon_{11}\phi_{,1}^{(1)},$$

$$\frac{1}{2}D_2^{(0)} = e_{25}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{26}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{22}\phi^{(1)} - \varepsilon_{23}\phi_{,3}^{(0)},$$

$$\frac{1}{2}D_3^{(0)} = e_{35}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{36}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{32}\phi^{(1)} - \varepsilon_{33}\phi_{,3}^{(0)},$$

$$T_1^{(1)} = \bar{c}_{11}u_{1,1}^{(1)} + \bar{c}_{13}u_{3,3}^{(1)} + \alpha_2\bar{c}_{14}u_{2,3}^{(1)} + \bar{e}_{11}\phi_{,1}^{(1)},$$

$$T_3^{(1)} = \bar{c}_{31}u_{1,1}^{(1)} + \bar{c}_{33}u_{3,3}^{(1)} + \alpha_2\bar{c}_{34}u_{2,3}^{(1)} + \bar{e}_{13}\phi_{,1}^{(1)},$$

$$T_4^{(1)} = \alpha_2\left(\bar{c}_{41}u_{1,1}^{(1)} + \bar{c}_{43}u_{3,3}^{(1)} + \alpha_2\bar{c}_{44}u_{2,3}^{(1)} + \bar{e}_{14}\phi_{,1}^{(1)}\right).$$

$$T_5^{(1)} = c_{55}\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) + \alpha_2 c_{56} u_{2,1}^{(1)} + e_{35}\phi_{,3}^{(1)},$$

$$T_6^{(1)} = \alpha_2\left[c_{65}\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) + \alpha_2 c_{66} u_{2,1}^{(1)} + e_{36}\phi_{,3}^{(1)}\right],$$

$$D_1^{(1)} = \bar{e}_{11} u_{1,1}^{(1)} + \bar{e}_{13} u_{3,3}^{(1)} + \alpha_2 \bar{e}_{14} u_{2,3}^{(1)} - \bar{\varepsilon}_{11}\phi_{,1}^{(1)},$$

$$D_2^{(1)} = e_{25}\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) + \alpha_2 e_{26} u_{2,1}^{(1)} - \varepsilon_{23}\phi_{,3}^{(1)},$$

$$D_3^{(1)} = e_{35}\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) + \alpha_2 e_{36} u_{2,1}^{(1)} - \varepsilon_{33}\phi_{,3}^{(1)}, \qquad (3)$$

where

$$\bar{c}_{pq} = c_{pq} - \frac{c_{p2}c_{2q}}{c_{22}}, \quad p, q = 1, 3, 4,$$

$$\bar{e}_{ip} = e_{ip} - \frac{e_{i2}e_{2p}}{e_{22}}, \quad i = 1, p = 1, 3, 4,$$

$$\bar{\varepsilon}_{ij} = \varepsilon_{ij} - \frac{e_{i2}e_{j2}}{c_{22}}, \quad i, j = 1.$$

$$\alpha_1 = \frac{\pi}{\sqrt{8}},$$

$$\alpha_2 = \frac{v}{\sqrt{\frac{c_{66}}{\rho}}}. \qquad (4)$$

In (2), $T_p^{(n)}$, n=0,1, p=1, 2, ..., 6 are the nth-order components of stress, $u_j^{(n)}$, j=1,2,3 are the nth-order components of mechanical displacement, $\phi^{(n)}$ are the nth-order electric potential, and

$$F_j^{(n)} = T_{2j}(x_2=b) - (-1)^n T_{2j}(x_2=-b),$$

$$D^{(n)} = D_2(x_2=b) - (-1)^n D_2(x_2=-b),$$

$$R = \frac{2b\rho'}{b\rho}, \qquad (5)$$

where $F_j^{(n)}$ and $D^{(n)}$ are the nth-order components of face-traction and face-charge, respectively, R is the ratio of the mass of symmetric platings per unit area to the mass of the plate per unit area, $2b$ and $2b'$ are the thickness of the plate and the platings, and ρ and ρ' are the mass densities of the plate and the plating, respectively.

In (5), $F_j^{(n)} = 0$ for faces free of traction and

$$D^{(0)} = 0,$$

$$D^{(1)} = \frac{2}{b}\left(e_{26}u_1^{(1)} + e_{22}u_2^{(1)} + e_{24}u_3^{(1)}\right) - \frac{2}{b}\varepsilon_{22}\phi_0 e^{i\omega t}, \qquad (6)$$

where A=0 for the plate faces in contact with free space or dielectric platings, A=1 for the plate faces in contact with electric conducting platings and driven by an alternative voltage $\phi(x_2 = \pm b, t) = \pm\phi_0 e^{i\omega t}$. In case the platings are shorted, we set $\phi_0 = 0$. Also in the case of plates without platings, we set $R = 0$.

We note that in (3) the constitutive relations for $T_1^{(1)}$, $T_3^{(1)}$, $T_4^{(1)}$ and $D_1^{(1)}$ are slightly modified by (4) as compared with those given in Ref. 3. In (3), we have allowed the free expansion of $u_2^{(2)}$ by setting $T_2^{(1)} = 0$ from which $u_2^{(2)}$ was solved and then substituted back to the remaining constitutive relations to have $u_2^{(2)}$ eliminated.

Equations (2) to (6) form a system of two-dimensional first-order governing equations on mechanical displacements $u_i^{(0)}$ and $u_i^{(1)}$, and electric potentials $\phi^{(0)}$ and $\phi^{(1)}$.

For AT-cut quartz, the cut-off frequencies of thickness-stretch vibrations and the thickness-shear vibrations in the x_2-x_3 plane are about 110% and 15%, respectively, higher than that of the thickness-shear vibrations in the x_2-x_1 plane.

For the vibrations with frequencies upto and including the fundamental thickness-shear frequencies, equations (2)-(6) may be simplified by eliminating these two higher modes.

To eliminate the thickness-stretch mode, we let

$$u_2^{(1)} = 0, \; F_2^{(1)} = 0,$$

$$T_2^{(0)} = 0, \; T_4^{(1)} = 0, \; T_6^{(1)} = 0. \qquad (7)$$

We see that $(2)_6$ is identically satisfied by (7). By substituting constitutive relations $(3)_2$, $(3)_{12}$, and $(3)_{15}$ into $(7)_2$, we obtain, respectively

$$\frac{\alpha_1}{b}u_2^{(1)} = -\frac{1}{c_{22}}\left[c_{21}u_{1,1}^{(0)} + c_{23}u_{3,3}^{(0)} + c_{24}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right)\right] - \frac{e_{12}}{c_{22}}\phi_{,1}^{(0)},$$

$$u_{2,3}^{(1)} = -\frac{1}{\alpha_2 \bar{c}_{44}}\left(\bar{c}_{41}u_{1,1}^{(1)} + \bar{c}_{43}u_{3,3}^{(1)} + \bar{e}_{14}\phi_{,1}^{(1)}\right),$$

and

$$u_{2,1}^{(1)} = -\frac{1}{\alpha_2 c_{66}}\left[c_{65}\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) + e_{36}\phi_{,3}^{(1)}\right]. \tag{8}$$

Substitution of above into the remaining relations in (3) gives the modified constitutive relations in which $u_2^{(1)}$ has been eliminated:

$$\frac{1}{2}T_1^{(0)} = \bar{c}_{11}u_{1,1}^{(0)} + \bar{c}_{13}u_{3,3}^{(0)} + \bar{c}_{14}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + \bar{e}_{11}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_3^{(0)} = \bar{c}_{31}u_{1,1}^{(0)} + \bar{c}_{33}u_{3,3}^{(0)} + \bar{c}_{34}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + \bar{e}_{13}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_4^{(0)} = \bar{c}_{41}u_{1,1}^{(0)} + \bar{c}_{43}u_{3,3}^{(0)} + \bar{c}_{44}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) + \bar{e}_{14}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_5^{(0)} = c_{55}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{56}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{35}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{25}\phi^{(1)},$$

$$\frac{1}{2}T_6^{(0)} = c_{65}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{66}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{36}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{26}\phi^{(1)},$$

$$\frac{1}{2}D_1^{(0)} = \bar{e}_{11}u_{1,1}^{(0)} + \bar{e}_{13}u_{3,3}^{(0)} + \bar{e}_{14}\left(u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}\right) - \bar{\varepsilon}_{11}\phi_{,1}^{(1)},$$

$$\frac{1}{2}D_2^{(0)} = e_{25}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{26}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{22}\phi^{(1)} - \varepsilon_{23}\phi_{,3}^{(0)},$$

$$\frac{1}{2}D_3^{(0)} = e_{35}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{36}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{32}\phi^{(1)} - \varepsilon_{33}\phi_{,3}^{(0)},$$

$$T_1^{(1)} = \hat{c}_{11}u_{1,1}^{(1)} + \hat{c}_{13}u_{3,3}^{(1)} + \hat{e}_{11}\phi_{,1}^{(1)},$$

$$T_3^{(1)} = \hat{c}_{31}u_{1,1}^{(1)} + \hat{c}_{33}u_{3,3}^{(1)} + \hat{e}_{13}\phi_{,1}^{(1)},$$

$$T_5^{(1)} = c_{55}'\left(u_{1,3}^{(1)} + u_{3,1}^{(1)}\right) + e_{35}'\phi_{,3}^{(1)},$$

$$D_1^{(1)} = \hat{e}_{11}u_{1,1}^{(1)} + \hat{e}_{13}u_{3,3}^{(1)} - \hat{\varepsilon}_{11}\phi_{,1}^{(1)},$$

$$D_2^{(1)} = e_{25}'\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) - \varepsilon_{23}'\phi_{,3}^{(1)},$$

$$D_3^{(1)} = e_{35}'\left(u_{3,1}^{(1)} + u_{1,3}^{(1)}\right) - \varepsilon_{33}'\phi_{,3}^{(1)}, \tag{9}$$

where

$$\hat{c}_{pq} = \bar{c}_{pq} - \frac{\bar{c}_{p4}\bar{c}_{4q}}{\bar{c}_{44}}, \quad p, q = 1, 3,$$

$$c_{55}' = c_{55} - \frac{c_{56}^2}{c_{66}},$$

$$\hat{e}_{ip} = \bar{e}_{ip} - \frac{\bar{e}_{i4}\bar{c}_{4p}}{\bar{c}_{44}}, \quad i = 1, p = 1, 3,$$

$$e_{ip}' = e_{ip} - \frac{e_{i6}c_{6p}}{c_{66}}, \quad i = 2, 3, p = 6,$$

$$\hat{\varepsilon}_{ij} = \bar{\varepsilon}_{ij} - \frac{\bar{e}_{i4}\bar{e}_{j4}}{\bar{c}_{44}}, \quad i = 1, j = 1,$$

$$\varepsilon_{ij}' = \varepsilon_{ij} - \frac{e_{i6}e_{j6}}{c_{66}}, \quad i, j = 2, 3. \tag{10}$$

We note that (9) and (2) with $(2)_6$ discarded, thus, form a system of governing equations on $u_i^{(0)}$, $u_\alpha^{(1)}$, $\phi^{(0)}$and $\phi^{(1)}$, where i=1, 2, 3 and α=1, 3. These equations are similar to Mindlin's equations derived by power series expansion[5].

To remove the thickness-shear vibrations in the x_2-x_3 plane, *i.e.* $u_3^{(1)}$, we follow Mindlin-Spencer's flexure-twist approximation[6]:

(a) In $(2)_7$, by neglecting the inertia term $\rho(1+2R)\ddot{u}_3^{(1)}$, allowing $u_3^{(1)}$ to have free development by letting $T_3^{(1)} = 0$, and setting $F_3^{(1)} = 0$ for the free faces, we have

$$\frac{\alpha_1}{b}T_4^{(0)} = T_{5,1}^{(1)},$$

$$u_{3,3}^{(1)} = -\frac{1}{\hat{c}_{33}}\left(\hat{c}_{31}u_{1,1}^{(1)} + \hat{e}_{13}\phi^{(1)}\right). \tag{11}$$

(b) In (9), by passing to the limit as the transverse shearing strain $u_{2,3}^{(0)} + \frac{\alpha_1}{b}u_3^{(1)}$ approaches to zero and the stiffness coefficient $\bar{c}_{44}$ approaches to infinity, so that

$T_4^{(0)}$ remain finite but indeterminant. Therefore , we have

$$u_{2,3}^{(0)} = -\frac{\alpha_1}{b}u_3^{(1)}, \tag{12}$$

and $T_4^{(0)}$ can be calculated in $(11)_1$.

By substituting $(11)_1$ into $(2)_2$ and letting $F_j^{(0)} = F_j^{(1)} = 0$ for the traction-free faces, we obtain the reduced field equation

$$T_{1,1}^{(0)} + T_{5,3}^{(0)} = 2\rho(1+R)\ddot{u}_1^{(0)},$$

$$T_{6,1}^{(0)} + \frac{b}{\alpha_1}T_{5,13}^{(1)} = \frac{2\rho}{\alpha_2^2}(1+R)\ddot{u}_2^{(0)},$$

$$T_{5,1}^{(0)} + T_{3,3}^{(0)} = 2\rho(1+R)\ddot{u}_3^{(0)},$$

$$D_{1,1}^{(0)} + D_{3,3}^{(0)} = 0,$$

$$T_{1,1}^{(1)} + T_{5,3}^{(1)} - \frac{\alpha_1}{b}T_6^{(0)} = \rho(1+2R)\ddot{u}_1^{(1)},$$

$$D_{1,1}^{(1)} + D_{3,3}^{(1)} - \frac{\alpha_1}{b}D_2^{(0)} + \frac{1}{b}D^{(1)} = 0. \tag{13}$$

And insertion of $(11)_2$ and (12) into (9) yields the constitutive relations in which both $u_2^{(1)}$ and $u_3^{(1)}$ are removed as follows:

$$\frac{1}{2}T_1^{(0)} = \bar{c}_{11}u_{1,1}^{(0)} + \bar{c}_{13}u_{3,3}^{(0)} + \bar{e}_{11}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_3^{(0)} = \bar{c}_{31}u_{1,1}^{(0)} + \bar{c}_{33}u_{3,3}^{(0)} + \bar{e}_{13}\phi_{,1}^{(0)},$$

$$\frac{1}{2}T_5^{(0)} = c_{55}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{56}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{35}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{25}\phi^{(1)},$$

$$\frac{1}{2}T_6^{(0)} = c_{65}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + c_{66}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) + e_{36}\phi_{,3}^{(0)} + \frac{\alpha_1}{b}e_{26}\phi^{(1)},$$

$$\frac{1}{2}D_1^{(0)} = \bar{e}_{11}u_{1,1}^{(0)} + \bar{e}_{13}u_{3,3}^{(0)} - \bar{\varepsilon}_{11}\phi_{,1}^{(1)},$$

$$\frac{1}{2}D_2^{(0)} = e_{25}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{26}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{22}\phi^{(1)} - \varepsilon_{23}\phi_{,3}^{(0)},$$

$$\frac{1}{2}D_3^{(0)} = e_{35}\left(u_{3,1}^{(0)} + u_{1,3}^{(0)}\right) + e_{36}\left(u_{2,1}^{(0)} + \frac{\alpha_1}{b}u_1^{(1)}\right) - \frac{\alpha_1}{b}\varepsilon_{32}\phi^{(1)} - \varepsilon_{33}\phi_{,3}^{(0)},$$

$$T_1^{(1)} = \tilde{c}_{11}u_{1,1}^{(1)} + \tilde{e}_{11}\phi_{,1}^{(1)},$$

$$T_5^{(1)} = c_{55}'\left(u_{1,3}^{(1)} - \frac{b}{\alpha_1}u_{2,31}^{(0)}\right) + e_{35}'\phi_{,3}^{(1)},$$

$$D_1^{(1)} = \tilde{e}_{11}u_{1,1}^{(1)} - \tilde{\varepsilon}_{11}\phi_{,1}^{(1)},$$

$$D_2^{(1)} = e_{25}'\left(u_{1,3}^{(1)} - \frac{b}{\alpha_1}u_{2,31}^{(0)}\right) - \varepsilon_{23}'\phi_{,3}^{(1)},$$

$$D_3^{(1)} = e_{35}'\left(u_{1,3}^{(1)} - \frac{b}{\alpha_1}u_{2,31}^{(0)}\right) - \varepsilon_{33}'\phi_{,3}^{(1)}, \tag{14}$$

where

$$\tilde{c}_{11} = \hat{c}_{11} - \frac{\hat{c}_{13}\hat{c}_{31}}{\hat{c}_{33}},$$

$$\tilde{e}_{11} = \hat{e}_{11} - \frac{\hat{e}_{13}\hat{e}_{31}}{\hat{e}_{33}},$$

$$\tilde{\varepsilon}_{11} = \hat{\varepsilon}_{11} - \frac{\hat{\varepsilon}_{13}\hat{\varepsilon}_{31}}{\hat{\varepsilon}_{33}}. \tag{15}$$

III. One-Dimensional Equations for Strip Resonators

In a strip resonator with four free edges, as shown in Fig. 1, the difficulty caused by the reflections of waves from the additional pair of boundaries at $x_3 = \pm c$ is the same as that even for the classical equations of the flexural vibrations of an isotropic rectangular plate, *i.e.* the solutions are not expressible in terms of a finite number of elementary functions.

However, since in a strip resonator the width is usually much smaller than the length, the components of mechanical displacement $u_i^{(0)}(i=1,2,3)$, $u_1^{(1)}$,and electric potentials ϕ_0 and ϕ_1 may be expanded further in terms of trigonometric functions of the width coordinate and, thus, an infinite set of one-dimensional equations can be deduced from the two-dimensional equations (13) and (14).

We let

$$u_i^{(0)} = \sum_{r=0}^{\infty} u_i^{(0r)}(x_1, t)\cos\frac{r\pi}{2}(1 - \psi_3),\ i = 1, 2, 3,$$

$$u_1^{(1)} = \sum_{r=0}^{\infty} u_1^{(1r)}(x_1, t)\cos\frac{r\pi}{2}(1 - \psi_3),$$

$$\phi^{(0)} = \sum_{n=0}^{\infty} \phi^{(0r)}(x_1, t)\cos\frac{r\pi}{2}(1 - \psi_3),$$

$$\phi^{(1)} = \sum_{n=0}^{\infty} \phi^{(1r)}(x_1, t)\cos\frac{r\pi}{2}(1 - \psi_3), \qquad (16)$$

where $r = 0, 1, 2, \cdots$, and $\psi_3 = x_3/c$.

By substituting (16) into (13) and multiplying these equations by $\cos r\pi(1 - \psi_3)/2$ and integrating with respect to ψ_3 from -1 to +1 in a manner similar to that employed in Lee, Syngellakis and Hou[3], we obtain

$$T_{1,1}^{(0s)} - \frac{s\pi}{2c}\bar{T}_5^{(0s)} + \frac{1}{c}\mathcal{T}_5^{(0s)} = \frac{2\rho}{\alpha_2^{2\delta_{0s}}}(1+R)(1 + \delta_{s0})\ddot{u}_1^{(0s)},$$

$$T_{6,1}^{(0s)} - \frac{s\pi}{2c}\frac{b}{\alpha_1}\bar{T}_{5,1}^{(1s)} + \frac{1}{c}\mathcal{T}_4^{(0s)} = 2\rho(1+R)(1 + \delta_{s0})\ddot{u}_2^{(0s)},$$

$$T_{5,1}^{(0s)} - \frac{s\pi}{2c}\bar{T}_3^{(0s)} + \frac{1}{c}\mathcal{T}_3^{(0s)} = \frac{2\rho}{\alpha_2^{2\delta_{0s}}}(1+R)(1 + \delta_{s0})\ddot{u}_3^{(0s)},$$

$$D_{1,1}^{(0s)} - \frac{s\pi}{2c}\bar{D}_3^{(0s)} + \frac{1}{c}\mathcal{D}_3^{(0s)} = 0,$$

$$T_{1,1}^{(1s)} - \frac{s\pi}{2c}\bar{T}_5^{(1s)} - \frac{\alpha_1}{b}T_6^{(0s)} + \frac{1}{c}\mathcal{T}_5^{(1s)}$$

$$= \rho(1+2R)(1 + \delta_{s0})\ddot{u}_1^{(1s)},$$

$$D_{1,1}^{(1s)} - \frac{s\pi}{2c}\bar{D}_3^{(1s)} - \frac{\alpha_1}{b}D_2^{(0s)} + \frac{1}{c}\mathcal{D}_3^{(1s)} + \frac{1}{b}D^{(1s)} = 0, \qquad (17)$$

where

$$T_p^{(ns)} = \int_{-1}^{1} T_p^{(n)}\cos\frac{s\pi}{2}(1 - \psi_3)d\psi_3,$$

$$\bar{T}_p^{(ns)} = \int_{-1}^{1} T_p^{(n)}\sin\frac{s\pi}{2}(1 - \psi_3)d\psi_3,$$

$$D_j^{(ns)} = \int_{-1}^{1} D_j^{(n)}\cos\frac{s\pi}{2}(1 - \psi_3)d\psi_3,$$

$$\bar{D}_3^{(ns)} = \int_{-1}^{1} D_3^{(n)}\sin\frac{s\pi}{2}(1 - \psi_3)d\psi_3,$$

$$D^{(1s)} = \int_{-1}^{1} D^{(1)}\cos\frac{s\pi}{2}(1 - \psi_3)d\psi_3.$$

$$\mathcal{T}_p^{(ns)} = T_p^{(n)}(x_3 = c) - (-1)^s T_p^{(n)}(x_3 = -c),$$

$$\mathcal{D}_3^{(ns)} = D_3^{(n)}(x_3 = c) - (-1)^s D_p^{(n)}(x_3 = -c), \qquad (18)$$

and n=0, 1, p=1, 3, 5, 6, and j=1, 2.

We note that in the present derivation of one-dimensional equations, correction factors α_1 and α_2 are introduced in the same manner as those in Ref. 3, and the following relations are also employed:

$$\int_{-1}^{1}\sin\frac{r\pi}{2}(1 - \psi_3)\sin\frac{s\pi}{2}(1 - \psi_3)d\psi_3 = \delta_{rs} - \delta_{r0}\delta_{s0},$$

$$\int_{-1}^{1}\cos\frac{r\pi}{2}(1 - \psi_3)\cos\frac{s\pi}{2}(1 - \psi_3)d\psi_3 = \delta_{rs} + \delta_{r0}\delta_{s0},$$

$$\int_{-1}^{1}\sin\frac{r\pi}{2}(1 - \psi_3)\cos\frac{s\pi}{2}(1 - \psi_3)d\psi_3 = B_{rs}$$

$$= \begin{cases} \dfrac{4r}{(r^2 - s^2)\pi}, & r + s = \text{odd} \\ 0, & r + s = \text{even}. \end{cases} \qquad (19)$$

By inserting (14) into (18) and making use of (19), we have the one-dimensional constitutive equations

$$T_1^{(0s)} = 2(1 + \delta_{0s})\left(\bar{c}_{11}u_{1,1}^{(0s)} + \bar{e}_{11}\phi_{,1}^{(0s)}\right)$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}}\alpha_r\frac{r^2}{r^2 - s^2}\bar{c}_{13}u_3^{(0r)},$$

$$T_3^{(0s)} = 2(1 + \delta_{0s})\left(\bar{c}_{31}u_{1,1}^{(0s)} + \bar{e}_{13}\phi_{,1}^{(0s)}\right)$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}}\alpha_r\frac{r^2}{r^2 - s^2}\bar{c}_{33}u_3^{(0r)},$$

$$T_5^{(0s)} = 2(1 + \delta_{0s})$$

$$\times\left[c_{55}u_{3,1}^{(0s)} + c_{56}\left(u_{2,1}^{(0s)} + \frac{\alpha_1}{b}u_1^{(1s)}\right) + \frac{\alpha_1}{b}e_{25}\phi^{(1s)}\right]$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\left(c_{55}u_1^{(0r)} + e_{35}\phi^{(0r)}\right),$$

$$T_6^{(0s)} = 2(1+\delta_{0s})$$

$$\times\left[c_{65}u_{3,1}^{(0s)} + c_{66}\left(u_{2,1}^{(0s)} + \frac{\alpha_1}{b}u_1^{(1s)}\right) + \frac{\alpha_1}{b}e_{26}\phi^{(1s)}\right]$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\left(c_{65}u_1^{(0r)} + e_{36}\phi^{(0r)}\right),$$

$$\bar{T}_3^{(0s)} = \frac{s\pi}{c}\bar{c}_{33}u_3^{(0s)}$$

$$+ \frac{8}{\pi}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s}{s^2 - r^2}\left(\bar{c}_{31}u_{1,1}^{(0r)} + \bar{e}_{13}\phi_{,1}^{(0r)}\right),$$

$$\bar{T}_5^{(0s)} = \frac{s\pi}{c}\left(c_{55}u_1^{(0s)} + e_{35}\phi^{(0s)}\right)$$

$$+ \frac{8}{\pi}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s}{s^2 - r^2}\left[c_{55}u_{3,1}^{(0r)} + \frac{\alpha_1}{b}e_{25}\phi^{(1r)}\right]$$

$$+ \frac{8}{\pi}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s}{s^2 - r^2}c_{56}\left(u_{2,1}^{(0r)} + \frac{\alpha_1}{b}u_1^{(1r)}\right),$$

$$D_1^{(0s)} = 2(1+\delta_{0s})\left(\bar{e}_{11}u_{1,1}^{(0s)} - \bar{\varepsilon}_{11}\phi_{,1}^{(0s)}\right)$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\bar{e}_{13}u_3^{(0r)},$$

$$D_2^{(0s)} = 2(1+\delta_{0s})$$

$$\times\left[e_{25}u_{3,1}^{(0s)} + e_{26}\left(u_{2,1}^{(0s)} + \frac{\alpha_1}{b}u_1^{(1s)}\right) - \frac{\alpha_1}{b}\varepsilon_{22}\phi^{(1s)}\right]$$

$$+ \frac{4}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\left(e_{25}u_1^{(0r)} - \varepsilon_{23}\phi^{(0r)}\right),$$

$$\bar{D}_3^{(0s)} = \frac{s\pi}{c}\left(e_{35}u_1^{(0s)} - \varepsilon_{33}\phi^{(0s)}\right)$$

$$+ \frac{8}{\pi}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s}{s^2 - r^2}\left[e_{35}u_{3,1}^{(0r)} - \frac{\alpha_1}{b}\varepsilon_{32}\phi^{(1r)}\right]$$

$$+ \frac{8}{\pi}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s}{s^2 - r^2}e_{36}\left(u_{2,1}^{(0r)} + \frac{\alpha_1}{b}u_1^{(1r)}\right),$$

$$T_1^{(1s)} = (1+\delta_{0s})\left(\tilde{c}_{11}u_{1,1}^{(1s)} + \tilde{e}_{11}\phi_{,1}^{(1s)}\right),$$

$$\bar{T}_5^{(1s)} = \frac{s\pi}{2c}\left[c_{55}'\left(u_1^{(1s)} - \frac{b}{\alpha_1}u_{2,1}^{(0s)}\right) + e_{35}'\phi^{(1s)}\right],$$

$$\bar{D}_1^{(1s)} = (1+\delta_{0s})\left(\tilde{e}_{11}u_{1,1}^{(1s)} - \tilde{\varepsilon}_{11}\phi_{,1}^{(1s)}\right),$$

$$\bar{D}_3^{(1s)} = \frac{s\pi}{2c}\left[e_{35}'\left(u_1^{(1s)} - \frac{b}{\alpha_1}u_{2,1}^{(0s)}\right) - \varepsilon_{33}'\phi^{(1s)}\right],$$

$$D^{(1s)} = \frac{2A}{b}\left[(1+\delta_{0s})e_{26}u_1^{(1s)} - 2\delta_{0s}\varepsilon_{22}\phi_0 e^{i\omega t}\right], \qquad (20)$$

where

$$f = \delta_{0s}\delta_{1r} = \begin{cases} 1, & \text{if } s=0,\ r=1 \\ 0, & \text{otherwise} \end{cases},$$

$$g = \delta_{1s}\delta_{0r} = \begin{cases} 1, & \text{if } s=1,\ r=0 \\ 0, & \text{otherwise} \end{cases}. \qquad (21)$$

Substitution of (20) into (17) yields 6s one-dimensional equations on the 6s variables($u_i^{(0s)}$, $u_1^{(1s)}$, $\phi^{(0s)}$ and $\phi^{(1s)}$)

$$(1+\delta_{0s})\left(\bar{c}_{11}u_{1,11}^{(0s)} + \bar{e}_{11}\phi_{,11}^{(0s)}\right)$$

$$-\left(\frac{s\pi}{2c}\right)^2\left(c_{55}u_1^{(0s)} + e_{35}\phi^{(0s)}\right)$$

$$+ \frac{2}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\bar{c}_{13}u_{3,1}^{(0r)}$$

$$- \frac{2}{c}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s^2}{s^2 - r^2}\left[c_{55}u_{3,1}^{(0r)} + \frac{\alpha_1}{b}c_{25}\phi^{(1r)}\right]$$

$$- \frac{2}{c}\sum_{r+s=\text{odd}} \alpha_1^g \frac{s^2}{s^2 - r^2}c_{56}\left(u_{2,1}^{(0r)} + \frac{\alpha_1}{b}u_1^{(1r)}\right) + \frac{1}{2c}\mathcal{F}_5^{(0s)}$$

$$= \rho(1+R)(1+\delta_{s0})\ddot{u}_1^{(0s)},$$

$$(1+\delta_{0s})\left[c_{65}u_{3,11}^{(0s)} + c_{66}\left(u_{2,11}^{(0s)} + \frac{\alpha_1}{b}u_{1,1}^{(1s)}\right) + \frac{\alpha_1}{b}e_{26}\phi_{,1}^{(1s)}\right]$$

$$-\left(\frac{s\pi}{2c}\right)^2\frac{b}{2\alpha_1}\left[c_{55}'\left(u_{1,1}^{(1s)} - \frac{b}{\alpha_1}u_{2,11}^{(0s)}\right) + e_{35}'\phi_{,1}^{(1s)}\right]$$

$$+ \frac{2}{c}\sum_{r+s=\text{odd}} \alpha_1^f \frac{r^2}{r^2 - s^2}\left(c_{65}u_{1,1}^{(0r)} + e_{36}\phi_{,1}^{(0r)}\right) + \frac{1}{2c}\mathcal{F}_4^{(0s)}$$

$$= \rho(1+R)(1+\delta_{s0})\ddot{u}_2^{(0s)},$$

$$(1+\delta_{0s})\left[c_{55}u_{3,11}^{(0s)} + c_{56}\left(u_{2,11}^{(0s)} + \frac{\alpha_1}{b}u_{1,1}^{(1s)}\right) + \frac{\alpha_1}{b}e_{25}\phi_{,1}^{(1s)}\right]$$

$$-\left(\frac{s\pi}{2c}\right)^2\bar{c}_{33}u_3^{(0s)}$$

$$+\frac{2}{c}\sum_{r+s=\text{odd}}\alpha_1^f\frac{r^2}{r^2-s^2}\left(c_{55}u_{1,1}^{(0r)}+e_{35}\phi_{,1}^{(0r)}\right)$$

$$-\frac{2}{c}\sum_{r+s=\text{odd}}\alpha_1^g\frac{s^2}{r^2-s^2}\left(\bar{c}_{31}u_{1,1}^{(0r)}+\bar{e}_{13}\phi_{,1}^{(0r)}\right)+\frac{1}{2c}\mathcal{F}_3^{(0s)}$$

$$=\frac{\rho}{\alpha_2^{2\delta_{0s}}}(1+R)(1+\delta_{s0})\ddot{u}_3^{(0s)},$$

$$(1+\delta_{0s})\left(\bar{e}_{11}u_{1,11}^{(0s)}-\bar{\varepsilon}_{11}\phi_{,11}^{(0s)}\right)-\left(\frac{s\pi}{2c}\right)^2\left(e_{35}u_1^{(0s)}-\varepsilon_{33}\phi^{(0s)}\right)$$

$$-\frac{2}{c}\sum_{r+s=\text{odd}}\alpha_1^g\frac{s^2}{s^2-r^2}\left[e_{35}u_{3,1}^{(0r)}-\frac{\alpha_1}{b}\varepsilon_{32}\phi^{(1r)}\right]$$

$$-\frac{2}{c}\sum_{r+s=\text{odd}}\alpha_1^g\frac{s^2}{s^2-r^2}e_{36}\left(u_{2,1}^{(0r)}+\frac{\alpha_1}{b}u_1^{(1r)}\right)$$

$$+\frac{2}{c}\sum_{r+s=\text{odd}}\alpha_1^f\frac{r^2}{r^2-s^2}\bar{e}_{13}u_{3,1}^{(0r)}+\frac{1}{2c}\mathcal{D}_3^{(0s)}=0,$$

$$(1+\delta_{0s})\left(\tilde{c}_{11}u_{1,11}^{(1s)}+\tilde{e}_{11}\phi_{,11}^{(1s)}\right)$$

$$-\left(\frac{s\pi}{2c}\right)^2\left[c_{55}'\left(u_1^{(1s)}-\frac{b}{\alpha_1}u_{2,1}^{(0s)}\right)+e_{35}'\phi^{(1s)}\right]$$

$$-(1+\delta_{0s})\frac{2\alpha_1}{b}$$

$$\times\left[c_{65}u_{3,1}^{(0s)}+c_{66}\left(u_{2,1}^{(0s)}+\frac{\alpha_1}{b}u_1^{(1s)}\right)+\frac{\alpha_1}{b}e_{26}\phi^{(1s)}\right]$$

$$-\frac{2}{c}\frac{2\alpha_1}{b}\sum_{r+s=\text{odd}}\alpha_1^f\frac{r^2}{r^2-s^2}\left(c_{65}u_1^{(0r)}+e_{36}\phi^{(0r)}\right)$$

$$+\frac{1}{2c}\mathcal{F}_5^{(1s)}=\rho(1+2R)(1+\delta_{s0})\ddot{u}_1^{(1s)},$$

$$(1+\delta_{0s})\left(\tilde{e}_{11}u_{1,11}^{(1s)}-\tilde{\varepsilon}_{11}\phi_{,11}^{(1s)}\right)$$

$$-\left(\frac{s\pi}{2c}\right)^2\left[e_{35}'\left(u_1^{(1s)}-\frac{b}{\alpha_1}u_{2,1}^{(0s)}\right)-\varepsilon_{33}'\phi^{(1s)}\right]$$

$$-(1+\delta_{0s})\frac{2\alpha_1}{b}$$

$$\times\left[e_{25}u_{3,1}^{(0s)}+e_{26}\left(u_{2,1}^{(0s)}+\frac{\alpha_1}{b}u_1^{(1s)}\right)-\frac{\alpha_1}{b}\varepsilon_{22}\phi^{(1s)}\right]$$

$$-\frac{2}{c}\frac{2\alpha_1}{b}\sum_{r+s=\text{odd}}\alpha_1^f\frac{r^2}{r^2-s^2}\left(e_{25}u_1^{(0r)}+\varepsilon_{23}\phi^{(0r)}\right)$$

$$+\frac{2A}{b}\left[(1+\delta_{0s})e_{26}u_1^{(1s)}-2\delta_{0s}\varepsilon_{22}\phi_0e^{i\omega t}\right]+\frac{1}{2c}\mathcal{D}_3^{(1s)}=0. \quad (22)$$

We see that (22), the 6s one-dimensional equations, are now replacing the two-dimensional equations (13) and (14) for the 6 two-dimensional variables $u_1^{(0)}$, $u_1^{(1)}$, $\phi^{(0)}$ and $\phi^{(1)}$.

IV. Vibrations of X-Length Strip Resonators of AT-cut Quartz

To examine the strong elastic coupling among the various modes, we (a) drop the piezoelectric coupling by setting e_{ip}=0, (b) neglect the weak elastic coupling due to c_{56} by setting c_{56}=0, (c) for the free edges at x_3=±c, set $\mathcal{F}_p^{(n)}=0$, and (d) retain only the modes with cut-off frequencies upto and including the fundamental thickness-shear frequencies. By expanding (22) for s=0, 1, 2, ... and implementing the above-mentioned considerations, we obtain the following four groups of coupled equations of motion:

1. Thickness-Shear(TSh) and Thickness-Length Flexure(tlL) Vibrations

The displacements corresponding to TSh and tlF modes are depicted in Fig. 2.

The displacement equations of motion are

$$\tilde{c}_{11}u_{1,11}^{(10)}-\frac{2\alpha_1}{b}c_{66}\left(u_{2,1}^{(00)}+\frac{\alpha_1}{b}u_1^{(10)}\right)=\rho(1+2R)\ddot{u}_1^{(10)},$$

$$c_{66}\left(u_{2,11}^{(00)}+\frac{\alpha_1}{b}u_{1,1}^{(10)}\right)=\frac{\rho}{\alpha_2^2}(1+R)\ddot{u}_2^{(00)}, \quad (23)$$

and the stress-displacement relations

$$T_6^{(00)}=4c_{66}\left(u_{2,1}^{(00)}+\frac{\alpha_1}{b}u_1^{(10)}\right),$$

$$T_1^{(10)}=2\tilde{c}_{11}u_{1,1}^{(10)}. \quad (24)$$

We choose the modes of vibrations to have the form

$$u_1^{(10)}=\sum_{r=1}^{2}A_{1r}\cos\xi_{rx}x_1e^{i\omega t},$$

$$u_2^{(00)}=\sum_{r=1}^{2}A_{2r}\sin\xi_{rx}x_1e^{i\omega t}, \quad (25)$$

which satisfy (23), provided

$$\begin{pmatrix} c_{66}^*\xi_{rx}^2-(1+R)\left(\frac{\Omega}{\alpha_2}\right)^2 & \frac{1}{\sqrt{2}}c_{66}^*\xi_{rx} \\ \sqrt{2}c_{66}^*\xi_{rx} & c_{11}^*\xi_{rx}^2+c_{66}^*-(1+2R)\Omega^2 \end{pmatrix}\begin{pmatrix} A_{1r} \\ A_{2r} \end{pmatrix}=0, \quad (26)$$

where the normalized frequency, wave number and elastic stiffness are defined by

$$\Omega = \frac{\omega}{\frac{\pi}{2b}\sqrt{\frac{c_{66}}{\rho}}}, \bar{\xi}_r = \frac{\xi_r}{\frac{\pi}{2b}},$$

$$c_{11}^{\bullet} = \frac{\tilde{c}_{11}}{c_{66}}, \quad c_{66}^{\bullet} = \frac{c_{66}}{c_{66}} = 1. \tag{27}$$

The vanishing of the determinants of the coefficient matrices of (26) gives the dispersion relations(r =1, 2) which yield two frequency branches as shown in Fig. 3.

For the traction-free ends of the strip, we require, at $x_1 = \pm a$

$$T_1^{(10)} = 0 \text{ and } T_6^{(00)} = 0. \tag{28}$$

Substitution of (25) into (24) and, in turn, into (28) results in

$$\sum_{r=1}^{2} \gamma_r \bar{\xi}_r \sin\xi_r a \, A_{2r} = 0,$$

$$\sum_{r=1}^{2} \left(\frac{\gamma_r}{\sqrt{2}} + \bar{\xi}_r\right) \cos\xi_r a \, A_{2r} = 0, \tag{29}$$

where

$$\gamma_r = \frac{A_{1r}}{A_{2r}}, \quad r = 1, 2, \tag{30}$$

are the amplitude ratios which can be computed from (26).

The vanishing of the determinant of the coefficients matrix of (29) gives the frequency equation which must be solved in conjunction with (26).

For computations, values of elastic stiffnesses referred to the plate axes for AT-cut quartz(θ=35.25°) are calculated from Bechmann's values[7] and they are given below.

$$[c_{pq}] = \begin{bmatrix} 86.74 & -8.26 & 27.15 & -3.65 & 0 & 0 \\ & 129.77 & -7.42 & 5.70 & 0 & 0 \\ & & 102.83 & 9.92 & 0 & 0 \\ & & & 38.61 & 0 & 0 \\ & & & & 68.81 & 2.53 \\ & & & & & 29.01 \end{bmatrix}$$

$\times 10^9$ N/m^2.

The correction factor α_2 , as it can be see in (4), depends on the phase velocity of the surface wave. For AT-cut quartz[2,3],

α_2 =0.9533.

Computational result of resonance frequency Ω as a function of the length-to-thickness ratio a/b, for a fixed width-to-thickness ratio c/b=3.78, is shown in Fig. 3. And the Ω vs. c/b curves, for a fixed a/b=16.0, are shown in Fig. 4. It is seen that the resonances are independent of the width of the strip. Therefore, the results is similar to that of straight-crested waves propagating in the x_1 direction[1].

2. First Twist-overtone of Thickness-shear(TSh) and Thickness-Length Flexure(tlF) Vibrations

The displacements associated with the first-overtone of the TSh-tlF vibrations are depicted in Fig. 6.

The displacement equations of motion are

$$c_{66}\left(u_{2,11}^{(01)} + \frac{\alpha_1}{b}u_{1,1}^{(11)}\right) - \frac{b}{2\alpha_1}\left(\frac{\pi}{2c}\right)^2 \times c_{55}'\left(u_{1,1}^{(11)} - \frac{b}{\alpha_1}u_{2,11}^{(01)}\right) = \rho(1+R)\ddot{u}_2^{(01)},$$

$$\tilde{c}_{11}u_{1,11}^{(11)} - \left(\frac{\pi}{2c}\right)^2 c_{55}'\left(u_1^{(11)} - \frac{b}{\alpha_1}u_{2,1}^{(01)}\right) - \frac{2\alpha_1}{b}c_{66}\left(u_{2,1}^{(01)} + \frac{\alpha_1}{b}u_1^{(11)}\right) = \rho(1+2R)\ddot{u}_1^{(11)}. \tag{31}$$

and the stress-displacement relations

$$T_6^{(01)} = 2c_{66}\left(u_{2,11}^{(01)} + \frac{\alpha_1}{b}u_1^{(11)}\right),$$

$$\bar{T}_5^{(11)} = \frac{\pi}{2c}c_{55}'\left(u_1^{(11)} - \frac{b}{\alpha_1}u_{2,1}^{(01)}\right),$$

$$T_1^{(11)} = \tilde{c}_{11}u_{1,1}^{(11)}. \tag{32}$$

These one-dimensional equations correspond to Mindlin-Spencer's two dimensional equations for the first twist-overtone of the thickness-shear and flexure vibrations[6].

Now, we choose the modes of the vibrations with the form

$$u_1^{(11)} = \sum_{r=1}^{2} A_{1r}\cos\xi_r x_1 e^{i\omega t},$$

$$u_2^{(01)} = \sum_{r=1}^{2} A_{2r}\sin\xi_r x_1 e^{i\omega t}, \tag{33}$$

which satisfy (32), provided

$$\begin{pmatrix} \left[1+\left(\frac{b}{c}\right)^2 c_{55}^{\bullet\bullet}\right]\xi_r^2-(1+R)\Omega^2 & \frac{1}{\sqrt{2}}\left[1-\left(\frac{b}{c}\right)^2 c_{55}^{\bullet\bullet}\right]\xi_r \\ \sqrt{2}\left[1-\left(\frac{b}{c}\right)^2 c_{55}^{\bullet\bullet}\right]\xi_r & c_{11}^{\bullet}\xi_r^2+1+c_{55}^{\bullet\bullet}\left(\frac{b}{c}\right)^2-(1+2R)\Omega^2 \end{pmatrix}\begin{pmatrix} A_{1r} \\ A_{2r} \end{pmatrix}=0. \tag{34}$$

where

$$c_{55}^{\bullet\bullet} = \frac{c_{55}'}{c_{66}}. \tag{35}$$

It can be seen from (34) that the dispersion relations depended on b/c. Dispersion curves for c/b=3.78 are computed and shown on Fig. 7.

For traction-free ends, we require[6], at $x_1=\pm a$

$$T_6^{(01)} - \frac{b}{\alpha_1}\left(\frac{\pi}{2c}\right)\bar{T}_5^{(11)}=0,$$

$$T_1^{(11)}=0. \tag{36}$$

Substitution of (31) into (36) leads to

$$\sum_{r=1}^{2} \gamma_r\bar{\xi}_r\sin\xi_r a A_{2r} = 0,$$

$$\sum_{r=1}^{2}\left(\frac{\gamma_r}{\sqrt{2}}A + B\bar{\xi}_r\right)\cos\xi_r a A_{2r} = 0, \tag{37}$$

where

$$\gamma_r = \frac{A_{1r}}{A_{2r}},\quad r = 1, 2,$$

$$A = c_{66}^{\bullet} - c_{55}^{\bullet\bullet}\left(\frac{b}{c}\right)^2,$$

$$B = c_{66}^{\bullet} + c_{55}^{\bullet\bullet}\left(\frac{b}{c}\right)^2. \tag{38}$$

In a similar manner, frequency spectra are computed from (37). Ω vs. a/b curves, for c/b=3.78, are shown in Fig. 8 and Ω vs. c/b curves, for a/b=16.0, are given in Fig. 9.

We note that results shown in Fig. 8 corresponds exactly to that obtained by Mindlin and Spencer[6].

From Fig. 9 we see that c/b has significant influence on the changes of resonance frequencies and its influence diminishes as c/b approaches to infinity.

3. Length-Extension(E), Width-Stretch(WSt) and Symmetric Width-Shear(s. WSh) Vbrations

The displacement corresponding to the E-WSt-s. WSh vibrations are depicted in Fig. 10.

The displacement equations of motion are

$$\bar{c}_{11}u_{1,11}^{(00)} + \frac{\alpha_1}{c}\bar{c}_{13}u_{3,1}^{(01)} = \rho(1+R)\ddot{u}_1^{(00)},$$

$$c_{55}u_{3,11}^{(01)} - \left(\frac{\pi}{2c}\right)^2\bar{c}_{33}u_3^{(01)} - \frac{2\alpha_1}{c}\bar{c}_{31}u_{1,1}^{(00)}$$

$$+\frac{2}{3c}(\bar{c}_{31}+4c_{55})u_{1,1}^{(02)} = \rho(1+R)\ddot{u}_3^{(01)},$$

$$\bar{c}_{11}u_{1,11}^{(02)} - \left(\frac{\pi}{c}\right)^2 c_{55}u_1^{(02)} - \frac{2}{3c}(\bar{c}_{31}+4c_{55})u_{3,1}^{(01)}$$

$$= \rho(1+R)\ddot{u}_1^{(02)}, \tag{39}$$

and the stress-displacement relations are

$$T_1^{(00)} = 4\left(\bar{c}_{11}u_{1,1}^{(00)} + \frac{\alpha_1}{c}\bar{c}_{13}u_1^{(01)}\right),$$

$$T_5^{(01)} = 2c_{55}\left(u_{3,1}^{(01)} + \frac{8}{3c}u_1^{(02)}\right),$$

$$T_1^{(02)} = 2\left(\bar{c}_{11}u_{1,1}^{(02)} - \frac{2}{3c}\bar{c}_{13}u_3^{(01)}\right). \tag{40}$$

We let

$$u_1^{(00)} = \sum_{r=1}^{3} A_{1r}\sin\xi_r x_1 e^{i\omega t},$$

$$u_3^{(01)} = \sum_{r=1}^{3} A_{2r}\cos\xi_r x_1 e^{i\omega t},$$

$$u_1^{(02)} = \sum_{r=1}^{3} A_{3r}\sin\xi_r x_1 e^{i\omega t}, \tag{41}$$

which satisfy (39), provided

$$\begin{pmatrix} c_{11}^{*}\bar{\xi}_r^2-(1+R)\Omega^2 & \frac{1}{\sqrt{2}}\frac{b}{c}c_{13}^{*}\bar{\xi}_r & 0 \\ \sqrt{2}\frac{b}{c}c_{13}^{*}\bar{\xi}_r & c_{55}^{*}\bar{\xi}_r^2+c_{33}^{*}\left(\frac{b}{c}\right)^2-(1+R)\Omega^2 & -\frac{4}{3\pi}\frac{b}{c}\left(c_{31}^{*}+4c_{55}^{*}\right)\bar{\xi}_r \\ 0 & -\frac{4}{3\pi}\frac{b}{c}\left(c_{31}^{*}+4c_{55}^{*}\right)\bar{\xi}_r & c_{11}^{*}\bar{\xi}_r^2+4c_{55}^{*}\left(\frac{b}{c}\right)^2-(1+R)\Omega^2 \end{pmatrix}\begin{pmatrix} A_{1r} \\ A_{2r} \\ A_{3r} \end{pmatrix}=0. \tag{42}$$

where

$$c_{55}^{*}=\frac{c_{55}}{c_{66}},\ c_{13}^{*}=\frac{\bar{c}_{13}}{c_{66}},\ c_{33}^{*}=\frac{\bar{c}_{33}}{c_{66}}. \tag{43}$$

The dispersion curves for c/b=3.78 are computed from (42) and plotted in Fig. 11 in which the real and imaginary parts of a complex branch are projected in real and imaginary planes, respectively.

For traction-free ends, we require, at $x_1=\pm a$

$$T_1^{(00)}=T_5^{(01)}=T_1^{(02)}=0. \tag{44}$$

Insertion of (41) into (44) leads to

$$\sum_{r=1}^{3}\left(c_{11}^{*}\gamma_{1r}\bar{\xi}_r+\frac{1}{\sqrt{2}}\frac{b}{c}c_{13}^{*}\gamma_{2r}\right)\cos\xi_r a\,A_{3r}=0,$$

$$\sum_{r=1}^{3}\left(\bar{\xi}_r-\frac{16}{3\pi}\frac{b}{c}\gamma_{2r}\right)\sin\xi_r a\,A_{3r}=0,$$

$$\sum_{r=1}^{3}\left(c_{11}^{*}\bar{\xi}_r-\frac{4}{3\pi}\frac{b}{c}c_{13}^{*}\gamma_{2r}\right)\cos\xi_r a\,A_{3r}=0, \tag{45}$$

where

$$\gamma_{sr}=\frac{A_{sr}}{A_{3r}},\quad s=1,2,\quad r=1,2,3. \tag{46}$$

From (45), Ω vs. a/b curves are computed and shown in Fig. 12 for c/b=3.78, and Ω vs. c/b curves are computed for a/b=16.0 and are shown in Fig. 13.

The width-stretch mode $u_3^{(01)}$, vibrating near the non-zero cut-off frequency where the complex frequency branch intersects with the real branch(see Fig. 11), was studied by Watanabe, Nakamura, and Shimizu[8] for developing the energy trapping resonators.

4. Width-Shear(WSh), Width-Length Flexural(wlF) and Anti-symmetric Width-Stretch(a. s. WSt) Vibrations

The displacements associated with the WSh-wlF-a. s. WSt vibrations are depicted in Fig. 14.

The displacement equations of motion are

$$c_{55}\left(u_{3,11}^{(00)}+\frac{\alpha_1}{c}u_{1,1}^{(01)}\right)=\frac{\rho}{\alpha_2^2}(1+R)\ddot{u}_3^{(00)},$$

$$\bar{c}_{11}u_{1,11}^{(01)}-\frac{2\alpha_1}{c}c_{55}\left(u_{3,1}^{(00)}+\frac{\alpha_1}{c}u_1^{(01)}\right)$$

$$+\frac{2}{3c}(c_{55}+4\bar{c}_{31})u_{3,1}^{(02)}=\rho(1+R)\ddot{u}_1^{(01)},$$

$$c_{55}u_{3,11}^{(02)}-\left(\frac{\pi}{c}\right)^2\bar{c}_{33}u_3^{(02)}-\frac{2}{3c}(c_{55}+4\bar{c}_{31})u_{1,1}^{(01)}$$

$$=\rho(1+R)\ddot{u}_3^{(02)}, \tag{47}$$

and the stress-displacement relations are

$$T_5^{(00)}=4c_{55}\left(u_{3,1}^{(00)}+\frac{\alpha_1}{c}u_1^{(01)}\right),$$

$$T_1^{(01)}=2\left(\bar{c}_{11}u_{1,1}^{(01)}+\frac{8}{3c}\bar{c}_{13}u_3^{(02)}\right),$$

$$T_5^{(02)}=2c_{55}\left(u_{3,1}^{(02)}-\frac{2}{3c}u_1^{(01)}\right). \tag{48}$$

We let

$$u_3^{(00)}=\sum_{r=1}^{3}A_{1r}\sin\xi_r x_1 e^{i\omega t},$$

$$u_1^{(01)}=\sum_{r=1}^{3}A_{2r}\cos\xi_r x_1 e^{i\omega t},$$

$$u_3^{(02)}=\sum_{r=1}^{3}A_{3r}\sin\xi_r x_1 e^{i\omega t}, \tag{49}$$

which satisfy (47), provided

$$\begin{pmatrix} c_{55}^{*}\bar{\xi}_r^2-(1+R)\left(\frac{\Omega}{\alpha_2}\right)^2 & \frac{1}{\sqrt{2}}\frac{b}{c}c_{55}^{*}\bar{\xi}_r & 0 \\ \sqrt{2}\frac{b}{c}c_{55}^{*}\bar{\xi}_r & c_{11}^{*}\bar{\xi}_r^2+c_{55}^{*}\left(\frac{b}{c}\right)^2-(1+R)\Omega^2 & -\frac{4}{3\pi}\frac{b}{c}\left(c_{55}^{*}+4c_{31}^{*}\right)\bar{\xi}_r \\ 0 & -\frac{4}{3\pi}\frac{b}{c}\left(c_{55}^{*}+4c_{31}^{*}\right)\bar{\xi}_r & c_{55}^{*}\bar{\xi}_r^2+4c_{33}^{*}\left(\frac{b}{c}\right)^2-(1+R)\Omega^2 \end{pmatrix}\begin{pmatrix} A_{1r} \\ A_{2r} \\ A_{3r} \end{pmatrix}=0. \tag{50}$$

For traction-free ends, we require, at $x_1=\pm a$

$$T_5^{(00)}=T_1^{(01)}=T_5^{(02)}=0. \tag{51}$$

By inserting (49) into (51), we obtain

$$\sum_{r=1}^{3}\left(\gamma_{1r}\bar{\xi}_r+\frac{1}{\sqrt{2}}\frac{b}{c}\gamma_{2r}\right)\cos\xi_r a\,A_{3r}=0,$$

$$\sum_{r=1}^{3}\left(c_{11}^{*}\gamma_{2r}\bar{\xi}_r-\frac{16}{3\pi}\frac{b}{c}c_{13}^{*}\right)\sin\xi_r a\,A_{3r}=0,$$

$$\sum_{r=1}^{3}\left(\bar{\xi}_r-\frac{4}{3\pi}\frac{b}{c}\gamma_{2r}\right)\cos\xi_r a\,A_{3r}=0, \tag{52}$$

where

$$\gamma_{sr}=\frac{A_{sr}}{A_{3r}},\quad s=1,2,\quad r=1,2,3. \tag{53}$$

Dispersion curves computed from (50) for AT-cut quartz strips with c/b=3.78 are shown in Fig. 15. From (50), we obtain the cut-off frequency for the a. s. WSt mode, for R=1,

$$\Omega=2\,\frac{b}{c}\left(\frac{\bar{c}_{33}}{c_{66}}\right)^{\frac{1}{2}}, \tag{54}$$

which has the value very close to 1.0(one) as it may be seen in Fig. 15.

We see from (54) that the resonance frequencies of a. s. WSt mode can be raised or lowered by adjusting the c/b ratio so that it may be separated from the resonances of the thickness-shear mode.

From (52), the Ω vs. a/b curves, for c/b=3.78, are computed and plotted in Fig. 16 and Ω vs. c/b curves, for a/b=16.0, are plotted in Fig. 17.

We, now, superimpose the Ω vs. a/b curves for the four groups of the coupled modes given in Figs. 4, 8, 12, and 16 into one as shown in Fig. 18. And the Ω vs. c/b curves by the superposition of the Figs. 5, 9, 13 and 17 are plotted in Fig. 19.

The predicted f•$2b$ vs. a/b curves for strip resonators with c/b=3.78 are compared with experimental data by Yamashita *et al*[4] as shown in Fig. 20, in which $f=\omega/2\pi$ is the circular frequency. We note that in Fig. 20 the predicted curves has been shifted downward by about 2.5 kHz•mm to match the fundamental thickness-shear frequencies.

In a similar manner, the computed f•$2b$ vs. c/b curves for strip resonators with a/b=16.0 are compared with experimental data by Yamashita *et al*[4] for a/b=15.7 as shown in Fig. 21. Again, a shift of coordinate is made to match the fundamental thickness-frequencies.

It may be seen from Fig. 20 that the complexity of the measured frequency spectrum near the thickness-shear frequencies(f•$2b$ =1654.6 kHz•mm) is mainly contributed by the fundamental thickness-shear(TSh-1) and the anti-symmetric width-stretch(a. s. WSt) modes and they are, in turn, coupled with tlF and wlF-WSh modes, respectively.

In Fig. 21, the four horizontal frequency branches, which are independent of c/b, can be identified, respectively, with predominantly tlF-18, TSh-1, TSh-3, and TSh-5 modes in the order of increasing frequencies.

In summary, by employing the presently derived one-dimensional governing equations, closed form solutions are obtained for which the traction-free conditions at four edges are accommodated.

Frequency equations for the four groups of coupled modes are obtained in analytical form in which both a/b and c/b appear as parameters. Therefore, the effect of a/b and c/b on the values of the resonance frequencies and the separation of frequencies of unwanted modes from that of the fundamental thickness-shear can be computed and analyzed.

Acknowledgment

The support to Ji Wang from Motorola Inc. through the University Partnership in Research Program is acknowledged.

References

[1] R. D. Mindlin and D. C. Gazis, " Strong resonances of rectangular AT-cut quartz plates", Proc. 4th U. S. National Congress of Applied Mechanics, 1962, pp. 305-310.

[2] P. C. Y. Lee and J. P. Hou, "Vibrations of doubly-rotated piezoelectric crystal strip with a pair of electrode-plated, traction-free edges", Proc. 39th Ann. Freq. Control Symposium, 1985, pp. 453-461.

[3] P. C. Y. Lee, S. Syngellakis, and J. P. Hou, "A two-dimensional theory for high-frequency vibrations of piezoelectric crystal plates with or without electrodes", J. Appl. Phys., **61**(4), pp. 1249-1262, Feb., 1987.

[4] S. Yamashita, N. Echigo, Y. Kawamura, A. Watanabe, and K. Kubota, "A 4.19 MHz beveled miniature rectangular AT-cut quartz resonator", Proc. 32nd Ann. Freq. Control Symposium, 1978, pp. 267-276.

[5] R. D. Mindlin, "High frequency vibrations of piezoelectric crystal plates", vol. 8, pp. 895-906, 1972.

[6] R. D. Mindlin and W. J. Spencer, "Anharmonic, thickness-twist overtones of thickness-shear and flexural vibrations of rectangular, AT-cut quartz plates", J. Acoust. Soc. Amer., **42**, pp. 1268-1277, 1967.

[7] R. Bechmann, "Elastic and piezoelectric constants of alpha-quartz", Phys. Rev., **110**, p. 1060, 1958.

[8] H. Watanabe, K. Nakamura, and H. Shimizu, "A new type of energy trapping caused by contributions from the complex branches of dispersion curves", IEEE Trans. on Sonics and Ultrasonics, vol. SU-28(4), pp. 265-270, July, 1981.

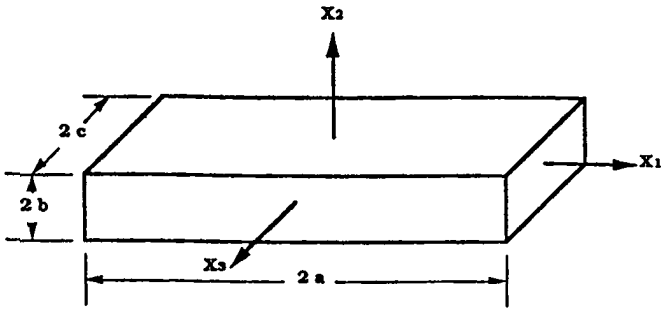


Fig. 1 An X-length strip resonator of AT-cut quartz

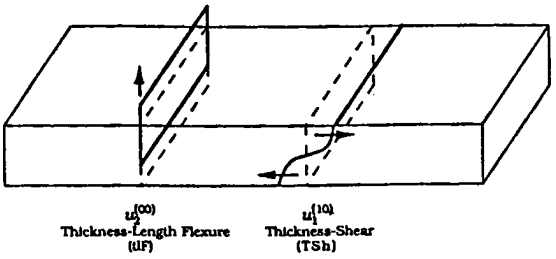


Fig. 2 Displacements associated with the thickness-shear and thickness-length flexure vibrations

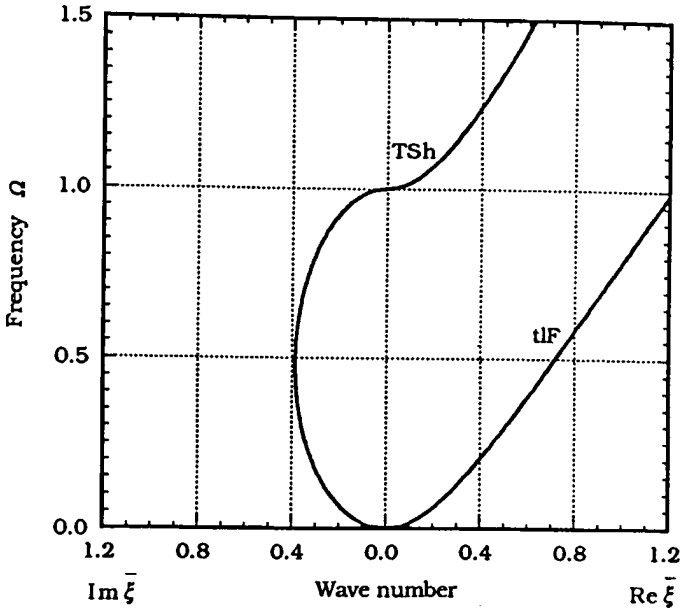


Fig. 3 Dispersion curves of coupled thickness-shear and thickness-length flexure vibrations of an AT-cut quartz strip with c/b=3.78

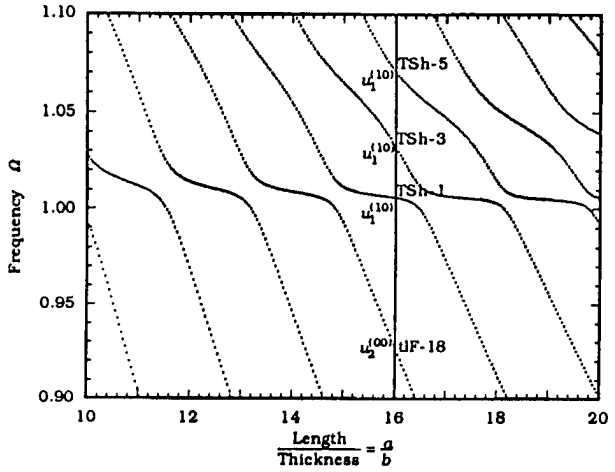


Fig. 4 Ω vs. a/b of coupled thickness-shear and thickness-length flexure vibrations in an AT-cut quartz strip with c/b=3.78

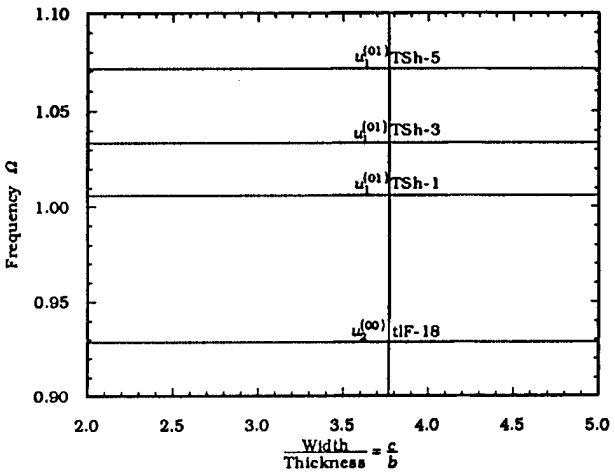


Fig. 5 Ω vs. c/b for coupled thickness-shear and thickness-length flexure vibrations in an AT-cut quartz strip with a/b=16.0

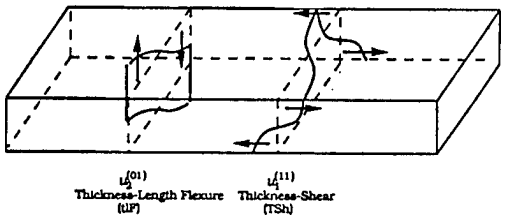


Fig. 6 Displacements associated with the first twist-overtone of thickness-shear and thickness-length flexure vibrations

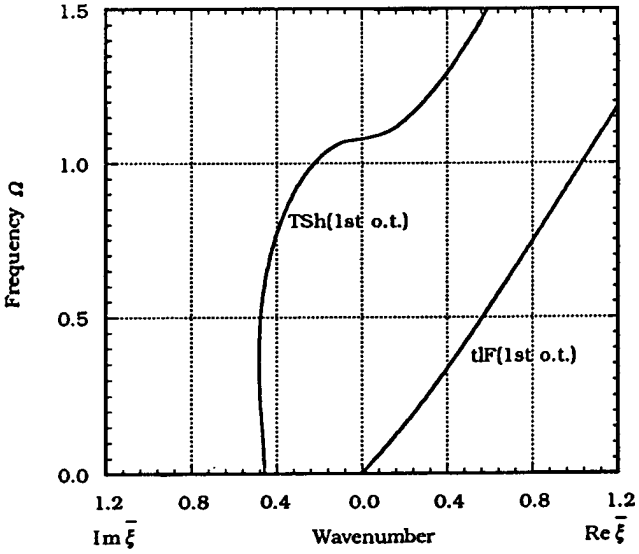


Fig 7 Dispersion curves of the coupled first twist-overtone of thickness-shear and thickness-length flexure vibrations of an AT-cut quartz strip with c/b=3.78

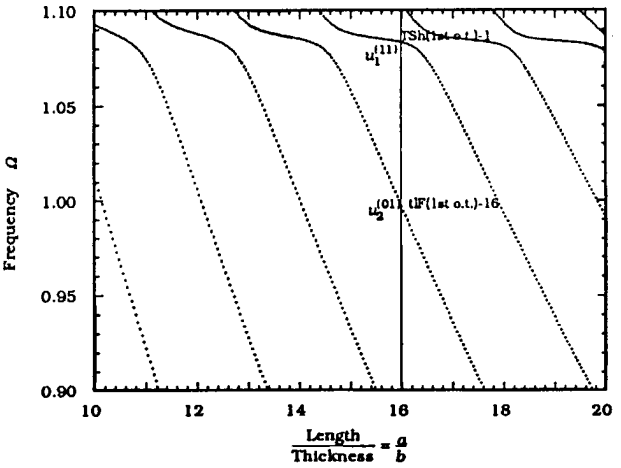


Fig. 8 Ω vs. a/b for the first twist-overtone of thickness-shear and thickness-length flexure vibrations in an AT-cut quartz strip with c/b=3.78

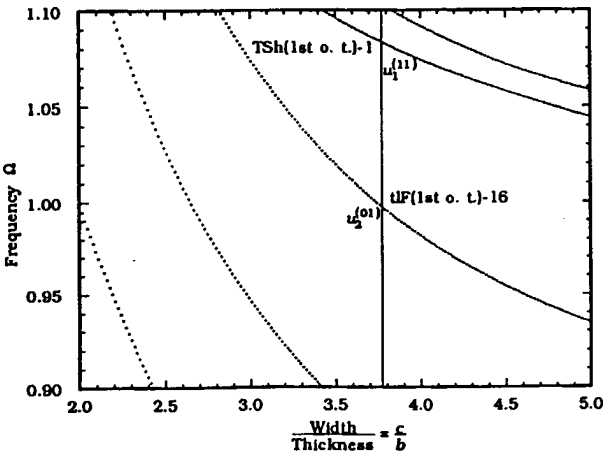


Fig. 9 Ω vs. c/b for the first twist-overtone of thickness-shear and thickness length flexure vibraions in an AT-cut quartz strip with a/b=16.0

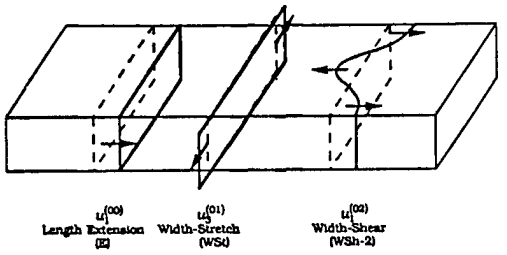


Fig. 10 Displacements associated with length-extension, width-stretch and width-shear vibrations

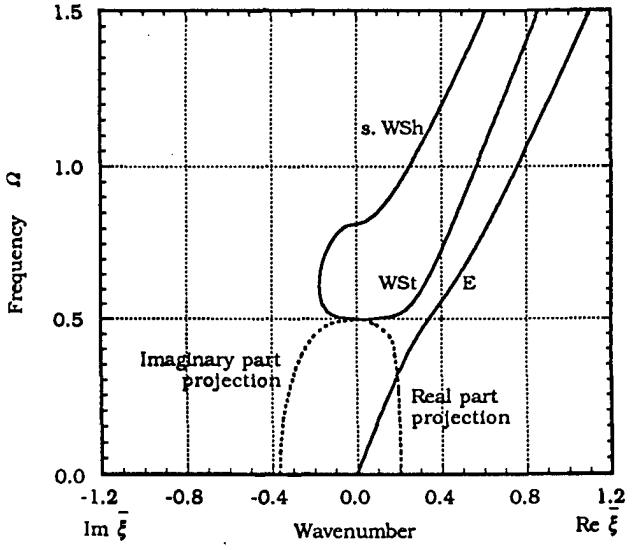


Fig. 11 Dispersion curves of the length-extension, width-stretch and symmetric width-shear vibrations of an AT-cut quartz with c/b=3.78

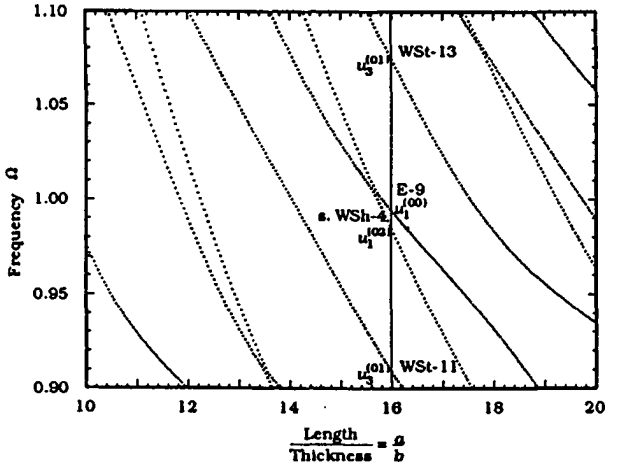


Fig. 12 Ω vs. a/b for coupled length-extension, width-stretch and s. width-shear vibrations in an AT-cut quartz strip with c/b=3.78

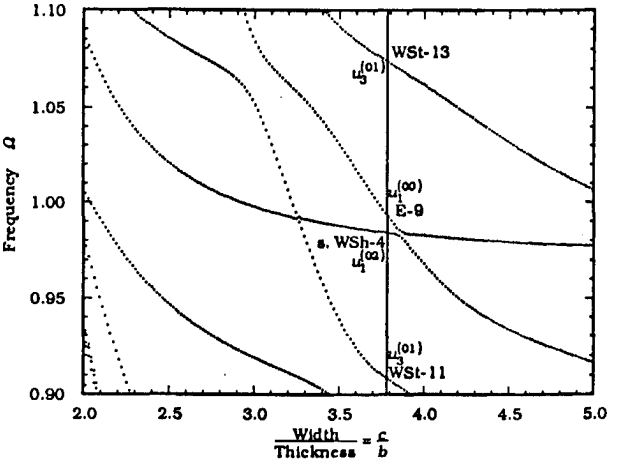


Fig. 13 Ω vs. c/b for the length-extension, width-stretch and s. width-shear vibrations in an AT-cut quartz strip with a/b=16.0

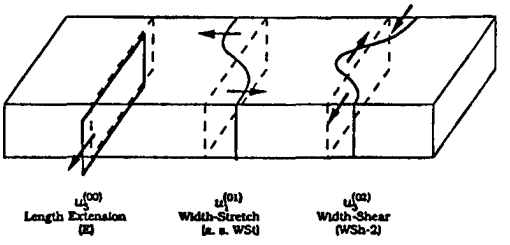


Fig. 14 Displacement associated with the width-shear, width-length flexure and anti-symmetric width-stretch vibrations

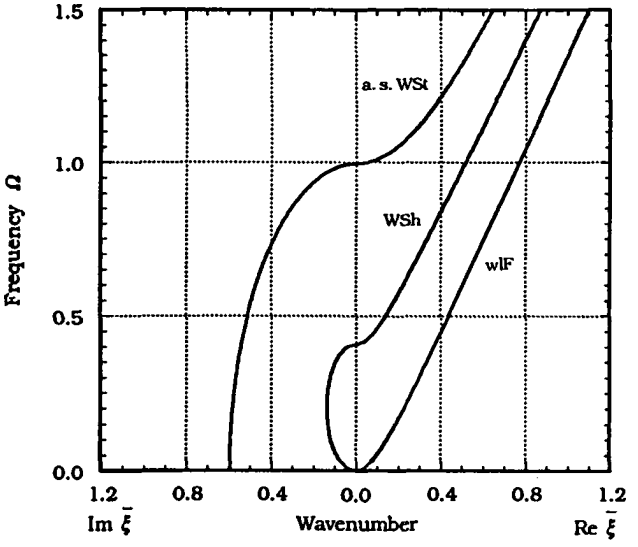


Fig. 15 Dispersion curves of the width-shear, width-length flexure and a. s. width-stretch vibrations of an AT-cut quartz strip with c/b=3.78

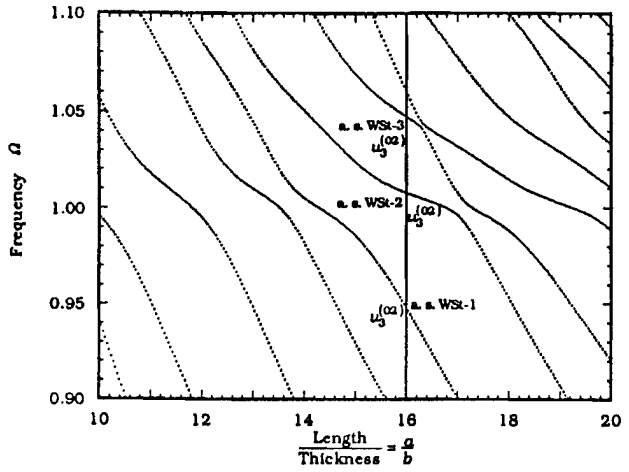


Fig. 16 Ω vs. a/b for the coupled width-shear, width-length flexure and anti-symmetric width-stretch vibrations in an AT-cut quartz strip with c/b=3.78

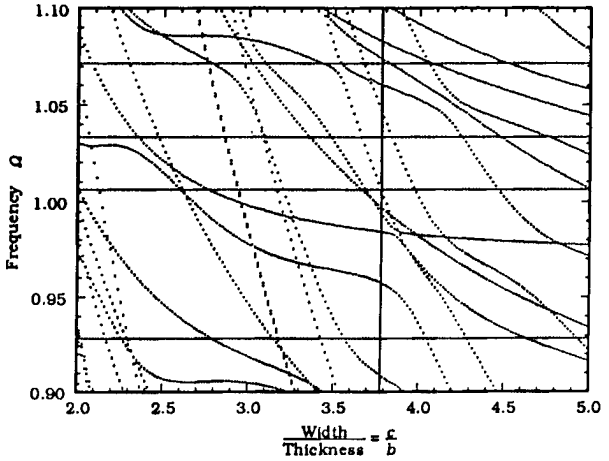


Fig. 19 Ω vs. c/b for the four groups of coupled modes in AT-cut quartz strips with a/b=16.0

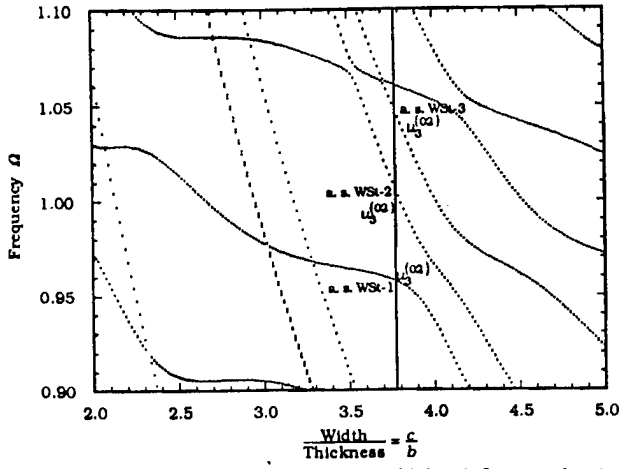


Fig. 17 Ω vs. c/b for the coupled width-shear, width-length flexure and anti-symmetric width-stretch vibrations in an AT-cut quartz strip with a/b=16.0

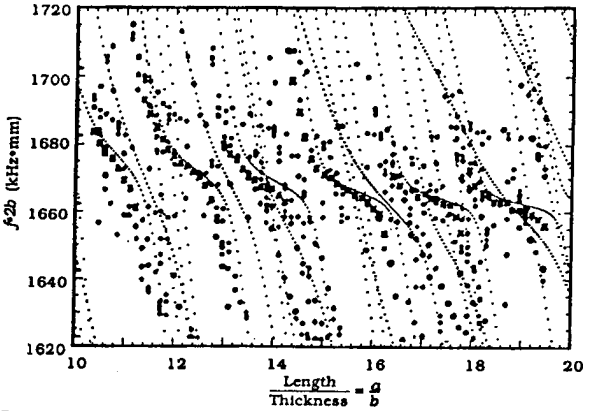


Fig. 20 Predicted f•2b vs. a/b for AT-cut quartz strips with c/b=3.78. The heavy dots and crosses are experimental data by Yamashita et al.

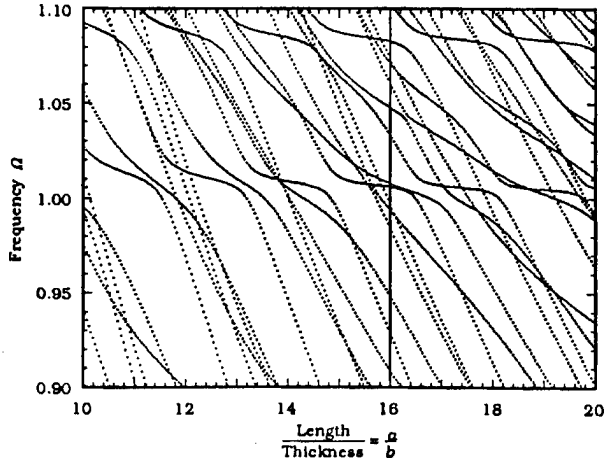


Fig. 18 Ω vs. a/b for the four groups of coupled modes in AT-cut quartz strips with c/b=3.78

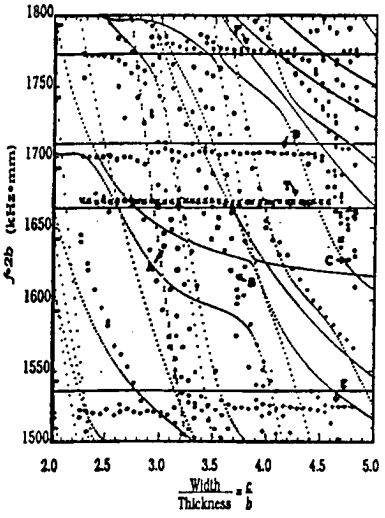


Fig. 21 Predicted f•2b vs. c/b for AT-cut quartz strips with a/b=16.0. The heavy dots and crosses are experimental data by Yamashita et al. for c/b=15.7.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LATERAL FIELD EXCITATION AND COUPLED MODES IN AT QUARTZ: SYNCHROTRON RADIATION X-RAY TOPOGRAPHY.

B.Capelle*,J.Detaint**,J.Schwartzel**,Y.Zheng*,A.Zarka*

*L.M.C.P. Université P.& M.Curie,Place Jussieu 75005 Paris France
**CNET dept.PAB/BAG/MCT/CMM 192,Av.H.Ravera 92220 Bagneux France

Abstract.

Using the synchrotron radiation delivered by the DCI storage ring at the LURE (Orsay, France), plano convex AT resonators excited by a lateral field were studied by the X-ray topography technique. It was observed different overtones and anharmonics of the three types of the thickness modes. Otherwise, coupled modes in thickness excited fundamental AT plane resonators are shown and few reasons aiding to their apparition are analysed.

Introduction.

In the lateral field plano convex AT resonators, it is possible by choice of the in plane direction of the field, to excite the overtones and the anharmonics of the three types of thickness modes, whereas only one thickness mode, the fast shear mode, is piezoelectrically excited by a strictly thickness field. But the question arises to determine whether the tridimensional modes existing in the lateral field resonators and in the thickness field resonators are identical. The use of the X-ray topography with the stroboscopic technique [1] permits to compare with a good precision the modes existing in these two kinds of resonators.

We have obtain, with the transmission and the reflexion Laue set-up, traverse topographs showing the three components of the displacement vector of different modes and could observe, on section topographs, the number of antinodal lines in the thickness of the resonator which corresponds to the overtone. Namely that we have could identify the modes by X-ray topography.

The topography, with the Laue technique, permits also to observe the coupled modes and to analyse their different components. More, the stroboscopic technique shows if the component are progressive or not.

I-The experimental set-up.

The transmission Laue set-up which uses the white X-ray beam delivered by the synchrotron is shown on the figure 1. The resonator is placed in the white beam and we obtain on the film, placed behind the crystal, a Laue pattern on which the crystal is simultaneously imaged with different diffraction vectors.The contrast observed depends of the relative position of the diffraction vector and the vibration displacement vector. In particular, the contrast due to the vibration vanishes on the topograph when these two vectors are perpendicular. To visualize a particular component of the displacement vector of a mode alone, we must choose a diffraction vector parallel to this component.

Two types of topographs can be obtained. The first one is a traverse topograph which is obtained when a large incident X ray beam is used, the second one is a section topograph when the incident beam is limited by a fine slit of about fifteen microns wide. With this second type of topograph we can see the section of the resonator and know how much antinodal lines we have in the thickness.

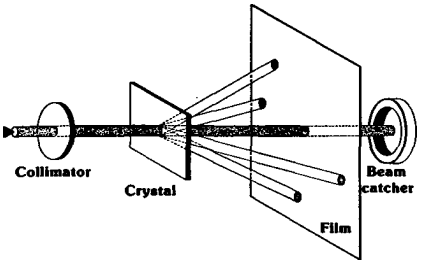


Figure-1- The transmission Laue set-up.

0-7803-0476-4/92 $3.00 © 1992 IEEE

To determine all the components of the displacement vector of a mode we must use also a diffraction vector perpendicular to the surface of the resonator. It is possible to have a diffraction vector like that, only with the reflexion Laue set-up (fig.2).

To use the stroboscopic experimental set-up the resonator must be designed to have a vibration resonant at a frequency equal to n times that of the X ray synchrotron radiation which is 3.169280 MHz.

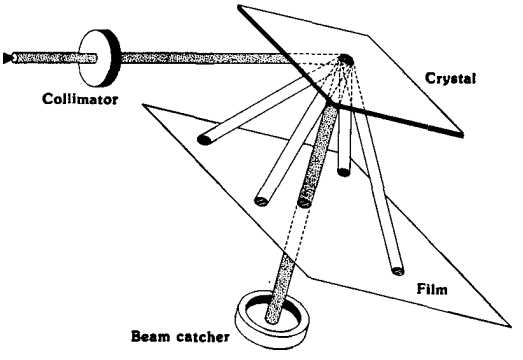


Figure-2- The reflexion Laue set-up.

II-One dimensional theory.

The figure 3 shows the axis used usually. The X, Y and Z axis are the natural ones; the X, Y' and Z' axis are the rotated axis corresponding to the AT cut, they are also named x_1, x_2 and x_3. To determine the three components u_1, u_2 and u_3 of a displacement vector **u** we must choose a diffraction vector parallel to x_1, x_2 and x_3 respectively. It is possible to have a diffraction vector parallel to x_1 or to x_3 with the transmission Laue set-up and a diffraction vector parallel to x_2 with the reflexion Laue set-up.

The one dimensional theory [2,3] of the AT cut resonators using the conventional thickness field excitation predicts (Table 1) the existence of three thickness modes: one piezoelectrically excited, the slow thickness shear mode or C mode, and two non piezoelectrically excited, the fast thickness shear mode or B mode and the thickness stretch mode or A mode. Namely that the coupling coefficient is equal to zero for the A and B modes and different to zero for the C mode

The displacement vector **u** for the C mode is exactly parallel to x_1 direction whereas the displacement vector for A and B modes makes an angle equal to about 3°.5 with the x_2 and x_3 directions respectively.

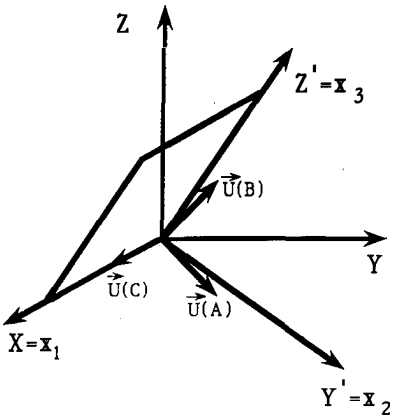
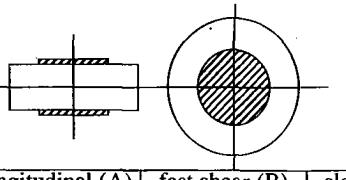


Figure-3- The different axis used usually. The X,Y and Z axis are the natural ones; the x_1, x_2 and x_3 axis are the rotated axis corresponding to the AT cut.

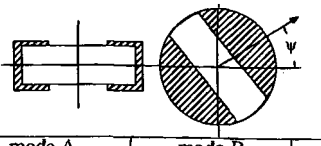


mode	longitudinal (A) (quasi)	fast shear (B) (quasi)	slow shear (C) (pure)
u (natural axis)	0.0 -0.77918 +0.62680	0.0 -0.62680 -0.77918	1.0 0.0 0.0
$V/2 = NF_a$ (kHz.mm)*	3503.6	1899.7	1660.8
k_{TE}	0	0	0.088

* $NF_r \# NF_a (1-4k_{TE}^2/\pi^2)$.

Table 1

In the case of AT cut resonators using lateral field excitation it is possible by choice of the in plane direction of the field to excite the overtones and the anharmonics of these three modes , then the correspondent coupling coefficient is different to zero. When the field is parallel to x_3 direction or Z' the C mode can be excited and when the field is parallel to x_1 direction the A and B modes can be excited, namely that with the lateral field excitation the coupling coefficients depend of the field (Table 2)[4].



	mode A quasi longitudinal	mode B quasi (fast) shear	mode C quasi (slow) shear
k_{LE} (field in the Z' direction)	0	0	0.06
k_{LE} (field in the X' direction)	0.064	0.062	0

Table 2

Some three dimensional models for the AT cut resonators with the thickness excitation have been proposed by different authors[5,6,7,8], they permit to describe with a good precision the modes and the characteristics of these resonators. The plano convex resonators excited by a lateral field have been studied since many years [9,10,11,12] on account of the potentialities to obtain ultra stable frequencies [13], but we have not the same comprehension of the real modes that for the thickness field excitation resonators.

III-Lateral field excitation :

a/ Parallel to the x_3 direction.

The fifth overtones of the slow thickness shear mode or C mode for plano convex resonators using lateral field excitation and thickness field excitation are imaged using the white beam transmission topography on the figures 4 and 5 respectively. These resonators

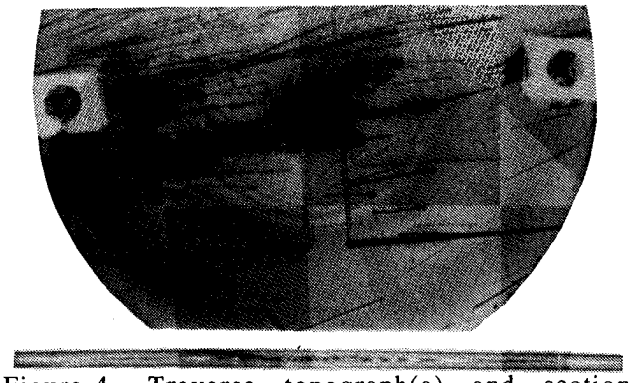


Figure-4- Traverse topograph(a) and section topograph(b) of the fifth overtones of the slow thickness shear mode for a plano convexe resonator using lateral field excitation with a low level of the excitation.

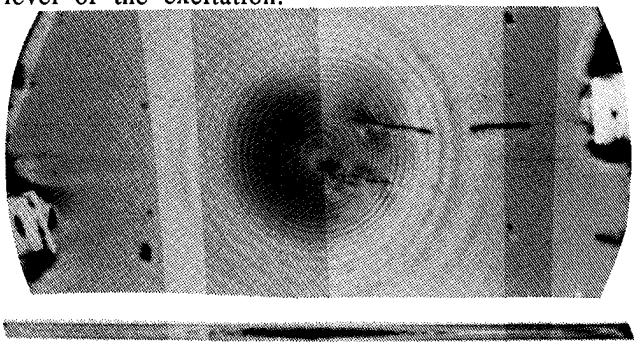


Figure-5- Traverse topograph(a) and section topograph(b) of the fifth overtones of the slow thickness shear mode for a plano convexe resonator using thickness field excitation with a low level of the excitation.

are designed to have a very high Q factor for this overtone (about 6.3 MHz). In fact the frequency is equal to twice that of the synchrotron radiation and these topographs are made with the stroboscopic technique.

With a low level of the excitation it appears extra Pendellösung fringes in the section topographs which correspond to elliptical fringes on the traverse topographs [14]. These elliptical fringes result from the integration of the previous extra Pendellösung fringes over the whole resonators surface and constitute equi-level curves of the vibration amplitude. On the figure 5 the elliptical fringes are particularly well defined and present nearly the theoretical shape because this resonator have only three dislocations. Indeed these fringes are very sensitive to the presence of crystal defects.

This mode presents for these two resonators the same x_1, x_2 and x_3 dependence like it is possible to see it on the traverse topographs for the x_1 and x_3 dependence and on the section topographs for the x_2 one.

When the level of the excitation is higher the extra Pendellösung fringes on the section topographs disappear, but the antinodal planes of the vibration in the thickness of the resonator appear. The section topograph (Fig.6) of the same lateral field resonator presents four black lines corresponding to the four antinodal planes of the fifth overtone of the C mode.



Figure-6- The same section topograph that on the figure 4 but with a level of excitation higher.

b/Parallel to the x_1 direction.

.B mode

It is possible to use another orientation of the lateral field to excite modes non piezoelectrically excited with a strictly thickness field. With the excitation field parallel to the x_1 direction we have excited the B mode. The topographs obtained with a diffraction vector parallel to the x_3 direction show on the figure 7 the u_3 component of different overtones.

The fundamental of the fast thickness shear mode (Fig.7a) is not trapped and presents a different lateral dependence.

The third and the fifth overtones (Fig.7b and 7d) are trapped and they have the same behaviour that the C mode. The section topographs (Fig.7c and 7e) show in the thickness of the resonator the two antinodal planes for the

third overtone and the four antinodal planes for the fifth overtone. Although the polarisation of the fast thickness shear mode is not exactly in the x_3 direction, but at an angle equal to about 3°.5, the topographs show a behaviour very close to that of a pure shear mode.

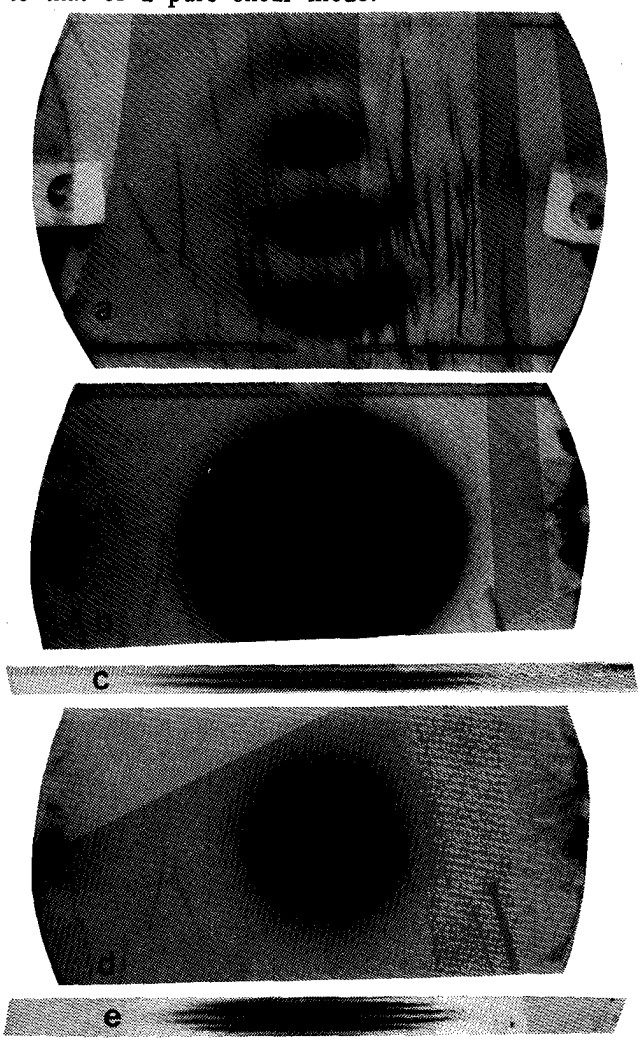
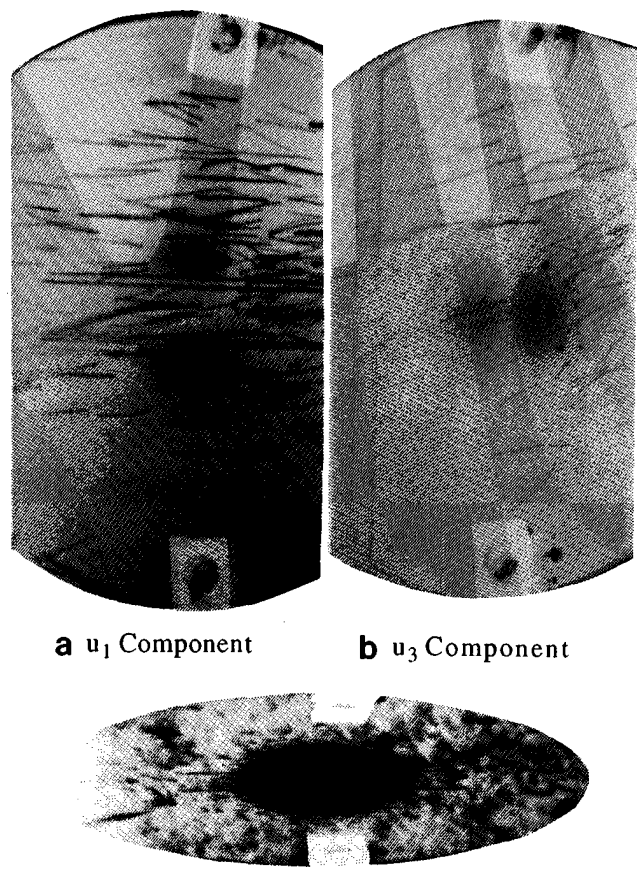


Figure-7- Topographs showing the u_3 component of different overtones of the fast shear mode.
a/The fundamental.
b-c/The third overtone.
d-e/The fifth overtone.

.A mode

The figure 8 shows three topographs obtained with a diffraction vector parallel to the x_1, x_2 and x_3 direction respectively, and displaying the three components of the third overtone of the A longitudinal mode which can be also excited by a lateral field parallel to the x_1 direction.

The topograph on the figure 8b obtained with the reflexion Laue set-up shows the main component u_2. This component seems to be trapped although a detailed examination indicates the existence of coupled plate modes which appear, on this type of topograph, like fine fringes. The two others topographs on the figure 8 are obtained with the transmission Laue set-up. The components u_1 and u_3 which appear to be antisymetrical, are not negligible and are quite trapped.



a u_1 Component **b** u_3 Component

c u_2 Component

Figure-8- Topographs obtained with a diffraction vector parallel to the x_1, x_2 and x_3 direction respectively, and displaying the three components of the third overtone of the A longitudinal mode.

Another example is shown on the figure 9. This is the fifth overtone of the A longitudinal mode obtained with the same conditions. Again the main component u_2 (Fig.9b) seems trapped in the center of the plate, but the energy is more confined that for the third overtone. The u_1

component (Fig.9a) extends up to the fixations in the x_1 direction and contains an important fraction of the acoustic energy. The u_3 component (Fig.9c) is confined in the center of the plate and has also a non negligible intensity.

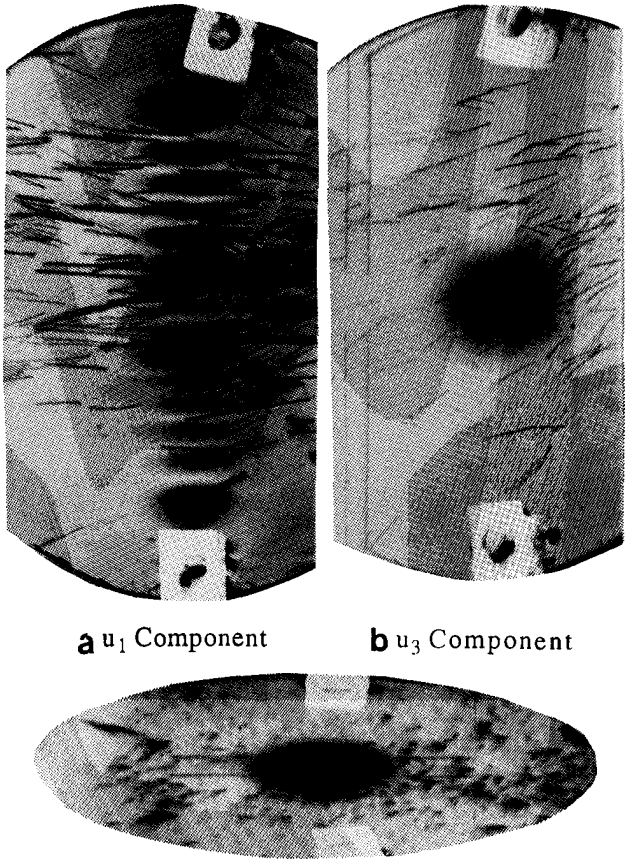


Figure-9- Topographs of the fifth overtone of the A longitudinal mode obtained with the same conditions that those on the figure 8.

.C mode : even overtone

The study of electrical response of a plano convex resonator using the lateral field excitation parallel to the x_1 direction has shown the existence of anharmonics of the second overtone of the C mode. The figure 10 shows at the 2.517 MHz a strong response corresponding to the 210 C mode. This mode is imaged on the figure 11 by X ray topography using a diffraction vector parallel to the x_1 direction. On the traverse topograph (Fig.11a) we can observe the nodal line in the x_3 direction and on the section topograph (Fig.11b) the antinodal line identifying the second overtone.

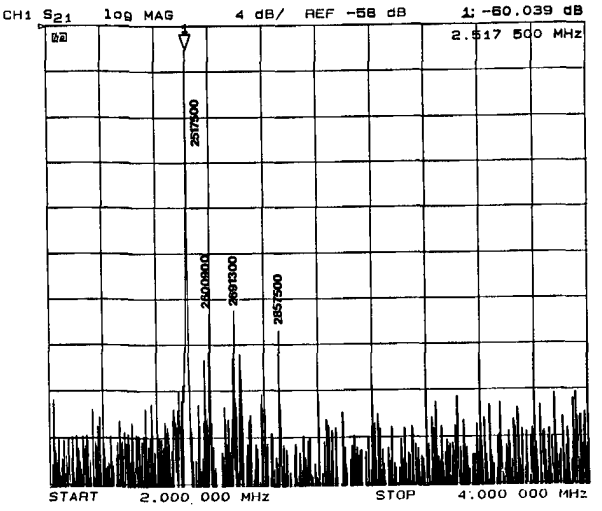


Figure-10- Electrical response corresponding to the even overtone 210 of the C mode, at 2.517 MHz, of a resonator using lateral field excitation parallel to the x_1 direction.

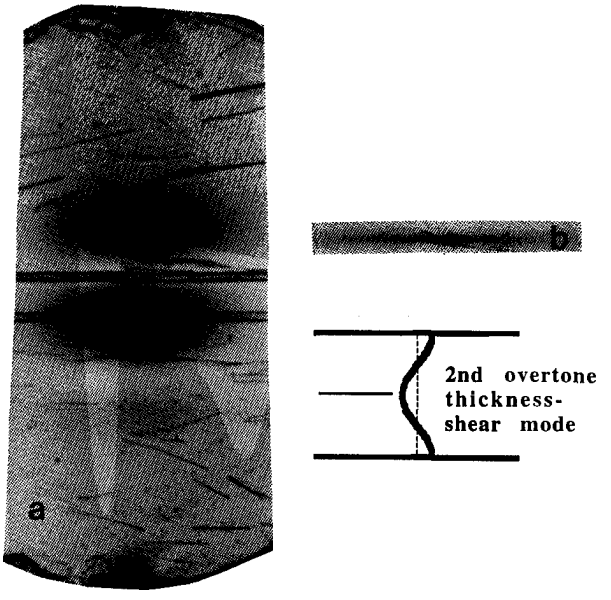


Figure-11- Traverse topograph(a) and section topograph(b) showing the even overtone 210 of the C mode of a plano convexe resonator using the lateral field excitation parallel to the x_1 direction.

Similarly we have observed in the same conditions the first antisymetrical anharmonic of the fourth overtone of the slow thickness shear mode (Fig.12). On the section topograph (Fig.12b) the three antinodal planes characteristic of the fourth overtone appear.

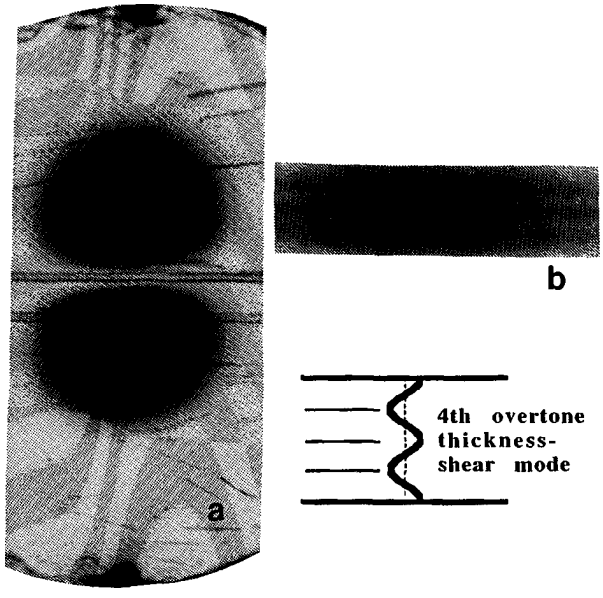


Figure-12-Traverse topograph(a) and section topograph(b) showing the even overtone 410 of the C mode of a plano convexe resonator using the lateral field excitation parallel to the x_1 direction.

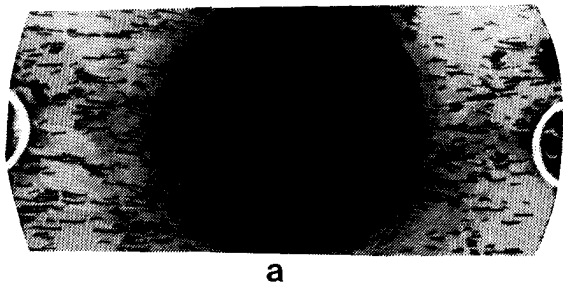
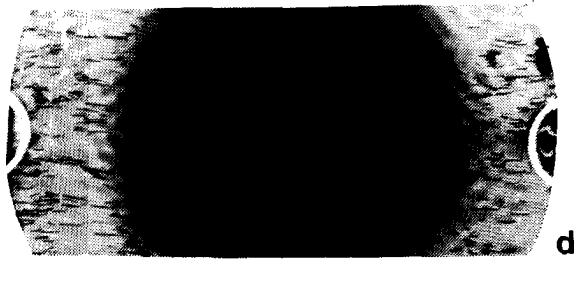


Figure-14-Stroboscopic topographs at 6.338 MHz. Low excitation : a)u_1 component phase=0°, b)u_3 component phase=0°, c)u_3 component phase=90° High excitation phase=0° : d)u_1 component, e)u_3 component



IV-The coupled modes.

Coupled plate modes in thickness excited fundamental AT resonators have been observed. These resonators are plane resonators which present, especially with the fundamental mode, much more effects that overtone plano-convex resonators.

The figure 13 shows the u_1 and u_3 components of thickness shear mode of a rectangular sample. The flexure coupled modes which appear are often observed in rectangular samples and some theoretical calculations have been published. With a diffraction vector parallel

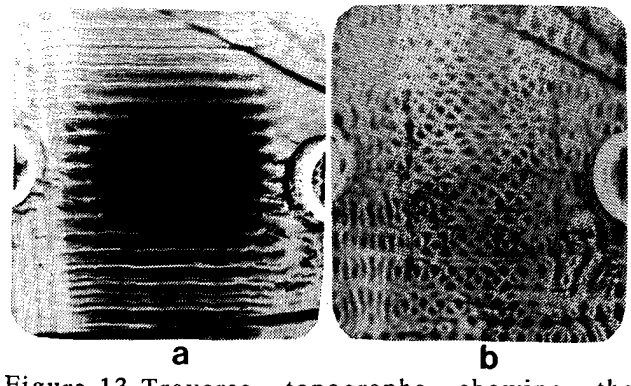
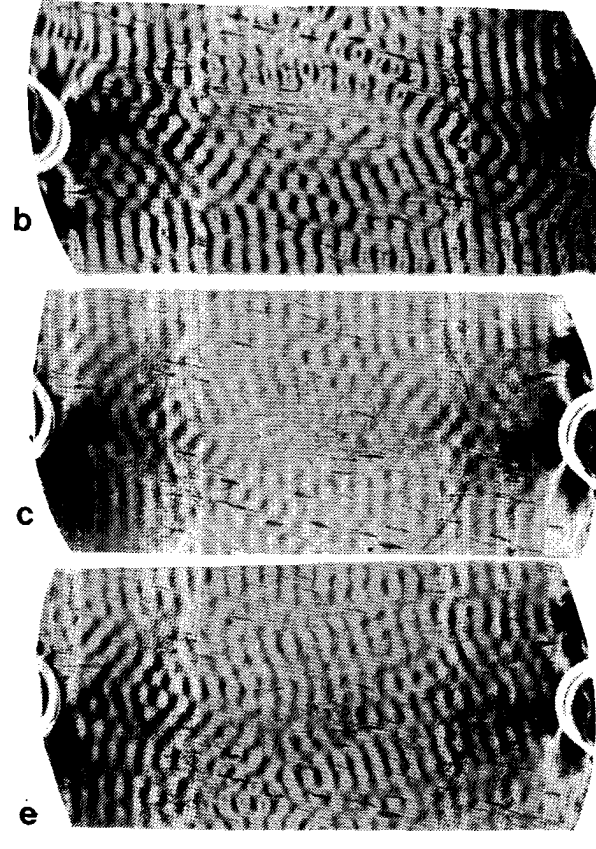


Figure-13-Traverse topographs showing the flexure coupled modes in a rectangular plane resonator.a)u_1 component, b)u_3 component.



to x_2 it is possible to see also an important corresponding u_2 component.

The topographs on the figure 14 are obtained with the stroboscopic set-up at 6.338 MHz. This is a plane resonator with a diameter equal to 15 mm made with synthetic quartz. A progressive u_3 component, which depends of the phase between the resonator current and the observation pulses (0° on Fig. 14b and 90° on Fig. 14c), is visible with a low excitation. With a high level of excitation it appears a progressive part of the u_1 component and the shape of the u_3 component becomes different due to a non linear behaviour.

The shape of the progressive part of modes is highly dependant of the mechanical conditions at the edge. On the stroboscopic topograph on the figure 15 a the resonator is simply fixed by two mounting clips and the mode presents an important progressive part. This resonator is made with natural quartz and shows only some growth bands which interfere with the acoustic mode (on the top on the left on the Fig.15a). With a thin layer of acoustic absorbent deposited on the edge of the resonator and with the same conditions of excitation, the progressive part is greatly attenuated (Fig.15b).

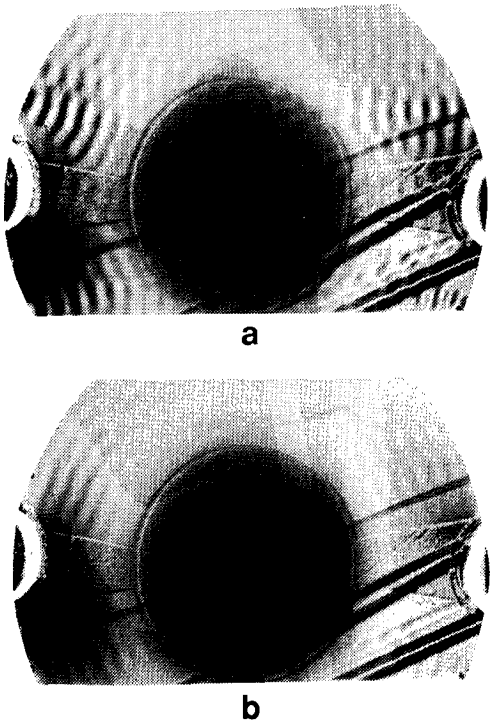


Figure-15-Stroboscopic topographs : without acoustic absorbant a) and with b).

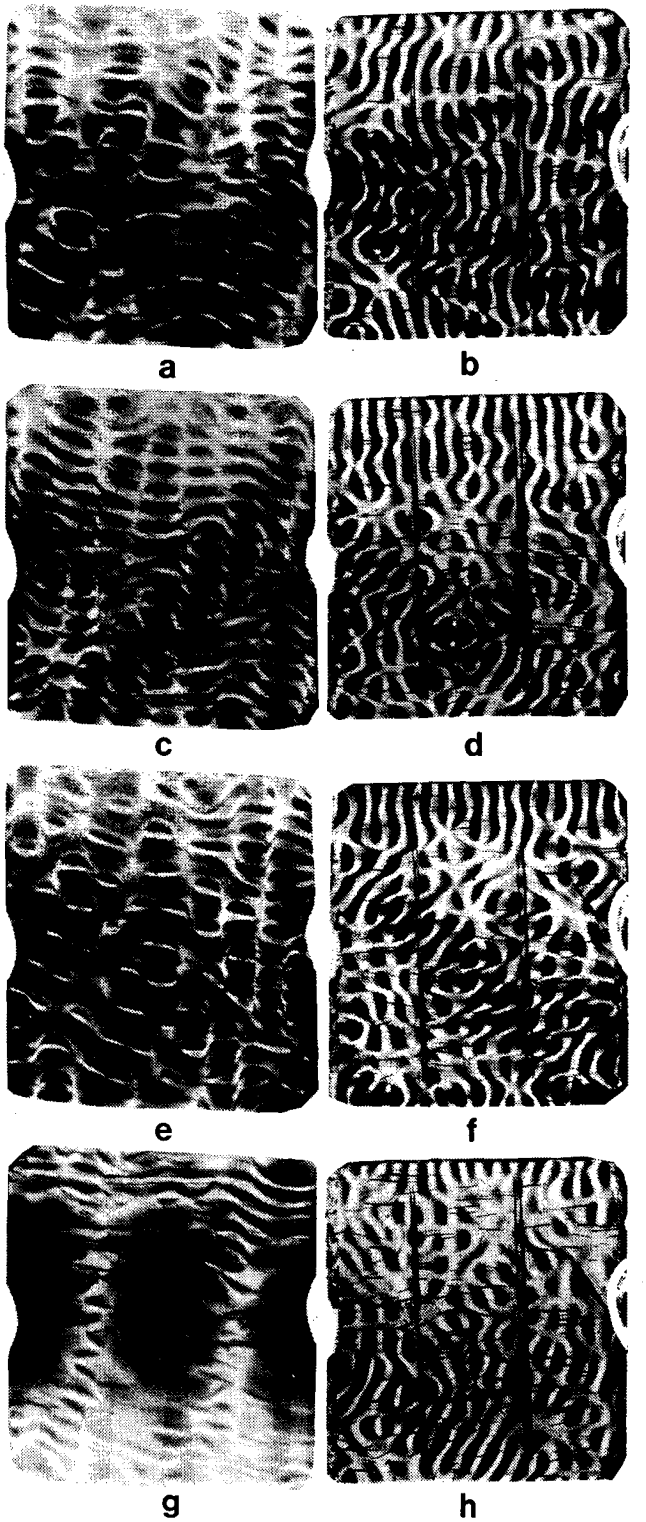


Figure-17-Different coupled plate modes appear when the fundamental mode is not trapped in a plane rectangular resonator.

a,c,e,g : u_1 component	b,d,f,h : u_3 component
a,b : 3.169 MHz	c,d : 3.172 MHz
e,f : 3.176 MHz	g,h : 3.253 MHz

The response of a plane rectangular resonator about 9x10 mm^2 operating at 3.169 MHz with design parameters such that the fundamental mode is not trapped is shown on the figure 16. As often observed in this case the mode decomposes into n modes by coupling with plate modes, probably due to the boundaries conditions at the edge of the sample. The figure 17 shows some of these n modes, the u_1 and the u_3 components are very intricate.

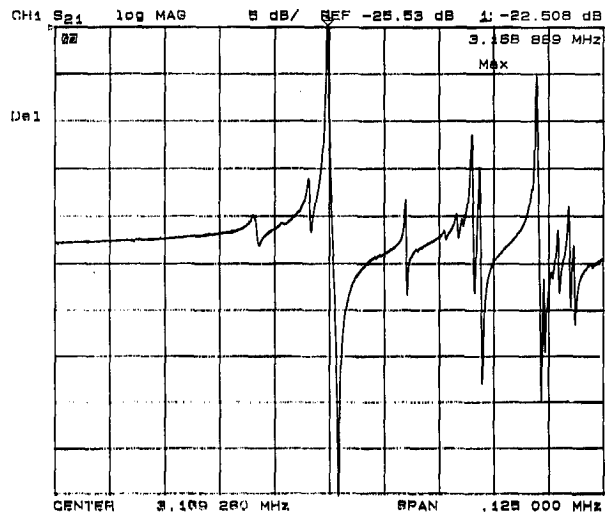


Figure-16-Electrical response of a rectangular plane resonator operating at the fundamental mode at 3.169 MHz.

Conclusion.

This has shown that it is possible to excite much more modes with the lateral field excitation that with the thickness field excitation. These modes can present intrinsic interest due to the values of the temperature coefficient and/or of the coupling coefficient. The new degree of freedom, given by the possibility of choosing the direction of the excitation field, permits interesting applications such as the suppression or excitation of one type of mode. This study can help to understand the coupled modes which appear in the thickness excitation resonators.

Intense coupling to plate modes appear very frequently in plane fundamental resonators. These plate modes have generally complicated tridimensional modes shapes and are very sensitive to the fixations of the resonators. Furthermore the coupled modes must play an important role in the environmental sensitivity.

References.

[1] A. ZARKA, B. CAPELLE, J. DETAINT & J. SCHWARTZEL. 41st Annual Frequency Control Symposium, 27-29 May 1987, Philadelphia, U.S.A..

[2] A.GLOWINSKI. Annales des télécommunications, tome 27, n°3-4, Mars-Avril 1972, pp 147-158.

[3] T. YAMADA & N. NIIZEKI. J.Appl.Phys., Vol. 41, August 1970,pp 3604-3609.

[4] E.R. HATCH & A. BALLATO. Procedings of the Ultrasonics Symposium,Oct.31-Nov.1-2, 1983, p512.

[5] D.S. STEVENS & H.F. TIERSTEN. J.Acoust.Soc.Am. **79** (6), June 1986, pp 1811-1826.

[6] DULMET Thèse 1991 Université de Besançon.

[7] R.C. PEACH. IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, Vol.35, n°5, Sept. 1988, pp 593-611.

[8] J.DETAINT, H. CARRU, J. SCHWARTZEL, B. CAPELLE & A. ZARKA. Proc. 42nd AFCS, 1988, pp 19-28.

[9] W. IANOUCHEWSKY. Annales Françaises de chronométrie, 31ième année 2ième série, tome XV, 1961, pp 41-51.

[10] R. BECHMANN. Proc. 14th AFCS, May-June 1960, pp 68-88.

[11] A.W. WARNER. Proc. 17th AFCS, May 1963, pp 248-266.

[12] R.C. SMYTHE & H.F. TIERSTEN. Proc. 41st AFCS, 1987, pp 311-313.

[13] J. DETAINT, H. CARRU, R. LEFEVRE, J. SCHWARTZEL, C. JOLY, B. CAPELLE, Y. ZHENG & A. ZARKA. Proc. 44th AFCS, 1990, pp 337-348.

[14] Y. ZHENG, A. ZARKA, B. CAPELLE, J. DETAINT & J. SCHWARTZEL. Acta Cryst., 1989, A45, pp275-285.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

NON-UNIFORM DISTRIBUTION OF MOTION INFLUENCE ON THE EFFECTIVE MASS-LOADING IN AT-CUT QUARTZ RESONATORS

I.Mateescu, E.Candet

Institute of Physics and Technology of Materials
PO Box MG-7, Bucharest-Magurele, ROMANIA

Abstract - It is shown that the harmonic dependence of the effective mass loading and piezo-electric coupling constant is strongly electrode material- dependent, and could be interpreted in the frame of a transmission line analog model with a slight non-uniformity of the distribution of the vibratory motion over the electroded area of the crystal.

Introduction

In spite of its respectable age as a material for piezoelectric devices, quartz is still focusing special attention as a basic material for crystal resonators. Such a device made of quartz enjoys, through a fortunate accident of Nature, a relatively simple mode spectrum, a near-zero temperature coefficient of frequency in the usual domain and a rather small electromechanical coupling factors, providing for fairly loose coupling between the external electrical circuit and its vibrating mass. This features background its applications in frequency and time control and a great variety of sensing techniques (pressure, thin films stressing, mass changing, etc.).

In order to imply a properly cut quartz plate of a certain geometry as part of any specifically design electrical circuit, metallic electrodes must be deposited onto its major surfaces, i.e. the plate must be mass-loaded.

Mass-loading the plate results in a series of effects [1] affecting mainly plate inertia, energy trapping, interfacial stress distribution and the waves propagation through the crystal. The electrode size, shape, thickness, uniformity and position directly affects both the static and motional equivalent circuit parameters and modes content. That is why appreciable effort is devoted to the problem of accurately measuring the mass-loading and appropriately modelling the real device in order to be able to fulfill any specific engineering application demands.

The present paper analyses some recent measurements of the mass-loading in AT-cut quartz resonators [2] in the frame of the currently accepted transmission line analog model [3], suggesting that the observed deviations of the deduced effective mass-loading and coupling coefficient values from the predicted ones, could be indicative for a non-uniform distribution of the vibratory motion over the electroded area of the resonators. The strong dependence upon the nature of the electrode material allows us to speculate about the influence the stresses built-up during the thin film electrode growth could have on the 'active area' parameter of the theory.

Theoretical survey

We shall briefly recall the thickness-mode (TM) vibration theory results for piezo-electric plates and their transmission line analog model consequences.

Recent numerical studies for the lateral dependence of the anharmonics in the vicinity of one overtone of a given TM and the influence of electrode stress will be shortly emphasized.

It was longly been recognized that different cuts of the quartz crystal offered different vibration mode patterns over a well established frequency domain [4]. For ex. AT-cut resonators vibrate mainly in the thickness-shear mode for 0.5 to 200 MHz frequency range, even if thickness-twist and flexure anharmonics are also present.

In this case, one usually assumes the so-called 'thickness-mode approximation' [5]: the plate is considered infinite in two dimensions (x_1,x_3) and lateral dependence of the displacement vector is

0-7803-0476-4/92 $3.00 © 1992 IEEE

ignored.
The basic equations are:

$$\frac{\partial \vec{T}}{\partial x_2}=\rho_Q \ddot{\vec{u}}$$

$$\vec{T}=\hat{c}\frac{\partial \vec{u}}{\partial x_2}-\vec{e}\frac{D_2}{e_{22}^S} \qquad (1)$$

$$\frac{\partial \phi}{\partial x_2}=\vec{e}^{\,*}\frac{\partial \vec{u}}{\partial x_2}-\frac{D_2}{e_{22}^S}$$

with the boundary conditions (TETM mode):

$$\vec{T}=-m\ddot{\vec{u}} \quad at \quad x_2=\pm h \qquad (2)$$

$$\phi=\mp\phi_0 e^{i\omega t} \quad at \quad x_2=\pm h$$

The symbols have the usual meaning and are found at the symbol list. Following [6], the real symmetric matrix $\hat{c}$ can be diagonalized and the eigenmodes of vibration are obtained. Defining the wave vector of the i-the eigenmode by

$$\eta_Q^{(i)}=\omega\cdot\left(\frac{\rho_Q}{\bar{c}_{66}^{(i)}}\right)^{1/2} \qquad (3)$$

and its coupling factor

$$(k^{(i)})^2=\frac{(e^0{}_i)^2}{e_{22}^S c^{(i)}} \qquad (4)$$

one obtains, after simple manipulations in (1), the solution in the anti-symmetrical form

$$u_i^0=\frac{e_i^0 D_2/e_{22}^S}{c_i\eta_Q^{(i)}\cos\eta^{(i)}h \; - \; m\omega^2\sin\eta^{(i)}h}\sin\eta^{(i)}x_2 \qquad (5)$$

and the impedance of the plate

$$Z=-i\frac{2h}{\omega e_{22}^S A}\cdot\left[1-\sum_{i=1}^{3}\frac{(k^{(i)})^2}{1-\mu\eta^{(i)}h\tan(\eta^{(i)}h)}\cdot\frac{\tan(\eta^{(i)}h)}{\eta^{(i)}h}\right] \qquad (6)$$

From (6) it is obvious that the resonant frequencies are obtained as solutions of:

$$\sum_{i=1}^{3}\frac{k^{(i)2}}{1-\mu\eta^{(i)}h\,\tan(\eta^{(i)}h)}\cdot\tan\frac{(\eta^{(i)}h)}{\eta^{(i)}h}=1 \qquad (7)$$

while anti-resonances occur at

$$f_A^{(n)i}=\frac{2n-1}{4h}\cdot(\bar{c}^{(i)}/\rho_Q)^{1/2} \qquad n=1,2,... \qquad (8)$$

For a single TM (for ex. thickness shear in AT-cut) and u = 0, the transmission-line analog of (6) is straightforward [3].

The equivalent circuit elements are:

$$C_0=\frac{e_{22}^S A}{2h} \qquad (9)$$

$$N=\frac{e^0 A}{2h} \qquad (10)$$

$$Z=A\rho v \; ; \quad \eta=\omega/v \qquad (11)$$

However, for the actal mass-loaded case, more realistic assumptions have to be considered. Recent models for the equivalent network of a single TM mass-loaded resonator [1] start with the following assumptions:
(i) the vibratory motion is distributed uniformly across the electroded area of the resonator
(ii) the lateral boundedness of the plate implies a difference between the electroded area A_e and the active area A_a.
In this case (6) transforms to

$$Z=-i\cdot\frac{2h}{\omega e_{22}^S A_e}\left[1-\frac{k_{eff}^2}{1-\mu_{eff}\eta_Q h\tan(\eta_Q h)}\cdot\frac{\tan(\eta_Q h)}{\eta_Q h}\right] \qquad (12)$$

and the equivalent circuit elements

$$C_0=\frac{e_{22}^S A_e}{2h} \; ; \; N=\frac{e^0 A_a}{2h} \; ; \; Z=A\rho v \; ; \; \eta=\omega/v \qquad (13)$$

where k^2 and μ have been replaced by 'effective'

$$k_{eff}^2=\frac{k^2}{b} \qquad (14)$$

$$\mu_{eff}=\mu b\frac{\tan(\eta_e t_e)}{\eta_e t_e} \; ; \quad \left(\mu=\frac{\rho_e t_e}{\rho_Q h}\right) \qquad (15)$$

with

$$b=\frac{A_e}{A_a} \qquad (16)$$

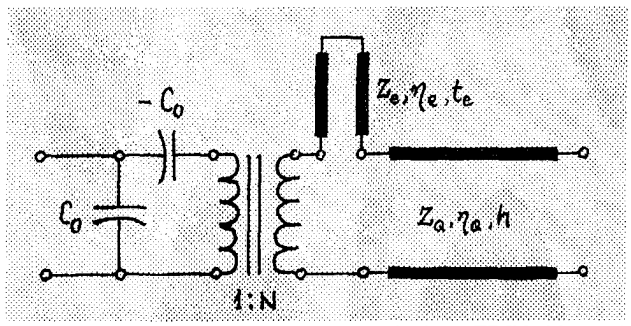


Figure 1. Equivalent network for loaded plate, TETM, bisected basis

Fig. 1 represents a pictorial of Eqn.(12).
Now, the equivalent of Eqn.(7) for the resonant

frequencies is

$$\frac{\tan X}{X}=\frac{1}{k^2}\ ;\ X=\frac{\pi}{2}\cdot\frac{f_{R0}^{(m)}}{f_{A0}^{(1)}} \qquad (17)$$

for the unloaded plate, and

$$\frac{\tan X}{X}=\frac{1}{k_{eff}^2+\mu_{eff}X^2}\ ;\ X=\frac{\pi}{2}\cdot\frac{f_{R\mu}^{(m)}}{f_{A0}^{(1)}} \qquad (18)$$

for the case when the electrodes are deposited. It is easy to see that in this case, the anti-resonant frequencies have to be determined from the following equation:

$$\frac{\tan X}{X}=\frac{1}{\mu_{eff}X^2}\ ;\ X=\frac{\pi}{2}\cdot\frac{f_{A\mu}^{(m)}}{f_{A0}^{(1)}} \qquad (19)$$

Eqns. (17)-(19) together with definitions (14)-(16), allow a method to prove the valability of the model: from the value of resonant frequencies of a series of harmonics for the un-loaded plate, one can determine the coupling factor with (17); from (19) and a series of anti-resonant frequencies of the loaded plate, on can determine the effective mass-loading; eqn. (18) and the series of resonant frequencies of the same loaded plate then gives the effective coupling constant. The effective values have to be consistent with those of definitions (15).

Experimental

We used circular AT-cut cultured Sawyer quartz plates, plane-parallel (PP) and plane-convex (PC) with θ = 35°17'.
Diameters of plates were 14 mm.
PP plates had 0.330 mm. while PC 0.373 mm. in thickness (200 mm. curvature radius for PC plates). The plates were beveled in order to avoid the occurrence of unwanted anharmonic vibration modes. Both PP and PC plates were carefully optically polished and surface cleaned. The series of resonance and anti-resonance frequencies (1,3,5,7th order) is then determined using a HP 4194A analyzer and the experimental setup indicated in [7]; the device allows one to determine the zero-gap critical frequency of the ideal resonator having perfectly conductive electrodes with a zero-mass without applying any stress on the plate simply by extrapolating the smoothed frequency-gap dependence.
Ag, Au and CrAu respectively were thermally evaporated in vacuum on both PP and PC plates, yielding 1500, 2000, 2500 and 3000 Å electrode thickness. The deposition was monitored in each run by a MIKI-MSV 1841 vibrational quartz thickness / rate monitor, but the actual thickness was checked on a witness by interferometric means [8]. The resonance and anti-resonance frequencies series were then measured for every mass-loading, using the same technique. Several plates per batch were mounted in fixtures in order to determine the electrical parameters [9],[13].

Results

I.Mass-loading and piezoelectric coupling

As mentioned above, Eqns. (17)-(19), (14)-(16), must be verified within the validity of the model.
Figure 2 presents a typical dependence of the theoretical effective mass-loading given by (15) for Ag (□), Au (+) and CrAu (◇) electrodes, 2000 Å thickness.

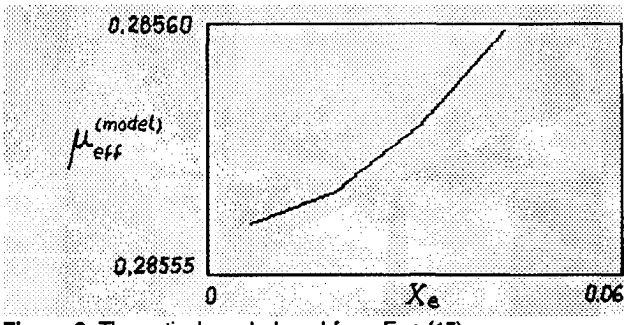


Figure 2. Theoretical μ_{eff} deduced from Eqn.(15):
(□) Ag electrodes 2000 Å (+) Au electrodes 2000 Å
(◇) CrAu electrodes 2000 Å

Specific densities and stiffness constants values for thin films of Ag and Au were those in ref. [10] and for CrAu in ref.[11]. This has to be compared with the

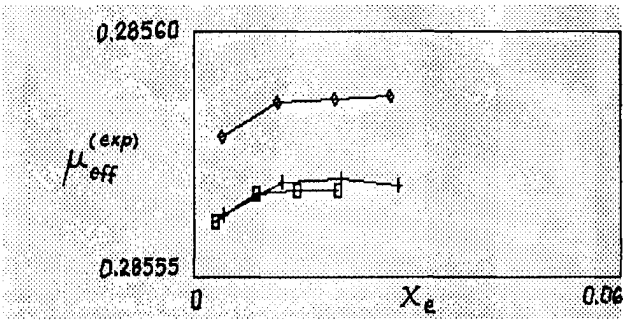


Figure 3. Measured μ_{eff} (Eqn.(19)) for
(□)Ag (+)Au (◇) CrAu electrodes

experimental effective mass-loading as determined from the loaded anti-resonance frequency series, with Eqn. (19), as presented in Figure 3.
One can easily observe that, apart from a shifting value, the experimental dependencies tend to saturate at high X_e ratios, even if at low X_e, the quadratic theoretical dependence is approached. The effect is further harmonic- and electrode- dependent. Figure 4 present the coupling coefficient dependence on harmonic, as deduced from (17).

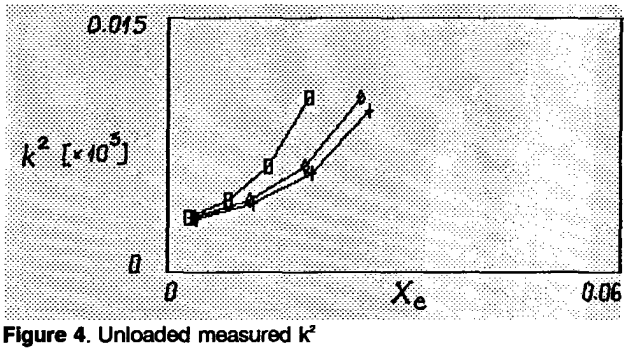


Figure 4. Unloaded measured k^2
(□)Ag (+)Au (◇)CrAu electrodes

The series of resonance frequencies for the loaded plates allows one to determine the effective coupling from Eqn. (18). The data are presented in Figure 5.

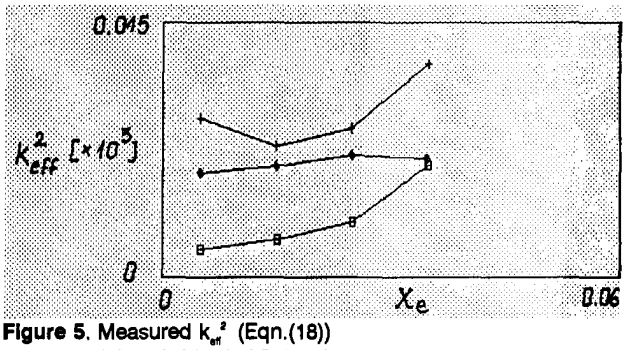


Figure 5. Measured k_{eff}^2 (Eqn.(18))
(□)Ag (+)Au (◇)CrAu electrodes

This time, both the deviations from the model theoretical predictions and the electrode material dependence are dramatical, indicative for a complex phenomenon (see bellow). Combining this results and Eqns. (14) and (16) one can deduce the parameter b, i.e. the variation of the parameter A_a of the model. The result are presented in Figures 6, normed to full-electroded area.

II.Electrical parameters

It is worth mentioning the influence various

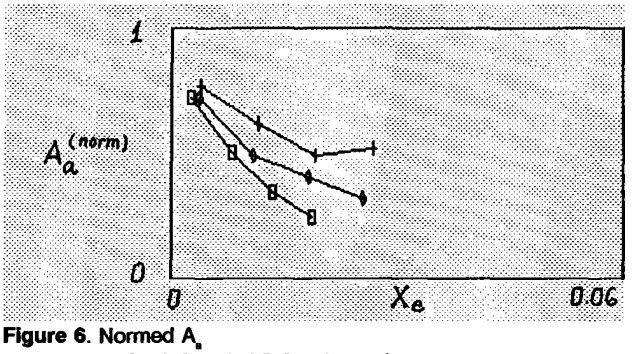


Figure 6. Normed A_a
(□)Ag (+)Au (◇)CrAu electrodes

parameters used by our study have on electrical equivalent characteristics dependences. We shall briefly summarize first what we might call 'usual dependences', namely:
(i) equivalent ohmic resistance R of the resonator - it is practically independent of weather the geometry of the plate was PP or PC and the nature of the metal used for electrodes. Its minimum corresponds to the fundamental resonance frequency (approx. 20 ohms for PP and 35 ohms for PC) and has a drop at approx. t_e=2000 Å that confirms the initial design of the resonator device.
(ii) dynamic equivalent capacitance C of the resonator- it is practically independent of mass-loading and the nature of the metal used for electrodes. Typical values are 22 fF for PP and 14 fF for PC in fundamental mode. It has a drastical reduction with harmonics, confirming the well-known relation

$$C=\frac{4Ae^{(0)2}}{m^2\pi^2hc}\ ;\quad m=1,3,5,... \qquad (20)$$

However, to fit experimental data with Eqn.20, we systematically found that the value of A should be closer to that of A_a. This phenomenon is very electrode-material dependent and is more pronounced in PP geometries.
In contrast to the two above-mentioned electrical characteristics, the electrical inductance L of the resonator seems to behave in a rather unusual way. Figure 7 shows typical behavior of L for both PP and PC plates with harmonics.
One can distinguish that depending on the nature of the metal used for electrodes, but especially in PC geometry, there is an increase of L from its fundamental value with harmonics.

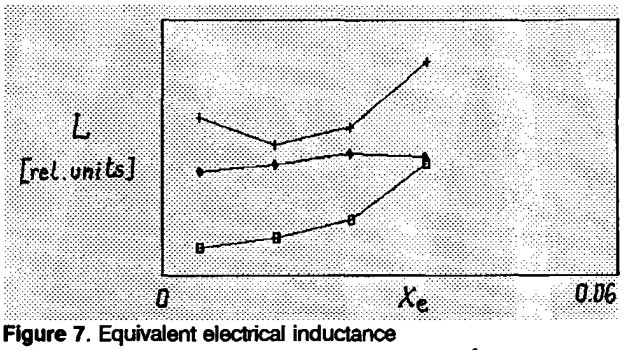


Figure 7. Equivalent electrical inductance (□)Ag (+)Au (◇)CrAu electrodes 2000 Å

Discussion

In order to explain both harmonic variations of the effective coupling and mass-loading and their consequences on the equivalent inductance, we suggest that, while retaining the transmission-line analog with lateral boundedness [1], the active area parameter A_a has to be a function of harmonics, as suggested by Figure 6. Recent numerical analysis on vibratory motion in resonators with finite lateral dimensions [12], support two characteristics of the above mentioned experimental data:

(i) strong energy-trapping of harmonics, depending on plate geometry, yielding sub-unitary values for the ratio between the electroded area and the area actually involved in the vibratory motion on a certain harmonic

(ii) variation of the effective volume of the crystal vibrating on a certain harmonic with the order of this harmonic

Moreover, electrode material-dependent behavior of A_a seems to point out the consequences stresses built-up during thin electrode-film deposition could have on the actual vibratory motion of the plate subjected to such anisotropic boundary conditions (Preliminary temperature-frequency shift determinations on these plates are indicative for such influences). It is therefore suggested that making a proper choice of the composite electrode material and carefully monitoring the deposition conditions (esp. substrate temperature), wold result in a strict control of the A_a for a certain harmonic and then would improve the electrical parameters of the resonator.

We think that, apart from the inherent limitations of the thickness-shear vibration model itself, different boundary conditions imposed by electrodes on the major surfaces of the quartz plate could lead to such non-uniform distribution of the vibratory motion, which in turn affects the equivalent electrical parameters, in the above-cited manner.

Conclusions

Measurements have been made of mass-loadings and electrical parameters on a number of plane-parallel and plane-convex quartz plate resonators. It was found that in order to explain the detailed differences between the theoretical transmission line analog model and experimental data, a non-uniformity induced by the electrodes of the vibratory motion of the crystal has to be considered. The description of the complex phenomena induced by the electrodes - plate interface requires further work on this subject.

Acknowledgements

The authors would like to thank Dr. A.Pasculescu, Dr. S.Frunza and S.Gane for the opportunity they kindly offered to computer-processing the huge amount of data involved in this work. All the staff of IFTM Piezoelectric Devices Dept. is acknowledged for specific contribution.

Symbols used in text

A_a	: Active area of vibration for a given harmonic
A_e	: Electroded area of the resonator
b	: A_e/A_a ratio
$\hat{c}$	: Stiffness matrix of the crystal
$\bar{c}$	: Single TM, diagonal element of $\hat{c}$
C_0	: Static equivalent capacitance of the resonator
C	: Dynamic equivalent capacitance of the resonator
D_2	: Electric displacement in x_2 direction
e	: Piezoelectric tensor acting on a surface normal to x_2
e^T	: Transposed of e
$f_{R0}^{(m)}$, $f_{A0}(m)$	: Unloaded resonant / anti-resonant frequency series (m=1,3,5,7)
$f_{Ru}^{(m)}$, $f_{Au}^{(m)}$	: Loaded resonant / anti-resonant frequency series (m=1,3,5,7)
h	: Half-thickness of the plate
k	: Un-loaded piezoelectric coupling constant
k_{eff}	: Loaded 'effective' piezoelectric coupling

L : Equivalent inductance of the resonator
m : Areal mass of the electrode
N : Ideal transformer ratio for the mechanical interface network of the transmission-line analog of the resonator
Q : Quality factor of the resonator
R : Equivalent resistance of the resonator
R_Q : Curvature radius of PC plates
t_e : Electrode thickness on one face
v : Wave velocity in crystal
X_{A0} : defined by Eqn. (19)
X_{R0} : defined by Eqn. (17)
X_{Au} : defined by Eqn. (19)
X_{Ru} : defined by Eqn. (18)
X_e : $\eta_e t_e$ product
Z_e : Equivalent motional impedance of the electrodes
Z_Q : Equivalent motional impedance of the crystal
ϵ_{22}^s : Dielectric permitivity of the crystal in x_2 direction
ϕ : Electric potential on the resonator plates
ϕ_e : Electrode diameter
ϕ_Q : Crystal diameter
η : Wave vector (η_Q in crystal, η_e in electrodes)
μ : Mass-loading
μ_{eff} : 'Effective' measured mass-loading
ρ_e : Electrode density
ρ_Q : Quartz density
ω : Angular frequency of the wave

References

[1] Kosinnski J.,Mallikarjun S., Ballato A.
'Mass- loading measurements of quartz crystal plates', Proc. 43rd ASFC, 1989, pp. 365

[2] Mateescu I.
'Effects of mass-loading on electrical parameters of AT-cut quartz resonators'
paper presented in the Poster Session of the 6th EFTF, Nordweijk, Holland 1992

[3] Ballato A.
'Transmission-line analogs for stacked piezo-electric crystal devices'
Proc. 26th ASFC, 1972, pp.86

[4] Mason W.P.
'Use of piezo-electric crystals and mechanical resonators in filters and oscillators'
Physical Acoustics, vol.IA (1964), pp. 335, Academic Press N.Y.

[5] Tiersten H.F.
'Linear piezoelectric plate vibrations'
Plenum Press N.Y. (1969), pp.87

[6] Yamada T., Niizeki N.
'A new formulation of piezoelectric plate thickness vibration'
Rev.El.Comm.Lab. 19 (1971), pp. 705

[7] Detaint J., Carru H., Schwartzel J., Jolly C., Capelle B., Zarka A.
'Air-gap determination of the cut-off frequency of un-electroded plates and effects of local thickness modifications in plane resonators'
Proc. 43rd Symp. on Freq. Control (1989),pp.563

[8] Chopra K.L.
'Thin film phenomena'
Mc-Graw Hill N.Y. (1969), pp.83

[9] Hafner E.
'Crystal resonators'
IEEE Trans.Son. Ultrason. SU-21 no.4 (1974), pp.220

[10] ShermannJ.H.Jr.
'Temperature coefficient of frequency shift arising from electrode film stress'
IEEE Trans. Son.Ultrason. 30 (1983), pp.164

[11] Weaver Ch.
'Adhesion of thin films'
J.Vac.Sci.Techn. 12 (1975), pp. 141

[12] Detaint J., Carru H., Schartzel J., Jolly C.
'Analysis of partially or totally contoured resonators with finite lateral dimensions'
Proc. 2nd EFTF (1988), pp.458

[13] Mateescu I.
'Monolithic quartz crystal filters'
Ph.D. Thesis, Bucharest Polytechnic Institute 1987

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

On Straight Crested Waves in a Third Overtone SC-Cut Quartz Resonator.

Y-K Yong and Z. Zhang

Dept. of Civil/Environmental Engineering, Rutgers University, P.O. Box 909, Piscataway, NJ 08855-0909

Abstract

Finite element matrix equations employing high frequency, piezoelectric plate equations are derived. The equations may be used for modeling third harmonic overtone of thickness-shear vibrations. A perturbation technique is developed to account for piezoelectric stiffening in the mechanical stiffness matrix. Results from the perturbation method compares well with the direct solution of the piezoelectric finite element equations. The technique will result in significant savings in computer memory and computational time. Numerical results for straight crested waves in a third overtone SC-cut quartz strip with and without electrodes are presented. Steady state response to an electrical excitation is calculated.

I. Introduction

We discuss in this paper the finite element formulation and results on straight crested waves in the vicinity of the third harmonic overtone of thickness shear, piezoelectric vibrations in a SC-cut quartz strip. The straight crested waves are spatially one-dimensional, but since the study is part of a wider study of third overtone SC-cut quartz resonators, two-dimensional finite element equations for piezoelectric plate vibrations are derived and presented. A one-dimensional finite element program is developed which could be modified subsequently to handle two-dimensional problems. The mechanical stiffness of the finite element matrix equations for the SC-cut quartz strip is a few orders of magnitude larger than the piezoelectric stiffness, and the piezoelectric coupling is weak. Hence, the piezoelectric effects could be modelled using a perturbation technique. A new perturbation method for piezoelectric vibrations is derived and proposed.

II. Plate Equations for High Frequency Piezoelectric Vibrations.

The piezoelectric plate equations for high frequency vibrations employed in the current finite element formulation were derived by Lee, Syngellakis and Hou [1]. The components of mechanical displacement and electric potential are expanded in an infinite series with their thickness-dependence expressed by trigonometric functions:

$$u_i(x_1, x_2, x_3, t) = \sum_{n=0}^{\infty} u_i^{(n)}(x_1, x_3, t) \cos\frac{n\pi}{2}(1-\xi)$$
$$\phi(x_1, x_2, x_3, t) = \sum_{n=0}^{\infty} \phi^{(n)}(x_1, x_3, t) \cos\frac{n\pi}{2}(1-\xi) \quad (1)$$

where $\xi = \frac{x_2}{b}$ is a nondimensional thickness coordinate of the plate. The terms $u_j^{(n)}$ and $\phi^{(n)}$ are functions of spatial coordinates x_1 and x_3 only and time t. We present without formal derivations the following field equations from reference [1]. The interested reader may obtain

0-7803-0476-4/92 $3.00 © 1992 IEEE

details of their derivations from the same reference. Some equations are given in a matrix form to facilitate the later development of finite element matrix equations.

IIa. Stress Equations of Motion.

$$T_{ij,i}^{(n)} - \frac{n\pi}{2b}\bar{T}_{2j}^{(n)} + \frac{1}{b}F_j^{(n)} = (1+\delta_{n0})\rho\ddot{u}_j^{(n)} \tag{2}$$

$$\text{where } T_{ij}^{(n)} = \int_{-1}^{1} T_{ij}\cos\left[\frac{n\pi}{2}(1-\xi)\right]d\xi \tag{3}$$

$$\text{and } \bar{T}_{2j}^{(n)} = \int_{-1}^{1} T_{2j}\sin\left[\frac{n\pi}{2}(1-\xi)\right]d\xi \tag{4}$$

are nth order stress. T_{ij} is the Cauchy stress tensor, ρ is the mass density, and δ_{n0} is a Kronecker delta. Summation is assumed for repeated indices. The nth order displacement is $u_j^{(n)}$, while the nth order face traction is

$$F_j^{(n)} = T_{2j}(b) - (-1)^n T_{2j}(-b). \tag{5}$$

IIb. Charge Equation of Electrostatics.

$$D_{i,i}^{(n)} - \frac{n\pi}{2b}\bar{D}_2^{(n)} + \frac{1}{b}D^{(n)} = 0 \tag{6}$$

$$\text{where } D_i^{(n)} = \int_{-1}^{1} D_i\cos\left[\frac{n\pi}{2}(1-\xi)\right]d\xi \tag{7}$$

$$\text{and } \bar{D}_2^{(n)} = \int_{-1}^{1} D_2\sin\left[\frac{n\pi}{2}(1-\xi)\right]d\xi \tag{8}$$

are nth order electric displacement, D_i is the electric displacement in the plate, and $D^{(n)}$ is the face charge defined as

$$D^{(n)} = D_2(b) - (-1)^n D_2(-b) \tag{9}$$

IIc. Strain-Displacement and Electric Field-Potential Relations.

The nth order strain-displacement and electric field-potential relations are:

$$\begin{aligned} S_{ij}^{(n)} &= \frac{1}{2}(u_{j,i}^{(n)} + u_{i,j}^{(n)}) \\ \bar{S}_{ij}^{(n)} &= \frac{n\pi}{4b}(\delta_{2i}u_j^{(n)} + \delta_{2j}u_i^{(n)}) \end{aligned} \tag{10}$$

$$\begin{aligned} E_i^{(n)} &= -\phi_{,i}^{(n)} \\ \bar{E}_i^{(n)} &= -\frac{n\pi}{2b}\delta_{2i}\phi^{(n)} \end{aligned} \tag{11}$$

where $S_{ij}^{(n)}$ and $\bar{S}_{ij}^{(n)}$ are symbols for the nth order strains while $E_i^{(n)}$ and $\bar{E}_i^{(n)}$ are the nth order electric field. The strains and electric fields may be written in matrix forms, namely,

$$\underline{S}^{(n)} = \begin{bmatrix} S_1^{(n)} \\ S_3^{(n)} \\ S_4^{(n)} \\ S_5^{(n)} \\ S_6^{(n)} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} & 0 & 0 \\ 0 & 0 & \frac{\partial}{\partial x_3} \\ 0 & \frac{\partial}{\partial x_3} & 0 \\ \frac{\partial}{\partial x_3} & 0 & \frac{\partial}{\partial x_1} \\ 0 & \frac{\partial}{\partial x_1} & 0 \end{bmatrix} \begin{bmatrix} u_1^{(n)} \\ u_2^{(n)} \\ u_3^{(n)} \end{bmatrix} = \underline{\partial}_a \underline{u}^{(n)}$$

$$\bar{\underline{S}}^{(n)} = \begin{bmatrix} S_2^{(n)} \\ S_4^{(n)} \\ S_6^{(n)} \end{bmatrix} = \begin{bmatrix} 0 & \frac{n\pi}{2b} & 0 \\ 0 & 0 & \frac{n\pi}{2b} \\ \frac{n\pi}{2b} & 0 & 0 \end{bmatrix} \begin{bmatrix} u_1^{(n)} \\ u_2^{(n)} \\ u_3^{(n)} \end{bmatrix} = \underline{\partial}_b \underline{u}^{(n)} \tag{12}$$

$$\underline{E}^{(n)} = \begin{bmatrix} E_1^{(n)} \\ E_3^{(n)} \end{bmatrix} = -\begin{bmatrix} \frac{\partial}{\partial x_1} \\ \frac{\partial}{\partial x_3} \end{bmatrix} \phi^{(n)} = -\underline{\partial}_c \phi^{(n)}$$

$$\bar{E}_2^{(n)} = -\frac{n\pi}{2b}\phi^{(n)} \tag{13}$$

The strain and electric fields of orders n less than

or equal to 3 may be written in a vector form using Eqns.(12) and (13):

$$\underline{S} = \begin{bmatrix} \underline{S}^{(0)} \\ \underline{S}^{(1)} \\ \underline{S}^{(2)} \\ \underline{S}^{(3)} \\ \overline{\underline{S}}^{(1)} \\ \overline{\underline{S}}^{(2)} \\ \overline{\underline{S}}^{(3)} \end{bmatrix} = \begin{bmatrix} \underline{\partial}_a & 0 & 0 & 0 \\ 0 & \underline{\partial}_a & 0 & 0 \\ 0 & 0 & \underline{\partial}_a & 0 \\ 0 & 0 & 0 & \underline{\partial}_a \\ 0 & \underline{\partial}_b & 0 & 0 \\ 0 & 0 & \underline{\partial}_b & 0 \\ 0 & 0 & 0 & \underline{\partial}_b \end{bmatrix}_{29x12} \begin{bmatrix} \underline{u}^{(0)} \\ \underline{u}^{(1)} \\ \underline{u}^{(2)} \\ \underline{u}^{(3)} \end{bmatrix}_{12x1} = \underline{\partial}_u \underline{u} \quad (14)$$

$$\underline{E} = \begin{bmatrix} \underline{E}^{(0)} \\ \underline{E}^{(1)} \\ \underline{E}^{(2)} \\ \underline{E}^{(3)} \\ \overline{\underline{E}}^{(1)} \\ \overline{\underline{E}}^{(2)} \\ \overline{\underline{E}}^{(3)} \end{bmatrix} = -\begin{bmatrix} \underline{\partial}_c & 0 & 0 & 0 \\ 0 & \underline{\partial}_c & 0 & 0 \\ 0 & 0 & \underline{\partial}_c & 0 \\ 0 & 0 & 0 & \underline{\partial}_c \\ 0 & \frac{\pi}{2b} & 0 & 0 \\ 0 & 0 & \frac{\pi}{b} & 0 \\ 0 & 0 & 0 & \frac{3\pi}{2b} \end{bmatrix}_{11x4} \begin{bmatrix} \phi^{(0)} \\ \phi^{(1)} \\ \phi^{(2)} \\ \phi^{(3)} \end{bmatrix}_{4x1} = -\underline{\partial}_\phi \underline{\phi} \quad (15)$$

In Eqns.(14) and (15), the dimensions of the vectors and matrices are indicated by subscripts *m x n* where *m* is the number of rows and *n* the number of columns.

IId. Kinetic Energy Density and Electric Enthalpy Density of the Plate.

The kinetic energy density and electric enthalpy density per unit area of the plate are

$$\overline{K} = \rho\left(\dot{u}_1^{(0)}\dot{u}_1^{(0)} + \frac{1}{\alpha_2^2}(\dot{u}_2^{(0)}\dot{u}_2^{(0)}) + \dot{u}_3^{(0)}\dot{u}_3^{(0)}\right) + \frac{1}{2}\rho\sum_{m=1}^{N}(\dot{u}_j^{(m)}\dot{u}_j^{(m)}) \quad (16)$$

$$\overline{H} = C_{ijkl}\left(\begin{array}{l} S_{ij}^{(0)}S_{kl}^{(0)} + \frac{1}{2}\alpha_2^{\mu_{ij}+\mu_{kl}}S_{ij}^{(1)}S_{kl}^{(1)} + \\ \alpha_1 B_{10}S_{ij}^{(0)}\bar{S}_{kl}^{(1)} + \sum_{m=3}^{N} B_{m0}S_{ij}^{(0)}\bar{S}_{kl}^{(m)} \end{array}\right)$$

$$-e_{kij}\left(\begin{array}{l} 2E_k^{(0)}S_{ij}^{(0)} + \alpha_2^{\mu_{ij}}E_k^{(1)}S_{ij}^{(1)} + \\ \alpha_1 B_{10}(\bar{E}_k^{(1)}S_{ij}^{(0)} + E_k^{(0)}\bar{S}_{ij}^{(1)}) + \\ \sum_{m=3}^{N} B_{m0}(\bar{E}_k^{(m)}S_{ij}^{(0)} + E_k^{(0)}\bar{S}_{ij}^{(m)}) \end{array}\right) -$$

$$\varepsilon_{kl}\left(E_k^{(0)}E_l^{(0)} + \alpha_1 B_{10}E_k^{(0)}\bar{E}_l^{(1)} + \sum_{m=3}^{N} B_{m0}E_k^{(0)}\bar{E}_l^{(m)}\right)$$

$$+\sum_{n=2}^{N}\left(\frac{1}{2}C_{ijkl}S_{ij}^{(n)}S_{kl}^{(n)} - e_{kij}S_{ij}^{(n)}E_k^{(n)}\right) +$$

$$\sum_{n=1}^{N}\left[\begin{array}{l} C_{ijkl}\left(\frac{1}{2}\bar{S}_{ij}^{(n)}\bar{S}_{kl}^{(n)} + \sum_{m=1}^{N} B_{mn}S_{ij}^{(n)}\bar{S}_{kl}^{(m)}\right) - \\ e_{kij}\left(\bar{S}_{ij}^{(n)}\bar{E}_k^{(n)} + \sum_{m=1}^{N} B_{mn}(S_{ij}^{(n)}\bar{E}_k^{(m)} + \bar{S}_{ij}^{(m)}E_k^{(n)})\right) \end{array}\right]$$

$$-\varepsilon_{kl}\left(\frac{1}{2}E_k^{(n)}E_l^{(n)} + \frac{1}{2}\bar{E}_k^{(n)}\bar{E}_l^{(n)} + \sum_{m=1}^{N} B_{mn}E_k^{(n)}\bar{E}_l^{(m)}\right) \quad (17)$$

where α_1 and α_2 are correction factors,

$\mu_{ij} = \cos^2(\frac{ij\pi}{2})$ and

$$B_{mn} = \begin{cases} \dfrac{4m}{\pi(m^2-n^2)}, & m+n \text{ odd} \\ 0, & m+n \text{ even} \end{cases} \quad (18)$$

The terms C_{ijkl}, e_{kij} and ε_{kl} are, respectively, the elastic stiffness coefficients, piezoelectric strain constants and dielectric permittivity.

IIe. Constitutive Relations.

The constitutive relations may be derived from the kinetic energy density and electric enthalpy density of the plate by using the following partial derivatives:

$$T_{ij}^{(n)} = \frac{\partial \overline{H}}{\partial S_{ij}^{(n)}}, \overline{T}_{ij}^{(n)} = \frac{\partial \overline{H}}{\partial \overline{S}_{ij}^{(n)}}$$
$$D_i^{(n)} = -\frac{\partial \overline{H}}{\partial E_i^{(n)}}, \overline{D}_i^{(n)} = -\frac{\partial \overline{H}}{\partial \overline{E}_i^{(n)}} \tag{19}$$

Eqns.(17) and (19) yield the following constitutive relations of Nth order approximation, where components $u_j^{(n)}$ and $\phi^{(n)}$ for $n > N$ are set to zero. The constitutive relations are written in a matrix form suitable for use in the finite element matrix equations.

$$\underline{T}^{(0)} = 2(\underline{C}_\alpha \underline{S}^{(0)} - \underline{e}_\alpha^T \underline{E}^{(0)}) + \alpha_1 B_{10}(\underline{C}_\beta \underline{\overline{S}}^{(1)} - \underline{e}_\beta^T \overline{E}_2^{(1)}) + \sum_{m=3}^{N} B_{m0}(\underline{C}_\beta \underline{\overline{S}}^{(m)} - \underline{e}_\beta^T \overline{E}_2^{(m)}) \tag{20}$$

$$\underline{T}^{(1)} = \underline{C}_\gamma \underline{S}^{(1)} - \underline{e}_\gamma^T \underline{E}^{(1)} + \sum_{m=2}^{N} B_{m1}(\underline{C}_\beta \underline{\overline{S}}^{(m)} - \underline{e}_\beta^T \overline{E}_2^{(m)}) \tag{21}$$

$$\underline{\overline{T}}^{(1)} = \underline{C}_\theta \underline{\overline{S}}^{(1)} - \underline{e}_\theta^T \overline{E}_2^{(1)} + \alpha_1 B_{10}(\underline{C}_\beta^T \underline{S}^{(0)} - \underline{e}_\kappa^T \underline{E}^{(0)}) + \sum_{m=2}^{N} B_{1m}(\underline{C}_\beta^T \underline{S}^{(m)} - \underline{e}_\kappa^T \underline{E}^{(m)}) \tag{22}$$

$$\underline{D}^{(0)} = 2(\underline{e}_\alpha \underline{S}^{(0)} + \underline{\varepsilon}_\alpha \underline{E}^{(0)}) + \alpha_1 B_{10}(\underline{e}_\kappa \underline{\overline{S}}^{(1)} + \underline{\varepsilon}_\beta \overline{E}_2^{(1)}) + \sum_{m=3}^{N} B_{m0}(\underline{e}_\kappa \underline{\overline{S}}^{(m)} + \underline{\varepsilon}_\beta \overline{E}_2^{(m)}) \tag{23}$$

$$\underline{D}^{(1)} = \underline{e}_\gamma \underline{S}^{(1)} + \underline{\varepsilon}_\alpha \underline{E}^{(1)} + \sum_{m=2}^{N} B_{m1}(\underline{e}_\kappa \underline{\overline{S}}^{(m)} + \underline{\varepsilon}_\beta \overline{E}_2^{(m)}) \tag{24}$$

$$\overline{D}_2^{(1)} = \underline{e}_\theta \underline{\overline{S}}^{(1)} + \varepsilon_{22} \overline{E}_2^{(1)} + \alpha_1 B_{10}(\underline{e}_\beta \underline{S}^{(0)} + \underline{\varepsilon}_\beta^T \underline{E}^{(0)}) + \sum_{m=2}^{N} B_{1m}(\underline{e}_\beta \underline{S}^{(m)} + \underline{\varepsilon}_\beta^T \underline{E}^{(m)}) \tag{25}$$

where

$$\underline{T}^{(n)} = \left[T_1^{(n)}\ T_3^{(n)}\ T_4^{(n)}\ T_5^{(n)}\ T_6^{(n)}\right]^T$$
$$\underline{\overline{T}}^{(n)} = \left[\overline{T}_2^{(n)}\ \overline{T}_4^{(n)}\ \overline{T}_6^{(n)}\right]^T \tag{26}$$
$$\underline{D}^{(n)} = \left[D_1^{(n)}\ D_3^{(n)}\right]^T$$

and

$$\underline{C}_\alpha = \begin{bmatrix} C_{11} & C_{13} & C_{14} & C_{15} & C_{16} \\ C_{13} & C_{33} & C_{34} & C_{35} & C_{36} \\ C_{14} & C_{34} & C_{44} & C_{45} & C_{46} \\ C_{15} & C_{35} & C_{45} & C_{55} & C_{56} \\ C_{16} & C_{36} & C_{46} & C_{56} & C_{66} \end{bmatrix}, \underline{C}_\beta = \begin{bmatrix} C_{12} & C_{14} & C_{16} \\ C_{23} & C_{34} & C_{36} \\ C_{24} & C_{44} & C_{46} \\ C_{25} & C_{45} & C_{56} \\ C_{26} & C_{46} & C_{66} \end{bmatrix}$$

$$\underline{C}_\gamma = \begin{bmatrix} C_{11} & C_{13} & \alpha_2 C_{14} & C_{15} & \alpha_2 C_{16} \\ C_{13} & C_{33} & \alpha_2 C_{34} & C_{35} & \alpha_2 C_{36} \\ \alpha_2 C_{14} & \alpha_2 C_{34} & \alpha_2^2 C_{44} & \alpha_2 C_{45} & \alpha_2^2 C_{46} \\ C_{15} & C_{35} & \alpha_2 C_{45} & C_{55} & \alpha_2 C_{56} \\ \alpha_2 C_{16} & \alpha_2 C_{36} & \alpha_2^2 C_{46} & \alpha_2 C_{56} & \alpha_2^2 C_{66} \end{bmatrix}$$

$$\underline{C}_\theta = \begin{bmatrix} C_{22} & C_{24} & C_{26} \\ C_{24} & C_{44} & C_{46} \\ C_{26} & C_{46} & C_{66} \end{bmatrix} \tag{27}$$

$$\underline{e}_\alpha^T = \begin{bmatrix} e_{11} & e_{31} \\ e_{13} & e_{33} \\ e_{14} & e_{34} \\ e_{15} & e_{35} \\ e_{16} & e_{36} \end{bmatrix}, \underline{e}_\beta^T = \begin{bmatrix} e_{21} \\ e_{23} \\ e_{24} \\ e_{25} \\ e_{26} \end{bmatrix}, \underline{e}_\gamma^T = \begin{bmatrix} e_{11} & e_{31} \\ e_{13} & e_{33} \\ \alpha_2 e_{14} & \alpha_2 e_{34} \\ e_{15} & e_{35} \\ \alpha_2 e_{16} & \alpha_2 e_{36} \end{bmatrix}$$

$$\underline{e}_\theta^T = \begin{bmatrix} e_{22} \\ e_{24} \\ e_{26} \end{bmatrix}, \underline{e}_\kappa^T = \begin{bmatrix} e_{12} & e_{32} \\ e_{14} & e_{34} \\ e_{16} & e_{36} \end{bmatrix}$$

$$\underline{\varepsilon}_\alpha = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{13} \\ \varepsilon_{13} & \varepsilon_{33} \end{bmatrix}, \underline{\varepsilon}_\beta = \begin{bmatrix} \varepsilon_{12} \\ \varepsilon_{23} \end{bmatrix}$$

We note that the stress and electric displacement vectors in Eqn.(26) contain only the components which are relevant to the stress equations of motion (2) and charge equation of electrostatics (6). These are the components which will appear in the finite element matrices. The superscript T represents the transpose. For constitutive relations of order n greater than one, there are no correction factors, and the following equations apply:

$$\underline{T}^{(n)} = (\underline{C}_{\alpha}\underline{S}^{(n)} - \underline{e}_{\alpha}^{T}\underline{E}^{(n)}) + \sum_{m=1}^{N} B_{mn}(\underline{C}_{\beta}\bar{\underline{S}}^{(m)} - \underline{e}_{\beta}^{T}\bar{\underline{E}}_{2}^{(m)}) \tag{28}$$

$$\bar{\underline{T}}^{(n)} = \underline{C}_{\theta}\bar{\underline{S}}^{(n)} - \underline{e}_{\theta}^{T}\bar{\underline{E}}_{2}^{(n)} + \sum_{m=0}^{N} B_{nm}(\underline{C}_{\beta}^{T}\underline{S}^{(m)} - \underline{e}_{\kappa}^{T}\underline{E}^{(m)}) \tag{29}$$

$$\underline{D}^{(n)} = (\underline{e}_{\alpha}\underline{S}^{(n)} + \underline{\varepsilon}_{\alpha}\underline{E}^{(n)}) + \sum_{m=1}^{N} B_{mn}(\underline{e}_{\kappa}\bar{\underline{S}}^{(m)} + \underline{\varepsilon}_{\beta}\bar{\underline{E}}_{2}^{(m)}) \tag{30}$$

$$\bar{D}_{2}^{(n)} = \underline{e}_{\theta}\bar{\underline{S}}^{(n)} + \varepsilon_{22}\bar{E}_{2}^{(n)} + \sum_{m=0}^{N} B_{nm}(\underline{e}_{\beta}\underline{S}^{(m)} + \underline{\varepsilon}_{\beta}^{T}\underline{E}^{(m)}) \tag{31}$$

III. Finite Element Formulation for Third Overtone Thickness Vibrations of SC-Cut Quartz Resonators.

For third overtone thickness vibrations, the order n of mechanical displacement components $u_j^{(n)}$ and electric potential $\phi^{(n)}$ must be at least three. We shall take $n \le 3$ since components of order four and above are essentially not excited at frequencies in the vicinity of the third overtone thickness shear vibrations. Two coupled variational equations suitable for finite element formulation are formed by multiplying eqns.(2) and (6) by virtual components $\delta u_j^{(n)}$ and $\delta\phi^{(n)}$, respectively and integrating the products with respect to A, the area of plate:

$$\sum_{n=1}^{3}\int_{A}\begin{pmatrix} T_{ij,i}^{(n)} - \frac{n\pi}{2b}\bar{T}_{2j}^{(n)} + \frac{1}{b}F_j^{(n)} \\ -(1+\delta_{n0})\rho\ddot{u}_j^{(n)}\end{pmatrix}\delta u_j^{(n)}\,dA = 0 \tag{32}$$

$$\sum_{n=1}^{3}\int_{A}(D_{i,i}^{(n)} - \frac{n\pi}{2b}\bar{D}_{2}^{(n)} + \frac{1}{b}D^{(n)})\,\delta\phi^{(n)}\,dA = 0 \tag{33}$$

By using the divergence theorem, integration by parts and relations in Eqns.(10) and (11), the variational equations yield

$$\sum_{n=1}^{3}\int_{A}\begin{pmatrix}\delta S_{ij}^{(n)}T_{ij}^{(n)} + \delta\bar{S}_{ij}^{(n)}\bar{T}_{ij}^{(n)} + \\ (1+\delta_{n0})\rho\delta u_j^{(n)}\ddot{u}_j^{(n)}\end{pmatrix}dA = \sum_{n=1}^{3}\left[\int_{A}(\frac{1}{b}\delta u_j^{(n)}F_j^{(n)})\,dA + \int_{C}(\delta u_j^{(n)}t_j^{(n)})\,dC\right] \tag{34}$$

$$\sum_{n=1}^{3}\int_{A}(\delta E_i^{(n)}D_i^{(n)} + \delta\bar{E}_2^{(n)}\bar{D}_2^{(n)})\,dA = -\int_{A}(\frac{1}{b}\delta\phi^{(n)}D^{(n)})\,dA - \int_{C}(\delta\phi^{(n)}q^{(n)})\,dC \tag{35}$$

where $\delta S_{ij}^{(n)}, \delta\bar{S}_{ij}^{(n)}, \delta E_i^{(n)}$ and $\delta\bar{E}_i^{(n)}$ are the variational equivalents of strains and electric fields in eqns.(10) and (11). The last integrals on the right hand side of Eqns.(34) and (35) are line integrals where $t_j^{(n)} = n_a T_{aj}^{(n)}$ and $q^{(n)} = n_a D_a^{(n)}$ on C are respectively the nth order edge traction and nth order edge charge. n_a are the unit direction vectors along the edge C. The equations are recast in a more useful matrix form, namely,

$$\int_{A}(\delta\underline{S}^{T}\underline{T} + \delta\underline{u}^{T}\underline{m}\ddot{\underline{u}})\,dA = \int_{A}(\delta\underline{u}^{T}\underline{F})\,dA + \int_{C}(\delta\underline{u}^{T}\underline{f})\,dC \tag{36}$$

$$\int_{A}(\delta\underline{E}^{T}\underline{D})\,dA = -\int_{A}(\delta\underline{\phi}^{T}\underline{Q})\,dA - \int_{C}(\delta\underline{\phi}^{T}\underline{q})\,dC \tag{37}$$

where the following vectors and matrices are defined:

$$\delta \underline{S} = \delta \begin{bmatrix} \underline{S}^{(0)} \\ \underline{S}^{(1)} \\ \underline{S}^{(2)} \\ \underline{S}^{(3)} \\ \overline{\underline{S}}^{(1)} \\ \overline{\underline{S}}^{(2)} \\ \overline{\underline{S}}^{(3)} \end{bmatrix}_{29x1}, \underline{T} = \begin{bmatrix} \underline{T}^{(0)} \\ \underline{T}^{(1)} \\ \underline{T}^{(2)} \\ \underline{T}^{(3)} \\ \overline{\underline{T}}^{(1)} \\ \overline{\underline{T}}^{(2)} \\ \overline{\underline{T}}^{(3)} \end{bmatrix}_{29x1} \tag{38}$$

$$\delta \underline{E} = \delta \begin{bmatrix} \underline{E}^{(0)} \\ \underline{E}^{(1)} \\ \underline{E}^{(2)} \\ \underline{E}^{(3)} \\ \overline{\underline{E}}_2^{(1)} \\ \overline{\underline{E}}_2^{(2)} \\ \overline{\underline{E}}_2^{(3)} \end{bmatrix}_{11x1}, \underline{D} = \begin{bmatrix} \underline{D}^{(0)} \\ \underline{D}^{(1)} \\ \underline{D}^{(2)} \\ \underline{D}^{(3)} \\ \overline{\underline{D}}_2^{(1)} \\ \overline{\underline{D}}_2^{(2)} \\ \overline{\underline{D}}_2^{(3)} \end{bmatrix}_{11x1} \tag{39}$$

$$\underline{m} = \begin{bmatrix} \underline{m}_0 & [0] & [0] & [0] \\ [0] & \underline{m}_1 & [0] & [0] \\ [0] & [0] & \underline{m}_1 & [0] \\ [0] & [0] & [0] & \underline{m}_1 \end{bmatrix}_{12x12} \tag{40}$$

$$\underline{m}_0 = \begin{bmatrix} 2\rho & 0 & 0 \\ 0 & \frac{2\rho}{\alpha^2} & 0 \\ 0 & 0 & 2\rho \end{bmatrix}_{3x3}, \underline{m}_1 = \begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{bmatrix}_{3x3}$$

$$\underline{F} = \begin{bmatrix} \underline{F}^{(0)} \\ \underline{F}^{(1)} \\ \underline{F}^{(2)} \\ \underline{F}^{(3)} \end{bmatrix}_{12x1}, \underline{f} = \begin{bmatrix} \underline{t}^{(0)} \\ \underline{t}^{(1)} \\ \underline{t}^{(2)} \\ \underline{t}^{(3)} \end{bmatrix}_{12x1} \tag{41}$$

where $\underline{F}^{(n)} = \frac{1}{b}[F_1^{(n)}, F_2^{(n)}, F_3^{(n)}]^T$

and $\underline{t}^{(n)} = [t_1^{(n)}, t_2^{(n)}, t_3^{(n)}]^T$

$$\underline{Q} = \frac{1}{b}\begin{bmatrix} D^{(0)} \\ D^{(1)} \\ D^{(2)} \\ D^{(3)} \end{bmatrix}_{4x1}, \underline{q} = \begin{bmatrix} q^{(0)} \\ q^{(1)} \\ q^{(2)} \\ q^{(3)} \end{bmatrix}_{4x1} \tag{42}$$

Substitution of the constitutive relations from section **II**e into the stress vector $\underline{T}$ and electric displacement vector $\underline{D}$ yields:

$$\begin{aligned} \underline{T} &= \hat{\underline{C}}\underline{S} - \hat{\underline{e}}^T\underline{E} \\ \underline{D} &= \hat{\underline{e}}\underline{S} + \hat{\underline{\varepsilon}}\underline{E} \end{aligned} \tag{43}$$

where

$$\hat{\underline{C}} = \begin{bmatrix} 2\underline{C}_\alpha & 0 & 0 & 0 & \alpha_1 B_{10}\underline{C}_\beta & 0 & B_{30}\underline{C}_\beta \\ 0 & \underline{C}_\gamma & 0 & 0 & 0 & B_{21}\underline{C}_\beta & 0 \\ 0 & 0 & \underline{C}_\alpha & 0 & B_{12}\underline{C}_\beta & 0 & B_{32}\underline{C}_\beta \\ 0 & 0 & 0 & \underline{C}_\alpha & 0 & B_{23}\underline{C}_\beta & 0 \\ \alpha_1 B_{10}\underline{C}_\beta^T & 0 & B_{12}\underline{C}_\beta^T & 0 & \underline{C}_\theta & 0 & 0 \\ 0 & B_{21}\underline{C}_\beta^T & 0 & B_{23}\underline{C}_\beta^T & 0 & \underline{C}_\theta & 0 \\ B_{30}\underline{C}_\beta^T & 0 & B_{32}\underline{C}_\beta^T & 0 & 0 & 0 & \underline{C}_\theta \end{bmatrix} \tag{44}$$

$$\hat{\underline{e}} = \begin{bmatrix} 2\underline{e}_\alpha & 0 & 0 & 0 & \alpha_1 B_{10}\underline{e}_\kappa & 0 & B_{30}\underline{e}_\kappa \\ 0 & \underline{e}_\gamma & 0 & 0 & 0 & B_{21}\underline{e}_\kappa & 0 \\ 0 & 0 & \underline{e}_\alpha & 0 & B_{12}\underline{e}_\kappa & 0 & B_{32}\underline{e}_\kappa \\ 0 & 0 & 0 & \underline{e}_\alpha & 0 & B_{23}\underline{e}_\kappa & 0 \\ \alpha_1 B_{10}\underline{e}_\beta & 0 & B_{12}\underline{e}_\beta & 0 & \underline{e}_\theta & 0 & 0 \\ 0 & B_{21}\underline{e}_\beta & 0 & B_{23}\underline{e}_\beta & 0 & \underline{e}_\theta & 0 \\ B_{30}\underline{e}_\beta & 0 & B_{32}\underline{e}_\beta & 0 & 0 & 0 & \underline{e}_\theta \end{bmatrix} \tag{45}$$

$$\underline{\varepsilon} = \begin{bmatrix} 2\underline{\varepsilon}_\alpha & 0 & 0 & 0 & \alpha_1 B_{10}\underline{\varepsilon}_\beta & 0 & B_{30}\underline{\varepsilon}_\beta \\ 0 & \underline{\varepsilon}_\alpha & 0 & 0 & 0 & B_{21}\underline{\varepsilon}_\beta & 0 \\ 0 & 0 & \underline{\varepsilon}_\alpha & 0 & B_{12}\underline{\varepsilon}_\beta & 0 & B_{32}\underline{\varepsilon}_\beta \\ 0 & 0 & 0 & \underline{\varepsilon}_\alpha & 0 & B_{23}\underline{\varepsilon}_\beta & 0 \\ \alpha_1 B_{10}\underline{\varepsilon}_\beta & 0 & B_{12}\underline{\varepsilon}_\beta & 0 & \varepsilon_{22} & 0 & 0 \\ 0 & B_{21}\underline{\varepsilon}_\beta & 0 & B_{23}\underline{\varepsilon}_\beta & 0 & \varepsilon_{22} & 0 \\ B_{30}\underline{\varepsilon}_\beta & 0 & B_{32}\underline{\varepsilon}_\beta & 0 & 0 & 0 & \varepsilon_{22} \end{bmatrix} \quad (46)$$

The constitutive matrix equations (43) are substituted into the weak forms of Eqns.(36) and (37):

$$\int_A (\delta \underline{S}^T \hat{\underline{C}} \underline{S} - \delta \underline{S}^T \hat{\underline{e}}^T \underline{E} + \delta \underline{u}^T \underline{m} \ddot{\underline{u}})\, dA = \int_A \delta \underline{u}^T \underline{F} dA + \int_C (\delta \underline{u}^T f)\, dC \quad (47)$$

$$\int_A (\delta \underline{E}^T \hat{\underline{e}} \underline{S} + \delta \underline{E}^T \hat{\underline{\varepsilon}} \underline{E})\, dA = -\int_A \delta \underline{\phi}^T \underline{Q} dA - \int_C (\delta \underline{\phi}^T q)\, dC \quad (48)$$

If each element has r number of nodes, the mechanical displacement and electric potential fields may be replaced by the linear combinations of finite element shape functions N_a (a = 1,2,3,...,r) with coefficients which represent the field values at the nodes. We use the symbols $\psi_a^{(n)}$ and $d_{ia}^{(n)}$ to denote the nodal values of the nth order electric potential $\phi^{(n)}$ and mechanical displacements $u_i^{(n)}$ at node a, respectively. Note that the nodal values are functions of time only. We write

$$\underline{\phi} = [\tilde{\underline{N}}_1, \tilde{\underline{N}}_2, ..., \tilde{\underline{N}}_r]_{4x4r} \begin{bmatrix} \underline{\Psi}_1 \\ \underline{\Psi}_2 \\ . \\ \underline{\Psi}_r \end{bmatrix}_{4rx1} = \tilde{\underline{N}} \underline{\Psi}$$

$$\text{where } \tilde{\underline{N}}_a = \begin{bmatrix} N_a & 0 & 0 & 0 \\ 0 & N_a & 0 & 0 \\ 0 & 0 & N_a & 0 \\ 0 & 0 & 0 & N_a \end{bmatrix}_{4x4}, \underline{\Psi}_a = \begin{bmatrix} \psi_a^{(0)} \\ \psi_a^{(1)} \\ \psi_a^{(2)} \\ \psi_a^{(3)} \end{bmatrix}_{4x1} \quad (49)$$

$$\underline{u} = \begin{bmatrix} \underline{u}^{(0)} \\ \underline{u}^{(1)} \\ \underline{u}^{(2)} \\ \underline{u}^{(3)} \end{bmatrix}_{12x1} = [\hat{\underline{N}}_1, \hat{\underline{N}}_2, ..., \hat{\underline{N}}_r]_{12x12r} \begin{bmatrix} \underline{d}_1 \\ \underline{d}_2 \\ . \\ \underline{d}_r \end{bmatrix}_{12rx1} = \hat{\underline{N}} \underline{d}$$

$$\text{where } \hat{\underline{N}}_a = N_a \underline{I}_{12x12} \quad (50)$$

$$\text{and } \underline{d}_a = \begin{bmatrix} \underline{d}_a^{(0)} \\ \underline{d}_a^{(1)} \\ \underline{d}_a^{(2)} \\ \underline{d}_a^{(3)} \end{bmatrix}_{12x1}, \underline{d}_a^{(n)} = \begin{bmatrix} d_{1a}^{(n)} \\ d_{2a}^{(n)} \\ d_{3a}^{(n)} \end{bmatrix}_{3x1}$$

The matrix $\underline{I}_{12x12}$ is a 12x12 identity matrix.

Using Eqns.(49) and (50), we write the strain and electric field vectors from Eqns.(14) and (15) as

$$\underline{S} = \partial_u \hat{\underline{N}} \underline{d}$$
$$\underline{E} = -\partial_\phi \tilde{\underline{N}} \underline{\Psi} \quad (51)$$

Eqn.(51) is substituted into Eqns.(47) and (48) to yield

$$\delta \underline{d}^T (\underline{K}_0 \underline{d} + \underline{P}^T \underline{\Psi} + \underline{M} \ddot{\underline{d}}) = \delta \underline{d}^T (\underline{F}_S + \underline{F}_C) \quad (52)$$

$$\delta \underline{\phi}^T (\underline{P} \underline{d} - \underline{R} \underline{\Psi}) = \delta \underline{\phi}^T (\underline{Q}_S + \underline{Q}_C) \quad (53)$$

where

$$\underline{K}_0 = \left[\int_A [\partial_u \hat{N}]^T \hat{C} [\partial_u \hat{N}]\, dA\right]_{12rx12r}$$
$$\underline{P} = \left[\int_A [\partial_\phi \tilde{N}]^T \hat{e} [\partial_u \hat{N}]\, dA\right]_{4rx12r}$$
$$\underline{R} = \left[\int_A [\partial_\phi \tilde{N}]^T \hat{\varepsilon} [\partial_\phi \tilde{N}]\, dA\right]_{4rx4r} \quad (54)$$
$$\underline{M} = \left[\int_A \hat{N}^T \underline{m} \hat{N} dA\right]_{12rx12r}$$

and

$$\underline{F}_S = \left[\int_A \hat{N}^T \underline{F} dA\right]_{12rx1}, \underline{F}_C = \left[\int_C \hat{N}^T \underline{t} dC\right]_{12rx1}$$
$$\underline{Q}_S = \left[\int_A \tilde{N}^T \underline{F} dA\right]_{4rx1}, \underline{Q}_C = \left[\int_C \tilde{N}^T \underline{t} dC\right]_{4rx1} \quad (55)$$

The matrices $\underline{K}_0$, $\underline{P}$, $\underline{R}$ and $\underline{M}$ are the element mechanical stiffness, piezoelectric coupling, dielectric stiffness and mass, respectively. The vectors $\underline{F}_S$, $\underline{F}_C$, $\underline{Q}_S$ and $\underline{Q}_C$ in Eqn.(55) are, respectively, the element face traction, edge traction, face charge and edge charge. Since the variational mechanical displacements and electric potentials in Eqns.(52) and (53) are arbitrary and linearly independent, the two equations yield the following finite element matrix equations for a piezoelectric vibration problem:

$$\underline{K}_0 \underline{d} + \underline{P}^T \underline{\Psi} + \underline{M} \ddot{\underline{d}} = \underline{F}_S + \underline{F}_C$$
$$\underline{P} \underline{d} - \underline{R} \underline{\Psi} = \underline{Q}_S + \underline{Q}_C \quad (56)$$

The element matrices and vectors in Eqns.(54) and (55) are assembled to form global matrices and vectors which represent the finite element model of the whole piezoelectric resonator. The assembly process may be found in numerous text books on the finite element method [2]. The assembled global finite element equations have the same form as Eqn.(56), hence in later sections, we shall refer to these equations as the finite element equations for piezoelectric vibrations. The dielectric matrix $\underline{R}$ is in general a singular matrix, unless one or more nodal electric potential are specified (essential boundary conditions).

IV. Perturbation Technique for Modeling Piezoelectric Effects on Mechanical Vibrations.

We consider without loss of generality the free vibration problem for the system of equations in Eqn.(56). This is done by setting the terms on the right hand side to zero and assuming simple harmonic motion ($\bar{\underline{d}} = \underline{d} e^{i\omega t}$, $\underline{\Psi} = \bar{\underline{\Psi}} e^{i\omega t}$) to obtain

$$\underline{K}_0 \bar{\underline{d}} + \underline{P}^T \bar{\underline{\Psi}} = \omega^2 \underline{M} \bar{\underline{d}}$$
$$\underline{P} \bar{\underline{d}} - \underline{R} \bar{\underline{\Psi}} = 0 \quad (57)$$

We may solve the second equation of Eqn.(57) for nodal electric potential $\bar{\underline{\Psi}}$ (assuming that the inverse of $\underline{R}$ exists) and substitute the result into the first equation to yield a general eigenvalue problem:

$$\tilde{\underline{K}} \bar{\underline{d}} = \lambda \underline{M} \bar{\underline{d}}$$
$$\text{where } \tilde{\underline{K}} = \underline{K}_0 + \underline{P}^T \underline{R}^{-1} \underline{P}, \text{ and } \omega^2 = \lambda \quad (58)$$

A serious disadvantage in forming the piezoelectrically stiffened, mechanical matrix $\tilde{\underline{K}}$ arises from its full matrix structure. That is, while the mechanical stiffness matrix $\underline{K}_0$ has a sparse matrix structure, the addition of the piezoelectric stiffening term $\underline{P}^T \underline{R}^{-1} \underline{P}$ renders the resulting matrix $\tilde{\underline{K}}$ full. This represents significant losses in memory and computation time from a numerical point of view, which is especially true for large finite element models. The piezoelectric stiffening term $\underline{P}^T \underline{R}^{-1} \underline{P}$ is however a few orders of magnitude smaller than the mechanical stiffness matrix $\underline{K}_0$, and hence Eqn.(58) lends itself to a classic application of perturbation methods.

A perturbation technique which yields rather accurate results for practical purposes and retains the sparse matrix structure is formulated. Let

$$\underline{P}^T\underline{R}^{-1}\underline{P} = \tau\underline{K}_1 \tag{59}$$

where τ is an arbitrary scalar perturbation parameter of the same order of magnitude as the estimated piezoelectric coupling constant of the resonator. The eigenpair $(\lambda_i, \bar{\underline{d}}_i)$ of Eqn.(58) is expanded in a power series of τ:

$$\begin{aligned}\bar{\underline{d}}_i &= \bar{\underline{d}}_i^{(0)} + \tau\bar{\underline{d}}_i^{(1)} + \tau^2\bar{\underline{d}}_i^{(2)} \\ \lambda_i &= \lambda_i^{(0)} + \tau\lambda_i^{(1)} + \tau^2\lambda_i^{(2)}\end{aligned} \tag{60}$$

Eqns.(59) and (60) are substituted into Eqn.(58) to give:

$$\begin{aligned}&(\underline{K}_0 + \tau\underline{K}_1)\,(\bar{\underline{d}}_i^{(0)} + \tau\bar{\underline{d}}_i^{(1)} + \tau^2\bar{\underline{d}}_i^{(2)}) = \\ &(\lambda_i^{(0)} + \tau\lambda_i^{(1)} + \tau^2\lambda_i^{(2)})\,\underline{M}\,(\bar{\underline{d}}_i^{(0)} + \tau\bar{\underline{d}}_i^{(1)} + \tau^2\bar{\underline{d}}_i^{(2)})\end{aligned} \tag{61}$$

which is expanded to yield

$$\begin{aligned}&(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(0)} + \\ &\tau\,[\,(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(1)} + (\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(0)}\,] + \\ &\tau^2\,[\,(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(2)} + (\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(1)} \\ &-\lambda_i^{(2)}M\underline{d}_i^{(0)}\,] = 0\end{aligned} \tag{62}$$

where the terms of order τ^3 or higher are dropped. Since the perturbation parameter τ is arbitrary, its coefficients must be zero:

$$(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(0)} = 0 \tag{63}$$

$$(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(1)} + (\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(0)} = 0 \tag{64}$$

$$(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(2)} + (\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(1)} = \lambda_i^{(2)}M\underline{d}_i^{(0)} \tag{65}$$

Eqn.(63) is the unperturbed, purely mechanical eigenvalue problem with sparse stiffness and mass matrices. The equation will yield the unperturbed solution $(\lambda_i^{(0)}, \bar{\underline{d}}_i^{(0)})$. Relations for calculating $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ are obtained by premultiplying Eqns.(64) and (65) respectively by $[\bar{\underline{d}}_i^{(0)}]^T$ and manipulating the resulting equations to give

$$\lambda_i^{(1)} = [\bar{\underline{d}}_i^{(0)}]^T\underline{K}_1\bar{\underline{d}}_i^{(0)} \tag{66}$$

$$\lambda_i^{(2)} = [\bar{\underline{d}}_i^{(0)}]^T\underline{K}_1\bar{\underline{d}}_i^{(1)} \tag{67}$$

Eqns.(64) and (65) may be rewritten in a form suitable for calculating the perturbation vectors $\bar{\underline{d}}_i^{(1)}$ and $\bar{\underline{d}}_i^{(2)}$:

$$\begin{aligned}&(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(1)} = \underline{v}_i^{(1)} \\ &\text{where } \underline{v}_i^{(1)} = -(\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(0)}\end{aligned} \tag{68}$$

$$\begin{aligned}&(\underline{K}_0 - \lambda_i^{(0)}\underline{M})\,\underline{d}_i^{(2)} = \underline{v}_i^{(2)} \\ &\text{where } \underline{v}_i^{(2)} = -(\underline{K}_1 - \lambda_i^{(1)}\underline{M})\,\underline{d}_i^{(1)} + \lambda_i^{(2)}M\underline{d}_i^{(0)}\end{aligned} \tag{69}$$

Infinitely many solutions exist for vectors $\bar{\underline{d}}_i^{(1)}$ and $\bar{\underline{d}}_i^{(2)}$ in Eqns.(68) and (69), respectively, since the vectors $\underline{v}_i^{(1)}$ and $\underline{v}_i^{(2)}$ are orthogonal to the eigenvector $\bar{\underline{d}}_i^{(0)}$ corresponding to $\lambda_i^{(0)}$ in the two equations [3]. Extra relations from the normalization of eigenvector $\bar{\underline{d}}_i$ with respect to the mass matrix $\underline{M}$ may be used to obtain unique solutions for $\underline{v}_i^{(1)}$ and $\underline{v}_i^{(2)}$:

$$\bar{\underline{d}}_i^T\underline{M}\bar{\underline{d}}_i = 1 \tag{70}$$

Upon substitution of Eqn.(60) into the equation, we obtain

$$\begin{aligned}&[\bar{\underline{d}}_i^{(0)}]^T\underline{M}\bar{\underline{d}}_i^{(0)} + \tau\left[2\,[\bar{\underline{d}}_i^{(0)}]^T\underline{M}\bar{\underline{d}}_i^{(1)}\right] + \\ &\tau^2\left[2\,[\bar{\underline{d}}_i^{(0)}]^T\underline{M}\bar{\underline{d}}_i^{(2)} + [\bar{\underline{d}}_i^{(1)}]^T\underline{M}\bar{\underline{d}}_i^{(1)}\right] = 1\end{aligned} \tag{71}$$

where terms of order τ^3 and above are neglected. The first term in Eqn.(71) represents the normalization of the unperturbed eigenvector $\bar{\underline{d}}_i^{(0)}$:

$$[\bar{\underline{d}}_i^{(0)}]^T\underline{M}\bar{\underline{d}}_i^{(0)} = 1 \tag{72}$$

We obtain from Eqns.(71) and (72) two additional relations

$$[\bar{d}_i^{(0)}]^T \underline{M} \bar{d}_i^{(1)} = 0 \tag{73}$$

$$[\bar{d}_i^{(0)}]^T \underline{M} \bar{d}_i^{(2)} = -\frac{1}{2}[\bar{d}_i^{(1)}]^T \underline{M} \bar{d}_i^{(1)} \tag{74}$$

which are employed with (N-1) equations of Eqns.(68) and (69), respectively, to get unique solutions for the perturbation vectors $\bar{d}_i^{(1)}$ and $\bar{d}_i^{(2)}$.

The proposed perturbation technique for piezoelectric vibration problems is similar to that presented by Lee and Yong for frequency-temperature behavior of quartz plates[4]. The technique is more general than the perturbation method proposed by Boucher, Lagier and Maerfeld[5] which requires a bandwidth of eigenpairs $(\lambda_i^{(0)}, \bar{d}_i^{(0)})$ to calculate the perturbation vectors. Results for the present technique are presented in the next section.

V. Results for Straight Crested Waves in Unelectroded and Electroded, Third Overtone SC-cut Resonators.

Figure 1 shows the reference coordinate system and dimensions for the plate and electrodes of an SC-cut quartz strip. Traction free conditions are assumed at the edges at $x_1 = \pm a$. The straight crested waves are assumed propagating in the x_1 direction. Straight crested waves are obtained by assuming that the partial derivatives of all field variables with respect to x_3 are equal to zero:

$$\frac{\partial(...)}{\partial x_3} = 0 \tag{75}$$

Since the straight crested waves are spatially one-dimensional, a four-node one-dimensional element with cubic shape functions is employed.

Piezoelectric vibrations with the mass effects of electrodes in thickness shear resonators excited by electric fields in the thickness direction may be modelled using the same finite element equations developed in the section **III**, but replacing the mass terms $\underline{m}_0$ and $\underline{m}_1$ in Eqn.(40) and face traction terms $F_j^{(n)}$ in Eqn.(5) by, respectively,

$$\underline{m}_0 = (1+R)\begin{bmatrix} 2\rho & 0 & 0 \\ 0 & \frac{2\rho}{\alpha^2} & 0 \\ 0 & 0 & 2\rho \end{bmatrix}_{3x3} \tag{76}$$

$$\underline{m}_1 = (1+2R)\begin{bmatrix} \rho & 0 & 0 \\ 0 & \rho & 0 \\ 0 & 0 & \rho \end{bmatrix}_{3x3}$$

where $R = \dfrac{2b'\rho'}{b\rho}$ is the ratio of the mass of both electrode platings per unit area to the mass of the crystal plate per unit area, and

$$F_j^{(n)} = T_{2j}(B) - (-1)^n T_{2j}(-B) \quad \text{where } B = b + 2b' \tag{77}$$

All frequencies, ω, reported in this paper are normalized frequencies, Ω, where

$$\Omega = \frac{\omega}{\omega_1} \text{ and } \omega_1^2 = \frac{\pi^2 C_{66}}{4b^2\rho}. \tag{78}$$

Va. Mechanical Vibrations.

The accuracy of the finite element program is first tested against analytical solutions for purely mechanical vibrations. Figures 2a and 2b show, respectively, the "exact" and finite element normalized frequency and mode shapes for the fundamental thickness shear mode of the SC-cut strip (a/b ratio = 19.99478) without electrodes. The numerical frequency is 0.03% lower than the "exact" frequency. Their mode shapes compare very well. We note that at this frequency the higher order displacements $u_j^{(n)}$ (for $n > 1$) are quiescent, which reconfirm the assumption that these higher order displacements can be

neglected when the frequency of interest is near the fundamental thickness-shear frequency.

Figure 3 shows the frequency spectrum in the vicinity of the third harmonic overtone of thickness-shear frequency at a/b ratios ranging from 18.0 to 20.0. The symbol TS-3,n represents the third overtone thickness-shear mode with n number of half waves of the predominant displacement component $u_1^{(3)}$. Figure 3a gives the mode shapes for TS-3,1. It is interesting to note that for this mode, all displacement components, except for $u_j^{(3)}, j = 1, 2, 3$, are quiescent.

Symbols En, FSn and Fn in Fig.3 represent, respectively, the extensional, face shear and flexural modes, where n is the number of half waves of their respective predominant displacement components. Due to the strong couplings of all displacement components for these modes, there is difficulty in stating unequivocally that the given mode is either flexural, extensional or face shear. To illustrate these strong couplings, we show in Fig. 3b the F67 mode (a/b = 18.5), where the magnitude of the zero order displacement components are similar to the first order components.

Vb. Results on Proposed Perturbation Technique for Piezoelectric Vibrations.

In order to compare the results on the perturbation technique developed in section **IV**, a second finite element program was written which solved Eqn.(58) directly, that is, $\tilde{\underline{K}}$ was stored as a full matrix. The table below show such a comparison for the fundamental thickness shear mode and its anharmonics with a/b = 19.99478.

TABLE 1. Comparison of results from perturbation method against direct solution of Eqn.(58)

Mode	Perturbation Method	Direct Solution	Error (%)
TS-1,1	0.9433721	0.9433723	-2×10^{-5}
TS-1,3	0.9756843	0.9756874	-3×10^{-4}
TS-1,5	0.9895339	0.9895293	$+5\times10^{-4}$

The perturbation results compare very well with the direct solution of Eqn.(58). Hence, the technique may be employed in large finite element models to enable substantial savings in memory and computational time. The comparison of eigenvectors are not shown due to their negligible differences.

Vc. Straight Crested Waves of Piezoelectric Vibrations in a Partially Electroded SC-cut Quartz Strip.

Figure 4a shows the frequency and mode shapes of the third overtone thickness-shear mode in a partially electroded SC-cut strip (a/b = 20.0, a'/b = 10.0, R = 0.01). The electrodes are in an open circuit. The electric potential $\phi^{(n)}, n = 0, 1, 2, 3$; and electric fields E_1 and E_2 at the top and mid-plane of plate are shown in Fig.4b. Due to energy trapping beneath the electrodes, the electric fields in the unelectroded sections of the plate are negligible.

Steady state vibrations driven by an alternating voltage $\phi_{\pm b} = 0.5e^{i\omega t}$ volts at the electrodes are calculated using Eqn.(56), where $\underline{F}_s, \underline{F}_c$ and $\underline{Q}_s$ are set to zero. Due to symmetry, the electric potentials $\phi^{(0)}$ and $\phi^{(2)}$ are set to zero. The second equation of Eqn.(56) is solved for electric potential $\underline{\bar{\psi}}$ which is then substituted into the first equation to yield an inhomogeneous matrix equation. The inhomogeneous equation is solved for $\bar{\underline{d}}$ the mode shape of vibration at the frequency of electrical excitation. Figure 5a shows the displacement mode shapes for Ω = 2.78350 in a plate with a/b = 20.0, a'/b = 10.0 and R = 0.01. The displacements are normalized with respect to the magnitude of elastic stiffness C_{66}. The excitation frequency is close to the natural frequency of third overtone thickness-shear mode, hence the mode shapes of

displacement components resemble the free vibration modal shapes at the particular natural frequency. The electric potential $\phi^{(n)}$, $n = 0, 1, 2, 3$; and electric fields E_1 and E_2 at the top and mid-plane of plate are shown in Fig.5b. The sharp drop of the electric potential at the electrode edges is reflected by the two "peaks" in the electric field E_1 at the top of plate.

Acknowledgment

This work was sponsored by the Army Research Office grant no. DAAL03-91-G-0018. We gratefully acknowledge their support.

References

1. P.C.Y. Lee, S. Syngellakis, and J.P. Hou, "A Two-Dimensional Theory for High-Frequency Vibrations of Piezoelectric Crystal Plates with or without Electrodes," *J. of Appl. Physics,* 61(4), 1987, pp. 1249-1262.
2. T.J.R. Hughes, The Finite Element Method, Linear Static and Dynamic Finite Element Analysis, Prentice-Hall, 1987, Chapter 2, pp. 57-108.
3. F.B. Hildebrand, Methods of Applied Mathematics, 2nd edition, Prentice-Hall, 1965, pp. 33-34.
4. P.C.Y. Lee and Y-K. Yong, "Temperature Derivatives of Elastic Stiffness Derived from the Frequency-Temperature Behavior of Quartz Plates," *J. of Appl. Physics,* 56(5), 1984, pp. 1514-1521.
5. D. Boucher, M. Lagier, and C. Maerfeld, "Computation of the Vibrational Modes for Piezoelectric Array Transducers Using a Mixed Finite Element-Perturbation Method," *IEEE Trans. on Sonics and Ultrasonics*, Vol. SU-28, No. 5, 1981, pp. 318-330.

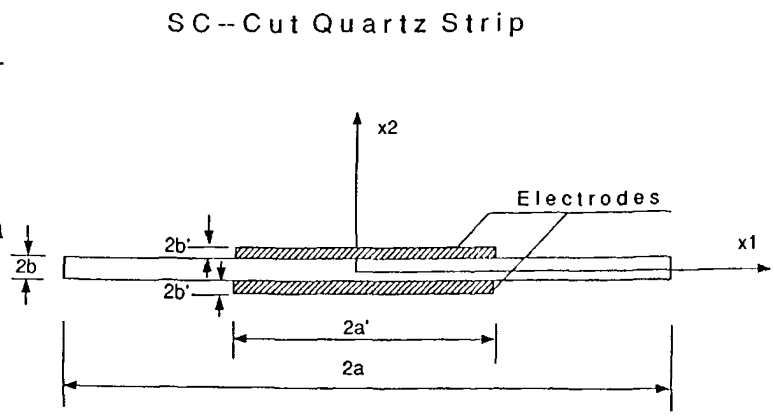


Fig.1 Reference coordinate system and plate dimensions of SC-cut quartz strip.

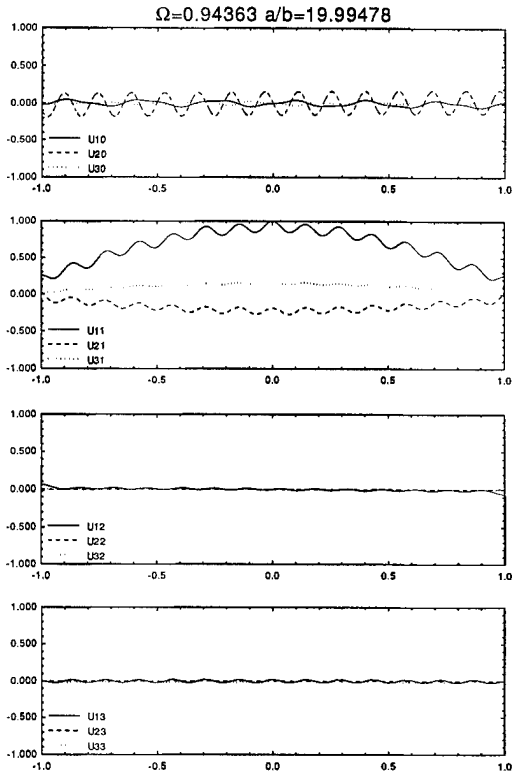


Fig.2a "Exact" frequency and mode shapes for the fundamental thickness-shear mode

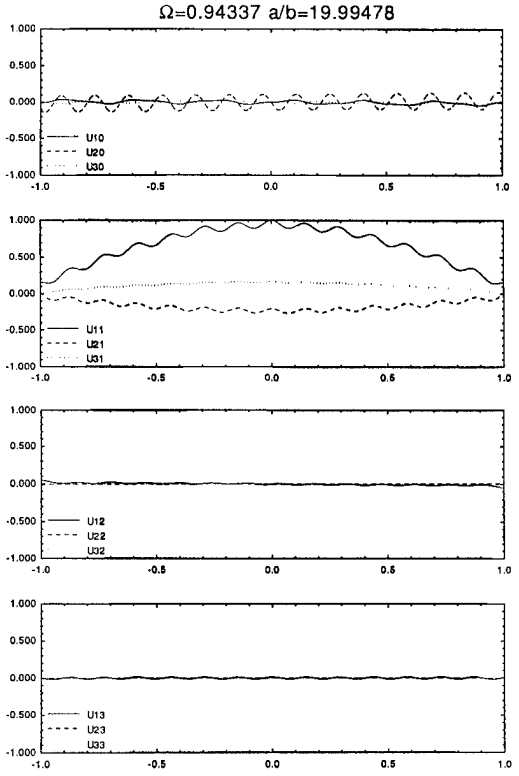


Fig.2b Numerical frequency and mode shapes for the fundamental thickness-shear mode.

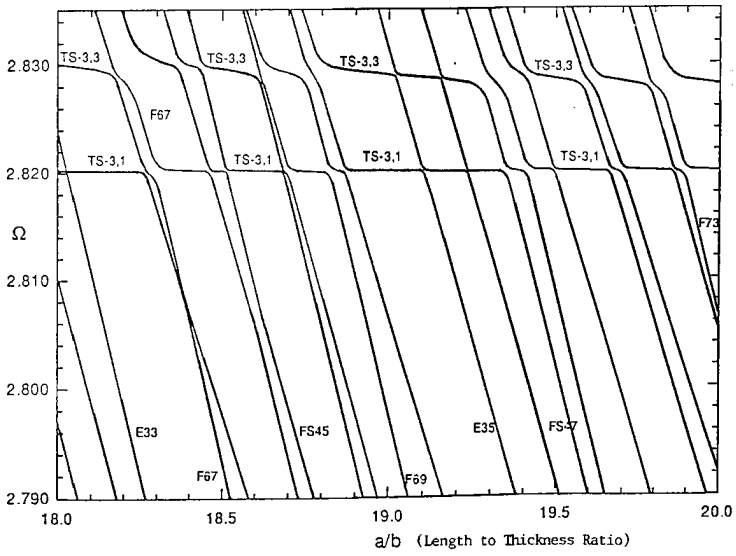


Fig.3 Frequency spectrum of straight crested waves in the vicinity of the third harmonic overtone of thickness-shear mode of the SC-cut quartz strip without electrodes.

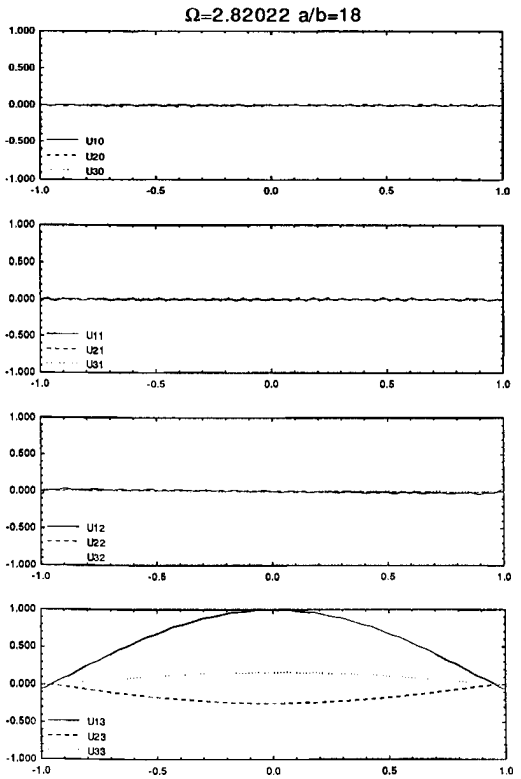


Fig.3a Frequency and mode shapes of the third harmonic overtone of thickness-shear mode from Fig.3.

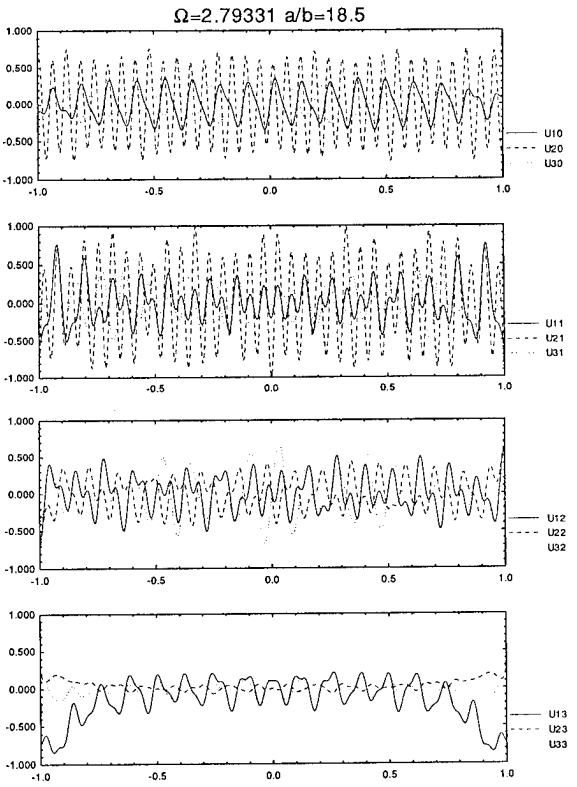


Fig.3b Frequency and mode shapes of a flexural mode F67 from the frequency spectrum of Fig.3.

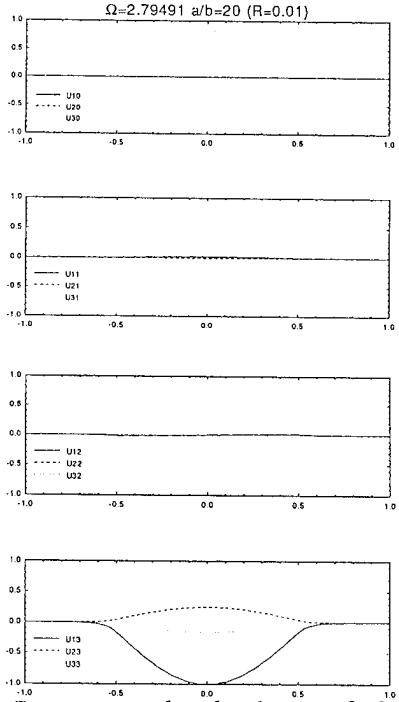


Fig.4a Frequency and mode shapes of the third overtone thickness-shear mode in a partially electroded SC-cut strip (a/b=20,a'/b=10,R=0.01)

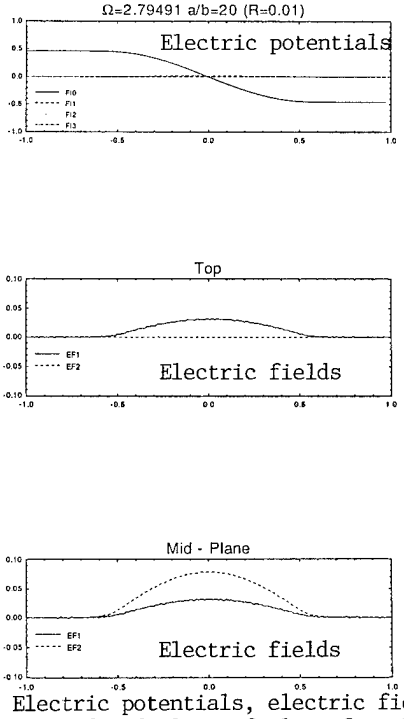


Fig.4b Electric potentials, electric fields at the top and mid-plane of plate from Fig.4b

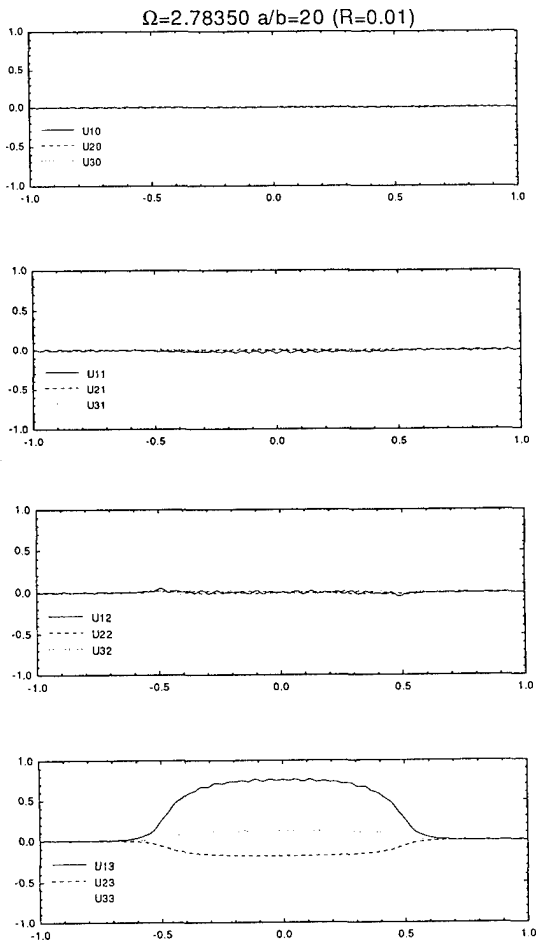


Fig.5a Steady state vibrations of the SC-cut strip with a/b=20, a'/b=10, R=0.01. Mode shapes of displacements.

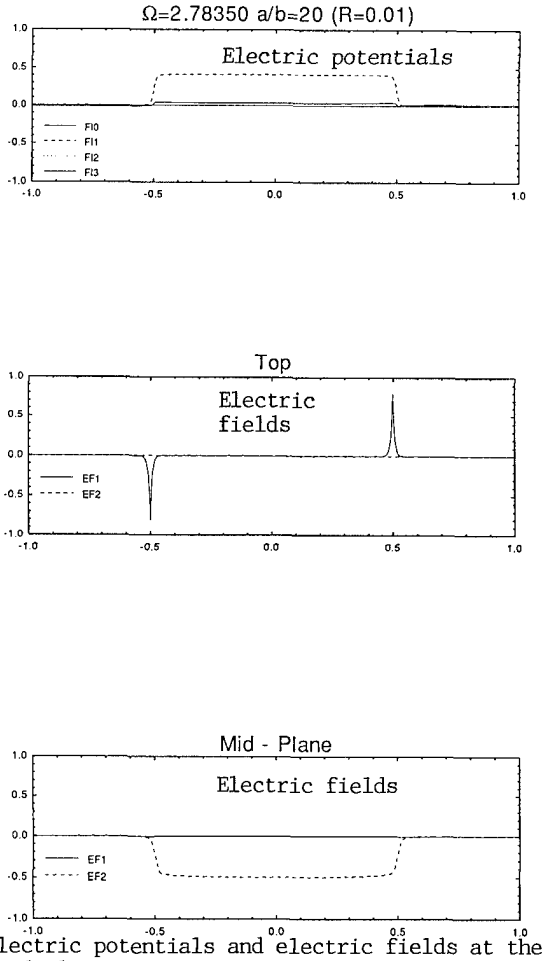


Fig.5b Electric potentials and electric fields at the top and mid-plane of plate, corresponding to the mode in Fig.5a.

FOUR-POINT SC-CUT CRYSTAL UNIT DESIGN CALCULATIONS FOR REDUCED ACCELERATION SENSITIVITY

LARRY D. CLAYTON and ERROL P. EERNISSE
Quartztronics, Inc., 1020 Atherton Dr., Bldg. C
Salt Lake City, UT 84123
(801) 266-6994 FAX (801) 266-7985

ABSTRACT

A solution technique developed to calculate frequency shifts for quartz resonators arising from mechanical stresses has been described in recent publications [1,2]. This solution technique utilizes the finite element method as a first step to calculate stress distributions produced in quartz resonators by external actions. Output from finite element solutions for crystal units is used along with thickness mode shapes in an algorithm which numerically integrates the perturbation integral derived by H.F. Tiersten [3] to calculate resonator frequency shifts. The solution technique is general in that any combination of mount and resonator geometries may be modeled. Any crystallographic orientation may be chosen and any load or combination of loads may be applied to the resonator. The frequency perturbation calculation includes movement of the mode with respect to stresses distributed throughout the resonator body. This solution technique has been used in the past to accurately calculate force-frequency coefficients for AT- and SC-cut quartz resonators [1]. The solution technique has also been used to calculate acceleration induced frequency shifts (i.e., gamma vectors) for AT- and SC-cut crystal units [1].

Recent efforts have been directed towards development of a readily manufacturable four point mount which minimizes acceleration sensitivity for 10MHz, 3rd, SC-cut crystal units. Gamma vector components have been systematically calculated for various mount shapes and for variations in the mode location and orientation of the resonator with respect to the mounts. Design criteria have been established for the mounts to improve crystal unit precision and reduce acceleration sensitivity. To minimize the effects of manufacturing tolerances, mounts should be designed for precise and repeatable fabrication to produce consistent and symmetric resonator boundary conditions.

A four point mount, called the horizontal strip mount in this report, satisfies the design criteria. Lower acceleration sensitivities are predicted for the horizontal strip mount design and optimum mode locations are in closer proximity to the resonator center over all mount orientations. Calculations have been performed which characterize variations in acceleration sensitivity with changes in critical dimensions for the horizontal strip mount. Optimum mount dimensions which minimize acceleration sensitivity have been calculated.

INTRODUCTION

Frequency instability in quartz crystal units arises from a variety of environmental effects including temperature, pressure, humidity, steady-state and time-varying inertial loads, magnetic and electric fields, concentrated loads, electrode stresses, and radiation [4-7]. In most cases frequency shifts induced by environmental effects result from mechanical stresses distributed throughout the active region of the quartz resonator. Stresses distributed throughout the resonator may arise from transient temperature gradients and from steady-state loads created by differences in thermal expansion between the quartz blank and the structure which supports it [8-15]. Resonator stresses may also arise from loads transmitted through the mounting structure originating with pressure acting on the crystal unit vacuum enclosure [16]. Electrode stresses are generated by the growth of thin films on the surfaces of the resonator during metallization processes and by differences in thermal expansion between the electrodes and the substrate [17-18]. Body forces arising from magnetic fields acting on structures made from magnetic materials which support the crystal blank induce stress in the resonator [19,20]. Stresses produced by concentrated diametrically opposed forces in the plane of quartz resonators at varying azimuthal angles have been the subject of numerous investigations [21-27]. This paper focuses primarily upon resonator stresses resulting from mount reactions to either steady-state or time-varying inertial loads typically characterized as either 2-g or acceleration sensitivity of quartz crystal units [28-44]. Time varying inertial loads generated by environmental vibrations at low frequencies are probably the most important source of frequency instability for quartz resonators.

Frequency instability resulting from environmental effects which induce mechanical stresses in quartz resonators may be analyzed using the solution technique described in References 1 and 2. Mechanical stresses distributed throughout the resonator are

0-7803-0476-4/92 $3.00 © 1992 IEEE

computed using Finite Element Analysis (FEA). Finite element models may be created which accurately characterize the mechanical behavior of quartz crystal units. Models may include full anisotropic material properties and virtually any combination of static and dynamic loads and crystal unit shape and boundary conditions. Element stresses and nodal displacements retrieved from finite element solutions for quartz crystal units are used in a program developed to calculate resonator frequency shifts [1,2]. Frequency shifts are calculated via numerical integration of the perturbation integral derived by H.F. Tiersten [3]. Recent studies suggest that the location of the thickness mode with respect to mechanical stresses distributed throughout the resonator as well as distortions in mode shape affect resonator frequency shift [34-39]. The formulation of the perturbation integral used in the solution technique from References 1 and 2 incorporates planar movement of the mode with respect to resonator stresses.

Work began in March of 1990 at Quartztronics on the Acceleration Insensitive Quartz Resonators or 2-g Project. The primary objective of this work has been to develop analytical, empirical, and manufacturing processes to measure and reduce the acceleration sensitivity of quartz thickness-shear mode disc resonators assembled into crystal units. Crystal units selected for this investigation contain a plano-convex 10MHz, 3rd, SC-cut resonator. The resonator is supported on four mounts distributed at equal angular intervals around its circumference. Spring-mass frequencies in excess of 3KHz were specified to insure that the crystal units withstand shock and vibration environments.

Several factors affect the acceleration sensitivity of crystal units containing thickness-shear mode quartz resonators. Primary factors include mode location in the plane of the resonator, stress patterns the mode interacts with, and the influence of mount and blank geometry and mount orientation upon that interaction. Secondary factors include mode location along the resonator thickness direction, stress patterns the mode interacts with, and the influence of mount and blank geometry and mount orientation upon that interaction. In addition to these factors, the effect that anisotropic material behavior of quartz has upon acceleration sensitivity must also be taken into consideration.

Efforts have focused upon the factors which have the dominant effect upon crystal unit acceleration sensitivity. These efforts included the development of a solution technique to calculate frequency shifts for quartz resonators arising from mechanical stresses. This solution technique has been used to investigate the effects of mode location, mount geometry, and mount orientation upon the acceleration sensitivity of quartz crystal units. Solution results have been used to identify a mount design and the corresponding mode location and mount orientation which significantly reduce the acceleration sensitivity predicted for a four-point SC-cut crystal unit. Features of the new mount design and its fabrication should improve the precision of manufactured crystal units reducing the large scatter in measured acceleration sensitivity.

SOLUTION TECHNIQUE

Frequency shifts for a resonant mechanical continuum subjected to external loading arise from elastic nonlinearities. Elastic nonlinearities are typically classified as either material or geometric. Material nonlinearity is characterized by non-linear stress-strain relationships for infinitesimal displacements and strains. Geometric nonlinearity arises from large displacements or large strains or both. External loading significantly alters the mechanical response of the structure and as a result nonlinear equilibrium equations which account for deformations must be employed. Frequency shifts for quartz resonators subjected to biasing fields are best calculated using the equation for the first perturbation of the eigenvalue derived by H.F. Tiersten [3]. Stress and displacement outputs from finite element solutions for quartz resonators are used in a program which numerically integrates the perturbation integral [1,2].

Differences between the unperturbed ω_M and perturbed ω eigenfrequencies for quartz resonators subjected to external loading may be expressed as

$$\Delta\omega = \omega_M - \omega = \frac{H_M}{2\omega_M}, \qquad (1)$$

where H_M represents the perturbation integral. The perturbation integral, simplified for purely elastic nonlinearities, is given by [45]

$$H_M = -\int_{V_o} \hat{c}_{L\gamma M\alpha} g^M_{\alpha,M} g^M_{\gamma,L} dV_o, \qquad (2)$$

where $g^M_{\alpha,M}$ and $g^M_{\gamma,L}$ are spatial gradients of the mass normalized unit eigenvectors for the Mth odd harmonic of the resonator pure thickness mode triad [46]. The effective elastic coefficient are given by

$$\hat{c}_{L\gamma M\alpha} = T_{LM}\delta_{\gamma\alpha} + c_{L\gamma M\alpha KN}E_{KN} + c_{L\gamma KM}w_{\alpha,K} + c_{LKM\alpha}w_{\gamma,K}, \qquad (3)$$

where T_{LM} and E_{KN} are second rank tensors containing static stress and strain components and $w_{\alpha,K}$ and $w_{\gamma,K}$ are the corresponding deformation gradients. The sixth rank tensor $c_{L\gamma M\alpha KN}$ contains third-order nonlinear elastic coefficients and $\delta_{\gamma\alpha}$ is the Kronecker delta. The last two terms on the right hand side of Equation 3 containing the deformation gradients contract to scalar quantities upon substitution into Equation 2 and the repeated indices may be rearranged without affecting the summation. Rearranging the indices and exploiting the symmetry of the second order elastic coefficient tensor $c_{LKM\alpha}$, Equation 3 may be rewritten as

$$\hat{c}_{L\gamma M\alpha} = T_{LM}\delta_{\gamma\alpha} + c_{L\gamma M\alpha KN}E_{KN} + 2c_{L\gamma KM}w_{\alpha,K} \ . \quad (4)$$

Deformation gradients are a combination of the symmetric strain tensor $E_{\alpha K}$ and the antisymmetric rotation tensor $\Omega_{K\alpha}$. Replacing $w_{\alpha,K}$ by the sum of the strain and rotation tensors, the effective elastic coefficient tensor becomes

$$\hat{c}_{L\gamma M\alpha} = (\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + c_{L\gamma M\alpha KN}S_{KNIJ} + 2c_{L\gamma KM}S_{\alpha KIJ})T_{IJ} + 2c_{L\gamma KM}\Omega_{K\alpha} \ , \quad (5)$$

where the strain tensors have been replaced by stress tensors using the following constitutive equation

$$E_{KN} = s_{KNLM}T_{LM} \ . \quad (6)$$

The effective elastic coefficients given by Equation 5 are used to formulate an algorithm to numerically integrate the perturbation integral. Substituting Equation 5 into Equation 2 the perturbation integral becomes

$$-H_M = \int_{V_o} [(\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + c_{L\gamma M\alpha KN}S_{KNIJ} + 2c_{L\gamma KM}S_{\alpha KIJ})T_{IJ} + 2c_{L\gamma KM}\Omega_{K\alpha}]g^M_{\alpha,M}g^M_{\gamma,L}dV_o \ . \quad (7)$$

Piecewise integration of the perturbation integral is performed numerically as follows

$$-H_M = \sum_{N=1}^{NEL} [(\delta_{LI}\delta_{MJ}\delta_{\gamma\alpha} + c_{L\gamma M\alpha KN}S_{KNIJ} + 2c_{L\gamma KM}S_{\alpha KIJ})T_{IJ}(N) + 2c_{L\gamma KM}\Omega_{K\alpha}(N)]g^M_{\alpha,M}(N)g^M_{\gamma,L}(N)\Delta V(N) \ , \quad (8)$$

using single quadrature points located at the centroids of elements constituting finite element models of the quartz resonators. The summation in Equation 8 is performed over *NEL*, the number of elements comprising a significant portion of the active region of the resonator mode. Components of the stress tensor $T_{IJ}(N)$ produced by external loading and computed at the center of each element N are retrieved from the finite element output files created for the resonator model. Components of the antisymmetric tensor $\Omega_{K\alpha}(N)$, for rigid body rotations about the centroid of each element, are computed using nodal displacements retrieved from the FEA solution. Element centroids and volumes $\Delta V(N)$ are computed using coordinates for nodes attached to each element used in the summation. Nodal coordinates are retrieved from files which echo data input to create the finite element model of the resonator. Independent experimental results confirmed the accuracy of frequency shifts calculated using Equation 8 and finite element models of quartz resonators [1,2]. Predicted variations in the force-frequency coefficient with azimuthal angle for AT- and SC-cut resonators agreed closely with measured results. The minimum acceleration sensitivity predicted for an AT-cut crystal unit modeled after those tested in Reference 37 is in good agreement with minimum acceleration sensitivities achieved experimentally by adjusting mode shape and location. Scatter in acceleration sensitivity calculated for movement of the mode off-center approximates measured scatter for the AT-cut crystal units prior to optimization [1,2].

ACCELERATION SENSITIVITY CALCULATED FOR SC-CUT CRYSTAL UNITS

Mount reactions produce mechanical stresses distributed throughout resonators subjected to either steady-state or time-varying inertial loads. Mechanical stresses in the active region change the resonant mode frequencies of quartz crystal oscillators. Frequency changes in proportion to the inertial load magnitude and depends upon resonator orientation in the acceleration field. Acceleration sensitivity of quartz crystal units is typically expressed as a vector quantity known as the gamma vector

$$\vec{\Gamma} = \Gamma_i\hat{i} + \Gamma_j\hat{j} + \Gamma_k\hat{k} \quad (9)$$

Gamma vector components Γ_i are fractional frequency shifts produced by unit acceleration in three orthogonal directions ($\hat{i}$, $\hat{j}$, $\hat{k}$). Fractional frequency shifts arising from acceleration in any direction are computed as the scalar product of the gamma vector with the acceleration vector

$$\frac{\Delta f}{f_o} = \vec{\Gamma}\cdot\vec{a} \quad (10)$$

where f_o is the frequency for no acceleration loading. Gamma vectors for precision resonators typically have magnitudes ranging from 10^{-9} to 10^{-10} per g.

Acceleration sensitivities for crystal units were calculated using the solution technique described above and in References 1 and 2 . Crystal units modeled for these calculations consist of a quartz thickness-shear mode disc resonator supported by four mounts of various geometries distributed at equal angular intervals around its circumference. The plano-convex resonator in the models is a 10MHz, 3rd overtone, SC-cut ($\phi = 22.4°, \theta = 34.3°$) with a diameter of 0.550" and a radius of curvature of 3/8 diopter. A thickness at the resonator origin of 0.021" produces a fundamental C-mode frequency of 3.336MHz ($N_o = 1811$ Hz-m). Calculations were performed to determine what effect mount geometry and orientation and mode location have upon acceleration sensitivity. Calculations were also performed to identify a mount geometry, mode location, and blank orientation which significantly reduce acceleration sensitivity and improve the precision of manufactured crystal units.

Acceleration sensitivities were initially calculated for the 10MHz, 3rd, SC-cut crystal unit depicted in Figure 1. The resonator is supported by "V"-clip mounts spaced at equal angular intervals around its circumference for these calculations. The "V"-clip mount is routinely used by manufacturers of crystal units. The crystal unit model depicted in Figure 2 consists of 3,008 8-node isoparametric hexahedral solid elements which incorporate full anisotropic material properties. There are 2,848 solid elements modeling the SC-cut resonator and 160 solid elements model the epoxy which joins the resonator to the mounts. Twelve 3-D beam elements model the 0.040"ϕ kovar header pins and 288 3-D shell elements model the 0.005" thick stainless steel mounts. The elements are interconnected by 4,269 nodes and nodal constraints couple the mount shell elements to elements modeling the pins and epoxy.

Gravitational acceleration was applied to the model along each of the orthogonal global axes depicted in Figure 2. Acceleration induced frequency shifts were calculated for the third overtone of the thickness-shear C-mode. Changes in the gamma vector and its components were calculated for variations in the resonator orientation with respect to the mounts and for changes in the mode location. Figure 3 depicts the sense of the mount azimuthal rotation ψ with respect to the crystallographic axes (X",Z") of the SC-cut resonator. Variations in the mount orientation were entered into the model through changes in the material properties of the anisotropic hexahedral elements used to model the resonator. Gamma vector components were computed for mount orientations ranging from $\psi = 0°$ to $\psi = 90°$ in 15° increments. Gamma vector components for mount orientations ranging from $\psi = -15°$ to $\psi = -90°$ were inferred from components for the positive orientations using symmetry. Calculations for the mode located at the center of the resonator were repeated for the mode off-center ± 0.1cm in the X- and Z-directions for each mount orientation. Off-center gamma vector components were used to linearly extrapolate optimum mode locations which minimize $\vec{\Gamma}$. Predicted variations in $\vec{\Gamma}$ with ψ for the "V"-clip mount and the mode located at the resonator center, off-center, and at the optimum location are presented in Reference 1. Moving the mode to the optimum location reduces acceleration sensitivity for each mount orientation. The largest reduction for the "V"-clip mount occurs for $\psi = 0°$ and by symmetry $\psi = \pm 90°$. Calculations indicate that moving the mode 0.23cm to the optimum location for these orientations reduces the gamma vector to a minimum of $0.012 X 10^{-10}/g$.

Acceleration induced frequency shifts initially calculated for SC-cut crystal units were based upon finite element models of ideal mount and resonator geometries. Manufacturing tolerances will result in imperfect mounts and resonators. Careful inspection of crystal units manufactured with "V"-clip mounts confirmed that significant deviations from ideal mount geometries do occur. To investigate what effect mount imperfections might have upon acceleration sensitivity, the FEA model of the SC-cut crystal unit assembled with "V"-clip mounts was modified. Both mounts along the X-axis in one model and along the Z-axis in another were skewed so that the attachment point of one mount was raised 0.015" with respect to its ideal location and the opposite mount was lowered 0.015". Gamma vector components were calculated for the skewed mount geometries and the optimum mount orientation $\psi = 0°; \pm 90°$. Calculation results for SC-cut crystal units with imperfect "V"-clip mounts presented in Reference 1 indicate that alterations in resonator stress distributions arising from imperfect and inconsistent mount geometries contribute to increased acceleration sensitivities and reduced precision for manufactured crystal units. To achieve improved performance for manufactured crystal units the mount geometry should be simple for precise and repeatable fabrication. Designs which require that the mount be bent for resonator insertion should be avoided. Mount design should minimize variations in the spacing, height, and alignment of the surfaces supporting the resonator. Mount attachment to the header should be precisely

controlled to avoid inconsistent alignment or skewing.

Acceleration sensitivity calculations performed for ideal FEA models of the SC-cut crystal unit with "V"-clip mounts were repeated for other mount geometries. Mounts studied include the platform mount and two angular extremes of the "V"-clip mount. These mount geometries, labeled the vertical and horizontal strip mounts, are depicted in Figure 4 along with drawings of the "V"-clip and platform mounts. Finite element models of four-point mount crystal units assembled using each of the mount geometries are depicted in Figure 5. Elements modeling the resonator have been suppressed in Figure 5 for clarity. Varying mount geometry affects mount stiffnesses and how the mounts react to inertial loading of the resonator. Angular extremes of the "V"-clip mount and the platform mount were modeled to characterize the effect mount geometry has upon acceleration sensitivity and to establish design criteria which minimize $\vec{\Gamma}$ over a broad range of mount orientations.

The mounts are located to one side of the nodal plane of vibration of the thickness-shear mode resonator for crystal units assembled with each of the mounts depicted in Figures 4 and 5. Prior studies [32, 39] have identified crystal unit asymmetry about the central (X",Z") plane of the resonator as an important factor contributing to increased acceleration sensitivity. Ideal crystal unit symmetry about the resonator plane has been sacrificed to achieve compact designs which are robust and readily manufacturable. An even number of mounts (4) has been selected to generate orthogonal planes of geometric symmetry normal to the central plane of the resonator for ideal crystal unit models. Mount geometries and orientations and mode locations which minimize acceleration sensitivity have been calculated to compensate for the inherent asymmetry of the crystal unit designs. Prior to acceleration sensitivity calculations, modal analyses performed for crystal units with each mount geometry confirmed that resonant frequencies for the lowest spring-mass modes would exceed 3KHz. Frequencies for the fundamental translational spring-mass modes for each crystal unit design are presented in Table 1.

Minimum acceleration sensitivities predicted for each mount geometry are listed in Table 2 along with the optimum mount orientations and mode locations. The minimum acceleration sensitivity is predicted for the "V"-clip mount and the optimum mode location in closest proximity to the resonator center is predicted for the horizontal strip mount geometry. To illustrate the potential advantages of each mount design, curves depicting variations in $\vec{\Gamma}$ with ψ for the mode located at the resonator center and each mount geometry are plotted in Figure 6 for comparison. Acceleration sensitivities for the mode centered are smallest over all mount orientations for the horizontal strip mount design. Significant reductions are predicted for the horizontal strip mount relative to the vertical strip mount design. Variations in $\vec{\Gamma}$ with ψ are similar for the "V"-clip and platform mounts when the mode is located at the resonator center. A second graph comparing acceleration sensitivities calculated for crystal units assembled with each mount geometry appears in Figure 7. Minimum acceleration sensitivities for the mode offset in the optimum direction 0.1cm or to the optimum location, whichever is less, are plotted for each mount orientation in this figure. Figure 7 illustrates the minimum $\vec{\Gamma}$ achieved for each mount and movement of the mode off-center a maximum of 0.1cm. Adjustment of mode location over this distance is practical for experimental techniques including selective mass loading [36] and selective spatial excitation using multiple electrodes [37]. The smallest acceleration sensitivities over all mount orientations are again predicted for the horizontal strip mount design. Plots of acceleration sensitivity versus ψ in Figures 6 and 7 indicate that, for the mode in close proximity to the resonator center, crystal units with horizontal strip mounts will be least sensitive to inertial loading.

Each mount geometry had substantially different effects upon the spring-mass resonant modes and acceleration sensitivities of the SC-cut crystal units. Comparison of crystal unit response characteristics for each mount helped identify additional design criteria which may be used to reduce acceleration sensitivity. Differences in mount stiffnesses are evident in the translational spring-mass mode frequencies for each crystal unit presented in Table 1. The vertical strip mounts are much stiffer in all three directions relative to the other mount designs producing higher translational mode frequencies. The horizontal strip mounts are the most compliant in all three directions with less deformation occurring in the resonator and more in the mounts. Lower acceleration sensitivities predicted for the horizontal strip mount design indicate that the mounts should be soft in all three directions while maintaining crystal unit spring-mass frequencies in excess of some predetermined minimum for structural integrity. The mounts should bend and twist allowing the resonator to translate and rotate rigidly. Horizontal strip mounts aligned with the in-plane acceleration direction are stiff axially and support the largest portion of the inertial load. Horizontal strip mounts on either side of the resonator rotated 90° from the in-plane acceleration bend and twist allowing the resonator to rotate. Vertical strip mounts aligned with the in-plane

acceleration are compliant and the mounts to either side support the largest portion of the inertial load. The rigid transverse mounts bend and twist the resonator in reaction to inertial loading distorting its shape. To reduce acceleration sensitivity the stress pattern produced by mounts reacting to in-plane acceleration should be tension changing to compression along the inertial axis as opposed to shear and bending which distort the resonator. Finally, the resonator is located closest to the mount attachment plane for the horizontal strip mount design indicating that a low center of gravity reduces moments of inertia and acceleration sensitivity for inherently asymmetric crystal unit designs.

HORIZONTAL STRIP MOUNT DESIGN CALCULATIONS

Acceleration performance predicted for crystal units modeled with horizontal strip mounts appears to be superior to crystal units modeled with vertical strip, "V"-clip, and platform mounts. Lower acceleration sensitivities are predicted for the range of mount orientations and optimum mode locations are in closer proximity to the resonator center. A final design for the horizontal strip mount evolved which incorporates the design criteria outlined in the preceding section and facilitates precise and repeatable fabrication. A drawing of this design appears in Figure 8. This design has several features which improve upon prior four-point mount designs. First, the four mounting surfaces and support structures are integrally formed from a single strip of 0.005" thick 302 stainless steel. The mounts are intimately joined by a cupped ring which rigidly fixes each mount with consistent boundary conditions. Precision metal stamping may be used to repeatably form the mounts quickly in limited quantities while maintaining extreme tolerances. Fixturing and processes to precisely attach the integral mounts to header bases should be straightforward. The integral mount design necessitates that electrical connections to the resonator electrodes be made independent of the mounts using fine wire conductors. Contrast this to prior four-point mount designs for which each mount is independently stamped and attached to a header lead. Individual mount shapes may be precisely formed but it is extremely difficult to maintain precise control of spacing and alignment for four mounts which are independently attached. Limitations of attachment processes significantly increase assembly tolerances contributing to increased acceleration sensitivities and increased scatter (i.e., reduced precision) for manufactured crystal units.

Numerous calculations were performed to accurately establish critical dimensions and an optimum orientation for the horizontal strip mount. Variations in mount geometry were modeled to identify trends in crystal unit response to acceleration and optimize mount dimensions. Gamma vector components were computed for variations in the mount height "H" and thickness "T" dimensions depicted in Figure 8. These calculations identified spatial separation between the resonator centroid and the plane containing the mount attachment points, characterized by "H", as a critical dimension affecting acceleration sensitivity. Plots of variations in $\vec{\Gamma}$ with ψ for H=0.015", H=0.025", and H=0.050" are depicted in Figure 9 for the mode centered and in Figure 10 for the optimum mode locations. These figures indicate that crystal units incorporating shorter mounts will be less sensitive to acceleration. Minimum acceleration sensitivities are predicted for the mount with H=0.025". An ideal combination of mount stiffnesses and reduced moments of inertia is achieved for this mount height.

Acceleration sensitivities were computed for the mount depicted in Figure 8 and orientations ranging from $\psi = 20°$ to $\psi = 40°$ and from $\psi = 65°$ to $\psi = 85°$ in 5° increments. Gamma vector components for corresponding negative mount orientations were inferred from symmetry. These angular intervals, which encompass two optimum orientations, were studied with the finer angular intervals to determine whether the optimum orientation which minimizes $\vec{\Gamma}$ is sensitive to minor orientation errors. Figure 11 depicts variations in $\vec{\Gamma}$ for refinements in mount orientation and the mode centered, off-center 0.1cm in the optimum direction, and at the optimum location. Optimum mount orientations remain unchanged at $\psi = -60°$ and by symmetry $\psi = +30°$.

Variations in mount dimensions affect mount stiffnesses and acceleration sensitivity for the SC-cut crystal unit. Figure 12 depicts variations in acceleration sensitivity with "H" for the optimum mount orientation ($\psi = +30°$) and the mode located at the resonator center. This figure illustrates how sensitive $\vec{\Gamma}$ is to changes in mount height for a given mount orientation and mode location. Minor deviations in the height dimension on the order of ± 0.002" may be tolerated without significant increases in $\vec{\Gamma}$. Calculations were also performed to determine how changes in mount thickness "T" affect $\vec{\Gamma}$. Figure 13 depicts variations in acceleration sensitivity with mount thickness for $\psi = +30°$ and the mode located at the resonator center. This figure illustrates how thickness tolerances affect acceleration sensitivity for a production crystal unit with a specific mount orientation and mode location. Mount

thickness should be precisely controlled to vary less than ± 0.0002". Figures 12 and 13 show that variations in mount dimensions arising from manufacturing tolerances alter mount stiffnesses and the optimum mount orientations and mode locations which minimize Γ. Horizontal strip mounts may be formed using precision metal stamping techniques to minimize deviations from the ideal mount shape depicted in Figure 8.

CONCLUSION

Calculations have been performed using the solutions technique described in References 1 and 2 to determine acceleration sensitivities of four-point SC-cut crystal units. Mounts modeled for these calculations included "V"-clip mounts, platform mounts, vertical and horizontal strip mounts. The mounts are located to one side of the central plane of the thickness-shear mode resonator in crystal units modeled for each mount geometry. Symmetry about the central plane of the resonator has been sacrificed to achieve crystal unit designs which are compact and readily manufacturable. Mount geometries and orientations and mode locations which minimize acceleration sensitivity have been calculated to compensate for the inherent asymmetry of the crystal unit designs. Results from these calculations revealed that in addition to mode shape and location, mount imperfections contribute significantly to increased acceleration sensitivity and reduced precision for manufactured crystal units.

Mount design criteria formulated as a result of these calculations have been used to identify a mount geometry best suited to improving the performance of SC-cut crystal units subjected to inertial loading. Lower acceleration sensitivities are predicted for the horizontal strip mount design over all mount orientations for the mode located in close proximity to the resonator center. Calculations were performed to determine the optimum mount dimensions, mode location, and blank orientation to minimize acceleration sensitivity for the horizontal strip mount design. A final design evolved which incorporates criteria conceived to minimize Γ and facilitate precise and repeatable fabrication.

Imperfections in mount, resonator, and crystal unit geometry arising from manufacturing and assembly tolerances are unavoidable. Deviations from ideal geometry and material imperfections will contribute to increased acceleration sensitivity and reduced precision for crystal units assembled with horizontal strip mounts at the optimum orientation. To compensate for manufacturing tolerances and achieve the minimum acceleration sensitivity predicted, experimental techniques to iteratively adjust mode shape and location must be employed. Strict adherence to precise fabrication and assembly coupled with mode shape and location adjustment should produce manufactured crystal units with reduced acceleration sensitivity and improved precision.

ACKNOWLEDGEMENTS

This work has been supported by the U.S. Army Laboratory Command Electronics Technology and Devices Laboratory under Contract No. DAAL01-90-C-009.

REFERENCES

1. L.D. Clayton and E.P. EerNisse, "Frequency Shift Calculations for Quartz Resonators," in Proceedings 45th Annual Symposium on Frequency Control, 1991, pp. 309-320.

2. L.D. Clayton and E.P. EerNisse, "Quartz Resonator Frequency Shifts Computed Using the Finite Element Method," Accepted for publication in the International Journal for Numerical Methods in Engineering.

3. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 1978, pp. 832-837.

4. A. Ballato, T. Lukaszek, and G. Iafrate, "Subtle Effects in High Stability Vibrators," in Proceedings 34th Annual Symposium on Frequency Control, 1980, pp. 431-444.

5. F. L. Walls, "Environmental Effects on the Medium and Long Term Frequency Stability of Quartz Oscillators," in Proceedings of the Second European Frequency and Time Forum, 1989, pp. 719-727.

6. J.J. Gagnepain, "Sensitivity of Quartz Oscillators to Environment: Characterization Methods and Pitfalls," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 37, 1990, pp. 347-354.

7. F.L. Walls and J.J. Gagnepain, "Environmental Sensitivities of Quartz Oscillators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, March 1992.

8. R. Holland, "Non Uniformly Heated Anisotropic Plates: I Mechanical Distortions and Relaxation," IEEE Transactions on Sonics and Ultrasonics, SU-21, July 1974, pp. 171-178.

9. R. Holland, "Non Uniformly Heated Anisotropic Plates: II Frequency Transients in AT and BT Quartz Plates," Ultrasonics Symposium Proceedings, 1974, pp. 592-598.

10. J.A. Kusters, "Transient Thermal Compensation for Quartz Resonators," IEEE Transactions on Sonics and Ultrasonics, July 1976, pp. 273-276.

11. E.P. EerNisse, "Calculations on the Stress Compensated (SC-Cut) Quartz Resonator," in Proceedings 30th Annual Symposium on Frequency Control, 1976, pp. 8-11.

12. J.A. Kusters, C. Adams, H. Yoshida, and J. Leach, "TTC's-Further Developments and Results," in Proceedings 31st Annual Symposium on Frequency Control, 1977, pp. 3-7.

13. A. Ballato and J. Vig, "Static and Dynamic Frequency-Temperature Behavior of Singly and Doubly Rotated, Oven-Controlled Quartz Resonators," in Proceedings 32nd Annual Symposium on Frequency Control, 1978, pp. 180-188.

14. B.K. Sinha and H.F. Tiersten, "Transient Thermally Induced Frequency Excursions in Doubly-Rotated Quartz Thickness Mode Resonators," in Proceedings 34th Annual Symposium on Frequency Control, 1980, pp. 393-402.

15. J.P. Valentin, G. Théobald, and J.J. Gagnepain, "Frequency Shifts Arising from In-Plane Temperature Gradient Distribution in Quartz Resonators," in Proceedings 38th Annual Symposium on Frequency Control, 1984, pp. 157-163.

16. F.L. Walls, "The Influence of Pressure and Humidity on the Medium and Long-Term Frequency Stability of Quartz Oscillators," in Proceedings 42nd Annual Symposium on Frequency Control, 1988, pp. 279-283.

17. E.P. EerNisse, "Simultaneous Thin Film Stress and Mass Change Measurements Using Quartz Resonators," Journal of Applied Physics, Vol. 43, No. 4, April 1972, pp. 1330-1337.

18. E.P. EerNisse, "Quartz Resonator Frequency Shifts Arising From Electrode Stress," in Proceedings 29th Annual Symposium on Frequency Control, 1975, pp. 1-4.

19. R. Brendel, C. El Hassani, M. Brunet, and E. Robert, "Magnetic Sensitivity of Quartz Oscillators," in Proceedings 43rd Annual Symposium on Frequency Control, 1989, pp. 268-274.

20. F. Deyzac, "Magnetic Sensitivity of Quartz Oscillators," in Proceedings 4th Annual European Forum on Time and Frequency, 1990, pp. 255-258.

21. A. Ballato, "Effect of Initial Stress on Quartz Plates Vibrating in Thickness Modes," in Proceedings 14th Annual Symposium on Frequency Control, 1960, pp. 89-114.

22. C.R. Mingins, L.C. Barcus, and R.W. Perry, "Reactions of Vibrating Piezoelectric Plate to Externally Applied Forces," in Proceedings 17th Annual Symposium on Frequency Control, 1963, pp. 51-87.

23. J.M. Ratajski, "Force-Frequency Coefficient of Singly Rotated Vibrating Quartz Crystals," IBM J. Res. Dev., 12, No. 1, January 1968, pp. 92-99.

24. C.R. Dauwalter, "The temperature Dependence of Force Sensitivity of AT-Cut Quartz Crystals," in Proceedings 26th Annual Symposium on Frequency Control, 1972, pp. 108-112.

25. A. Ballato, E.P. EerNisse, and T. Lukaszek, "The Force-Frequency Effect in Doubly Rotated Quartz Resonators," in Proceedings 31st Annual Symposium on Frequency Control, 1977, pp. 8-16.

26. E.P. EerNisse, T. Lukaszek, and A. Ballato, "Variational Calculation of Force-Frequency Constants of Doubly Rotated Quartz Resonators," IEEE Transactions on Sonics and Ultrasonics, SA-25, No. 3, May 1978, p. 132.

27. E.P. EerNisse, "Temperature Dependence of the Force Frequency Effect for the AT-, FC-, SC-, and Rotated X-Cuts," Proceedings 34th Annual Symposium on Frequency Control, 1980, pp. 426-430.

28. A.W. Warner and W.L. Smith, "Quartz Crystal Units and Precision Oscillators for Operation in Severe Mechanical Environments," in Proceedings 14th Annual Symposium on Frequency Control, 1960, pp. 200-216.

29. A.D. Ballato, "Resonators Compensated for Acceleration Fields," in Proceedings 33rd Annual Symposium on Frequency Control, 1979, pp. 322-336.

30. P.C.Y. Lee and K-M Wu, "Effects of Acceleration on the Resonance Frequencies of Crystal Plates," in Proceedings 30th Annual Symposium on Frequency

Control, 1976, pp. 1-7.

31. R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 35, No. 3, May 1988, pp. 297-305.

32. Y.S. Zhou and H.F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Quartz Cover Plates Supported by Clips," in Proceedings 45th Annual Symposium on Frequency Control, 1991, pp. 298-308.

33. J.R. Vig, et. al., "The Effect of Acceleration on Precision Frequency Sources," LABCOM R+D Tech. Report SLCET-TR-91-3, March 1991, available from U.S. Army, Electronics Technology and Devices Laboratory, Fort Monmouth, NJ.

34. E.P EerNisse, R.W. Ward, and O.L. Wood, "Acceleration Induced Frequency Shifts in Quartz Resonators," in Proceedings 43rd Annual Symposium on Frequency Control, 1989, pp. 388-395.

35. E.P. EerNisse, L.D. Clayton, and M.H. Watts, "Distortions of Thickness Shear Mode Shapes in Plano-Convex Quartz Resonators with Mass Perturbations," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 37, No. 6, Nov. 1990, pp. 571-576.

36. E.P. EerNisse, et. al., "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 37, No. 6, Nov. 1990, pp. 566-570.

37. R.C. Smythe and W.H. Horton, "Adjustment of Resonator G-Sensitivity by Circuit Means," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 437-443.

38. P.C.Y. Lee and X. Gyo, "Acceleration Sensitivity of Crystal Resonators Affected by Mass and Location of Electrodes," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 468-473.

39. Y.S. Zhou and H.F. Tiersten, "On the Influence of A Fabrication Imperfection on the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," in Proceedings 44th Annual Symposium on Frequency Control, 1990, pp. 452-467.

40. P.C.Y. Lee and K-M Wu, "The Influence of Support Configuration on the Acceleration Sensitivity Quartz Resonator Plates," in Proceedings 31st Annual Symposium on Frequency Control, 1977, pp. 29-34.

41. P.C.Y. Lee and M.S.H Tang, "Acceleration Effect on the Thickness Vibrations of Doubly Rotated Crystal Resonators," in Proceedings 41st Annual Symposium on Frequency Control, 1987, pp. 277-281.

42. R. Besson, "A New Piezoelectric Resonator Design," in Proceedings 30th Annual Symposium on Frequency Control, 1976, pp. 78-83.

43. R. Besson, J.J. Gagnepain, D. Janiaud, and M. Valdois, "Design of a Bulk Wave Quartz Resonator Insensitive to Acceleration," in Proceedings 33rd Annual Symposium on Frequency Control, 1979, pp. 337-345.

44. R. Besson and U.R. Peier, "Further Advances on B.V.A. Quartz Resonators," in Proceedings 34th Annual Symposium on Frequency Control, 1980, pp. 175-182.

45. H.F. Tiersten and D.V. Shick, "An Analysis of the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," in Proceedings 43rd Annual Symposium on Frequency Control, 1989, pp. 396-404.

46. 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, 1986, pp. 1811-1826.

Table 1. SC-Cut Crystal Unit Modal Analysis

Mount	Mode	Mode Shape	Frequency (Hz)
Platform Mount	1	X-Translation	3049.52
	2	Z-Translation	3049.95
	4	Y-Translation	11,878.8
"V"-Clip Mount	1	X-Translation	4156.53
	2	Z-Translation	4167.02
	3	Y-Translation	5822.25
Horizontal Strip Mount	1	Y-Translation	3139.98
	2	X-Translation	3806.69
	3	Z-Translation	3817.89
Vetrical Strip Mount	1	X-Translation	9745.98
	2	Z-Translation	9831.07
	3	Y-Translation	16306.1

Table 2. Minimum Predicted Acceleration Sensitivities for Each Mount Geometry

Mount Geometry	Minimum $\lvert\vec{\Gamma}\rvert X10^{-10}/g$	Optimum Mount Orientation ψ	Optimum Mode Location (cm)
"V" Clip	0.012	0° ; ±90°	0.233
Platform	0.20	-15° ; +75°	0.4
Vertical Strip	0.647	-75° ; +15°	0.4
Horizontal Strip	0.22	-60° ; +30° -15° ; +75°	0.028

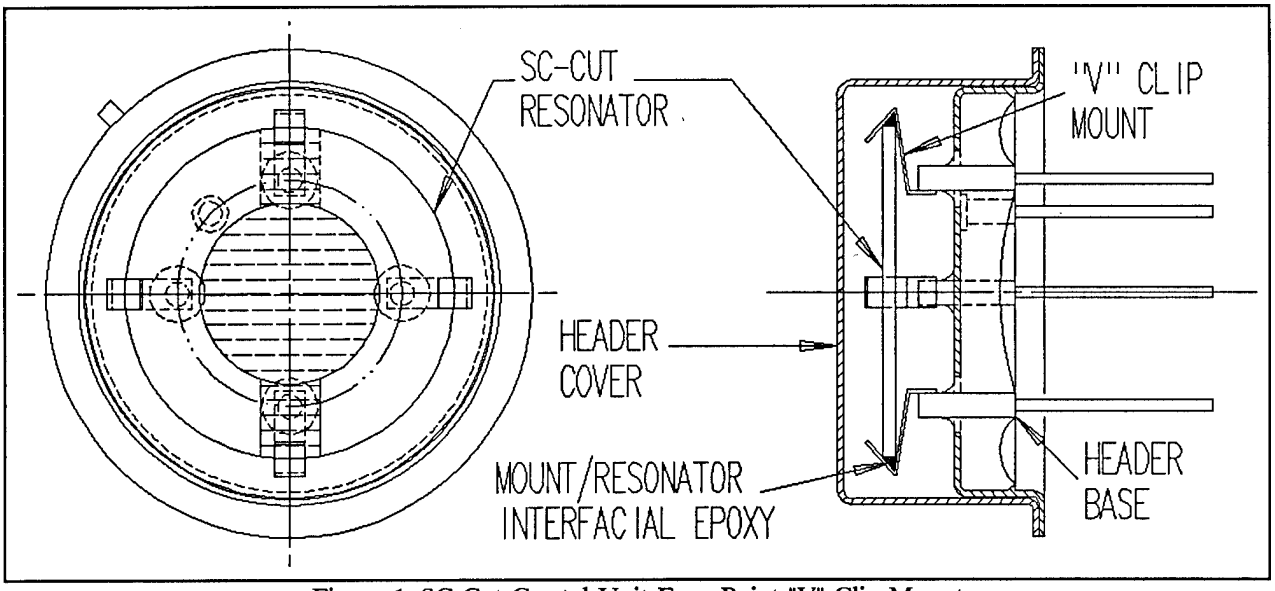


Figure 1. SC-Cut Crystal Unit Four-Point "V"-Clip Mount

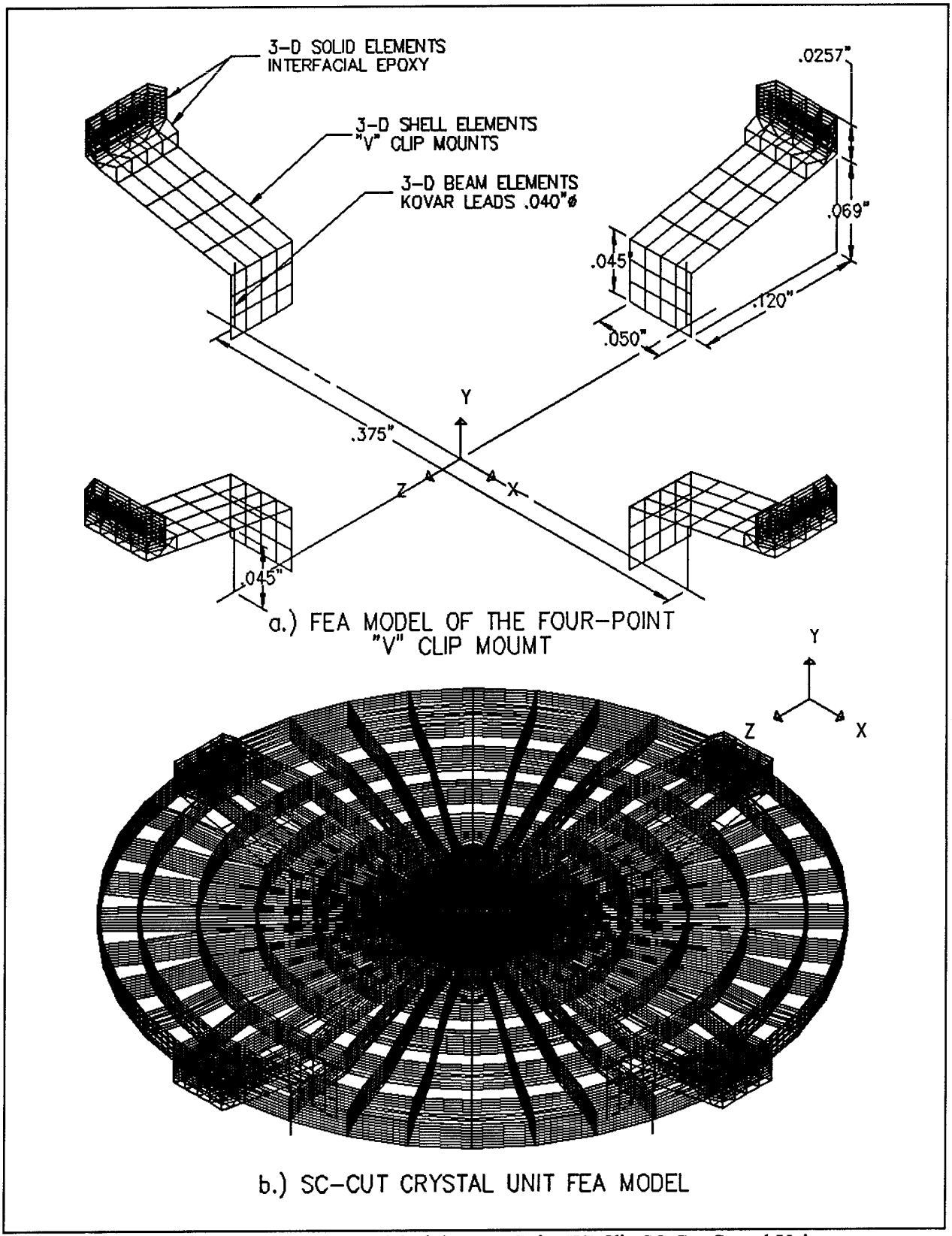


Figure 2. Finite Element Model of the Four-Point "V"-Clip SC-Cut Crystal Unit

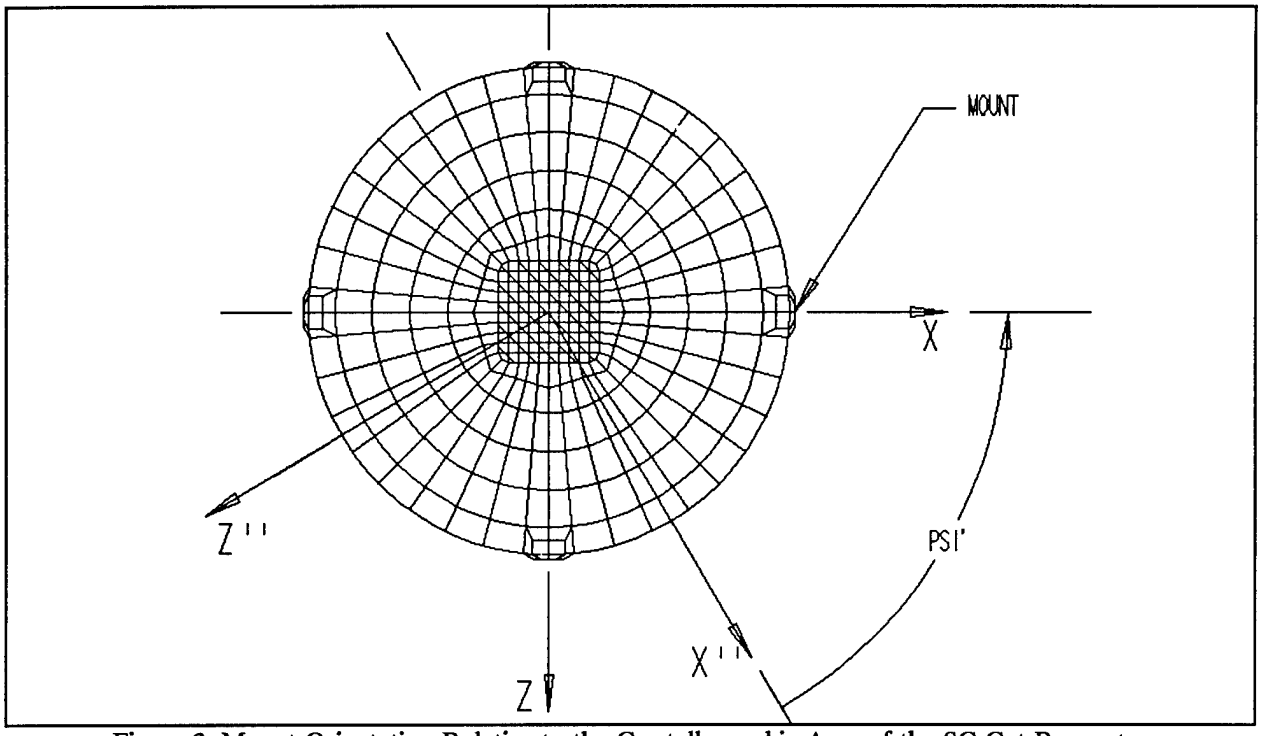


Figure 3. Mount Orientation Relative to the Crystallographic Axes of the SC-Cut Resonator

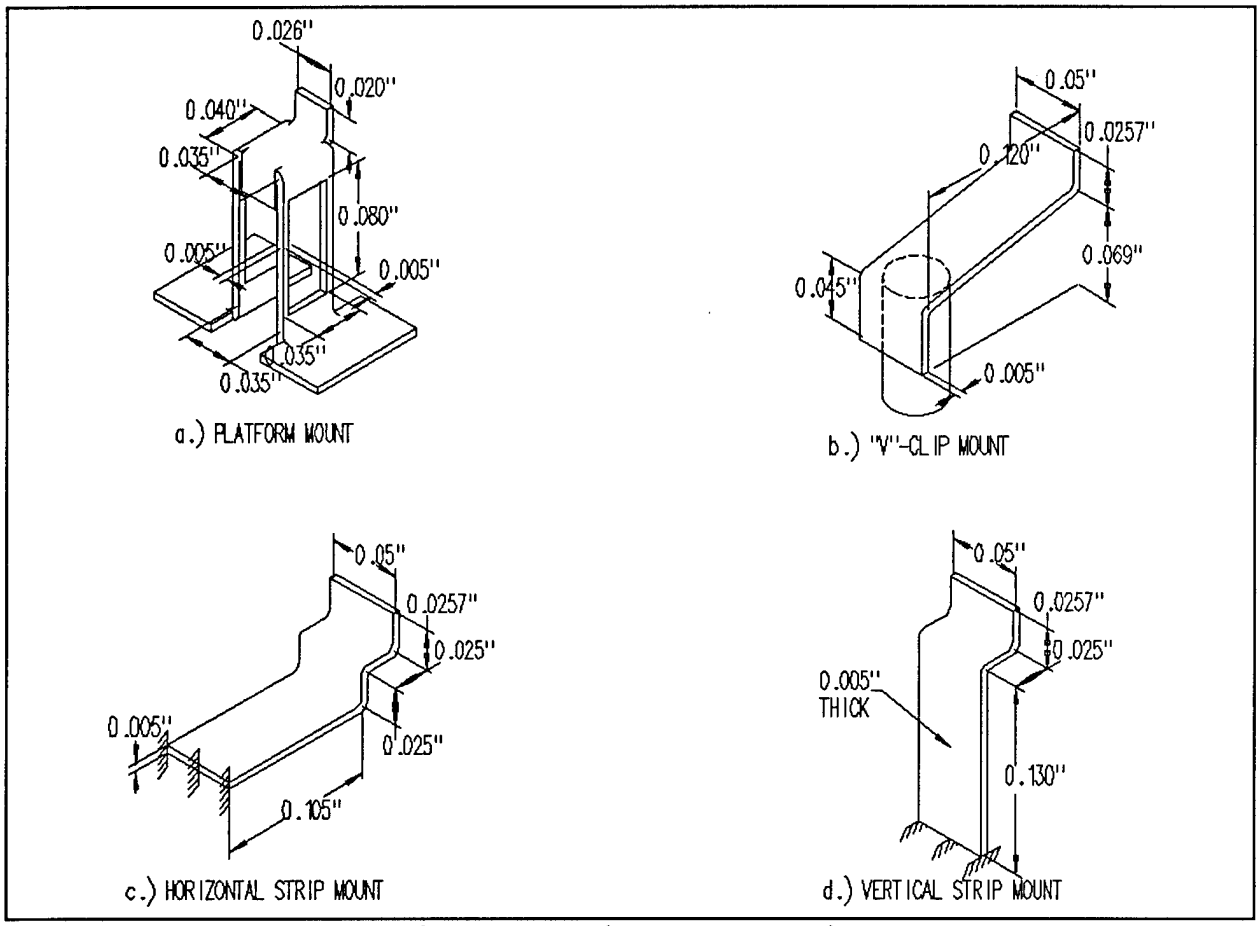


Figure 4. Four-Point Mount Geometries

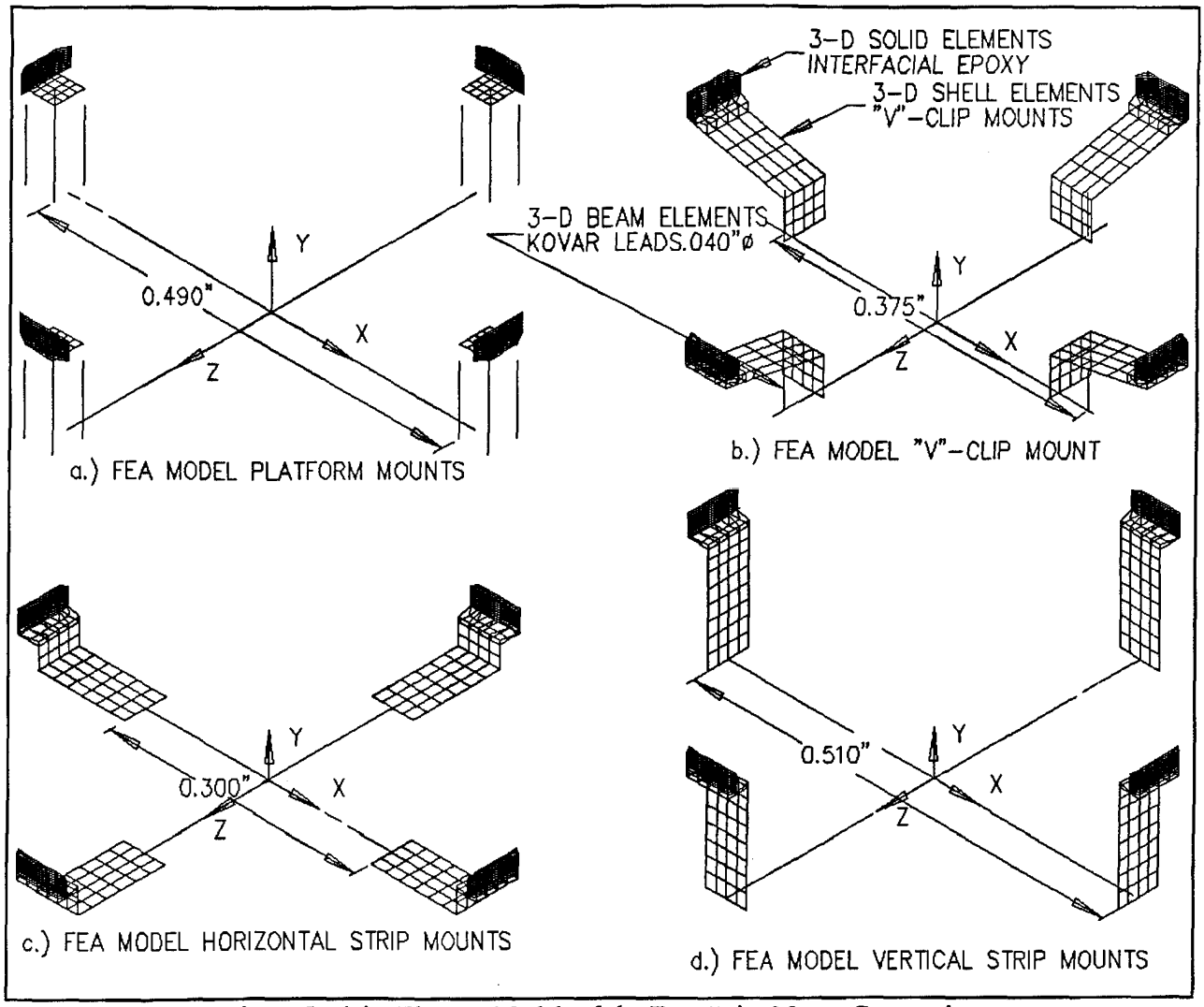


Figure 5. Finite Element Models of the Four-Point Mount Geometries

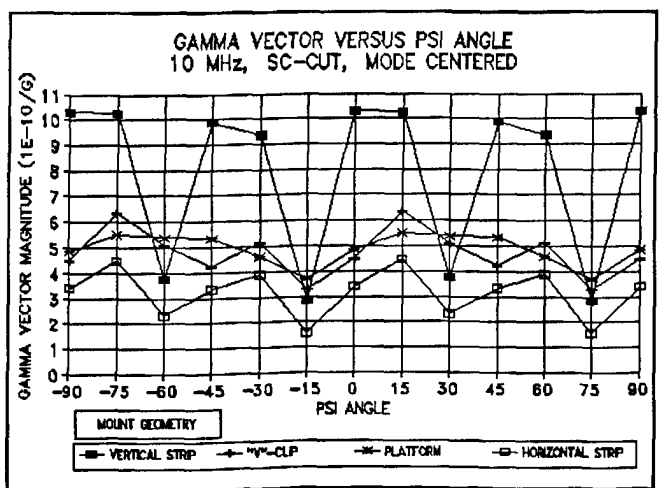


Figure 6. Acceleration Sensitivities for the Mode Centered and Each Mount Geometry

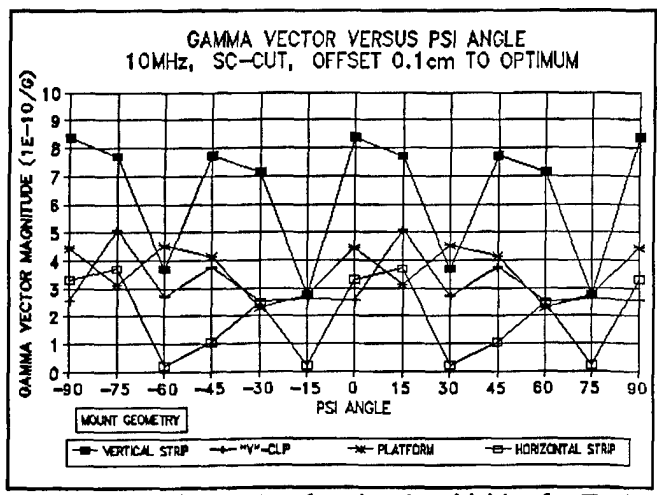


Figure 7. Minimum Acceleration Sensitivities for Each Mount and the Mode Off-Center a Maximum of 0.1cm

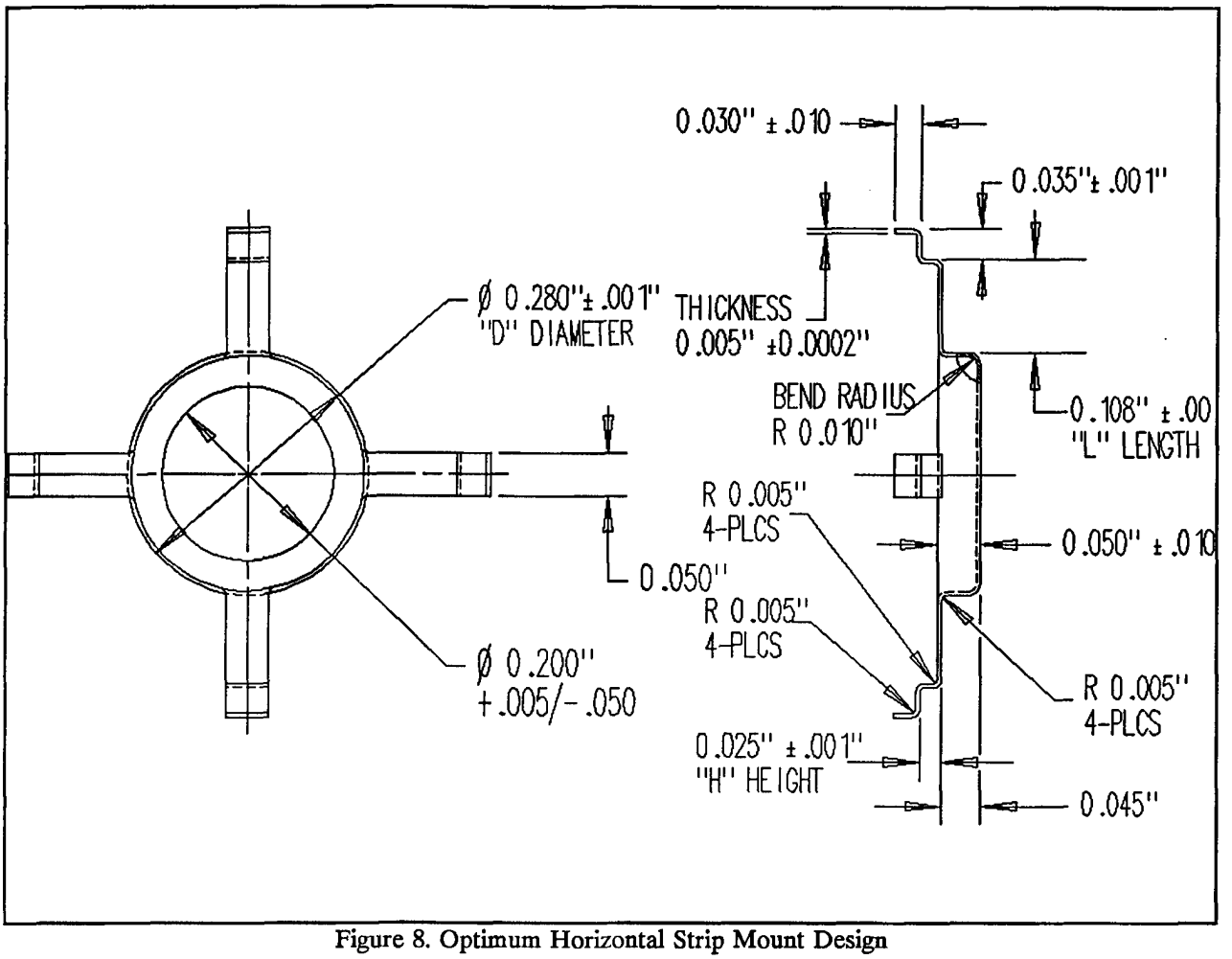


Figure 8. Optimum Horizontal Strip Mount Design

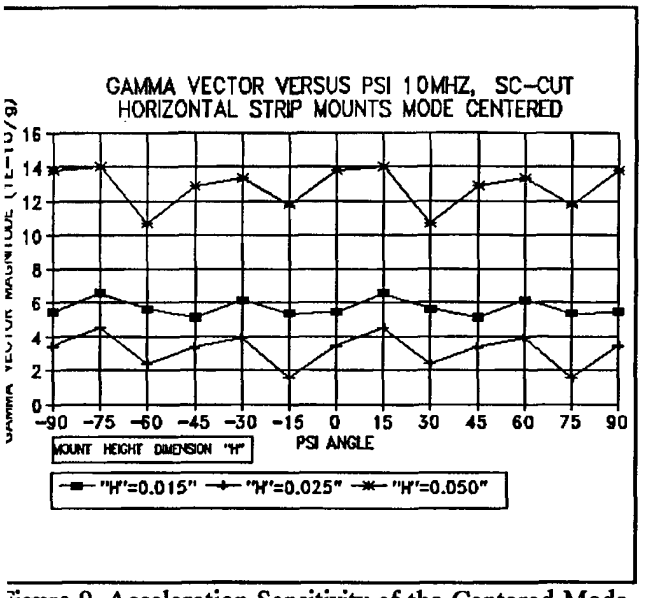


Figure 9. Acceleration Sensitivity of the Centered Mode for Each Mount Height "H"

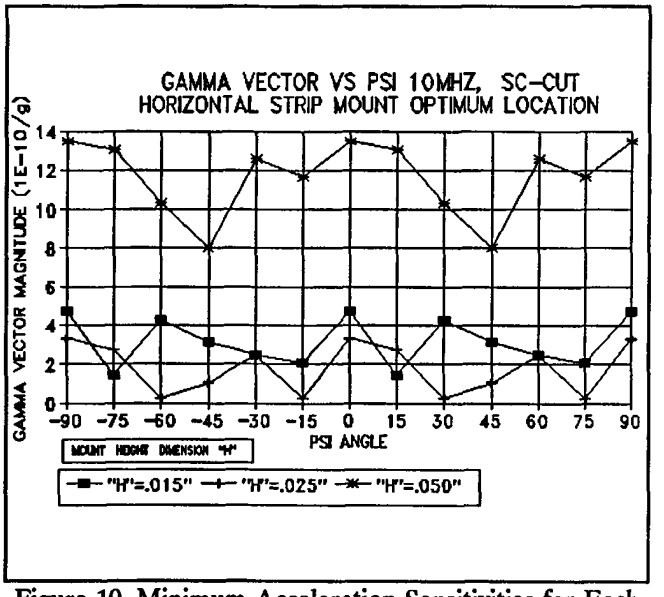


Figure 10. Minimum Acceleration Sensitivities for Each Mount Height "H"

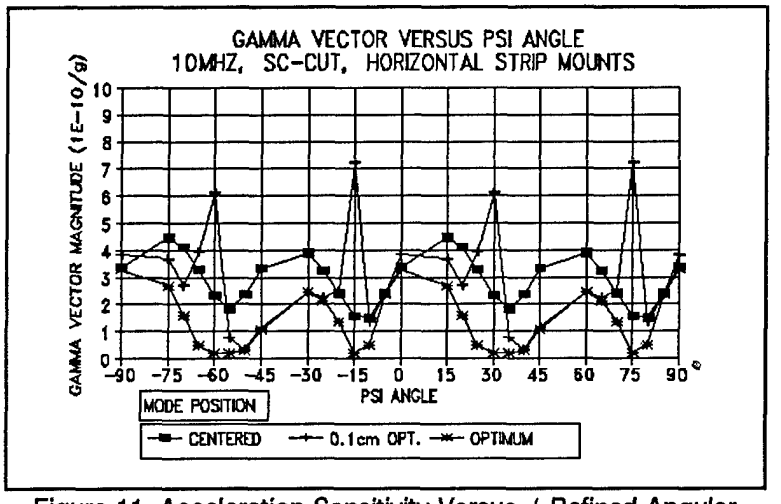


Figure 11. Acceleration Sensitivity Versus ψ Refined Angular Increments for the SC-Cut Horizontal Strip Mount Crystal Unit

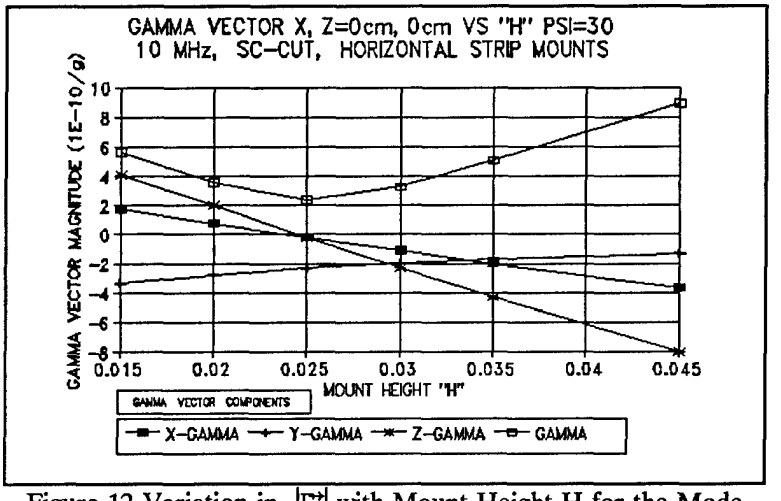


Figure 12 Variation in $|\Gamma^{\rightarrow}|$ with Mount Height H for the Mode Centered and $\psi = +30°$

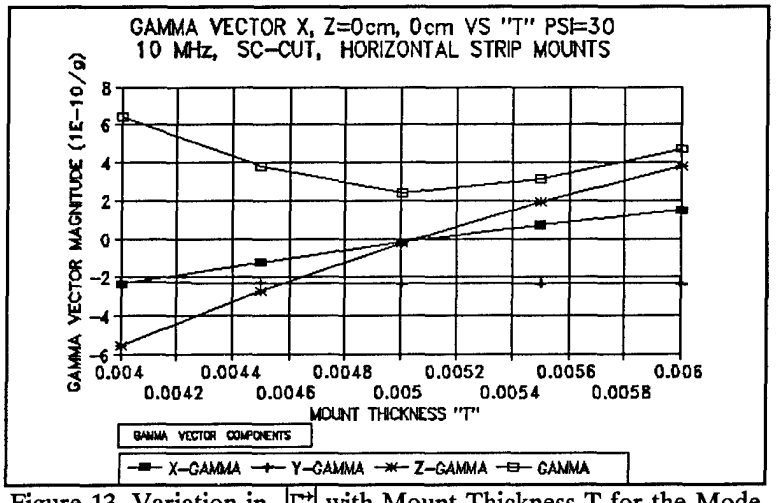


Figure 13. Variation in $|\Gamma^{\rightarrow}|$ with Mount Thickness T for the Mode Centered and $\psi = +30°$

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LOW TEMPERATURE LIMITATION ON THE QUALITY FACTOR OF QUARTZ RESONATORS

A. El Habti, F. Bastien

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

Abstract

The quality factor (Q) for different resonators at several overtones has been determined between 1.5 K and 300 K. These measurements form the basis for an improved interpretation of the limitation on Q. A significant consequence is an improved understanding of the relationship between Q and random frequency fluctuations. The curves of 1/Q = f(T) show the usual features between 20 K and 300 K. These include the sodium ion peak at 55 K as well as a peak at 20 K. Hovewer, the region of principal interest lies between 1.5 K and 20 K. At very low temperature, a plateau is always present. If for a given resonator, the 1/Q value of this plateau is sub tracted from the experimental values, the residual 1/Q is almost a linear function of T^4. This variation obeys the Landau-Rumer theory of acoustic wave absorption caused by phonon-phonon interaction in the crystal. Thus we conclude that between 6 K and 20 K the main limitation in Q is due to the crystal itself, but the plateau is not cause by intrinsic crystal properties. Data for different overtones permit the elimination, at least for the 7th overtone, of damping effects of the crystal supports. Measurements have also been carried out on crystals with surface having different roughness characteristics. These roughness characteristics have been measured using optical interferometry. With well polished, good crystal we measured, for 7 th overtone at 3 K a loaded Q of 25 million corresponding to an unloaded Q of 33 million. Better control of surface roughness and crystal quality could lead to a Q of about 100 million in the liquid helium region. This value would be for a 11 MHz, 7 th overtone crystal having a diameter of 15 mm.

Introduction

This work has been done to gain a better understanding of the limitations of quartz resonators. The resonator quality factor Q is measured as a function of temperature. The features of the curve $1/Q=f(T)$ are associated with different damping. These measurements have been done for different overtones. The results have provided a method that make the interpretation of limiting factors for the quality factor easier. The previous studies [1, 2, 3] on quartz resonators at very low temperatures have not included systematic studies of the physical causes limiting the quality factor.

In this paper, we describe the experimental device and the method of measurement, and, subsequently, the results.

Measurement device

Temperatures between 1.5 K and room temperature were obtained using a liquid helium cryostat. Between 4.2 K and room temperature the regulation is done by a heating resistor. Between 1.5 K and 4.2K, the regulation is performed using the gas pressure above the liquid helium.

The temperature is measured from 1.4 K to 100 K by a calibred germanium sensor and from 50 K to room temperature by a calibred silicon sensor. The precision is about 3×10^{-2} K at 4.2 K.

In fact, a precision of 0.1 K is sufficient for the present purpose. Obviously, the sensor is not the quartz itself. We have to evaluate the difference between sensor temperature and the quartz crystal which is enclosed in a capsule under vacuum.

Let us assume that the quartz is linked to the resonator can by two thin metal strips 5mm long with 0.5 mm^2 cross-section having a conductivity of 2 watts/m/K. We obtain for 10 µW power a temperature difference of 0.1 K. A similar computation for the quartz itself gives an even smaller temperature gradient. As a conclusion we can say that the temperature gradient is small and does not influence the measurement. We have also checked that the resistivity of the silver paste used for the connections is very low even at very low temperatures ($< 0.01\ \Omega$).

The quality factor measurement is done by exciting the quartz crystal at the overtone frequency and then by observation of the damping of the free oscillation. After analog rectification signal is processed by a numerical oscilloscope and a computer which give directly the value of Q. The temperature is also obtained automatically but the frequency adjustment is still manual. In spite of this manual adjustment it is possible to perform precise and fast measurements during the temperature ramp. Before presenting the results, we have had to evaluate the lowest Q that can be measured by this device.

0-7803-0476-4/92 $3.00 © 1992 IEEE

For a typical quartz crystal with a 15 mm diameter we have, for the 7 th overtone, a motional resistance of 945.5 ohms at room temperature for a $Q = 1.16 \times 10^6$. Therefore, with a load resistance of about 20 Ω we can easily measure a Q of about 10^8. In conclusion, our results are not limited by the device itself.

For a given excitation of the quartz we find, for the different measurements of Q, a very good precision of about 2 %. Unfortunately, the non linearity of the quartz and the rectifier do not permit the assumption of a value of Q independent of excitation conditions with a precision of better than about 10%. This difficulty is not a problem because this variation does not change the features of the $1/Q=f(T)$ curves.

The quartz crystals themselves are either AT or SC cut inside metal cans under vacuum. Using plano-convex-contoured quartz allows energy trapping. The electrodes are deposited on the quartz itself.

Experimental results

In Fig. 1, we give an example of $1/Q$ as a function of odd overtones from 1 to 9 at room temperature. The curve shows a minimum for the 5th overtone, generally used at room temperature. For lower overtones the $1/Q$ increase is due to increased damping of the resonator/support structure. For higher overtones $1/Q$ is almost proportional to the frequency F ; that is to say $Q \times f$ = constant. For a bulk acoustic wave this corresponds to an absorption which varies as the square of the frequency.

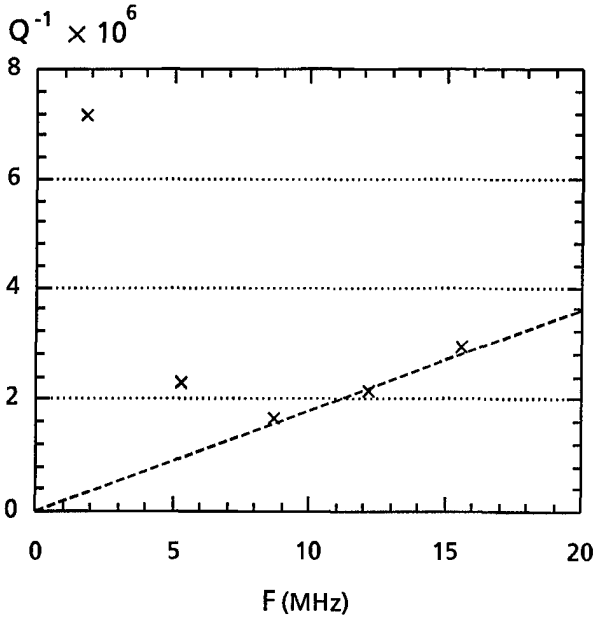


Fig 1 : *1/Q for different overtones of an AT- cut quartz crystal.*

X *Experimental Results (for T = 300 K)*

--- $1/Q \propto F^2$

In Fig. 2, we give an example of $1/Q=f(T)$. We observe an increase of $1/Q$ when T decreases from 300 K to 20 K with a peak at 55 K. Below 20 K the curve decreases rapidly toward a plateau between 4 K and 1.5 K. In order to understand this result let us analyze the damping in the mounting structures. Due to plano-convex geometry the energy trapping provides a fast decrease of damping as the overtone order increases. At room temperature the resonator is designed to have an optimal diameter. That is to say the smallest diameter is used which allow weak damping due to the mounting structures at the 3rd overtone compared with total damping. Nevertheless the use of a higher overtone can reduce the mounting structure damping.

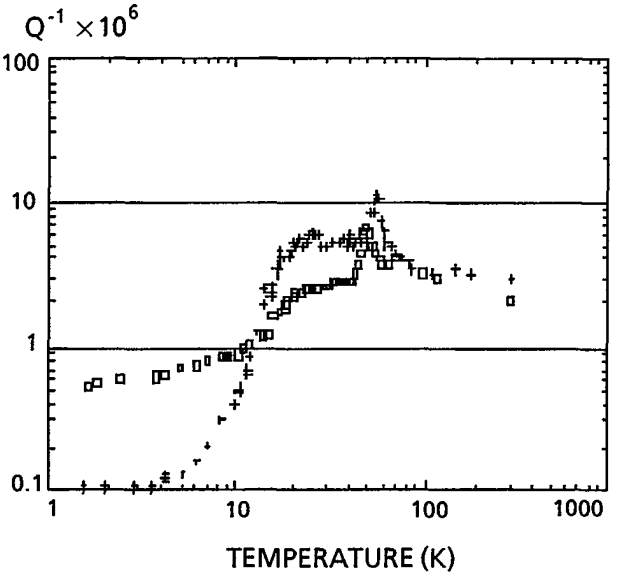


Fig 2 : *1/Q as a function of temperature for an SC-cut quartz crystal of 15 mm diameter.*

□ *3th overtone (9,99 MHz).*

\+ *7th overtone (23,23 MHz).*

Let us evaluate this decrease as a function of successive overtones. For this evaluation we use the simplified theory of Wilson [4]. The wave amplitude as a function of distance (x) from the center is given by

$$A(x) = H_{n-1}\left(\frac{x}{a_1}\right) exp\left(-\frac{x^2}{2a_1^2}\right)$$

where

$$a_1^4 = \frac{R\, b^3\, C_{11}}{m^2\, \pi^2\, C_{66}}$$

H_{n-1} is the Hermite polynomial of order n-1

m is the overtones order
R is the curvature radius
b is the quartz thickness at the center
and C_{11} and C_{66} are the stiffness coefficients for quartz.

The damping is assumed to be a function of amplitude $A(r)$, r being the quartz radius. Futhermore for mode m11 we get $n=1$ and $H_0=1$. Therefore as a first approximation, the damping is proportional to $exp(-Km)$, K being a constant and m the order of the overtone. Then we obtain

$$\frac{1}{Q_7} = \frac{1}{Q_3}\,\frac{exp(-7)}{exp(-3)} = 0.02\,\frac{1}{Q_3}$$

Q_3 and Q_7 are, respectively, the quality factors for the 3rd overtone and the 7th overtone. From the value of $1/Q$ at room temperature as a function of overtone order, we can deduce that the losses due to the mounting structure reach a maximum of 10^{-6} for the 3rd overtone. Using the value of Q obtained for the 3rd overtone, we obtain a value of 2×10^{-8} for the 7th overtone. That is to say, the losses are negligible, even at the lowest temperatures. Another way to see that the mounting losses are negligible at low temperatures is to see that the 7th overtone can give a better Q-value than the 5th overtone at low temperatures.

In conclusion, we can neglect the losses due to the support structure, at least for the 7th overtone.

The $1/Q$ peak near 55 K is due to the presence of impurities like Na^+. According to the generally-admitted theory, some atoms of Si^{+4} are replaced by Al^{+3}. To obtain electrical neutrality, a Na^+ ion associates with the Al^{+3} ion. This ion is trapped in a double potential well with a potential barrier of V_0. Then the theory [5] gives $1/Q$ varying as

$$\left(\frac{\omega_o\,\tau}{1+\omega_o^2\,\tau^2}\right)$$

where ω_0 is the acoustic wave angular frequency and τ is the relaxation time for reaching equilibrium after the system has been disturbed by the acoustic wave. Futhermore,

$$\frac{1}{\tau} = \frac{1}{\tau_o}\,e^{-V_o/kT}\;;$$

T being the absolute temperature and k the Boltzmann constant. Consequently, the maximum absorption is at $\omega.\tau = 1$. For a frequency changing from ω_1 to ω_2 the temperature corresponding to the absorption peak changes from T_1 to T_2. From the above, we find the relation

$$\omega_1\,e^{-V_o/kT_2} = \omega_2\,e^{-V_o/kT_1}\,.$$

According to the previous relationship, with $V_0 = 0.05$ eV, $f_1 = 10$ MHz and $T_1 = 52$ K (experimental values) we find for $f_2 = 23.2$ Mhz a value of $T_2 = 56.25$ K. This last value compares favorably with 56 K as found in the experiment.

Below a maximum at around 20 K, the $1/Q = f(T)$ curve decreases very rapidly to reach a plateau (see Figs. 2, 3 and 4).
If the plateau value is substracted from the curve, we get a straight line for $Ln(1/Q) = f(lnT)$ (see Fig.5). The slope of this line gives the relation $1/Q \propto T^4$. For very low temperatures, the dependence as a function of T is the value from Landau-Romer theory [6, 7] for acoustic wave absorption by phonon-phonon interaction into the crystal itself. Between 4 K and room temperature, according to the theory, the intrinsic absorption by the crystal itself presents a maximum. This maximum absorption is a function of the product of angular frequency multiplied by the phonon relaxation time τ. Knowing $\tau = f(T)$ gives the temperature corresponding to the maximum as the frequency varies. Let us compute the position of the maximum for $f = 1000$ MHz taking into account the fact that the maximum is $T = 22$ K for $f = 23$ MHz. It is possible to show [5] that

$$\tau \simeq \frac{3K}{C\,\overline{V}^2}$$

where K is the thermal conductivity, C is the heat capacity and

$$\frac{3}{\overline{V}^3} = \frac{1}{V_\ell^3} + \frac{2}{V_t^3}$$

(V_ℓ is the longitudinal wave phase velocity and V_t is the transverse wave phase velocity).

Consequently, as a first approximation, τ is proportional to K/C. Using values of $K(T)$ and $C(T)$ given by the literature [3] we obtain $\tau(T)$.

The relation $\omega_1\tau(T_1) = \omega_2\tau(T_2)$ gives $T = 55$ K for $f = 1\,000$ MHz. This result is very close to the value obtained for the maximum temperature, $T = 58$K, given by [8].

Futhermore, this experimental result gives a T^4 law down to about $T = 18$ K for $f = 1\,000$ MHz. Applying the same relation gives, for $f = 11$ MHz, a corresponding temperature of about 2 K.

Therefore the T^4 law seems verified at least down to 2 K.

In conclusion, for $1.5\text{ K} < T < 15$ K we can take the relation $1/Q = \alpha+\beta T^4$. The term βT^4 comes from the intrinsic absorption of the quartz. Now we wish to understand the origin of the constant term.

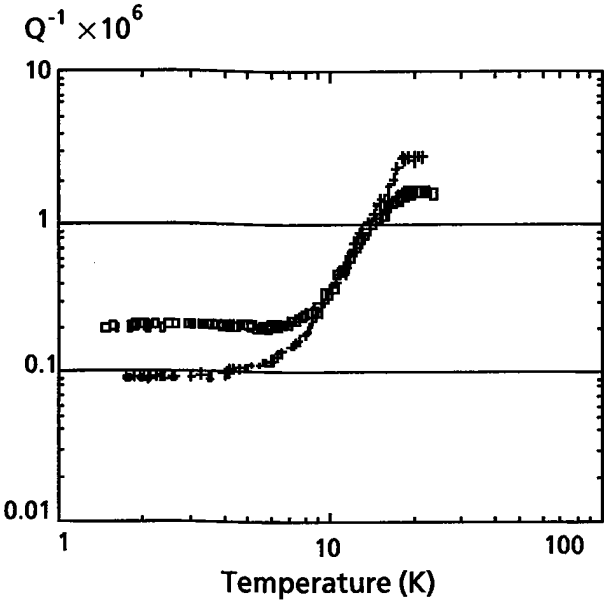


Fig. 3 : *1/Q as a function of temperature for a mechanically polished quartz crystal (SC-cut; 15 mm diameter).*

□ *3th overtone (5,1 MHz).*

\+ *7th overtone (11,8 MHz).*

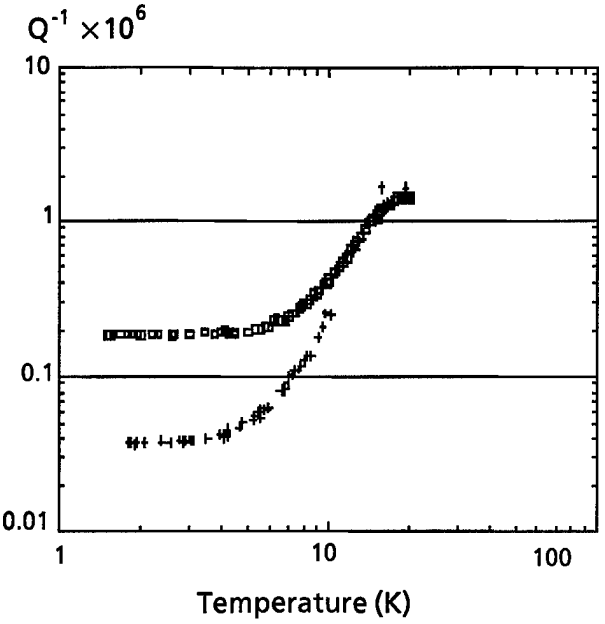
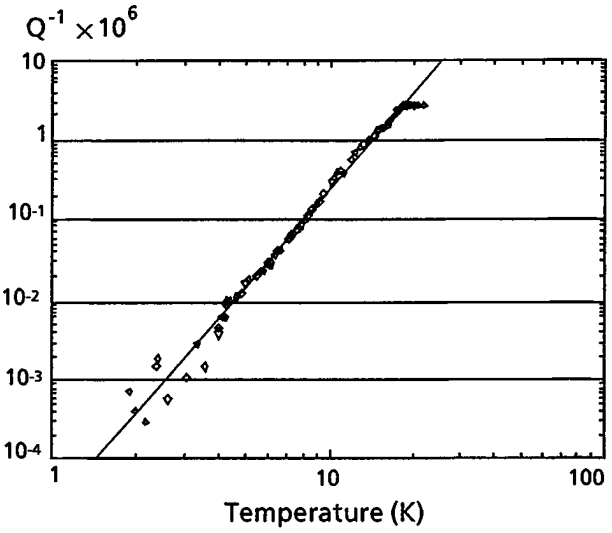


Fig. 4 : *As fig 3 but with an other mechanically polished quartz crystal.*

□ *3th overtone (5,1 MHz).*

\+ *7th overtone (11,8 MHz).*



◇ *Value of the low-temperature plateau has been substracted from the experimental result.*

— T^4 *Variation.*

Fig. 5 : *7th overtone for the quartz used for Fig. 3*

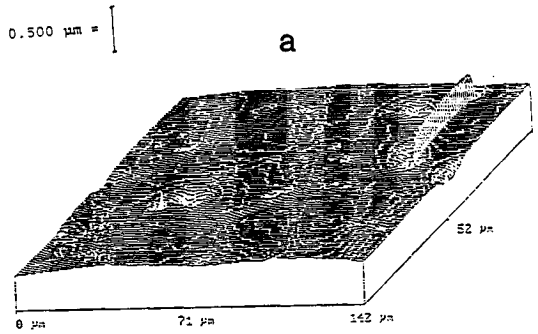
Very-low-temperature test for quartz resonator with different surface treatment

Use of the 7th overtone rules out damping by the mounting structure. The most probable origin of damping is in the resonator surface condition. In order to verify this hypothesis we have compared results obtained by resonators which are as identical as possible but which have different surface treatments. There are three type of surface treatments, namely :

- good mechanical polishing (category a),
- chemical polishing (category b), or
- a crude mechanical polishing (category c).

This polishing operation has been performed on an industrial basis. Control of the surface condition can only be done afterward. We removed the resonator from its can and we performed a surface test using an optical interferometric method. We have obtained a qualitative value for the surface roughness. Two examples of the image obtained are given in Fig. 6.

The quality of the polishing in these three categories can be characterized by the parameters R_a and R_z ; R_a being the average amplitude of the roughness on the observed surface ; R_z is the average of the maximum amplitude observed for 25 identical elements on the total observed surface.



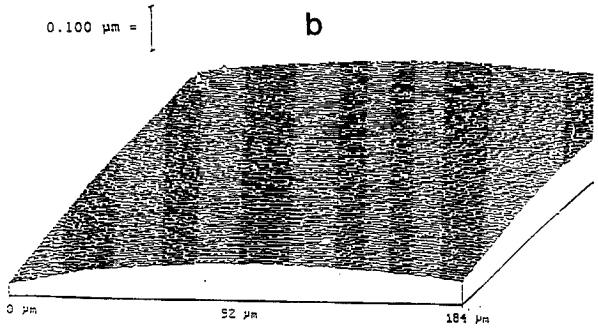


Fig. 6a : *Surface profile obtained by optical interferometric measurements. The element of 142 × 152 µm corresponds to a mechanically polished quartz crystal. The sharp crest at the upper right corner is an artefact of the measurement technique.*

Fig. 6b : *Same as for 6a but with a different quartz sample*

For category (a), R_a = 0.03 µm, R_z = 0.2 µm. For category (b), R_a = 0.4µm and R_z = 2 µm ; and for category (c), R_a is estimated to be 2 µm. The poor surface quality of category (c) did not permit a useful estimate for R_z. The results obtained for Q show some dispersion. In order to classify the quality of the resonator we adopted the following criterium. A resonator is juged good if it satisfied at least two out of three of the following conditions :
- unloaded Q factor for the 7th overtone at liquid helium temperature (Q^7_{He}) greater than 12×10^6;
- unloaded Q factor for the 3rd overtone at liquid helium temperature (Q^3_{He}) greater than 11×10^6;
- unloaded Q factor at room temperature (Q^3_{Room}) greater than 2×10^6.

A quartz crystal is classified as poor if we obtain at least two of the three following conditions:

$$Q^7_{He} < 3 \times 10^6 , Q^3_{He} < 2 \times 10^6 \text{ and } Q^3_{Room} < 1.5 \times 10^6.$$

The following table summarizes our results :

Category	total number of resonators	Number of resonators classified as good	Number of resonators classified as poor
a	7	6	1
b	4	0	2
c	4	0	2

In categories b and c, we found two resonators which fell in between the classifications of good and poor. The quartz crystal from category (a) that was classified as poor shows much scratching. If we neglect this rather special case, the previous results mean that a resonator with good polishing without scratches can easily give a $Q > 12 \times 10^6$ (with a 20 Ω load we get $Q > 10^7$ for f = 11 MHz). That is to say we obtain a product of $f \times Q > 10^{14}$.

The poor quality for chemically polished quartz is due to the fact that the polishing not only removed the disturbed layer, but also induced large surface irregularities.
Nevertheless, for category (a) resonators, values of Q show a broad dispersion as follows :

$$12 \times 10^6 < Q^3_{He} < 45 \times 10^6$$

$$11 \times 10^6 < Q^7_{He} < 50 \times 10^6$$

without a clear correlation with the mechanical polishing. We conclude that there exists another cause for temperature-independent damping. The most probable cause is the presence of a damaged region due to mechanical polishing. Unfortunately, we do not have at this time a suitable experimental method for controlling the superficial layer of the quartz crystal.

In order to evaluate the order of magnitude of the influence of the disturbed region, we have done the following calculation. The intrinsic attenuation coefficient corresponding to a Q factor of $Q = 10^8$ is about $\alpha_Q = 10^{-4}$ m^{-1}. In an amorphous quartz crystal at 4 K, the attenuation coefficient is α_A = 2 m^{-1} [9]. Therefore, as long as the absorption is considered, one millimeter of quartz crystal corresponds to a thickness of $h = \alpha_Q / \alpha_A \, 10^{-3}$ m of amorphous quartz. We find a thickness of 50 Å which is a plausible thickness for the disturbed layer.

Conclusion

The previous results show that the intrinsic quality factor due to crystal attenuation is about 10^8 at 4.2 K and 10^9 at 1.5K. The practical Q is between 10^7 and 4×10^7 for a good mechanically-polished quartz at 11 MHz. Furthermore, a very good chemical-mechanical polishing process should be able to produce a reproducible Q between 5×10^7 and 1×10^8 at 4.2 K.

As a consequence, the noise of the corresponding oscillator should decrease as Q increases for two reasons : a) increased filtering of the noise not due to

the resonator itself, and b) decrease in the resonator noise.

Regarding the resonator, several cases are to be considered. First, the attenuation is due to bulk wave propagation. Therefore there exists a relationship between a field X and the consequent linear results Y. The Kramers-Kronig relation may be applied to the complex susceptibility K. If, furthermore, the complex susceptibility is governed by only one time constant, the power noise density $S_y(f)$ varies as $1/Q^4$ [10]. But if the damping is linked to another cause of attenuation, for example wave diffraction on the surface irregularities or in the supports, the relationship $Sy(f) \propto 1/Q^4$ has no reason to hold. The previous remarks explain the fact that a curve of $S_y(f) = f(Q)$ shows an approximate $1/Q^4$ decrease but a large dispersion of S_y for a given Q [10].

It is difficult to compare directly results obtained at low temperatures with those obtained at room temperature because the main damping is generally not due to the same cause.

Nevertheless, with a very good quality quartz crystal and a very good surface treatment, Q should approach the intrinsic quartz crystal limit at low temperatures. For such a case the relation $S_y(f) \propto 1/Q^4$ should hold, and consequently the noise should also become extremely low.

Acknowledgements

The authors are grateful to Dr. G. Tribillon from the 'Laboratoire d'Optique de l'Université de Besançon' for the interferometric measurement of the quartz surface roughness. The authors wish to thank G. Marianneau for helpful discussions.

References

[1] J. J. Suter, "Acoustic Loss Phenomena in Alpha Quartz Crystal Resonators Over the 1.4-77 K Temperature Range". J. Appl. Phys., **63**, 5594-5595, 1988.

[2] G. Goujon, "Mesures des Fluctuations de Fréquence des Résonateurs à Quartz", Thesis, (number 107), Université de Franche-Comté (France) (1980).

[3] G. Robichon, J. Groslambert, J. J. Gagnepain, "Frequency Stability of Quartz Crystal at Low Temperatures : Preliminary Results" Proc. 38th Ann. Freq. Cont. Symp., pp. 201-205, 1984.

[4] C. J Wilson, J Phys. D. Appl. Phys. **7**, 2449-2454, 1974.

[5] W.P. Mason."Effect of Impurities and Phonon Prcesses on the Ultrasonic Attennuation of Germanium crystal Quartz and Silicon" in Physical Acoustics, edited by Warren P. Mason, Vol III part B, pp. 235-287, Academic Press New York, 1965.

[6] L. Landau, G. Rumer, Physik Z Sowjet-Union **11**,18, 1937.

[7] P.G. Klemens, "Effect of Thermal and Phonon Prcesses on Ultrasonic Attennuation" in Physical Acoustics, edited by Warren P. Mason, Vol III, part B, pp. 201-234, Academic Press New York, 1965.

[8] H. E. Bömmel. K. Dransfeld. Phys. Rev. **117**, 5, 1245-1252, 1960.

[9] S. Hunklinger, W. Arnold, "Ultrasonic Properties of Glasses at Low Temperatures" In Physical Acoustics, edited by Warren P. Mason and R.N. Thurston, Vol XII, Academic Press New York, 1976.

[10] J. J. Gagnepain, J. Ubersfeld, G. Goujon and P. Handel, "Relation Between 1/f Noise and Q Factor in Quartz Resonators at Room and Low Temperatures, First Interpretation", Proc. 35th Ann. Freq. Cont. Symp., pp. 476-483, 1981.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

The effect of the temperature on the vibration frequencies of plane-convex plate of quartz resonators

S.M. Bauer, S.B. Filippov, B.N. Semenov, P.E.Tovstik, Y.L. Vorokhovsky
Sankt-Petersburg State University, "Morion" Mfg. Co., Russia, Sankt-Petersburg

Abstract

The approximate analytical formulas for eigenvalues and vibration modes for homogeneous anisotropic plates with varied thickness are obtained. It's supposed that the vibration modes with one or some semi-waves of deformations in the thickness direction of plate are localized near the point of maximal thickness. By the perturbation method the influence of the irregular heating on the working frequency of the quartz resonator is studied.

1. Introduction

At present time the quartz piezoelectric resonators with shift vibrations in the thickness directions are widely used. These resonators are manufactured in the forms of the round or rectangular plates [1]. In order to concentrate the vibrations energy in the plate center one or both sides of the resonator plates are made convex. The vibration modes localized near the point of maximal thickness and the corresponding frequencies are approximately determined for AT-cut quartz resonator in [2].

In this paper the part of the spectrum for the free high frequency vibrations of anisotropic plates with varied thickness are studied by the method of asymptotic integration of three-dimensional dynamic equations of elasticity theory. The eigenvalues and the vibration modes are represented in the forms of a power series for the rational degrees of a small geometrical parameter which is equal to the maximal thickness of a plate divided by the radius of its convex surface. The frequencies and the localized vibration modes do not practically depend on the boundary conditions.

The influence of the temperature on the working frequency of the quartz resonator has been studied for the uniform heating [1]. In this work we got the expressions for the eigenvalues and the vibration modes which permit us to examine the nonuniform heating of the plates. The general formula is obtained for the estimation following four basic temperature reason factors on the variation of the working frequency - the effect of the thermal dependence of the elastic and piezoelastic modulus of quartz; the influence of the nonlinear dependence of strains and stresses which includes the influence the third-order elastic modulus [1]; the influence of the initial stresses in the plate caused by the nonuniform heating or by other reasons and the influence of the thermal strains.

2. The high-frequencies free vibrations of anisotropic plates with varied thickness

The three-dimensional equations of the harmonic

0-7803-0476-4/92 $3.00 © 1992 IEEE

vibrations of anisotropic plates in Cartesian coordinates x_1, x_2, x_3 have the form [3]

$$\sigma_{ij,j} + \rho\omega^2 u_i = 0, \quad \sigma_{ij} = E_{ijkl} u_{k,l}, \qquad (2.1).$$

where σ_{ij} - the stresses, u_k - the components of the displacement, E_{ijkl} - the components of elastic modulus tensor, ρ - the density, ω - the vibration frequency.

In (2.1) and later Latin indices i, j, k, l, q, p are equal to 1, 2, 3 and the Greek indices to - 1, 2.

Let one plate side be plane and the other side be the sphere of radius R. Then the plate thickness can be expressed approximately as $h(x_1, x_2) = h_0 - 1/2R^{-1}(x_1^2 + x_2^2)$. The both sides of the plate are supposed to be free, i.e.

$$\sigma_{i3} = 0 \quad \text{for} \quad x_3 = 0,$$

$$\sigma_{i3} - \frac{\partial h}{\partial x_1}\sigma_{i1} - \frac{\partial h}{\partial x_2}\sigma_{i2} = 0 \quad \text{for} \quad x_3 = h(x_1, x_2). \qquad (2.2)$$

We consider the vibration modes, which exponentially come down when the value of $x_1^2 + x_2^2$ comes up. Further such vibration modes are called localized. This modes and corresponding frequencies practically do not depend on the boundary conditions.

The solution of the equations (2.1), (2.2) is found as power series for degrees of the small geometrical parameter $\mu = (h_0/R)^{1/4}$:

$$u_{in} = U_i^p \left(v_{pn}^0 + \mu v_{pn}^1 + \mu^2 v_{pn}^2 + \ldots\right),$$

$$\omega_{in} = \frac{n\pi}{h_0}\sqrt{\frac{\lambda_i}{\rho}}\left(1 + \mu^2\beta_2 + \mu^4\beta_4 + \ldots\right), n = 1, 2, \ldots, \qquad (2.3)$$

where λ_i - the eigenvalues and U_i^p - the normalized eigenvectors of matrix E_{i3k3}.

Let us consider the subset of frequencies ω_{in} when index i is constant. Without any restrictions we can assume $i = 3$. Two groups of conditions must be fulfilled for the existence of the localized vibration modes u_{3n}:

$$\lambda_3 n^2 \neq \lambda_\beta m^2, \quad m = 1, 2, \ldots; \quad a_\alpha > 0, \qquad (2.4)$$

here a_1, a_2 - the eigenvalues of matrix $a_{\alpha\beta}$, and

$$a_{\alpha\beta} = C_{3\alpha3\beta} + \frac{1}{4\lambda_3}B_{3\alpha3}B_{3\beta3} + \frac{1}{d_\gamma}B_{3\alpha\gamma}B_{\gamma\beta3} - a_{\alpha\beta}^\gamma F_\gamma,$$

$$a_{\alpha\beta}^\gamma = \frac{4\lambda_3}{k_\gamma}\left(\frac{C_{33\gamma\alpha}}{\lambda_3} - \frac{B_{3\alpha\gamma}}{d_\gamma}\right)\left(\frac{C_{\gamma33\beta}}{\lambda_\gamma} + \frac{B_{\gamma\beta3}}{d_\gamma}\right),$$

$$B_{i\alpha j} = C_{i\alpha j3} + C_{i3j\alpha}, \quad C_{qjpl} = E_{ijkl}U_i^p U_k^q,$$

$$d_\gamma = \lambda_3 - \lambda_\gamma, \quad k_\gamma = n\pi\sqrt{\frac{\lambda_3}{\lambda_\gamma}},$$

$$F_\gamma = -\tan(k_\gamma/2) \quad \text{if} \quad n = 2m,$$

$$F_\gamma = 1/\tan(k_\gamma/2) \quad \text{if} \quad n = 2m+1.$$

The first terms of the series (1.3) for frequencies and vibration modes are found in the case when conditions (2.4) are fulfilled

$$\omega_{3n}^{(m_1,m_2)} = \frac{n\pi}{h_0}\sqrt{\frac{\lambda_3}{\rho}}\left\{1 + \frac{\mu^2}{n\pi}\left[\sqrt{\frac{a_1}{\lambda_3}}(m_1 + 1/2) + \sqrt{\frac{a_2}{\lambda_3}}(m_2 + 1/2)\right] + 0(\mu^4)\right\}, \qquad (2.5)$$

$$u_{3n}^{(m_1,m_2)} = U_3^p\left[v_{pn}^0 + \mu v_{pn}^1 + 0(\mu^2)\right], \qquad (2.6)$$

where

$$v_{\alpha n}^0 = 0, \quad v_{3n}^0 = V(t_1, t_2)\cos(n\pi z),$$

$$V(t_1, t_2) = H_{m_1}(c_1 t_1)H_{m_2}(c_2 t_2)\exp\left[G(t_1, t_2)\right],$$

$$G(t_1, t_2) = -1/2\left(c_1^2 t_1^2 + c_2^2 t_2^2\right), c_\alpha = \sqrt{n\pi}\left(\lambda_3 a_\alpha^{-1}\right)^{1/4},$$

$$t_1 = y_1\cos(\beta) - y_2\sin(\beta), \quad t_2 = y_1\sin(\beta) + y_2\cos(\beta),$$

$$\tan(2\beta) = \frac{2a_{12}}{a_{22} - a_{11}}, \quad y_\alpha = \frac{x_\alpha}{R\mu^3}, \quad z = \frac{x_3}{h},$$

$H_m(z)$ - Hermit's polynomial of m power and functions v^1_{pn} are linear combinations of derivatives $\partial V/\partial y_\alpha$ the coefficients of which depend on z.

3. The effect of the heating on the vibration frequencies of solids

Let the distribution of the temperature T in the initial state be known:

$$T^\circ = T_\circ + \theta^\circ(x_1), \tag{3.1}$$

and the change of the temperature θ° is satisfied the stationary equation of thermal conductivity:

$$\lambda_{ij}\theta^\circ_{,ij} = 0 \quad \left(\text{or} \quad \lambda_{ij}\frac{\partial^2\theta^\circ}{\partial x_i \partial x_j} = 0\right) \tag{3.2}$$

and corresponding boundary conditions on the surface of solid. Here λ_{ij} – the coefficients of thermal conductivity.

The displacements u°_i ,the strains ϵ°_{ij} and the stresses σ°_{ij} in the initial state are satisfied equations [4]:

$$((\delta_{ki} + u^\circ_{i,k})\sigma^\circ_{jk})_{,j} = 0, \quad \sigma^\circ_n = 0 \quad \text{on} \quad \partial V, \tag{3.3}$$

$$\epsilon^\circ_{ij} = e^\circ_{ij} + \frac{1}{2}u^\circ_{m,i}u^\circ_{m,j}, \quad e^\circ_{ij} = \frac{1}{2}(u^\circ_{i,j} + u^\circ_{j,i}). \tag{3.4}$$

The perturbated state is represented as a small vibrations near the initial state. The displacements u_i, the strains ϵ_{ij} , the stresses σ_{ij} and the temperature T in the perturbated state are represented as

$$u_i = u^\circ_i + u^*_i, \quad \epsilon_{ij} = \epsilon^\circ_{ij} + \epsilon^*_{ij},$$
$$\sigma_{ij} = \sigma^\circ_{ij} + \sigma^*_{ij}, \quad T = T_\circ + \theta^\circ + \theta^*, \tag{3.5}$$

where the additional values connected with vibrations are marked by the sign $*$. The vibrations are supposed to be small enough the values with $*$ to be linear on u^*_i terms. The equations of motion in the perturbated state have a form

$$((\delta_{ki} + u_{i,k})\sigma_{jk})_{,j} + \rho\omega^2 u^*_i = 0, \quad \sigma_n = 0 \quad \text{on} \quad \partial V, \tag{3.6}$$

where ρ is the material density in the natural state. Subtracting (3.3) from (3.6) and keeping only linear on u^*_i terms one may obtain

$$((\delta_{ki} + u^\circ_{i,k})\sigma^*_{jk} + \sigma^\circ_{jk}u^*_{i,k})_{,j} + \rho\omega^2 u^*_i = 0,$$
$$\sigma^*_n = 0 \quad \text{on} \quad \partial V. \tag{3.7}$$

Substituting into (3.7) the equations of termopiezoelasticity with due regard for (3.4) after neglecting the terms of the second order on ϵ°_{ij} we get the expression

$$(C_{ijkl}(\theta^\circ)u^*_{k,l} - e_{kij}(\theta^\circ)\Phi_{,k})_{,j} + \rho\omega^2 u^*_i$$
$$+(C_{ijkl}(\theta^\circ)u^\circ_{m,k}u^*_{m,l} + C_{ijkl}(\theta^\circ)u^\circ_{i,k}u^*_{l,m}$$
$$+C_{ijklmn}u^\circ_{m,n}u^*_{k,l} + \sigma^\circ_{jk}u^*_{i,k})_{,j} = 0,$$
$$(e_{kij}(\theta^\circ)u^*_{i,j} + g_{kl}(\theta^*)\Phi_{,l})_{,k} = 0 \tag{3.8}.$$

Here $\Phi(x_i)$ is the potential; $E_k = \Phi_{,k}$ the tensity of the electric field; e_{kij}-the piezoelectric constants; C_{ijklmn}- the third-order elastic modulus.

Let $\omega_\circ$ be one of the vibration frequencies of solid in natural state, and v_i, ϕ- corresponding vibration mode. Let us suppose, that the eigenvalue $\omega_\circ$ is the simple one, i.e. only the single vibration mode corresponds to this eigenvalue and there are no other eigenvalues near $\omega_\circ$. Let ω be the corresponding eigenvalue near the initial state and

$$\omega = \omega_\circ + \Delta\omega = \omega_\circ(1 + \xi), \quad \xi = \frac{\Delta\omega}{\omega_\circ} \tag{3.9}.$$

The value of ξ is obtained by the perturbation method. Let us represent the unknown vibration

modes in the form:

$$U_i = v_i + \Delta U_i, \quad \Phi = \phi + \Delta\Phi \qquad (3.10).$$

The value of ξ is found from the existence condition for the solution of the system of the equations determining the correction terms ΔU_i, and $\Phi = \phi + \Delta\Phi$ and may be represented as:

$$\xi = \xi_\theta + \xi_\epsilon + \xi_\sigma + \xi_3 \qquad (3.11).$$

The terms in the right side describe consecutively

- the effect of the thermal dependence of the elastic modulus of quartz.

- the influence of the thermal strains,

- the influence of the initial stresses in the plate caused by nonuniform heating or by other reasons,

- the influence of the nonlinear dependence strains and stresses which includes the influence the third-order elastic modulus.

Let us note, that the value of ξ does not depend on the density ρ.

Conclusion

The analytical formulas for quartz resonators of arbitrary orientation obtained in this work allow to determine the frequencies and the localized modes of shift vibrations. The frequency corrections caused by the nonuniform heating is also found.

The equations (2.1) and boundary conditions (2.2) do not assume piezoelastic effect. If one take into account the piezoelastic effect the vibrations frequencies approximately change by 0.1%. In the same time the frequency corrections (which is of order of μ^2 according formula (2.5) are about 1%. That is why it has a sense to consider the piezoelectric effect only simultaneously with the evaluation of the frequency corrections of the order of μ^4.

In this work we consider only the plane-spherical plates. But the proposed asymptotic method allows to study the vibrations of a plate of arbitrary form. The only condition is that the plate thickness has a smooth extremum in some point.

References

[1] I.Zelenka, Piezoelectric resonators on the acoustic volume and surface waves. Moscow: "Mir", 1990, 584 p.(in Russian).

[2] S.H.Slavov, "Modes of vibration, motion inductance and resonance intervol circular convex AT-cut bevelled design trapped energy quartz resonators", Appl. Phys., A40, pp. 59-65, 1986.

[3] W.Nowacki, Theory of elasticity. Moscow: "Mir", 1975. 872 p. (in Russian).

[4] K.F.Chernykh, Non-linear theory of elasticity, Leningrad: "Mashinostroenie", 1986, 336p. (in Russian).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

INVESTIGATION OF ELECTRODE-QUARTZ INTERFACE

P.I.Asseev, V.B.Grouzinenko

Scientific-Research Institute "Fonon"
Moscow, Russia

Abstract

The results of investigation of electrode-quartz interface by using RMA method for Au- Ag- and Cu-electrodes are presented. Possible reasons for disturbed equilibrium state of double layer existence are considered (e.g. under high levels of drive of the crystal unit or flashover phenomenon and electrode erosion). The opinion of the authors is expressed with regard to the mechanism for lowering crystal unit motional resistance caused by current overdrive.

Introduction

The electrode in a quartz crystal unit is a part of a resonance system, but the interface boundary of two contacting phases (those of electrode and of quartz) can be considered as two-dimensional structural defects. It can be expected that the local changes in the adhesion bond of the electrode with the quartz surface cause amendments of crystal unit parameters.

This paper describes the results of investigation of electrode-quartz interface with respect to Sd-metals (Au, Ag, Cu).

The quartz-electrode bond strengthening could occur owing to the diffusion redestribution of atoms, as a result of which the interphase interface boundary turns into eroded transient layer. However, the investigation of possible Au-, Ag- and Cu-diffusion in quartz made by RMA method in the region of elements traces does not confirm the diffusion contribution into the adhesion bond forming. Hence, the mechanism of the adhesion bond is dominated directly by the electrode-quartz interface[1,2].

The interest to Sd-metals is connected with the fact that these metals do not form a transient oxide layer as distinct from metals having high specific heat values of oxidation, such as Ti, Cr, V. The existing adhesion bond of Au, Ag and Cu with quartz, caused by dispersion interaction is very weak and cannot explain the real values of electrode adhesion of such metals to quartz. It was supposed earlier, that the quartz surface while forming an adhesion bond electrode-quartz to be an ideal one as for the case of crystal split in vacuum. We feel that this does not correspond to the reality.

It was experimentally proven that on the real surface of a quartz crystal element a hydroxide layer always remains bonded by oxigen with the silicium atom after various cleaning techniques. The existence of hydroxide layer on the quartz surface is the reason of possibility for the adhesion bond to originate between the metal condensate atom and surface OH-group, the bond being defined by electrostatic interaction, i.e. by an electron transfer from the metal atom to the proton of an acid group [3] . The existence of a charged state during its absorption on a surface of the dielectric has been confirmed by a number of authors. It is known that for hydroxylated quartz surface the dipole moment of a hydroxyl is equal to 2.5D as in the case of freely rotating hydroxyl group[4]The solution of Schrödinger equation for electron eigenvalues in the dipole field shows that with a dipole value of $P \geq 1.625D$, at least one bonded electron state exists in its field, the transition of conductivity electrons is possible from the metal into a two-dimensional crystal consisting of dipoles of hydroxyl groups[5].Hence,at the interface boundary a double electrical layer originates dominating the adhesion bond of electrode-quartz. If a strong electric

0-7803-0476-4/92 $3.00 © 1992 IEEE

field is formed in the proximity to the metal electrode, then as a result of a tunnel transition across the potential barrier, the electrons leave the metal. In our case the potential barrier is formed under the action of a mirror reflexion forces potential and of a potential coming from hydroxyl layer of quartz. If the quartz surface is completely covered by hydrophobic impurities or, on the contrary, the surface is absolutely clean (without hydroxyl), then in both cases no transition of electrons from metal to quartz occurs [6].

In other words, from the point of view of the adhesion strength the "superclean" quartz surface is as bad as the surface covered by impurities.

The authors have developed programs and calculated probable values of the adhesion bond strength charge component F_{ad} versus density of hydroxyl groups Z_{OH} on the quartz surface and versus the number "n" of metal atoms reacting with these groups. The calculation results are given in Table 1.

Table 1

Dependence of the adhesion bond strength F_{ad}(kg mm^{-2}) on the density Z_{OH}(cm^{-2}) and the number of metal atoms "n"

Z_{OH}(cm^{-2})	n	F_{ad}(kg mm^{-2})
$6\ 10^{14}$	$5\ 10^{13}$	18
$6\ 10^{14}$	$1.2\ 10^{13}$	1.5
$6\ 10^{14}$	$4\ 10^{12}$	0.16
$6\ 10^{14}$	$9\ 10^{11}$	0.008
$6\ 10^{14}$	$1\ 10^{11}$	0.00011

The density value $Z_{OH} = 6\ 10^{-4}$ cm^{-2} was chosen in accordance with the experimental data of [7]. Our experimental values of F_{ad} were equal to:

- for polished crystal elements not subjected to ion cleaning process:

  0.01 ÷ 0.05 kg/mm^2;

- for crystal element metallized after ion cleaning in vacuum:

  0.2 ÷ 2.0 kg/mm^2.

The spread in the experimental data obtained while carrying out the program can be explained by the significant porosity of the metal film. This assumption is confirmed by the microphotographs of silver electrode areas with the electrode thickness of 150 nm. The essential lowering of F_{ad}-value spread was attained by using specially designed evaporator of a new type which practically eliminated porosity in the film. The number of local surface contaminations present was decreased owing to the control of the ion cleaning conditions.

If the adeobatic conditions of double electrical layer existence are disturbed at the local electrode regions, the potential difference appearing within these regions between the electrode and the quartz surface causes transfer of electron density from the hydroxyl groups to the metal. For the crystal unit this means weakening adhesion strength at the corresponding local electrode areas. The equilibrium state of the double layer can be disturbed, for example, while excitation of a quartz crystal unit. In a number of cases (under high levels of drive) the flashover and the electrode erosion was observed. This confirms indirectly an electrostatic adhesion nature.

The electron density transfer is not obligatory occurs involving mechanism of flashover. Under softer conditions, when the barrier width w_b is changing less than an order of magnitude, the electron density transfer can be affected by using the mechanism of cold emission across the internal electrode-quartz interface. This occurs under the electric field values $E \sim 10^6$V/cm, which is fulfilled in our case. The barrier width w_b and the adhesion strength, respectively, can change along the electrode area and will correlate with the amplitude of vibration with normal displacement components U_n depending on the quartz crystal drive level.

For the quartz crystal unit this correlation can be interpreted as the local disdamping of acoustic vibrations and lowering its motional resistance accordingly: - the effect called "current overdrive". One can assume that further investigation of the interface electrode-quartz behaviour in the acoustic field of a crystal unit will enable to explain such phenomena as 1/f noise, "sleeping" effect etc.

Conclusion

1. The diffusion component does not contribute to the adhesion bond of the electrode with the polished quartz surface treated under glow discharge conditions.

2. After glow discharge treatment of the crystal element the concentration of hydroxyl groups is approximately equal to $6\ 10^{14}$ cm^{-2} and ensures a sufficient value of silver- and copper-quartz adhesion bond strength.

3. Incomplete cleaning of the quartz surface from hydrocarbon contaminations and complete removing of hydroxyl groups from its surfaces weakens the metal-quartz adhesion bond strength.

4. For strengthing the metal-quartz adhesion bond it is necessary to improve the evaporator design (to improve its optical density) and the technological ion-beam cleaning process by excluding photopolymerization hydrocarbons on the quartz surfaces.

References

[1] Asseev P.I. "Diffusion of metals in AT-cut quartz crystal plate", Electronic Engineering, 1983, series 5, issue 1(50), p.57-59.

[2] Vargina S.A. "Strength and energy characteristics of metal-quartz adhesion", Preprint

[3] Mjassnikov I.A. et al. "On aggregation state of the adsorbed Sd metal atoms on oxides", Journal of Physical Chemistry, 1973, No.9, p.2349-2353.

[4] Mironov S.L. at al. "Adsorption potential and electrophysical properties of silicium dioxide", Ukrainer Journal on Physics, 1979, v.24, No.7, p.983-989.

[5] Turner J., Anderson V., Fox K. "Ground state energy eigenvalues and eigenvectors for an electron in an electric dipole field", Phys. Rev., 1968, v.174, No.1, p.81-89

[6] Gorlov Y.I. et al. "On the mechanism of dehydratation of silica", Theoretical and Experimental Chemistry, 1980, v.16, No.3, p.333-338.

[7] Kisselyov V.F. "Surface phenomena in semiconductors and dielectrics", Publishing House "Science", 1972.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

A THEORETICAL STUDY OF QUARTZ CRYSTAL RESONATORS FOR CHEMICAL SENSORS, VIBRATING IN C MODE

Mitsuo NAKAZAWA*, Makoto TAKEUCHI*, Toyosaka MORIIZUMI**, Hiroo NIIYAMA**, Arthur BALLATO***
and Theodore LUKASZEK***

Shinshu University*(500 Wakasato, Nagano-shi 380, JAPAN), Tokyo Institute of Technology**(O-okayama, Meguro-ku, Tokyo 152, JAPAN), Electronics Technology and Devices Laboratory***(Fort Monmouth, NJ 07703-5601, USA)

Abstract

The cut angles and their governing equation for quartz crystal resonators vibrating in c mode, having reliable frequency-temperature characteristics, under high temperature region, such as 200 °C are derived from the improved analytical theory with piezoelectric effect. Lateral field excitation is also discussed.

1. Introduction

As we know, single crystal alpha-quartz is one of the most famous materials used for resonators, because it has excellent properties in thermal, elastic, piezoelectric and chemical fields. Hence, recently, the quartz crystal resonators have been investigated for chemical sensors based on the mass loading effect. Those concerned with the design of piezoelectric oscillators have been consistently sought to discover new crystal cuts which have reliable frequency-temperature characteristics in the high temperature region. As for those research it has been eagerly carried out by Kahan, 1982 that there are AK cut quartz crystal resonators. However, if the quartz crystal cuts calculated by Kahan are fabricated with conventional thickness-field or with perpendicular excitation in doubly rotated cuts both the b and c modes will be excited in many cases.

Accordingly, it is the objects of the present research to provide an improvement for the stress and temperature compensated in piezoelectric resonators and to provide an improvement in quartz crystal resonators which operate exclusively in the c mode or slow quasi-shear mode of vibration. It is a further object of the present research to provide a precision equation to predict the cut angles of the quartz crystal resonators with reliable frequency-temperature characteristics.

2. Theory

The right-handed rectangular Cartesian coordinate system (0-X_1, X_2, X_3) in relation to the quartz crystal axes is defined as shown in Fig.1. If a point (x_1, x_2, x_3) in a thin quartz plate is described in the polar coordinate system ($0-s, \theta', \phi'$), as shown in Fig.1, we obtain the following relationships:

$$x_1 = s \sin\theta' \cos\phi', \quad x_2 = s \sin\theta' \sin\phi', \quad x_3 = s \cos\theta' \tag{1}$$

The resonance frequency equation for thickness-vibration modes in a thin ideal crystal plate are given by

$$f = \frac{n}{2y_0}\left(\frac{c^{(m)}}{\rho}\right)^{1/2} \tag{2}$$

for m=1,2,3, and n=1,3,..., where ρ and y_0 are the mass density and the thickness of the crystal plate, respectively, and $c^{(m)}$ represent the eigen values for each thickness-vibration mode m and are solutions of the following equation:

$$\left| c_{ijkl} m_j m_k - c^{(m)} \delta_{il} \right| = 0, \tag{3}$$

where

$$c_{ijkl} = c^E_{ijkl} + (m_q \epsilon^S_{qj} m_j)^{-1} (e_{hij} m_h)(m_h e_{hkl}) \tag{4}$$

with

c^E_{ijkl} :elastic stiffness constants at constant electric fields for i,j,k,l=1,2,3.
e_{hij} :piezoelectric constants for h,i,j=1,2,3,
ϵ^S_{qj} :dielectric constants at constant strain for q,j=1,2,3,
δ_{il} :Kronecker's delta for i,l=1,2,3, and
m_j, m_q: i th or q th component of unit normal vector perpendicular to the plate plane for j,q=1,2,3.

In eq. (2), y_0, ρ, and $c^{(m)}$ are, in general, functions of temperature T and these temperature coefficients for y_0 and $c^{(m)}$ depend on the azimuth ϕ' and colatitude θ' in Fig.1. Hence, Taylor expansion for f around reference temperature T_0 is approximately expressed as follows:

$$f(T) = f(T_0)\left[1 + \alpha(T - T_0) + \tfrac{1}{2}\beta(T - T_0)^2 + \tfrac{1}{6}\gamma(T - T_0)^3\right], \tag{5}$$

where α, β, and γ are the first-, second-, and third-order temperature coefficients of frequency, respectively, and are defined as

$$\alpha = \frac{1}{f(T_0)}\left(\frac{\partial f}{\partial T}\right)_{T_0}, \quad \beta = \frac{1}{f(T_0)}\left(\frac{\partial^2 f}{\partial T^2}\right)_{T_0}, \quad \gamma = \frac{1}{f(T_0)}\left(\frac{\partial^3 f}{\partial T^3}\right)_{T_0}. \tag{6}$$

Equations (2)-(6) given above are solved by using the values of elastic stiffness constants, piezoelectric constants, mass density, and their temperature coefficients, etc.

If we expand the frequency of quartz crystal resonators by Taylor series around reference polar angles and reference temperature, the following

0-7803-0476-4/92 $3.00 © 1992 IEEE

equation is obtained:

$$f(\phi_0+h,\ \theta_0+k,\ T_0+l)=f(\phi_0,\ \theta_0,\ T_0)+\sum_{n=1}^{n}\frac{1}{n!}\left(h\frac{\partial}{\partial\phi}+k\frac{\partial}{\partial\theta}+l\frac{\partial}{\partial T}\right)^n f(\phi_0,\ \theta_0,\ T_0)+R_n, \quad \cdots (7)$$

where
$h=\Delta\phi$, $k=\Delta\theta$, $l=\Delta T$, and R_n is residual term.
From eq. (7) we have

$$\begin{aligned}&f(\phi_0+h,\ \theta_0+k,\ T_0+l)\\&\fallingdotseq f(\phi_0,\ \theta_0,\ T_0)\left[1+\left(\alpha l+\frac{1}{2}\beta l^2+\frac{1}{6}\gamma l^3\right)\right.\\&+\left(d_h h+\frac{1}{2}d_{hh}h^2+\frac{1}{6}d_{hhh}h^3\right)\\&+\left(d_k k+\frac{1}{2}d_{kk}k^2+\frac{1}{6}d_{kkk}k^3\right)\\&+(d_{hk}hk+d_{kl}kl+d_{lh}lh)\\&+\frac{1}{2}\left(d_{hhk}h^2k+d_{kkl}k^2l+d_{kll}kl^2+d_{llh}l^2h\right.\\&\left.\left.+d_{lhh}lh^2+2d_{hkl}hkl+d_{hkk}hk^2\right)+\cdots\right], \quad (8)\end{aligned}$$

where

$$\alpha=d_l=\left[\frac{1}{f}\frac{\partial f}{\partial T}\right]_{\phi_0,\ \theta_0,\ T_0},$$

$$\beta=d_{ll}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial T^2}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$\gamma=d_{lll}=\left[\frac{1}{f}\frac{\partial^3 f}{\partial T^3}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_h=\left[\frac{1}{f}\frac{\partial f}{\partial\phi}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{hh}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial\phi^2}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{hhh}=\left[\frac{1}{f}\frac{\partial^3 f}{\partial\phi^3}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_k=\left[\frac{1}{f}\frac{\partial f}{\partial\theta}\right]_{\phi_0,\ \theta_0,\ T_0},\quad d_{kk}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{kkk}=\left[\frac{1}{f}\frac{\partial^3 f}{\partial\theta^3}\right]_{\phi_0,\ \theta_0,\ T_0},$$

$$d_{hk}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial\phi\partial\theta}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{kl}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta\partial T}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{lh}=\left[\frac{1}{f}\frac{\partial^2 f}{\partial T\partial\phi}\right]_{\phi_0,\ \theta_0,\ T_0}$$

$$d_{hhk}=\left[\frac{1}{f}\frac{\partial^3 f}{\partial\phi^2\partial\theta}\right]_{\phi_0,\ \theta_0,\ T_0}, \quad \cdots \quad (9)$$

From eq. (8) the condition having the turning point with respect to the temperature change l is as follows:

$$\frac{\partial f(\phi_0+h,\ \theta_0+k,\ T_0+l)}{\partial l}\fallingdotseq$$

$$f(\phi_0,\ \theta_0,\ T_0)\left\{\frac{1}{2}d_{kkl}k^2+(d_{kl}+l\,d_{kll}+h\,d_{hkl})k\right.$$

$$\left.+\left(d_{lh}+l\,d_{llh}+\frac{1}{2}h\,d_{lhh}\right)h+\left(\alpha+\beta l+\frac{1}{2}\gamma l^2\right)\right\}=0. \quad (10)$$

Now we set as

$$h=\phi-\phi_0,\quad k=\theta-\theta_0,\quad l=T-T_0. \quad (11)$$

Hence, from equations (10) and (11) we have

$$\begin{aligned}&\frac{1}{2}d_{kkl}(\theta-\theta_0)^2+\\&(d_{kl}+d_{kll}(T-T_0)+d_{hkl}(\phi-\phi_0))(\theta-\theta_0)\\&+\left(d_{lh}+d_{llh}(T-T_0)+\frac{1}{2}d_{lhh}(\phi-\phi_0)\right)(\phi-\phi_0)\\&+\left(\alpha+\beta(T-T_0)+\frac{1}{2}\gamma(T-T_0)^2\right)=0. \quad (12)\end{aligned}$$

Eq. (12) shows the relationships for the turnover temperatures T vs the cut angles ϕ and θ of the quartz crystal resonators.

3. Results and discussion

Now let us consider that a plane wave is propagated with the wave vector k in thin quartz crystal plate, as shown in Fig.1. If we standardize the elastic stiffness constants, piezoelectric strain constants, dielectric constants, and density of the plate to room temperature 20°C, and then substitute these values into equations (2), (3), and (4), and changing the values for the polar angles, we obtain the frequency-temperature characteristics.

Fig.2 shows the contour map as functions of polar angles for the first order frequency-temperature coefficients of quartz crystal resonators vibrating in c mode. Sinha's locus of zero frequency-stress coefficients is also shown. In the figure we can find AT-, SC-, and NLSC- cuts.

For the sake of providing an improvement for the stress and temperature compensated in quartz crystal resonators in the high temperature region, the calculations based on the present theory described in equations (1)-(6) have been carried out by us. Tables 1 and 2 show the results. Table 1 shows the cut angles, modes, velocities, frequency-constants, frequency-temperature coefficients, and turnover temperatures. Table 2 shows cut angles for the normal direction and the plate direction, modes, displacement vector components with respect to fundamental axes of quartz crystal, deflection angles of displacement vectors, and coupling coefficients.

Fig.3 shows plate orientations with respect to crystallographic quartz axes.

In Tables 1 and 2 we see that the family of doubly rotated cut (YXwl) ϕ/θ = 30°/32° resembles the 30°V cut, and that the family of doubly rotated cut (YXwl) ϕ/θ = 19°/-28° closes to the SK cuts found by Kosinski.

Figs. 4 and 5 show lateral field coupling coefficients as functions of the angles psi with respect to the X_2'' for all three modes which are cut in (YXwl) ϕ/θ =19°/-28° and (YXwl) ϕ/θ =30°/32°, respectively. In the figures we see the c mode lateral coupling coefficient has a maximum of approximately 9.5% for the 19°/-28° cut which is approximately double that of the thickness coupling coefficient. This means that the excitation is

greater by a factor of 4. In the case of the 30°/32° cut the c mode lateral coupling coefficient has a maximum of about 10.5% as opposed to 2% for the thickness coefficient, so that excitation strength is more than 25 times greater.

Now in our experiment it is confirmed that the turn-over temperature of frequency vibrating in c mode for the (YXwl) 30°/32° quartz crystal resonator is near to 138°C. On the other hand, in the theory it was about 203°C. It is thought that the difference is due to the non-linear effects of temperatures and cut angles.

Therefore, the new equation (12) is proposed to govern the turn-over temperatures vs the cut angles. For example, we can predict that the rotated Y cut quartz crystal resonator in the (YXwl) ϕ / θ =0°/36.3° guarantees the turn-over temperature in the vicinity of 200°C.

4. Conclusion

From our theoretical research it is found that the realization of stress-compensated quartz resonators with reliable frequency-temperature characteristics in the high temperature region can be predicted.

Acknowledgement

The authors express their sincere thanks to President of San-ei Denshi Kogyo Co., Japan, Mr. Hisao Karasawa, and Drs. and nurses for the Koseiren Shinonoi and Matsushiro Hospitals for their valuable and devoted support.

The authors also wish to thank Messrs. Michiaki Takagi, Seiko-Epson Co., Hideki Yamaguchi in Miyota Co., and Akio Chiba in Nihon Dempa Kogyo Co., Japan for their help and suggestions.

References

1) A. Kahan: 36th FCS, pp.170-180, 1982.
2) A. Ballato: Physical Acoustics, chap.5, edited by W.P. Mayson, Academic Press, 1977.
3) J.A. Kosinski: United State Patent, No.4,754,187 , Jun. 28, 1988.
4) W.G. Cady: Piezoelectricity, Vol. 1, Dover Publications, 1964.
5) M. Aruga: Bull. Tokyo Inst. Tech., Vol. A-2, pp. 88-182, 1956 (in Japanese).
6) M. Nakazawa, A. Ballato, and T. Lukaszek: IEEE Trans on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-34, No.2, pp.270-277, 1987.

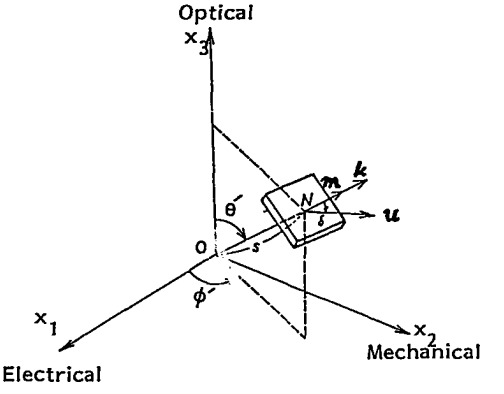


Fig.1 A thin quartz crystal plates cut in any direction.

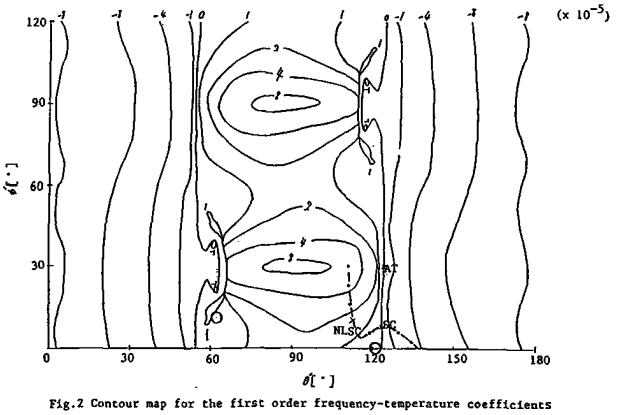


Fig.2 Contour map for the first order frequency-temperature coefficients of quartz crystal resonators(c mode). Symbol —· — : Sinha's locus of zero frequency-stress coefficients.

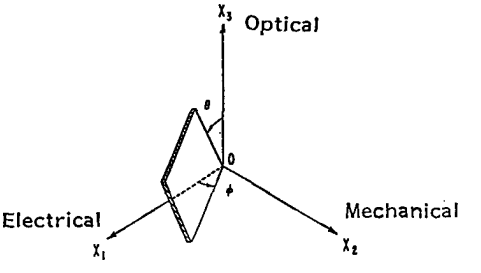


Fig.3 Plate orientations with respect to crystallographic quartz axes.

TABLE 1 Cut angles, modes, velocities, frequency constants , frequency-temperature coefficients and turnover temperatures

ϕ'/θ'	mode	v	fy_0	α	β	γ	T_1	T_2
0 /122	a	6.546835	3.273418	-54.98	-16.90	-14.00		
	b	4.069113	2.034557	-21.46	-4.79	0.15		
	✱ c	3.812836	1.906418	5.00	-0.31	-2.66	202.6	-185.9
0 /120	a	6.511207	3.255604	-52.86	-16.48	-13.95		
	b	4.091629	2.045815	-17.84	-4.40	-0.45		
	c	3.794232	1.897116	5.21	-0.10	-2.34	227.7	-196.3
0 /124	a	6.578222	3.209111	-57.07	-17.29	-14.02		
	b	4.079479	2.039739	-22.53	-5.02	-0.004		
	c	3.805197	1.902598	1.70	-0.79	-2.29	112.1	-141.1
9 /62	a	6.178267	3.089134	-58.43	-17.09	-13.64		
	b	4.283146	2.141573	-9.69	-3.98	-2.29		
	c	4.077992	2.038996	6.72	-1.06	-4.08	177.4	-189.3
9 /68	a	6.022241	3.011120	-50.82	-15.37	-13.20		
	b	4.530844	2.265422	-0.35	-4.37	-5.00		
	c	3.905595	1.952797	5.99	-0.43	-2.56	218.9	-212.0
11 /62	a	6.107132	3.053566	-60.95	-17.46	-13.59		
	b	4.310432	2.155216	-9.52	-4.08	-2.45		
	✱ c	4.156102	2.078051	8.91	-0.98	-4.54	197.7	-200.9
(deg.)		10^5 cm/s	MHz-mm	10^{-6} (1/°C)	10^{-8} $(1/°C)^2$	10^{-10} $(1/°C)^3$	(°C)	(°C)

TABLE 2 Cut angles, modes, displacement vector components and coupling coefficients

ϕ'/θ'	ϕ/θ	mode	A_1	A_2	A_3	δ	k
0 /122	30 /32	b	0.602	0.200	0.773	84.21	6.0
		c	-0.352	0.936	0.032	108.38	2.0
11 /62	19 /-28	b	-0.668	-0.181	0.722	105.72	5.5
		c	0.037	0.961	0.275	71.16	4.3

normal direction (deg.)	plate direction (deg.)	displacement vector components with respect to fundamental axes of quartz crystal	deflection angle of displacement vector (deg.)	coupling coefficient (%)

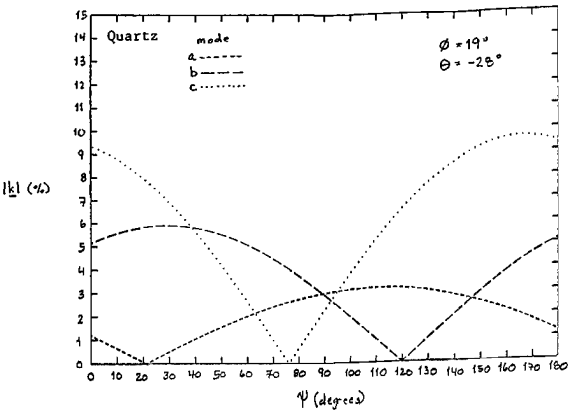


Fig.4 Lateral field coupling coefficients as functions of the angle psi with respect to the X_2" for all three modes (YXwl 19°/-28°).

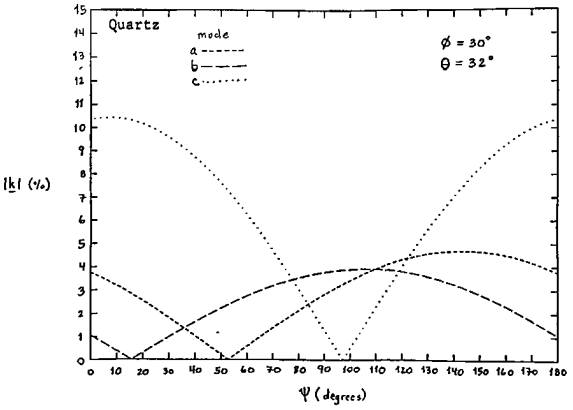


Fig.5 Lateral field coupling coefficients as functions of the angle psi with respect to the X_2" for all three thickness modes (YXwl 30°/32°).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

AN ANALYSIS OF THE IN-PLANE ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS STIFFENED BY IDENTICAL TOP AND BOTTOM QUARTZ COVER PLATES SUPPORTED BY CLIPS

Y.S. Zhou and H.F. Tiersten

Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute, Troy, NY 12180-3590

Abstract

It has been shown that both the normal and resultant in-plane acceleration sensitivities of contoured quartz resonators vanish for perfect symmetry of the resonator with respect to the support system. It has further been shown 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 which significantly reduces the linear increase in acceleration sensitivity with fabrication error 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 the normal acceleration sensitivity of this configuration was treated and a sharp minimum in the linear increase in the normal acceleration sensitivity with fabrication error was found for a particular case. In this work an analysis of the linear increase in the in-plane acceleration sensitivity for the stiffened configuration is performed. Results are presented which show the influence of differences in the properties of the clips and a mispositioning of either a clip or the mode shape for a range of thickness of the cover plates relative to the active plate.

1. Introduction

Since a perfectly symmetric contoured quartz resonator plus support system has been shown[1,2] to have zero acceleration sensitivity but is impossible to build, it is desirable to have a support configuration in which the linear increase in acceleration sensitivity with fabrication error is as small as possible. To this end we have been investigating a stiffened structure, which significantly reduces the linear increase in sensitivity with fabrication error. 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.

Recently, in work[3] on the normal acceleration sensitivity of this configuration a particular orientation and aspect ratio of the rectangular SC-cut quartz plate, which had been found in an earlier treatment of a simpler configuration, was shown to have a sharp minimum in the linear increase in the normal acceleration sensitivity with fabrication error. In this work an analysis of the linear increase in the in-plane acceleration sensitivity of the stiffened configuration is performed. In order to perform the analysis for the biasing state for this stiffened structure, the variational principle for the extension of thin anisotropic plates, in which all conditions appear as natural conditions, is extended to include the influence of the cover plates, the active region, the connecting sidewalls and the clips. The very important shearing deformation in the connecting regions and the sidewalls and the associated strain energy are included in the description. The biasing state is determined by means of our variational approximation procedure[4,5] using the extended form of the variational principle. The calculated extensional biasing states are employed in the existing perturbation equation[6] along with the mode shapes[7] of the contoured resonator to calculate the linear increase in the in-plane acceleration sensitivity with fabrication error. The calculations reveal that the linear increase in the in-plane acceleration sensitivity with fabrication error is quite sensitive to the spring constants and with proper selection can be kept quite low. As a consequence, the optimum orientation and aspect ratio found in the normal case has a resultant increase in the acceleration sensitivity with fabrication error four-to-five times lower than the square plate. This implies that the circle is an undesirable shape because of its inherent isotropy.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency obtained from the perturbation analysis[6] mentioned in the Introduction may be written in 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, and

0-7803-0476-4/92 $3.00 © 1992 IEEE

$$H_{\mu} = - \int_{V} \hat{K}^{n}_{L\gamma} g^{\mu}_{\gamma,L} \, dV , \qquad (2.2)$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g^{μ}_{γ} denotes the normalized mechanical displacement vector, and $\hat{K}^{n}_{L\gamma}$ denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g^{μ}_{γ}, and is given by

$$\hat{K}^{n}_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} , \qquad (2.3)$$

where

$$\hat{c}_{L\gamma M\alpha} = T^{1}_{LM} \delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN} E^{1}_{KN} + c_{2L\gamma KM} w_{\alpha,K} + c_{LKM\alpha} w_{\gamma,K} , \qquad (2.4)$$

and

$$T^{1}_{LM} = c_{2LMKN} E^{1}_{KN} , \quad E^{1}_{KN} = \frac{1}{2} (w_{K,N} + w_{N,K}) . \qquad (2.5)$$

The quantities T^{1}_{LM}, E^{1}_{KN} and w_K denote the static biasing stress, strain and displacement field, respectively. The coefficients c_{2LMKN} and $c_{3L\gamma M\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g^{μ}_{γ} and $\hat{f}^{\mu}$ is defined by

$$g^{\mu}_{\gamma} \equiv \frac{u^{\mu}_{\gamma}}{N_{\mu}} , \quad \hat{f}^{\mu} = \frac{\tilde{\varphi}^{\mu}}{N_{\mu}} , \quad N^{2}_{\mu} = \int \rho u^{\mu}_{\gamma} u^{\mu}_{\gamma} \, dV , \qquad (2.6)$$

where u^{μ}_{γ} and $\tilde{\varphi}^{\mu}$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\hat{K}^{\ell}_{L\gamma} = c_{2L\gamma M\alpha} u_{\alpha,M} + e_{ML\gamma} \hat{\varphi}_{,M} ,$$
$$\hat{\mathcal{D}}^{\ell}_{L} = e_{LM\gamma} u_{\gamma,M} - \epsilon_{LM} \hat{\varphi}_{,M} , \qquad (2.7)$$

$$\hat{K}^{\ell}_{L\gamma,L} = \rho \ddot{u}_{\gamma} , \quad \hat{\mathcal{D}}^{\ell}_{L,L} = 0 , \qquad (2.8)$$

subject to the appropriate boundary conditions, and ρ is the mass density. Equations (2.7) are the linear piezoelectric constitutive relations and (2.8) are the stress equations of motion and charge equation of electrostatics, respectively. The upper cycle notation for many dynamic variables and the capital Latin and lower case Greek index notation is being employed for consistency with Ref.6 as is the remainder of the notation in this section.

The substitution of (2.3) in (2.2) yields

$$H_{\mu} = - \int_{V} \hat{c}_{L\gamma M\alpha} g^{\mu}_{\alpha,M} g^{\mu}_{\gamma,L} \, dV . \qquad (2.9)$$

Since g^{μ}_{α} denotes the normalized mode shape in the contoured resonator and from (2.4) $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_{μ} can be evaluated when the resonator mode shape and biasing state are known.

3. Preliminary Considerations

Since the plates are thin, for in-plane acceleration the equations of anisotropic plane stress apply. For anisotropic plane stress the stresses K_{2L} vanish and only K_{AB} exist, where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2. The equations for anisotropic plane stress may be written in the form

$$K_{AB,A} - \rho a_{B} = 0 , \qquad (3.1)$$

where a_B denotes the acceleration in the plane of the plate and the anisotropic constitutive equations for plane stress may be written in the form

$$K_{AB} = \gamma_{ABCD} E_{CD} , \quad E_{CD} = \frac{1}{2} (w_{C,D} + w_{D,C}) , \qquad (3.2)$$

and γ_{ABCD} are Voigt's anisotropic plate elastic constants, which are given by

$$\gamma_{RS} = C_{RS} - C_{RW} C^{-1}_{WV} C_{VS} , \quad R,S = 1,3,5; \quad W,V = 2,4,6 . \qquad (3.3)$$

From the relaxation of the stresses K_{2L}, we have

$$E_{W} = - C^{-1}_{WV} C_{VS} E_{S} , \qquad (3.4)$$

which enables us to obtain all the strains E_{KL} when a solution is known as well as the rotation Ω_{13} from the relation

$$\Omega_{13} = \frac{1}{2} (w_{3,1} - w_{1,3}) . \qquad (3.5)$$

However, the rotations Ω_{2A} cannot be obtained from a solution because of the relaxation of the K_{2L}. Nevertheless, the flexural type plate rotations Ω_{2A} are taken to vanish in the case of uncoupled anisotropic plane stress, which is the kinematic uncoupling assumption for anisotropic plane stress. Accordingly, we have

$$\Omega_{2A} = 0 , \qquad (3.6)$$

and now with (3.2) - (3.5), (3.6) and

$$w_{K,L} = E_{KL} + \Omega_{LK} , \qquad (3.7)$$

we have the three-dimensional displacement gradients for the extensional biasing deformation field, which is required for the evaluation of $\hat{c}_{L\gamma M\alpha}$ in (2.4).

A plan view and cross-section of the stiffened structure is shown in Fig.1 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. The static solution for the biasing state is referred to the unprimed coordinate system only.

4. Analysis of Sidewall

A cross-section of the sidewall is shown in Fig.2. On account of the stubbiness of the cross-section of a sidewall, we model its resistance to the normal displacements w as a variable shear strip. In accordance with the assumption of plane stress discussed in Section 3, the cover plates and active plate have a uniform extensional stress distribution as shown in the free-body diagrams in Fig.3. The shearing stresses are shown at the mid-planes of the connecting materials and the X-axis is directed up from the mid-plane of the region connecting the active plate with the top cover plate. From the free-body diagrams of the symmetric active plate and the top cover plate, respectively, we obtain

$$K_{nn} 2h = 2\tau_o^a t_f, \quad [\hat{\underset{\sim}{K}}_{nn}] 2\hat{h} = \tau_o^C t_f, \tag{4.1}$$

where

$$[\hat{\underset{\sim}{K}}_{nn}] = \hat{K}_{nn}^{+} - \hat{K}_{nn}^{-}. \tag{4.2}$$

Similarly, from the appropriate free-body diagrams at arbitrary positions in the top cover plate and active plate, respectively, we obtain

$$\tau_o^C t_f - [\hat{\underset{\sim}{K}}_{nn}](X - h_f) = \hat{\tau} t_f,$$
$$\tau_o^a t_f - K_{nn}(-X - h_f) = \tau t_f. \tag{4.3}$$

Clearly, the shearing stress-strain relations in the respective regions are given by

$$\tau_o = \mu^f \frac{dw}{dX}, \quad \hat{\tau} = \hat{\mu} \frac{d\hat{w}}{dX}, \quad \tau = \mu \frac{dw}{dX}. \tag{4.4}$$

Substituting from (4.1) and (4.3) into (4.4), integrating and applying appropriate continuity conditions[8], we obtain

$$\hat{w}^C = [\hat{\underset{\sim}{K}}_{nn}] G^C + w_o^{fC}, \quad w^a = -K_{nn} G^a + w_o^{fa}, \tag{4.4}$$

where

$$G^C = \frac{1}{\hat{\mu} t_f}\left(\frac{3}{2}\hat{h}^2 + 2\hat{r}_\mu \hat{h} h_f\right),$$
$$G^a = \frac{1}{\mu t_f}\left(\frac{h^2}{2} + r_\mu h h_f\right), \tag{4.5}$$
$$\hat{r}_\mu = \hat{\mu}/\mu^f, \quad r_\mu = \mu/\mu^f, \tag{4.6}$$

and $\hat{w}^C$ and w^a are the shearing displacements at the mid-planes of the top cover and active plates, respectively. The shearing displacements w_o^{fC} and w_o^{fa} just above and below the mid-plane of the region connecting the top cover plate with the active plate should be equal, but will be treated as a constraint condition with a Lagrange multiplier in the variational principle in Section 5. It is not necessary to treat the bottom cover plate explicitly because of the assumed symmetry of the configuration.

The foregoing gives the relative normal displacement between the active plate and the cover plates for the sidewalls. It is clear from the geometry that the relative tangential displacement between the active plate and the cover plates for the sidewall may be taken to vanish because the sidewall is essentially rigid in the tangential direction. This tangential condition will also be introduced as a constraint condition with a Lagrange multiplier in the variational principle in Section 5.

The energy density per unit length of the sidewall will be needed for the variational principle in the next section. To this end we write

$$U^e = U^{CT} + U^{fT} + U^a + U^{fB} + U^{CB} \tag{4.7}$$

where U^e is the energy density per unit length of the sidewall and the meaning of the individual terms in (4.7) is clear from the superscripts. Typical expressions for the energy densities are given by

$$U^{CT} = t_f \frac{\hat{\mu}}{2} \int \left(\frac{d\hat{w}}{dX}\right)^2 dX,$$
$$U^a = t_f \frac{\mu}{2} \int \left(\frac{dw}{dX}\right)^2 dX, \tag{4.8}$$

and the others are clear from the notation.

5. Unconstrained Variational Principle for Extension of Box Structure

In this section we present the unconstrained variational principle for static extension of the box structure, a plan view and cross-section of which is shown in Fig.1. This principle is required for the determination of the extensional biasing state for the configuration shown in Fig.1 using the approximation procedure we employ. Before proceeding with the variational principle it is advisable to discuss the configuration in some detail. As already noted, the structure consists of the active biconvex quartz resonator, which is attached to identical top and bottom quartz cover plates by means of small sidewalls around the periphery. The cover plates have identical holes in them in order to provide access to the active resonator. The holes are identical as are the cover plates in order to achieve the desired symmetry. The mounting clips are attached to the cover plates without touching the active plate.

For the purpose of obtaining the variational principle we do not consider the rectangle shown in the plan view of Fig.1, but rather the asymmetric plan view shown in Fig.4, in which the outside edge of the cover plates consists of a smooth curve with outward unit normal N_A. In addition, the cover plates have an inner edge of an opening consisting of a curve similar to the outer curve. Portions of the outside edge of the cover plates are supported while others are left free. Since

the supports are of spring type, the entire outside edges of the cover plates may be considered to be subject to prescribed tractions, which are natural type conditions. Hence, in accordance with our usual nomenclature the outside edges of the cover plates are denoted $\hat{c}_o^N$. Since the inside edges of the cover plates are subject to prescribed tractions, which vanish in the problem to be treated, the inside edges of the cover plates are denoted $\hat{c}_i^N$.

The active plate is attached to the identical top and bottom cover plates by the connecting material of thickness $2h_f$ that makes up a portion of the sidewall, as discussed in Section 4. The curve representing the sidewall in the variational principle will be denoted c^a. The sidewall both contains energy and has constraint conditions that will be introduced as zeros times Lagrange multipliers, as noted in Section 4.

As usual, in order that all variations may be regarded as independent when constraints exist, we introduce each constraint as a zero times a Lagrange multiplier[9,10] in the variational principle. Then, in accordance with the above discussion, the unconstrained variational principle for static extension of the box structure may be written in the form

$$\delta\Big[2h\int_{S^a}(\bar{f}_A^a w_A - U^a)dS + 2(2\hat{h})\int_{\hat{S}^C}(\bar{f}_A^C\hat{w}_A - \hat{U}^C)\,dS - \oint_{c^a}[U^e + 2\lambda^n(w_o^{fC} - w_o^{fa})]\,ds + 2(2\hat{h})\oint_{c^N=\hat{c}_o+\hat{c}_i^N}(\bar{t}_s\hat{w}_s + \bar{t}_n\hat{w}_n)\,ds + 2h\oint_{c^a}\lambda^S(\hat{w}_S^C - w_S^a)\,ds\Big] = 0, \qquad (5.1)$$

where S^a and $\hat{S}^C$ denote the areas of the active plate and the cover plates, respectively, $\bar{f}_A^a$ and $\bar{f}_A^C$ denote the body forces per unit area applied to the active plate and cover plates, respectively, $\bar{t}_S$ and $\bar{t}_n$ denote the tractions applied tangential and normal to the cover plates, respectively, and the $\hat{w}_S$ and $\hat{w}_n$ are the corresponding displacements,

$$U^a = \frac{1}{2}\gamma_{ABCD}w_{A,B}w_{C,D}\,,$$
$$\hat{U}^C = \frac{1}{2}\gamma_{ABCD}\hat{w}_{A,B}\hat{w}_{C,D}\,, \qquad (5.2)$$

are the energy densities of the active and cover plates, respectively, and λ^n and λ^S are the Lagrange undetermined multipliers for the normal and tangential constraint conditions of the sidewall. From the analysis of the sidewall in Section 4 for later use in this section we may write[8]

$$\delta U^e + 2\lambda^n(\delta w_o^{fC} - \delta w_o^{fa}) + 2\delta\lambda^n(w_o^{fC} - w_o^{fa}) = \delta\mathscr{W} + 2[\underset{\sim}{\hat{K}}_{nn\sim}]\,2\hat{h}\,\delta\hat{w}^C - K_{nn}2h\,\delta w^a\,, \qquad (5.3)$$

where

$$\lambda^n = \frac{1}{2}(\tau_o^a + \tau_o^C),\quad \delta\mathscr{W} = (K_{nn}h - [\underset{\sim}{\hat{K}}_{nn\sim}]2\hat{h})(\delta w_n^a + \delta\hat{w}_n^C + \delta K_{nn}G^a - \delta[\underset{\sim}{\hat{K}}_{nn\sim}]\,G^C) + (\hat{w}_n^C - w_n^a - [\underset{\sim}{\hat{K}}_{nn\sim}]\,G^C - K_{nn}G^a)(\delta K_{nn}h + \delta[\underset{\sim}{\hat{K}}_{nn\sim}]\,2\hat{h})\,. \qquad (5.4)$$

Taking the variations in (5.1), integrating by parts and using the surface divergence theorem[11], we obtain

$$2h\int_{S^a}(\bar{f}_A^a + K_{BA,B})\delta w_A\,dS + 2(2\hat{h})\int_{\hat{S}^C}(\bar{f}_A^C + \hat{K}_{BA,B})\delta\hat{w}_A\,dS + 2h\oint_{c^a}[(-K_{nn}+K_{nn})\delta w_n^a + (-K_{ns} - \lambda^s)\delta w_s^a + \delta\lambda^s(\hat{w}_s^C - w_s^a)]\,ds + 2(2\hat{h})\oint_{c_o^N}\Big[\Big(-\hat{K}_{nn}^+ + \bar{t}_n\Big)\delta\hat{w}_n^C + \Big(-\hat{K}_{ns} + \frac{h}{\hat{h}}\lambda^s + \bar{t}_s\Big)\delta\hat{w}_s^C\Big]ds + 2(2\hat{h})\oint_{c_i^N}[(\bar{t}_n - \hat{K}_{nn})\delta\hat{w}_n^C + (\bar{t}_s - \hat{K}_{ns})\delta\hat{w}_s^C]\,ds - \oint_{c^a}\delta\mathscr{W}\,ds = 0\,. \qquad (5.5)$$

Since on account of the use of Lagrange multipliers with the constraint conditions all variations may be regarded as independent, we obtain

$$K_{BA,B} + \bar{f}_A^a = 0,\quad \hat{K}_{BA,B} + \bar{f}_A^C = 0, \qquad (5.6)$$

$$\lambda_s = -K_{ns},\ \hat{w}_s^C - w_s^a = 0,\ \hat{K}_{nn}^+ = \bar{t}_n,$$
$$2\hat{h}\hat{K}_{ns} + hK_{ns} = 2\hat{h}\bar{t}_s,\ K_{nn}h - [\underset{\sim}{\hat{K}}_{nn\sim}]\,2\hat{h} = 0,$$
$$\hat{w}_n^C - w_n^a - [\underset{\sim}{\hat{K}}_{nn\sim}]\,G^C - K_{nn}G^a = 0,\ \text{on } c_o, \qquad (5.7)$$

$$\hat{K}_{nn} = \bar{t}_n,\ \hat{K}_{ns} = \bar{t}_s \text{ on } c_i, \qquad (5.8)$$

which consist of the differential equations in the active plate and the cover plates in (5.6), the natural- and constraint-type edge conditions and the expressions for the tangential Lagrange multiplier λ_s and $\hat{K}_{nn}^+$ along the sidewall in (5.7) and the natural edge conditions along the holes in the cover plates in (5.8). The expression for the normal Lagrange multiplier λ_n has already been given in $(5.4)_1$ with (4.1).

Thus it is clear that the variational principle (5.1) with unconstrained variations yields the differential equations (5.6) for both the active plate and the cover plates, the natural edge conditions $(5.7)_{4,5}$ and (5.8) and the constraint-type edge conditions $(5.7)_{2,6}$. In addition, the Lagrange multipliers λ_s and λ_n have been expressed in terms of the variables of the

problem. Consequently, $\delta\lambda_s$ and $\delta\lambda_n$ may be obtained from $(5.7)_1$ and $(5.4)_1$ with (4.1) and substituted[8] in (5.5), which may then be used to obtain an approximate solution for the extensional biasing state without any apriori conditions[12,13] on the approximating functions.

6. Extension of Box Structure Supported by Clips

A plan view and cross-section of the box structure is shown in Fig.1 along with the unprimed coordinate system, which is relevant in this section. From Eqs.(5.6) for acceleration a_B in the plane, the equations of plane stress for the active plate and the cover plates take the respective forms

$$K_{AB,A} - \rho a_B = 0, \quad \hat{K}_{AB,A} - \hat{\rho} a_B = 0. \tag{6.1}$$

Since the inner edges around the openings in the cover plates are free and the outer edges are free except at the clips, where they are spring supported, and the active plate is attached to the cover plates by means of sidewalls around the periphery, from Eqs.(5.7) and (5.8) the edge conditions take the form

$$\hat{K}_{11} = 0, \ \hat{K}_{13} = 0 \text{ at } X_1 = \pm a_c,$$

$$\hat{K}_{33} = 0, \ \hat{K}_{31} = 0 \text{ at } X_3 = \pm b_c,$$

$$2\hat{h}\hat{K}_{13} + hK_{13} = 0, \ 2\hat{h}\hat{K}_{11} + hK_{11} = 0, \text{ at } X_1 = \pm a, \text{ E.C.},$$

$$2\hat{h}\hat{K}_{13} + hK_{13} = \mp 2\hat{h}k_{1s}\hat{w}_3, \ 2\hat{h}\hat{K}_{11} + hK_{11} = \mp 2\hat{h}k_{1n}\hat{w}_1,$$

$$\text{at } X_1 = \pm a, \text{ A.C.},$$

$$2\hat{h}\hat{K}_{31} + hK_{31} = 0, \ 2\hat{h}\hat{K}_{33} + hk_{33} = 0, \text{ at } X_3 = \pm b, \text{ E.C.},$$

$$2\hat{h}\hat{K}_{31} + hK_{31} = \mp 2\hat{h}k_{3s}\hat{w}_1, \ 2\hat{h}\hat{K}_{33} + hK_{33} = \mp 2\hat{h}k_{3n}\hat{w}_3,$$

$$\text{at } X_3 = \pm b, \text{ A.C.},$$

$$\hat{w}_1 - w_1 = \mp G_1^{C}\hat{K}_{11} \pm G_1^{a}K_{11} \text{ at } X_1 = \pm a, \ \hat{w}_3 - w_3 =$$

$$\mp G_3^{C}\hat{K}_{33} \pm G_3^{a}K_{33} \text{ at } X_3 = \pm b, \text{ E.C.},$$

$$\hat{w}_1 - w_1 \pm (G_1^{C}\hat{K}_{11} - G_1^{a}K_{11}) + G_1^{C}k_{1n}\hat{w}_1 = 0$$

$$\text{at } X_1 = \pm a, \text{ A.C.},$$

$$\hat{w}_3 - w_3 \pm (G_3^{C}\hat{K}_{33} - G_3^{a}K_{33}) + G_3^{C}k_{3n}\hat{w}_3 = 0$$

$$\text{at } X_3 = \pm b, \text{ A.C.},$$

$$\hat{w}_3 - w_3 = 0 \text{ at } X_1 = \pm a, \ \hat{w}_1 - w_1 = 0 \text{ at } X_3 = \pm b, \tag{6.2}$$

where k_{1s}, k_{1n}, k_{3s} and k_{3n} denote the shearing and normal spring constants of the clips along the edges normal to X_1 and X_3, respectively, and we have introduced the convention that E.C. and A.C. refer to the same edge conditions "Except at the Clips" and "At the Clips," respectively.

Since as already indicated the problem defined in (6.1) and (6.2) cannot be solved exactly, an approximation procedure is employed. To this end we first transform the inhomogeneities from the differential equations (6.1) into the edge conditions by writing

$$w_A = w_A^* + w_A^S, \ \hat{w}_A = \hat{w}_A^* + \hat{w}_A^S, \tag{6.3}$$

where w_A^* and $\hat{w}_A^*$ are polynomials in X_1 and X_3 satisfying the differential equations (6.1) and depending on the acceleration a_B, as well as satisfying the edge conditions $\hat{w}_1^* - w_1^* = 0$ at $X_3 = \pm b$. The substitution of (6.3) into (6.1) yields

$$\gamma_{ABCD} w_{C,DA}^S = 0, \ \gamma_{ABCD}\hat{w}_{C,DA}^S = 0. \tag{6.4}$$

The further substitution of (6.3) with (3.2) into (6.2) yields the transformed edge conditions, which for brevity we do not write here[8].

As a solution of (6.4) we write

$$w_C^S = \sum_m \beta_{Cm} e^{i(\nu_m X_1 + \eta_m X_3)},$$

$$\hat{w}_C^S = \sum_m \hat{\beta}_{Cm} e^{i(\hat{\nu}_m X_1 + \hat{\eta}_m X_3)}, \tag{6.5}$$

where $\nu_m = N_m\pi/2\sigma a = \hat{\nu}_m$, N_m denotes integers not necessarily spaced by 1 and σ is chosen to be an irrational number (here π) in order that neither $\cos \nu_m a$ or $\sin \nu_m a$ vanish. The substitution of (6.5) into (6.4) yields

$$\left[\gamma_{11}\nu_m^2 + 2\gamma_{15}\nu_m\eta_m + \gamma_{55}\eta_m^2\right]\beta_{1m} + \left[\gamma_{35}\eta_m^2 + (\gamma_{13}+\gamma_{55})\eta_m\nu_m + \gamma_{15}\nu_m^2\right]\beta_{3m} = 0,$$

$$\left[\gamma_{15}\nu_m^2 + (\gamma_{13}+\gamma_{55})\eta_m\nu_m + \gamma_{35}\eta_m^2\right]\beta_{1m} + \left[\gamma_{33}\eta_m^2 + 2\gamma_{35}\eta_m\nu_m + \gamma_{55}\nu_m^2\right]\beta_{3m} = 0, \tag{6.6}$$

for each ν_m. Eqs.(6.6) constitute a system of two linear homogeneous algebraic equations in β_{1m} and β_{3m}, which yields nontrivial solutions when the determinant of the coefficients vanishes. The vanishing of the determinant yields four roots $\tilde{\eta}_m^{(p)}$ $(p = 1,2,3,4)$ for each ν_m with four sets of amplitude ratios $\beta_{1m}^{(p)}:\beta_{3m}^{(p)}$. The roots and ratios are independent of m, and clearly from (6.4) and (6.5) we have $\hat{\eta}_m = \eta_m$ and $\hat{\beta}_{Cm} = \beta_{Cm}$ since $\hat{\nu}_m = \nu_m$. Hence, the solution may be written in the form

$$w_A^S = \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} D_m^{(p)}\beta_A^{(p)} e^{i\eta_m^{(p)} X_3},$$

$$\hat{w}_A^S = \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} \hat{D}_m^{(p)}\beta_A^{(p)} e^{i\eta_m^{(p)} X_3}, \tag{6.7}$$

where the $D_m^{(p)}$ and $\hat{D}_m^{(p)}$ are amplitude coefficients still to be determined. By satisfying the homogeneous edge conditions $\hat{w}_1^S - w_1^S = 0$ at $X_3 = \pm b$ for each m, we obtain[8] $\hat{D}_m^3$ and $\hat{D}_m^4$ in terms of $D_m^{(p)}$, $\hat{D}_m^{(1)}$ and $\hat{D}_m^{(2)}$, which reduces the number of unknown coefficients from 8 to 6 for each m and permits us to write the approximate solution functions in the form

$$w_A^S = \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} \tilde{D}_m^{(p)} \beta_A^{(p)} e^{i\eta_m^{(p)} X_3},$$

$$\hat{w}_A^S = \sum_m e^{i\nu_m X_1} \sum_{\ell=1}^{6} \hat{D}_m^{(\ell)} \sum_{p=1}^{4} \hat{g}_{\ell m}^{(p)} \beta_A^{(p)} e^{i\eta_m^{(p)} X_3}, \quad (6.8)$$

where $\tilde{D}_m^{(p)} = D_m^{(p)}$, $\tilde{D}_m^{(5)} = \hat{D}_m^{(1)}$, $\tilde{D}_m^{(6)} = \hat{D}_m^{(2)}$.

Since the solution functions in (6.3) satisfy the differential equations (6.1), only the edge conditions which have not been satisfied remain in the form of the variational principle in (5.5). Substituting from (6.3) and (6.8) and the homogeneous edge conditions which have been satisfied exactly by each term in the solution sums into (5.5) and employing $(5.4)_2$, $(5.7)_1$ and the fact that $\bar{t}_s$ and $\bar{t}_n$ vanish on the inside edges and on the outside edges of the cover plates where there are no clips, and the spring conditions, which appear in the conditions denoted A.C. in (6.2), where there are clips, we obtain[8]

$$\begin{aligned}
&4\hat{h} \int_{-b_c}^{b_c} \left[(\hat{K}_{11}\delta\hat{w}_1 + \hat{K}_{13}\delta\hat{w}_3) \right]_{X_1=-a_c}^{X_1=a_c} dX_3 + \\
&4\hat{h} \int_{-a_c}^{a_c} \left[(\hat{K}_{33}\delta\hat{w}_3 + \hat{K}_{31}\delta\hat{w}_1) \right]_{X_3=-b_c}^{X_3=b_c} dX_1 \\
&- \int_{-b}^{b} \Big[2h(\hat{w}_3 - w_3)\delta K_{13} + (4\hat{h}\hat{K}_{13} + 2hK_{13})\delta\hat{w}_3 \\
&+ (K_{11}h + 2\hat{h}\hat{K}_{11})\delta(\hat{w}_1 + w_1 \pm G_1^a K_{11} \pm G_1^{C\hat{}} K_{11})_{X_1=a} \\
&+ (\hat{w}_1 - w_1 \mp G_1^a K_{11} \pm G_1^{C\hat{}} K_{11})\delta(hK_{11} - 2\hat{h}\hat{K}_{11}) \Big]_{X_1=-a} dX_3 \\
&- \int_{-a}^{a} \Big[(4\hat{h}\hat{K}_{31} + 2hK_{31})\delta\hat{w}_1 + (K_{33}h + 2\hat{h}\hat{K}_{33})\delta(\hat{w}_3 + w_3 \\
&\pm G_3^a K_{33} \pm G_3^{C\hat{}} K_{33}) + (\hat{w}_3 - w_3 \mp G_3^a K_{33} \\
&\pm G_3^{C\hat{}} K_{33})\delta(hK_{33} - 2\hat{h}\hat{K}_{33}) \Big]_{X_3=-b}^{X_3=b} dX_1 \\
&- \int_{f-t}^{f+t} \Big[\pm 4\hat{h}k_{1s}\hat{w}_3\delta\hat{w}_3 + (K_{11}h + 2\hat{h}\hat{K}_{11})G_1^C k_{1n}\delta\hat{w}_1 \\
&+ 2\hat{h}k_{1n}\hat{w}_1\delta(\pm w_1 + G_1^a K_{11} + G_1^{C\hat{}} K_{11}) + 2\hat{h}k_{1n}(\pm w_1 \\
&+ G_1^a K_{11} - G_1^C K_{11})\delta\hat{w}_1 + G_1^C k_{1n}\hat{w}_1\delta(hK_{11} \\
&- 2\hat{h}\hat{K}_{11}) \Big]_{X_1=-a,\, f=f_3}^{X_1=a,\, f=f_1} dX_3 \\
&- \int_{f-t}^{f+t} \Big[\pm 4\hat{h}k_{3s}\hat{w}_1\delta\hat{w}_1 + (K_{33}h + 2\hat{h}\hat{K}_{33})G_3^C k_{3n}\delta\hat{w}_3 \\
&+ 2\hat{h}k_{3n}\hat{w}_3\delta(\pm w_3 + G_3^a K_{33} + G_3^{C\hat{}} K_{33}) \\
&+ 2\hat{h}k_{3n}(\pm w_3 + G_3^a K_{33} - G_3^C \hat{K}_{33})\delta\hat{w}_3 + \\
&+ G_3^C k_{3n}\hat{w}_3\delta(hK_{33} - 2\hat{h}\hat{K}_{33}) \Big]_{X_3=-b,\, f=f_4}^{X_3=b,\, f=f_2} dX_1 = 0,
\end{aligned} \quad (6.9)$$

where f_i $(i = 1,2,3,4)$ denotes the mispositioning of a clip as shown in Fig.1. The notation $\Big]_{X_1=-a,\, f=f_3}^{X_1=a,\, f=f_1}$ means the quantity evaluated at a minus the quantity evaluated at -a, with the attendant value of f_i used for the integral evaluated at the appropriate position. The δK_{AB}, $\delta\hat{K}_{AB}$ are obtained from (3.2) with w_C^S and $\hat{w}_C^S$ instead of w_C and $\hat{w}_C$ since the variation of prescribed quantities vanishes. Substituting from (6.3) with (6.8) into (6.9) and performing the integrations, we obtain

$$\sum_{n=-N}^{N} \sum_{k=1}^{6} \left[\sum_{m=-N}^{N} \sum_{\ell=1}^{6} a_{mn}^{\ell k} \tilde{D}_m^{(\ell)} + b_n^{(k)} \right] \delta\tilde{D}_n^{(k)} = 0. \quad (6.10)$$

The $a_{mn}^{\ell k}$ and b_n^k are very lengthy expressions, which provide no additional understanding and were left in the computer where they were generated. Since the $\delta\tilde{D}_n^{(k)}$ are arbitrary, we obtain

$$\sum_{m=-N}^{N} \sum_{\ell=1}^{6} a_{mn}^{\ell k} \tilde{D}_m^{(\ell)} = -b_n^{(k)} \quad \begin{matrix} k = 1 - 6, \\ n = \pm 1, \pm 2, \ldots \pm N, \end{matrix} \quad (6.11)$$

which constitute 12N inhomogeneous linear algebraic equations in the 12N unknowns $\tilde{D}_m^{(\ell)}$, the inversion of which gives the approximate solution.

7. The Mode Shape in the Contoured Resonator

A schematic diagram of the contoured crystal resonator along with the associated coordinate system is shown in Fig.5. 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]

$$u_{1n} = \sin\frac{n\pi X_2}{2h}\, u_{noo} e^{i\omega_{noo} t}, \quad n = 1,3,5, \ldots, \quad (7.1)$$

where

$$u_{noo} = e^{-\alpha_n(\hat{X}_1^2/2)-\beta_n(\hat{X}_3^2/2)}, \qquad (7.2)$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3M'_n}, \quad \beta_n^2 = \frac{n^2\pi^2\hat{c}(1)}{8Rh_o^3P'_n}, \qquad (7.3)$$

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. From Eq.(97) of Ref.7 in this work we have

$$\hat{X}_A = R_{AB}(X_B - d_B), \qquad (7.4)$$

where, from Eq.(98)_1 of Ref.7

$$R_{AB} = \begin{Vmatrix} \cos\hat{\beta}_n & -\sin\hat{\beta} \\ \sin\hat{\beta}_n & \cos\hat{\beta}_n \end{Vmatrix}, \qquad (7.5)$$

and $\hat{\beta}$ is defined in Eq.(98)_2 of Ref.7. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2\hat{c}(1)}{4h_o^2\rho}\left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \times \left(\sqrt{\frac{M'_n}{\hat{c}(1)}} + \sqrt{\frac{P'_n}{\hat{c}(1)}}\right)\right], \qquad (7.6)$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.7.

In the in-plane case being considered in this work only the dominant displacement field in (7.1) is needed. Consequently, the equivalent trapped energy mode employed in Refs.1 and 3 is not required here.

8. Small Offset of Mode Center

In this section we obtain a very useful result for the limiting case of small offset of the mode center. From Fig.1 it is clear that we may write the transformation

$$X_A = d_A + X'_A, \qquad (8.1)$$

and from (6.3) and (6.8), we may write

$$w_A = w_A^* + \sum_m e^{i\nu_m X_1} \sum_{p=1}^{4} \hat{D}_m^{(p)}\beta_A^{(p)} e^{i\eta_m^{(p)}X_3}, \qquad (8.2)$$

where w_A^* are polynomials in X_1 and X_3. The substitution of (8.1) in (8.2) enables us to write

$$w_A = w_A^*(d_B, X'_B) + \sum_m h^{(m)} e^{i\nu_m X'_1} \sum_{p=1}^{4} \hat{D}_m^{(p)}\beta_A^{(p)}\bar{g}_p^{(m)} e^{i\eta_m^{(p)}X'_3}, \qquad (8.3)$$

where

$$h^{(m)} = e^{i\nu_m d_1}, \quad \bar{g}_p^{(m)} = e^{i\eta_m^{(p)}d_3}. \qquad (8.4)$$

Since the perturbation integral in (2.9) vanishes for $d_A = 0$, the substitution of (8.3) with (8.4) into (2.9) and expansion to first order in d_A yields, with the aid of (2.1),

$$(\Delta\omega/\omega)^D = \Gamma^D = C_A^D d_A, \qquad (8.5)$$

where the superscript D(= 1,3) represents the 1 and 3 components of the in-plane acceleration sensitivity vector. The C_A^D are complicated expressions which are not terribly revealing and have been calculated.

Since it is of interest to know the maximum value of the magnitude of the in-plane acceleration sensitivity $\Gamma_{(p)}$ for a given $|\underset{\sim}{d}|$, we write

$$\Gamma_{(p)}^2 = \Gamma^D\Gamma^D = C_A^D d_A C_B^D d_B = C_A^D C_B^D Q_A Q_B d^2, \qquad (8.6)$$

where

$$Q_1 = \cos\theta, \quad Q_3 = \sin\theta, \qquad (8.7)$$

and θ is the angle between the X_1-axis and $\underset{\sim}{d}$. Since the maximum value of the magnitude of the in-plane acceleration sensitivity per unit error d is the quantity of actual interest here, we define $\gamma_{(p)}$ by

$$\gamma_{(p)}^2 = \Gamma_{(p)}^2/d^2, \qquad (8.8)$$

which may be maximized by

$$\partial\gamma_{(p)}^2/\partial\theta = 0, \quad \partial^2\gamma_{(p)}^2/\partial\theta^2 < 0. \qquad (8.9)$$

The result is given by

$$\gamma_{(p)}^2\Big)_{max} = \frac{1}{2}\left[C_A^D C_A^D + \Delta_{(p)}\right], \qquad (8.10)$$

where

$$\Delta_{(p)} = \sqrt{(C_1^D D_1^D - C_3^D C_3^D)^2 + 4(C_1^D C_3^D)^2}. \qquad (8.11)$$

The foregoing includes the influence of d on the in-plane acceleration sensitivity only. Since the influence of d on the resultant acceleration sensitivity including the normal acceleration sensitivity is of primary practical importance, we extend the treatment to include the influence of d on the normal acceleration sensitivity which was calculated in earlier work[3]. This is accomplished simply by writing

$$(\Delta\omega/\omega)^L = \Gamma^L = C_A^L d_A, \qquad (8.12)$$

in place of (8.5) because D takes the values 1 and 3 only and L takes the values 1, 2 and 3. We then define the magnitude of the resultant acceleration sensitivity $\Gamma_{(r)}$ for a given $|\underset{\sim}{d}|$ by

$$\Gamma^2_{(r)} = \Gamma^L\Gamma^L = C^L_A d_A C^L_B d_B = C^L_A C^L_B Q_A Q_B d^2. \quad (8.13)$$

As before, we define the magnitude of the resultant acceleration sensitivity per unit error d by $\gamma_{(r)}$ thus

$$\gamma^2_{(r)} = \Gamma^2_{(r)}/d^2. \quad (8.14)$$

Clearly then $\gamma^2_{(r)})_{max}$ is given by

$$\gamma^2_{(r)})_{max} = \frac{1}{2}\left[C^L_A C^L_A + \Delta_{(r)}\right], \quad (8.15)$$

where

$$\Delta_{(r)} = \sqrt{(C^L_1 C^L_1 - C^L_3 C^L_3)^2 + 4(C^L_1 C^L_3)^2}. \quad (8.16)$$

Calculations of $\gamma_{(p)})_{max}$ and $\gamma_{(r)})_{max}$ are presented and discussed in the next section.

Since the direction and magnitude of the mispositioning of any clip is fixed, the C_f compose vectorially, i.e.,

$$C^2_f = C^L_f C^L_f, \quad (8.17)$$

where analogous to (8.5), we have

$$(\Delta\omega/\omega)^L_{f_i} = C^L_{f_i} f_i, \quad (\Delta\omega/\omega)_{f_i} = C_{f_i} f_i. \quad (8.18)$$

9. Calculated Results

From Section 7 we now know g^μ_γ and from Sections 2 and 6 we know $\hat{c}_{L\gamma M\alpha}$ for in-plane acceleration, i.e., extension of the box structure. 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 C^D_A defined in (8.5), $\gamma_{(p)})_{max}$ in (8.10), $\gamma_{(r)})_{max}$ in (8.15) and $C^D_{f_i}$ and C_{f_i} defined in (8.18). In the calculation of $\gamma_{(r)}$ and C_{f_i} we employ the calculations of C^2_A and $C^2_{f_i}$ from Ref.3. 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. In earlier work[1,2] C was referred to as an offset coefficient. In more recent work[3] the name was changed to a mispositioning coefficient. The symbols $\gamma_{(p)}$, $\gamma_{(r)}$ and $|C^L_A|$ are generically represented by C, which denotes the mispositioning coefficient for the mode center, and C_{f_i} denotes the mispositioning coefficient for the ith clip shown in Fig.1. 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.6.

Calculations of the different C and C_{f_i} have been performed using the known values of the second order[14] and the third order[15] elastic constants of quartz. In Fig. 7 we have plotted the mispositioning coefficients for acceleration in the X_1-direction as a function of the ratio of the thickness of a cover plate to the active plate for an SC cut with an orientation of the support rectangle $\beta = -50^\circ$ and an aspect ratio $a/b = 2.0$. It can be seen from the figure that the mispositioning coefficients vary very little as the thickness of the cover plates increases. The figure also shows that the mispositioning of the mode center is more deleterious than the mispositioning of the clips. The orientation of the support rectangle $\beta = -50^\circ$ is used because it was found in earlier work[1] that for a simply-supported active plate a sharp minimum in C was found at this orientation for an aspect ratio $a/b = 2.0$. In addition, it was shown[3] recently that the sharp minimum at this orientation and aspect ratio existed for normal acceleration of the box structure. This orientation of the support rectangle of an SC cut is maintained in all calculations presented in this work. The ratio $a_c/a = 0.5$ for the openings in the cover plates is also maintained except for the last figure. The mispositioning coefficients for acceleration in the X_3-direction for the same case have been plotted in Fig. 8, which also shows that they change very little with the thickness of the cover plates. The curves exhibit essentially the same type of behavior as in Fig. 7.

In Fig. 9 we have plotted a number of resultant mispositioning coefficients as a function of the ratio of the thickness of a cover plate to the active plate for $a/b = 2.0$. The only curve obtained completely from the calculations performed in this work is for $\gamma_{(p)}$ because it depends only on in-plane calculations. All the other curves depend also on the calculations for the normal case[16] performed in Ref.3. The curve for $|C^2|$ is taken completely from Fig.7 of Ref.3 and is for normal acceleration only. This curve, which contains low values throughout, is the main reason this support orientation and aspect ratio are important for minimizing the influence of fabrication error on acceleration sensitivity. The $\gamma_{(r)}$, $|C_{f1}|$ and $|C_{f_2}|$ curves are the resultant curves that have practical significance. They show that for $\hat{h}/h$ around 2 the influence of mispositioning errors

in the location of the mode center and clips are all quite low. The results of similar calculations for a/b = 1.8 are shown in Fig.10, which indicates that this aspect ratio, although not as good as a/b = 2.0, is still quite good. Figure 11 presents the same results for a square plate and indicates that the increase in acceleration sensitivity with fabrication error for a square plate is much worse than for a/b = 2.0 or 1.8. This implies that a circle is not a good shape for minimizing the influence of fabrication error on acceleration sensitivity and that an appropriately selected ellipse would be much better.

It should be noted that the calculations that habe been presented in Figs. 7- 11 are not for identical clips on both sides as shown in Fig.1, but for the dimensioning of the clips shown in Fig.12. When identical clips are used on both faces, the influence of mispositioning is about four times as great. In Fig.13 we show the influence of changes in the dimension of the large clips on the resultants $\gamma_{(p)}$ and $\gamma_{(r)}$. The figure shows that cutting the size of the large clip in half from that of the best results obtained increases the influence of a mispositioning of the mode center on the acceleration sensitivity by a factor of 2. This indicates that the dimensioning of the clips still has to be studied further.

Figure 14 shows the mispositioning coefficients as a function of the size of the openings in the cover plates for the thickness ratio considered, i.e., h/h = 2.0. The curves show that there is moderate change in the coefficient for the mispositioning of the mode center with hole size, but that there is virtually no change in the coefficients for the mispositioning of the clips over a reasonable range of hole size.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No.DAAL03-91-C-0120 and the National Science Foundation under Grant No.MSS-8908188.

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, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips," Proceedings of the 46th Annual Symposium on Frequency Control, U.S. Army Electronics Technology and Devices Laboratory, Fort Monmouth, New Jersey, and Institute of Electrical and Electronics Engineers, New York, IEEE Cat. No.91CH2965-2, 298 (1991).

4. H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," J. Appl. Phys., 64, 4334 (1988).

5. 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).

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. For more detail see Y.S. Zhou and H.F. Tiersten, "On the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Quartz Cover Plates Supported by Clips," to be issued as a technical report, Rensselaer Polytechnic Institute. Troy, New York.

9. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., 9, 1445 (1968).

10. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4.

11. L. Brand, Vector and Tensor Analysis (Wiley, New York, 1947), p.222.

12. L.V. Kantorovich and V.I. Krylov, Approximate Methods of Higher Analysis (Interscience Publ. Inc., New York and P. Noordhoff Ltd., Groningen, The Netherlands, transl. by C.D. Benster from 3rd Russian ed., 1964), pp.258-260, 272-273, 279-281.

13. L. Collatz, The Numerical Treatment of Differential Equations, transl. by P.C. Williams (Springer-Verlag, Berlin, 1960), 2nd ed., pp.202-207 and 213-216.

14. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

15. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

16. The curves taken from Ref.3 are for different clip dimensions than were used in the in-plane calculations in this work. We do not believe that the normal acceleration sensitivities will be very sensitive to clip size. However, this remains to be examined.

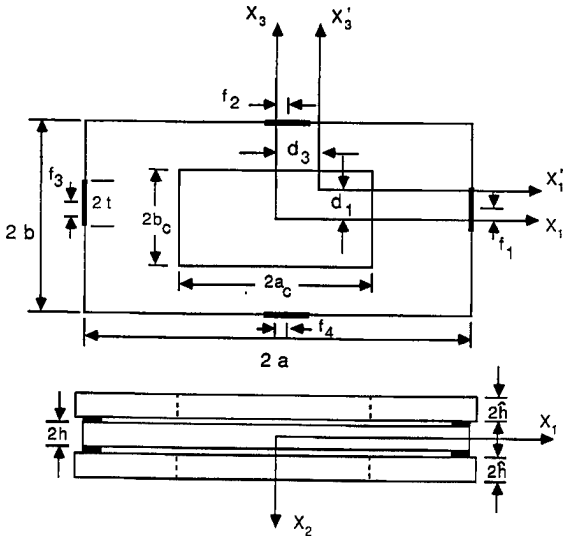


Figure 1 Plan View and Cross-Section of Box Structure Showing Active Plate, Cover Plates and Sidewalls and Displaced Location of Mode Center

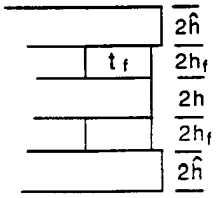


Figure 2 Cross-section of Sidewall

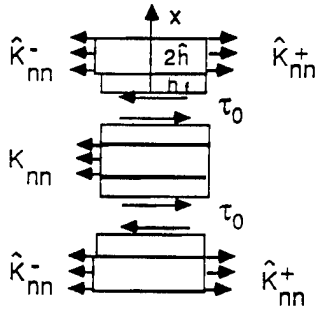


Figure 3 Free-body Diagrams of Sidewall

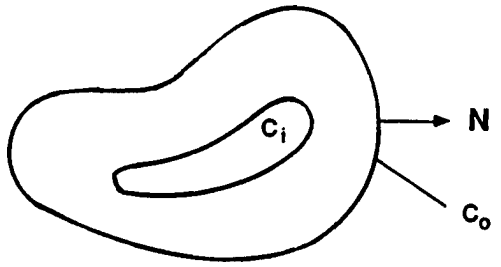


Figure 4 Plan View of Asymmetric Configuration with Arbitrary Curved Edges for Derivation of Variational Principle

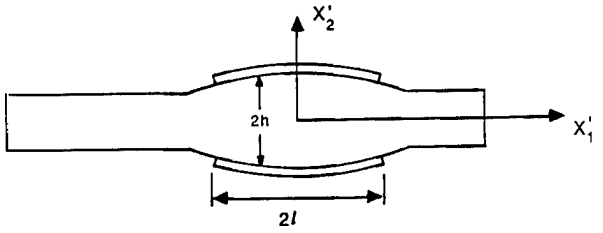


Figure 5 Cross-Section of Biconvex Resonator

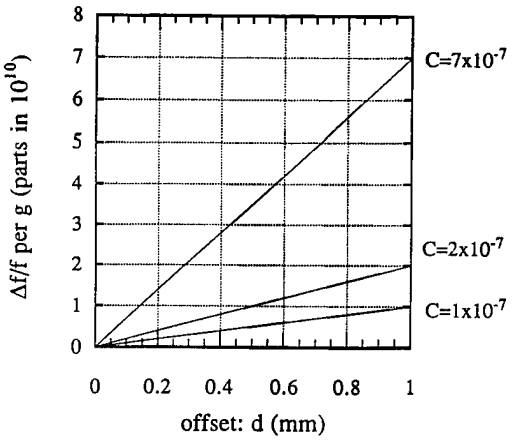


Figure 6 Acceleration Sensitivity Versus Mispositioning Dimension for Different Values of the Mispositioning Coefficient

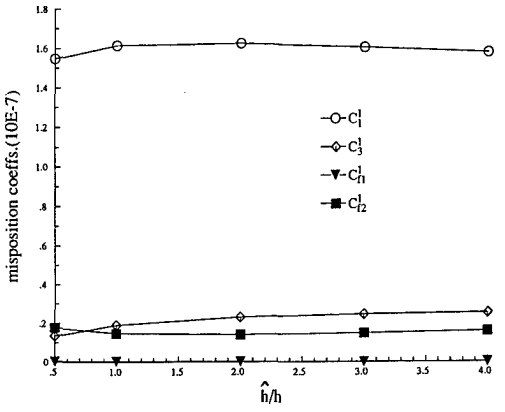


Figure 7 Mispositioning Coefficient Versus Thickness-Ratio for X_1-Acceleration. Orientation β = - 50° and a/b = 100 mm^2 are maintained in all calculations.

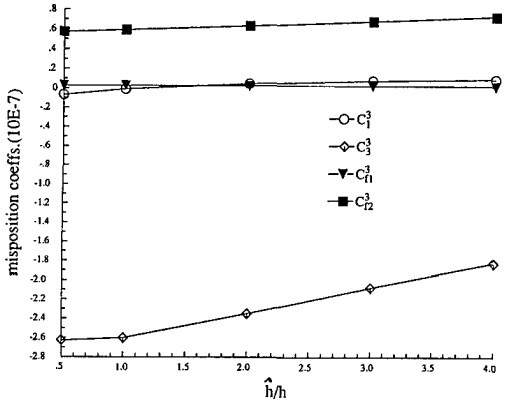


Figure 8 Mispositioning Coefficient Versus Thickness Ratio for X_3-Acceleration

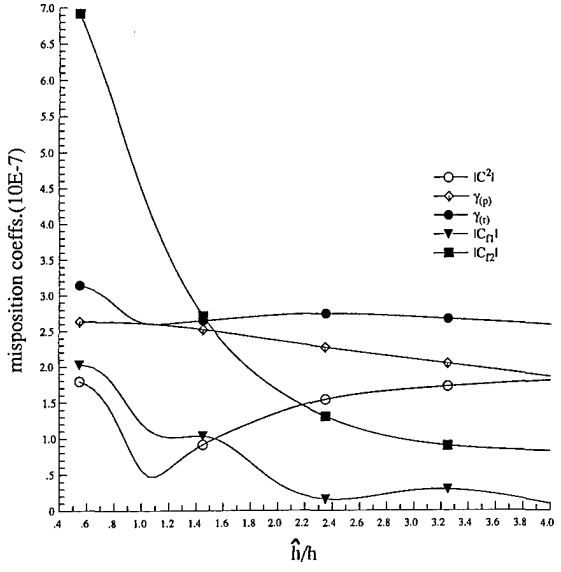


Figure 9 Resultant Mispositioning Coefficient Versus Thickness Ratio for a/b = 2.0

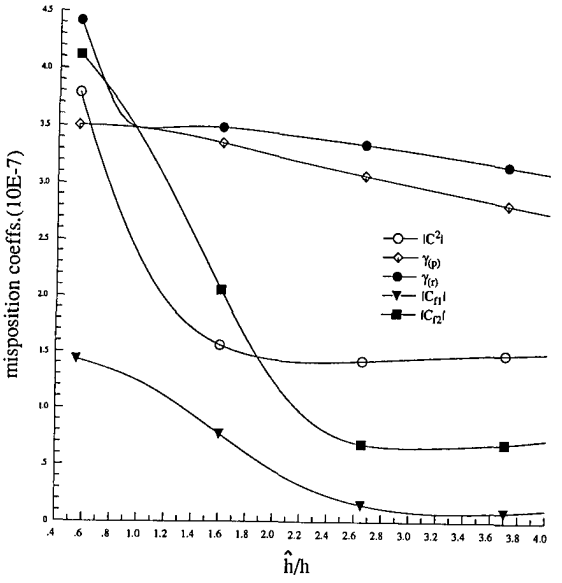


Figure 10 Resultant Mispositioning Coefficients Versus Thickness Ratio for a/b = 1.8

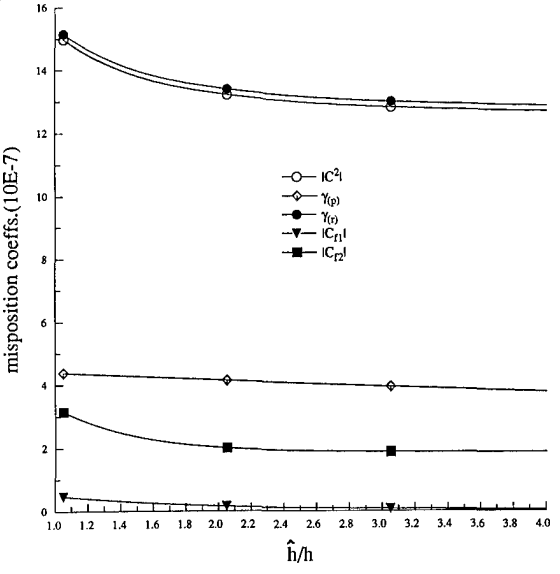


Figure 11 Resultant Mispositioning Coefficients Versus Thickness-Ratio for a/b = 1.0 (Square Plate)

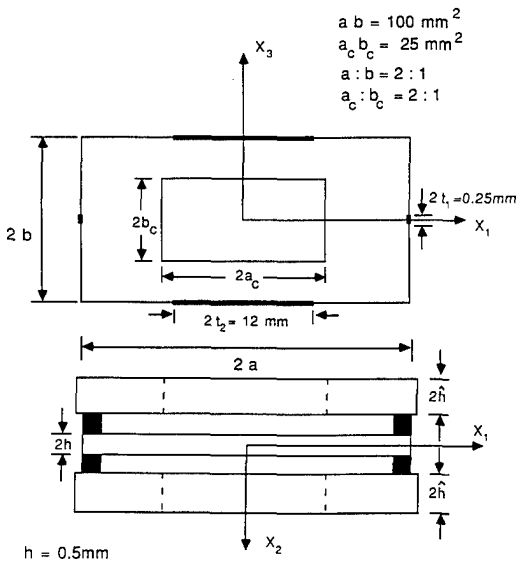


Figure 12 Plan View and Cross-Section of Box Structure Showing Desirable Dimensioning of Clips for In-Plane Calcutions Presented

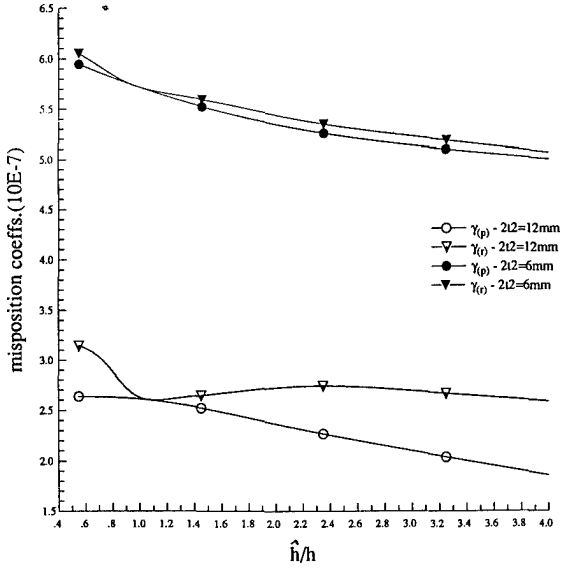


Figure 13 Resultant Mispositioning Coefficients Versus Thickness Ratio for Undesirable Reduced Size of Large Clips

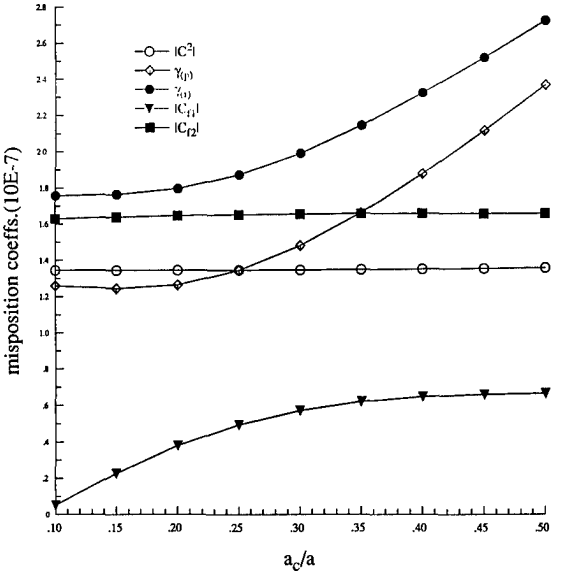


Figure 14 Resultant Mispositioning Coefficients Versus Opening Ratio for Thickness Ratio h/h = 2.0

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

VIBRATIONS OF CIRCULAR DISK DIELECTRIC RESONATORS

P. C. Y. Lee and J. S. Yang

Department of Civil Engineering & Operations Research
Princeton University, Princeton, N. J. 08544

Abstract

Two-dimensional equations for guided EM waves propagating in a dielectric plate surrounded by free space were recently derived from the three-dimensional Maxwell's equations. These equations are employed for the study of vibrations of circular disk dielectric resonators embedded in free space. Closed form solutions are obtained for which the continuity conditions of the tangential E and H fields at the end faces and the lateral cylindrical surface are accommodated. It is found that the solutions represent two types of modes: (1) the modes in which the electric field is transverse to the axis of the disk (TE modes) and (2) the modes in which the magnetic field is transverse to the axis of the resonator (TM modes). Frequency equations are solved and resonance frequencies are computed as functions of the diameter-to-thickness ratio and for various of the refractive index of the resonator. Predicted values are compared with experimental data and results by other approximate methods.

I. Introduction

Dielectric resonators of finite extent embedded in free space were investigated by Richtmyer in 1939. He obtained closed form solutions for a sphere and a circular ring of small circular cross section.[1] Roots of the frequency equation of the sphere were computed by Yee[2] approximately and by Gastine, Courtois, and Dormann exactly.[3]

One of the effective methods to facilitate a systematic study of the vibrations in plate resonators is, first, to derive a system of two-dimensional approximate governing equations and, then, to solve these equations for closed form, analytical or numerical solutions. This kind of approach has been employed in deriving the approximate equations of motion for elastic plates from the three-dimensional equation of elasticity for a long time and by many investigators.[4-8]

In an earlier paper,[9] two-dimensional governing equations of successively higher-order approximations for guided EM waves in an isotropic dielectric plate surrounded by free space are deduced from the three-dimensional Maxwell's equations by expanding the EM 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. A single system of governing equations is obtained by further satisfying the continuity conditions of the EM field at the interfaces between the plate and free space.

In the present paper, these equations are employed for the study of vibrations of circular disk dielectric resonators embedded in free space. Frequency equations are obtained for the resonant frequencies of the resonators. It is found that the solutions represent two types of modes, one type does not have axial component of electric field while the other type has no axial magnetic field. Frequency equations are solved and resonant frequencies are computed as functions of the diameter-to-thickness ratio and for various values of the refractive index of the resonator. The results agree reasonably well with experimental data and results by other approximate methods.

II. Two – Dimensional Equations

We consider a circular dielectric disk of radius a, thickness $2b$, permittivity ϵ and permeability μ_o which is embedded in the infinite space of permittivity ϵ_o and permeability μ_o as shown in Fig. 1. We note that the circular disk will be referred ei-

0-7803-0476-4/92 $3.00 © 1992 IEEE

ther to the rectangular coordinates x_i, $i = 1, 2, 3$ or the cylindrical coordinates r, θ, z as they are also shown in Fig. 1.

The two-dimensional equations for the trapped or guided waves in an isotropic dielectric plate surrounded by free space which have taken into account the interaction of the field in the plate with the evanescent field in free space by satisfying the continuity conditions at the faces of the plate are[9]

$$
\begin{aligned}
&(1+\delta_{p0})k(A^{(p)}_{3,13} - A^{(p)}_{1,33}) + (\frac{p\pi}{2b})^2 A^{(p)}_1 \\
&\quad - \hat{n}^2 A^{(p)}_{a,a1}(1+\delta_{p0}) - \frac{p\pi}{2b}(1-\hat{n}^2)A^{(p-1)}_{2,1} \\
&\quad + \frac{1}{\bar{\eta}b}\sum_{m=0}^{\infty}\hat{\delta}_{mp}(k-1)A^{(m)}_{a,a1} \\
&\quad + \frac{1}{\bar{\eta}b}\sum_{m=0}^{\infty}\hat{\delta}_{mp}(-A^{(m)}_{1,aa} + \bar{\eta}^2 A^{(m)}_1) \\
&\quad + \frac{\hat{n}^2}{c_o^2}(1+\delta_{p0})A^{(p)}_{1,tt} + \frac{1}{\bar{\eta}bc_o^2}\sum_{m=0}^{\infty}\hat{\delta}_{mp}A^{(m)}_{1,tt} = 0 \ , \\
&\frac{p}{2b}\pi A^{(p)}_{a,a} - A^{(p-1)}_{2,aa} + \frac{\hat{n}^2}{c_o^2}A^{(p-1)}_{2,tt} = 0 \ , \\
&(1+\delta_{p0})k(A^{(p)}_{1,13} - A^{(p)}_{3,11}) + (\frac{p\pi}{2b})^2 A^{(p)}_3 \\
&\quad - \hat{n}^2 A^{(p)}_{a,a3}(1+\delta_{p0}) - (\frac{p\pi}{2b})(1-\hat{n}^2)A^{(p-1)}_{2,3} \\
&\quad + \frac{1}{\bar{\eta}b}\sum_{m=0}^{\infty}\hat{\delta}_{mp}(k-1)A^{(m)}_{a,a3} \\
&\quad + \frac{1}{\bar{\eta}b}\sum_{m=0}^{\infty}\hat{\delta}_{mp}(-A^{(m)}_{3,aa} + \bar{\eta}^2 A^{(m)}_3) \\
&\quad + \frac{\hat{n}^2}{c_o^2}(1+\delta_{p0})A^{(p)}_{3,tt} + \frac{1}{\bar{\eta}bc_o^2}\sum_{m=0}^{\infty}\hat{\delta}_{mp}A^{(m)}_{3,tt} = 0 \ ,
\end{aligned}
\tag{1}
$$

where $p = 0, 1, 2, ...$ is the order of the theory, $\hat{n} = (\frac{\epsilon}{\epsilon_o})^{\frac{1}{2}}$ is the refractive index, $c_o = (\epsilon_o\mu_o)^{-\frac{1}{2}}$ is the speed of light in free space, $\bar{\eta}$ is a quantity describing the exponential decay rate of the fields as a function of the distance from the face of the plate, k is a correction factor, δ_{mp} is the Kronecker delta, and $\hat{\delta}_{mp} = \frac{1}{2}[1 + (-1)^{m+p}] = 1$ for $m+n =$ even and $= 0$ for $m+n =$ odd.

We note $A^{(p)}_i$ are the two-dimensional pth-order vector potential which are the coefficients of the following expansions of the vector potential A_i

$$
\begin{aligned}
A_a &= \sum_{p=0}^{\infty} A^{(p)}_a(x_1, x_3, t)cos\frac{p\pi}{2}(1 - \frac{x_2}{b}) \ , \\
A_2 &= \sum_{p=0}^{\infty} A^{(p)}_2(x_1, x_3, t)sin\frac{p+1}{2}\pi(1 - \frac{x_2}{b}) \ ,
\end{aligned}
\tag{2}
$$

where $a = 1, 3$.

By substituting (2) into the field potential relations

$$
\begin{aligned}
E_i &= -A_{i,t} \ , \\
B_i &= \epsilon_{ijk}A_{k,j} \ ,
\end{aligned}
\tag{3}
$$

we have

$$
\begin{aligned}
E_a &= \sum_{p=0}^{\infty} E^{(p)}_a cos\frac{p\pi}{2}(1-\psi) \ , \\
E_2 &= \sum_{p=0}^{\infty} E^{(p)}_2 sin\frac{p+1}{2}\pi(1-\psi) \ , \\
B_a &= \sum_{p=0}^{\infty} B^{(p)}_a sin\frac{p+1}{2}\pi(1-\psi) \ , \\
B_2 &= \sum_{p=0}^{\infty} B^{(p)}_2 \cos\frac{p\pi}{2}(1-\psi) \ ,
\end{aligned}
\tag{4}
$$

where

$$
\begin{aligned}
E^{(p)}_i &= -A^{(p)}_{i,t} \ , \\
B^{(p-1)}_1 &= \frac{p}{2b}\pi A^{(p)}_3 - A^{(p-1)}_{2,3} \ , \\
B^{(p)}_2 &= k(A^{(p)}_{1,3} - A^{(p)}_{3,1}) \ , \\
B^{(p-1)}_3 &= -\frac{p}{2b}\pi A^{(p)}_1 + A^{(p-1)}_{2,1} \ ,
\end{aligned}
\tag{5}
$$

are the two-dimensional E and B field-potential relations. We note that in (5), when $p = 0$, $B^{(-1)}_a = A^{(-1)}_2 = 0$.

By defining the two-dimensional pth-order components of the H and D fields as

$$
\begin{aligned}
H^{(p)}_a &= \int_{-1}^{1} H_a sin\frac{p+1}{2}\pi(1-\psi)d\psi \ , \\
H^{(p)}_2 &= \int_{-1}^{1} H_2 cos\frac{p\pi}{2}(1-\psi)d\psi \ , \\
D^{(p)}_a &= \int_{-1}^{1} D_a cos\frac{p\pi}{2}(1-\psi)d\psi \ , \\
D^{(p)}_2 &= \int_{-1}^{1} D_2 sin\frac{p+1}{2}\pi(1-\psi)d\psi \ ,
\end{aligned}
\tag{6}
$$

and substituting (4) into the constitutive relations

$$H_i = B_i/\mu_o \ , \qquad D_i = \epsilon E_i \ , \tag{7}$$

and, in turn, into (6), we obtain the two-dimensional D, H field-potential relations

$$\begin{aligned} H_1^{(p-1)} &= \frac{1}{\mu_o}(\frac{p}{2b}\pi A_3^{(p)} - A_{2,3}^{(p-1)}) \ , \\ H_2^{(p)} &= \frac{1}{\mu_o}k(1+\delta_{p0})(A_{1,3}^{(p)} - A_{3,1}^{(p)}) \ , \\ H_3^{(p-1)} &= \frac{1}{\mu_o}(-\frac{p}{2b}\pi A_1^{(p)} + A_{2,1}^{(p-1)}) \ , \\ D_a^{(p)} &= -\epsilon(1+\delta_{p0})A_{a,t}^{(p)} \ , \\ D_2^{(p-1)} &= -\epsilon A_{2,t}^{(p-1)} \ . \end{aligned} \tag{8}$$

Again, we note that $H_a^{(-1)} = D_2^{(-1)} = 0$.

In the following, the interactions between $A_i^{(p)}$ of different orders will be neglected. This means that in (1) only the term $m = p$ will be kept for the summation of m from 0 to ∞. Earlier results[9] show that the interactions among different orders of $A_i^{(p)}$ are weak and the omission of them affects the dispersion relations very little. Hence (1) are reduced to

$$\begin{aligned} &\alpha A_{a,bb}^{(p)} - (\frac{\pi}{2b})^2 \kappa A_a^{(p)} - \gamma A_{b,ba}^{(p)} \\ &\qquad - (\frac{\pi}{2b})\delta A_{2,a}^{(p-1)} = \frac{\beta}{c_o^2}\ddot{A}_a^{(p)} \ , \\ &A_{2,aa}^{(p-1)} - \frac{\pi}{2b}pA_{b,b}^{(p)} = \frac{\hat{n}^2}{c_o^2}\ddot{A}_2^{(p-1)} \ , \end{aligned} \tag{9}$$

where

$$\begin{aligned} \alpha &= [(1+\delta_{p0}) + \frac{1}{\bar{\eta}b}]k \ , \\ \beta &= (1+\delta_{p0})\hat{n}^2 + \frac{1}{\bar{\eta}b} \ , \\ \gamma &= (1+\delta_{p0})(k-\hat{n}^2) + (k-1)\frac{1}{\bar{\eta}b} \\ &= \alpha - \beta \ , \\ \kappa &= p^2 + \frac{4b\bar{\eta}}{\pi^2} \ , \\ \delta &= p(\hat{n}^2 - 1) \ , \\ \bar{\eta}^2 &= \xi^2 - \omega^2/c_o^2 \ . \end{aligned} \tag{10}$$

Once the solutions of $A_i^{(p)}$ are obtained from (9) the distribution of E and B fields in the region $r < a$ and $|x_2| < b$ can be calculated from (5) and (4). In the upper half of the free space, i.e. in $r < a$ and $x_2 > b$, the distribution of the E and B fields are[9]

$$\bar{E}_i = E_i^+ e^{-\bar{\eta}(x_2-b)} \ , \qquad \bar{B}_i = B_i^+ e^{-\bar{\eta}(x_2-b)} \ , \tag{11}$$

where the two-dimensional field-potential relations are given by

$$\begin{aligned} E_i^+ &= -A_{i,t}^+ \ , \\ B_1^+ &= -\bar{\eta}A_3^+ - A_{2,3}^+ \ , \\ B_2^+ &= k(A_{1,3}^+ - A_{3,1}^+) \ , \\ B_3^+ &= \bar{\eta}A_1^+ + A_{2,1}^+ \ . \end{aligned} \tag{12}$$

In (12), the A_i^+ are related to $A_i^{(p)}$ through the continuity conditions[9] by

$$\begin{aligned} A_a^+ &= A_a^{(p)} \ , \\ A_2^+ &= \frac{p\pi}{2}\hat{n}^2 A_2^{(p-1)} - [(1+\delta_{p0})b\hat{n}^2 + \frac{1}{\bar{\eta}}]A_{a,a}^{(p)} \ . \end{aligned} \tag{13}$$

For the distribution of E and B fields in the lower half of the free space, i.e. for $r < a$ and $x_2 < b$:

$$\bar{E}_i = E_i^- e^{\bar{\eta}(x_2+b)} \ , \qquad \bar{B}_i = B_i^- e^{\bar{\eta}(x_2+b)} \ , \tag{14}$$

where

$$\begin{aligned} E_i^- &= -A_{i,t}^- \ , \\ B_1^- &= \bar{\eta}A_3^- - A_{2,3}^- \ , \\ B_2^- &= k(A_{1,3}^- - A_{3,1}^-) \ , \\ B_3^- &= -\bar{\eta}A_1^- + A_{2,1}^- \ , \end{aligned} \tag{15}$$

and

$$A_a^- = (-1)^p A_a^+ \ , \qquad A_2^- = (-1)^{p-1} A_2^+ \ . \tag{16}$$

We note that once the E and B fields are ob-

tained from (4), (11), and (14), the corresponding D and H fields can be calculated from (7).

Governing equations for the vector potential in the free space outside the cylindrical region, i.e. for $r > a$, can be obtained from (9) and (10) by setting the correction factor $k = 1$ and the refractive index $\hat{n} = 1$. Hence, we have, for $r > a$

$$\begin{aligned} &\bar{\alpha}A^{(p)}_{a,bb} - (\frac{\pi}{2b})^2\bar{\kappa}A^{(p)}_a - \bar{\gamma}A^{(p)}_{b,ba} \\ &\quad - (\frac{\pi}{2b})\bar{\delta}A^{(p-1)}_{2,a} = \frac{\bar{\beta}}{c_o^2}\ddot{A}^{(p)}_a \ , \\ &A^{(p-1)}_{2,aa} - \frac{\pi}{2b}pA^{(p)}_{b,b} = \frac{1}{c_o^2}\ddot{A}^{(p-1)}_2 \ , \end{aligned} \tag{17}$$

where

$$\begin{aligned} &\bar{\alpha} = (1+\delta_{p0}) + \frac{1}{\bar{\eta}b} \ , \\ &\bar{\beta} = (1+\delta_{p0}) + \frac{1}{\bar{\eta}b} = \bar{\alpha} \ , \\ &\bar{\gamma} = 0 \ , \\ &\bar{\kappa} = p^2 + \frac{4b\bar{\eta}}{\pi^2} \ , \\ &\bar{\delta} = 0 \ , \\ &\bar{\eta}^2 = \xi^2 - \omega^2/c_o^2 \ . \end{aligned} \tag{18}$$

In (18), the expression for $\bar{\eta}$ is the same as that in (14) so that the exponential decay rate of the exterior fields in the x_2 direction are the same as those of the interior fields. We note that the terms with coefficients $\bar{\gamma}$ and $\bar{\delta}$, which are zero, are retained in (17) to keep (9) and (17) in the same form so that the result presented in the next section can be applied to both.

III. Invariant Form

To transform the equations in the preceding section from the rectangular coordinates to the cylindrical coordinates as shown in Fig.1, we define the two-dimensional vector potential associated with pth-order equations (9) by

$$\hat{\mathbf{A}} = \hat{\mathbf{A}}_t + \hat{\mathbf{A}}_z \ , \tag{19}$$

where $\hat{\mathbf{A}}_t$ and $\hat{\mathbf{A}}_z$ are the transverse and normal parts of $\hat{\mathbf{A}}$ with respect to the x_3x_1 plane and are given by

$$\begin{aligned} &\hat{\mathbf{A}}_t = A^{(p)}_3\mathbf{e}_3 + A^{(p)}_1\mathbf{e}_1 = A^{(p)}_r\mathbf{e}_r + A^{(p)}_\theta\mathbf{e}_\theta \ , \\ &\hat{\mathbf{A}}_z = A^{(p-1)}_2\mathbf{e}_2 = A^{(p-1)}_z\mathbf{e}_z \ . \end{aligned} \tag{20}$$

By the use of the vector identity $\nabla\times\nabla\times\hat{\mathbf{A}}_t = \nabla\nabla\cdot\hat{\mathbf{A}}_t - \nabla^2\hat{\mathbf{A}}_t$ and recalling $\gamma = \alpha - \beta$, (9) can be written in the invariant form as follows:

$$\begin{aligned} &-\alpha\nabla\times\nabla\times\hat{\mathbf{A}}_t + \beta\nabla(\nabla\cdot\hat{\mathbf{A}}_t) - (\frac{\pi}{2b})^2\kappa\hat{\mathbf{A}}_t \\ &\quad - (\frac{\pi}{2b})\delta\nabla(\hat{\mathbf{A}}_z\cdot\mathbf{e}_z) = \frac{\beta}{c_o^2}\ddot{\hat{\mathbf{A}}}_t \ , \\ &\nabla^2(\hat{\mathbf{A}}_z\cdot\mathbf{e}_z) - \frac{\pi}{2b}p\nabla\cdot\hat{\mathbf{A}}_t = \frac{\hat{n}^2}{c_o^2}(\ddot{\hat{\mathbf{A}}}_z\cdot\mathbf{e}_z) \ . \end{aligned} \tag{21}$$

where $\nabla = \mathbf{e}_3\frac{\partial}{\partial x_3} + \mathbf{e}_1\frac{\partial}{\partial x_1} = \mathbf{e}_r\frac{\partial}{\partial r} + \frac{1}{r}\mathbf{e}_\theta\frac{\partial}{\partial\theta}$ is the two-dimensional gradient operator.

The field-potential relations (5) for $r < a$, $|x_2| < b$ become

$$\begin{aligned} &\hat{\mathbf{E}}_t = -\hat{\mathbf{A}}_{t,t} \ , \quad \hat{\mathbf{E}}_z = -\hat{\mathbf{A}}_{z,t} \ , \\ &\hat{\mathbf{B}}_t = \frac{p\pi}{2b}\mathbf{e}_z\times\hat{\mathbf{A}}_t + \nabla\times\hat{\mathbf{A}}_z \ , \\ &\hat{\mathbf{B}}_z = k\nabla\times\hat{\mathbf{A}}_t \ , \end{aligned} \tag{22}$$

where

$$\begin{aligned} &\hat{\mathbf{E}}_t = E^{(p)}_3\mathbf{e}_3 + E^{(p)}_1\mathbf{e}_1 \ , \\ &\hat{\mathbf{E}}_z = E^{(p-1)}_2\mathbf{e}_2 \ , \\ &\hat{\mathbf{B}}_t = B^{(p-1)}_3\mathbf{e}_3 + B^{(p-1)}_1\mathbf{e}_1 \ , \\ &\hat{\mathbf{B}}_z = B^{(p)}_2\mathbf{e}_2 \ . \end{aligned} \tag{23}$$

For the fields in the region of $r < a$ and $z > b$, (12) and (13) can be written as

$$\begin{aligned} &\mathbf{E}^+_t = -\mathbf{A}^+_{t,t} \ , \quad \mathbf{E}^+_z = -\mathbf{A}^+_{z,t} \ , \\ &\mathbf{B}^+_t = -\bar{\eta}\mathbf{e}_z\times\mathbf{A}^+_t + \nabla\times\mathbf{A}^+_z \ , \\ &\mathbf{B}^+_z = \nabla\times\mathbf{A}^+_t \ , \end{aligned} \tag{24}$$

and

$$\begin{aligned} &\mathbf{A}^+_t = \hat{\mathbf{A}}_t \ , \\ &\mathbf{A}^+_z\cdot\mathbf{e}_z = \frac{p\pi}{2}\hat{n}^2\mathbf{A}_z\cdot\mathbf{e}_z \\ &\quad - [(1+\delta_{p0})b\hat{n}^2 + \frac{1}{\bar{\eta}}]\nabla\cdot\hat{\mathbf{A}}_t \ . \end{aligned} \tag{25}$$

Expressions similar to (25) can be obtained from (15). We note that in (22) and (24), $(\ )_{,t} \equiv \partial(\ )/\partial t$, where t is the time.

The field-potential relations (22), (24) and

(25) which are given in the invariant form can now be easily transformed in the cylindrical coordinate system

$$\begin{aligned}
\hat{E}_r &= -\hat{A}_{r,t} \ , \\
\hat{E}_\theta &= -\hat{A}_{\theta,t} \ , \\
\hat{E}_z &= -\hat{A}_{z,t} \ , \\
\hat{B}_r &= -\frac{p\pi}{2b}\hat{A}_\theta + \frac{1}{r}\frac{\partial \hat{A}_z}{\partial \theta} \ , \\
\hat{B}_\theta &= \frac{p\pi}{2b}\hat{A}_r - \frac{\partial \hat{A}_z}{\partial r} \ , \\
\hat{B}_z &= k[\frac{1}{r}\frac{\partial (r\hat{A}_\theta)}{\partial r} - \frac{1}{r}\frac{\partial \hat{A}_r}{\partial \theta}] \ ,
\end{aligned} \tag{26}$$

for $r < a$, $|x_2| < b$, and

$$\begin{aligned}
E_r^+ &= -\hat{A}_{r,t} \ , \\
E_\theta^+ &= -\hat{A}_{\theta,t} \ , \\
E_z^+ &= -\hat{A}_{z,t} \ , \\
B_r^+ &= \bar{\eta}\hat{A}_\theta + \frac{1}{r}\frac{\partial A_z^+}{\partial \theta} \ , \\
B_\theta^+ &= -\bar{\eta}\hat{A}_r - \frac{\partial A_z^+}{\partial r} \ , \\
B_z^+ &= k[\frac{1}{r}\frac{\partial (rA_\theta^+)}{\partial r} - \frac{1}{r}\frac{\partial A_r^+}{\partial \theta}] \ ,
\end{aligned} \tag{27}$$

where

$$\begin{aligned}
&A_r^+ = \hat{A}_r \ , \quad A_\theta^+ = \hat{A}_\theta \ , \\
&A_z^+ = \frac{p\pi}{2}\hat{n}^2\hat{A}_z \\
&\quad - [(1+\delta_{p0})b\hat{n}^2 + \frac{1}{\bar{\eta}}][\frac{1}{r}\frac{\partial (r\hat{A}_r)}{\partial r} + \frac{1}{r}\frac{\partial \hat{A}_\theta}{\partial \theta}] \ .
\end{aligned} \tag{28}$$

IV. TE Modes

For TE modes, the electric field is transverse to the normal of the faces of the resonator and hence $A_z = 0$. It can be seen from $(21)_2$ that when $A_z = 0$ and $p \neq 0$,

$$\nabla \cdot \hat{\mathbf{A}}_t = 0 \ , \tag{29}$$

and $(21)_1$ reduces to

$$\begin{aligned}
&-\alpha \nabla \times \nabla \times \hat{\mathbf{A}}_t \\
&\quad -(\frac{\pi}{2b})^2 \kappa \hat{\mathbf{A}}_t = \frac{\beta}{c_o^2}\ddot{\hat{\mathbf{A}}}_t \ .
\end{aligned} \tag{30}$$

By applying the coordinate transformation to the cylindrical coordinates and considering time harmonic waves $\hat{\mathbf{A}}_t(r,\theta,t) = \hat{\mathbf{A}}_t(r,\theta)e^{i\omega t}$, (30) can be written as

$$\begin{aligned}
&\alpha\frac{1}{r}\frac{\partial U}{\partial r} - (\frac{\pi}{2b})^2\lambda\hat{A}_r = 0 \ , \\
&\alpha\frac{\partial U}{\partial r} + (\frac{\pi}{2b})^2\lambda\hat{A}_\theta = 0 \ ,
\end{aligned} \tag{31}$$

where

$$\begin{aligned}
&U = \mathbf{e}_z \cdot (\nabla \times \hat{\mathbf{A}}_t) = \frac{1}{r}\frac{\partial (r\hat{A}_\theta)}{\partial r} - \frac{1}{r}\frac{\partial \hat{A}_r}{\partial \theta} \ , \\
&\lambda(\frac{\pi}{2b})^2 = c_o{}^{-2}\beta\omega^2 - \kappa(\frac{\pi}{2b})^2 \ ,
\end{aligned} \tag{32}$$

and ω is the angular frequency.

In similar manner, (29) is transformed to

$$V = \nabla \cdot \hat{\mathbf{A}}_t = \frac{1}{r}\frac{\partial (r\hat{A}_r)}{\partial r} + \frac{1}{r}\frac{\partial \hat{A}_\theta}{\partial \theta} = 0 \ . \tag{33}$$

Insertion of $\hat{A}_r$ and $\hat{A}_\theta$ from (31) into $(32)_1$ leads to the governing equation on U

$$\alpha(\frac{\partial^2 U}{\partial r^2} + \frac{1}{r}\frac{\partial U}{\partial r} + \frac{1}{r^2}\frac{\partial^2 U}{\partial \theta^2}) + (\frac{\pi}{2b})^2\lambda U = 0 \tag{34}$$

from which we obtain the solution

$$U(r,\theta) = C\frac{\pi}{2b}J_n(\xi r)cosn\theta \ , \tag{35}$$

where n are integers, J_n is the the Bessel function of the first kind of order n, and

$$\xi^2 = \frac{\lambda}{\alpha}(\frac{\pi}{2b})^2 \ . \tag{36}$$

By substituting (35) back into (31) and demanding that (33) be satisfied, we have

$$\begin{aligned}
&\hat{A}_r(r,\theta) = -n\xi^{-1}\frac{\pi}{2b}C\frac{1}{\xi r}J_n(\xi r)sin(n\theta) \ , \\
&\hat{A}_\theta(r,\theta) = -\xi^{-1}\frac{\pi}{2b}CJ_n'(\xi r)cos(n\theta) \ , \\
&\hat{A}_z(r,\theta) = 0 \ .
\end{aligned} \tag{37}$$

By inserting (37) into (26) and (27) , and the

results, in turn, into (4) and (11), respectively, we obtain the spatial distribution of the E and B fields inside the cylindrical region $r < a$:

For $|z| < b$

$$
\begin{aligned}
E_r =& i\omega n\xi^{-1}\frac{\pi}{2b}C\frac{1}{\xi r}J_n(\xi r)sin(n\theta) \\
& cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
E_\theta =& i\omega\xi^{-1}\frac{\pi}{2b}CJ'_n(\xi r)cos(n\theta) \\
& cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
E_z =& 0 , \\
B_r =& \xi^{-1}\frac{\pi}{2b}\frac{p\pi}{2b}CJ'_n(\xi r)cos(n\theta) \\
& sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
B_\theta =& -n\xi^{-1}\frac{\pi}{2b}\frac{p\pi}{2b}C\frac{1}{\xi r}J_n(\xi r)sin(n\theta) \\
& sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
B_z =& k\frac{\pi}{2b}CJ_n(\xi r)cos(n\theta) \\
& cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} .
\end{aligned}
\tag{38}
$$

For $z > b$

$$
\begin{aligned}
E_r =& i\omega n\xi^{-1}\frac{\pi}{2b}C\frac{1}{\xi r}J_n(\xi r)sin(n\theta) \\
& e^{-\bar{\eta}(z-b)}e^{i\omega t} , \\
E_\theta =& i\omega\xi^{-1}\frac{\pi}{2b}CJ'_n(\xi r)cos(n\theta) \\
& e^{-\bar{\eta}(z-b)}e^{i\omega t} , \\
E_z =& 0 \\
B_r =& -\bar{\eta}\xi^{-1}\frac{\pi}{2b}CJ'_n(\xi r)cos(n\theta) \\
& e^{-\bar{\eta}(z-b)}e^{i\omega t} , \\
B_\theta =& \bar{\eta}n\xi^{-1}\frac{\pi}{2b}C\frac{1}{\xi r}J_n(\xi r)sin(n\theta) \\
& e^{-\bar{\eta}(z-b)}e^{i\omega t} , \\
B_z =& k\frac{\pi}{2b}CJ_n(\xi r)cos(n\theta) \\
& e^{-\bar{\eta}(z-b)}e^{i\omega t} .
\end{aligned}
\tag{39}
$$

For $r > a$ and $|z| < b$, we have the same equations as those in (31) to (34) except α, β, κ are replaced by $\bar{\alpha}$, $\bar{\beta}$, $\bar{\kappa}$ which are given in (18), and λ replaced by $\bar{\lambda}$ which is defined by

$$
\bar{\lambda}(\frac{\pi}{2b})^2 = c_o{}^{-2}\bar{\beta}\omega^2 - \bar{\kappa}(\frac{\pi}{2b})^2 . \tag{40}
$$

Therefore, the governing equation for U is

$$
\bar{\alpha}(\frac{\partial^2 U}{\partial r^2}+\frac{1}{r}\frac{\partial U}{\partial r}+\frac{1}{r^2}\frac{\partial^2 U}{\partial\theta^2})+(\frac{\pi}{2b})^2\bar{\lambda}U = 0 , \tag{41}
$$

from which we may obtain the decaying solutions

$$
U(r,\theta) = \bar{C}\frac{\pi}{2b}K_n(\xi r)cos(n\theta) , \tag{42}
$$

where n are integers, K_n is the modified Bessel function of the second kind of order n, and

$$
\bar{\xi}^2 = -\frac{\bar{\lambda}}{\bar{\alpha}}(\frac{\pi}{2b})^2 . \tag{43}
$$

From (42) we obtain, in a similar manner, the vector potential

$$
\begin{aligned}
\hat{A}_r =& n\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta) , \\
\hat{A}_\theta =& \bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}K'_n(\bar{\xi}r)cos(n\theta) , \\
\hat{A}_z =& 0 ,
\end{aligned}
\tag{44}
$$

the components of the E and B fields, for $r > a$, $|x_2| < b$

$$
\begin{aligned}
E_r =& -i\omega n\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta) \\
& cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
E_\theta =& -i\omega\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}K'_n(\bar{\xi}r)cos(n\theta) \\
& cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
E_z =& 0 , \\
B_r =& -\bar{\xi}^{-1}\frac{\pi}{2b}\frac{p\pi}{2b}\bar{C}K'_n(\bar{\xi}r)cos(n\theta) \\
& sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} , \\
B_\theta =& n\bar{\xi}^{-1}\frac{\pi}{2b}\frac{p\pi}{2b}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta) \\
& sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} ,
\end{aligned}
\tag{45}
$$

$$B_z = \frac{\pi}{2b}\bar{C}K_n(\bar{\xi}r)cos(n\theta)$$
$$cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t} ,$$

and, for $r > a$, $x_2 > b$

$$E_r = -i\omega n\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t} ,$$

$$E_\theta = -i\omega\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}K_n'(\bar{\xi}r)cos(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t} ,$$

$$E_z = 0$$

$$B_r = \bar{\eta}\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}K_n'(\bar{\xi}r)cos(n\theta) \qquad (46)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t} ,$$

$$B_\theta = -\bar{\eta}n\bar{\xi}^{-1}\frac{\pi}{2b}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t} ,$$

$$B_z = \frac{\pi}{2b}\bar{C}K_n(\bar{\xi}r)cos(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t} .$$

According to a uniqueness theorem derived for the two-dimensional equations[10] and for $E_z = 0$, we require that $\hat{H}_z$ and $\hat{E}_r$ be continuous at $r = a$. By substituting (37) and (44) into (8) and imposing the contonuity conditions at $r = a$, we have

$$\begin{bmatrix} kJ_n(\xi a) & K_n(\bar{\xi}a) \\ \xi^{-1}J_n'(\xi a) & -\bar{\xi}^{-1}K_n'(\bar{\xi}a) \end{bmatrix}\begin{bmatrix} C \\ \bar{C} \end{bmatrix} = 0 . \qquad (47)$$

V. *TM* Modes

By taking the divergence of $(21)_1$ and letting

$$V = \nabla\cdot\mathbf{A}_t ,$$
$$W = \hat{\mathbf{A}}_z\cdot\mathbf{e}_z , \qquad (48)$$

for the time harmonic modes, we may express (21) in terms of V and W as

$$-\beta\nabla^2 V - \omega^2 c_o^{-2}\beta V$$
$$+(\frac{\pi}{2b})^2\kappa V + \frac{\pi}{2b}\delta\nabla^2 W = 0 , \qquad (49)$$
$$-\nabla^2 W - \omega^2 c_o^{-2}\hat{n}^2 U + \frac{\pi}{2b}pV = 0 .$$

For TM modes, the magnetic field is transverse to the normal of the face of the resonator, i.e. $B_z^{(p)} = k(A_{1,3}^{(p)} - A_{3,1}^{(p)}) = 0$ or

$$U = \mathbf{e}_z\cdot(\nabla\times\hat{\mathbf{A}}_t) = 0 . \qquad (50)$$

In cylindrical coordinates, (48) to (50) become

$$-\beta[\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial V}{\partial R}) + \frac{1}{r^2}\frac{\partial^2 V}{\partial\theta^2}]$$
$$-\frac{\beta}{c_o^2}\omega^2 V + (\frac{\pi}{2b})^2\kappa V$$
$$+\frac{\pi}{2b}\delta[\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial W}{\partial R}) + \frac{1}{r^2}\frac{\partial^2 W}{\partial\theta^2}] = 0 , \qquad (51)$$
$$-[\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial W}{\partial R}) + \frac{1}{r^2}\frac{\partial^2 W}{\partial\theta^2}]$$
$$-\frac{\hat{n}^2}{c_o^2}\omega^2 W + \frac{\pi}{2b}pV = 0 ,$$

and

$$U = \frac{1}{r}\frac{\partial}{\partial r}(r\hat{A}_\theta) - \frac{1}{r}\frac{\partial\hat{A}_r}{\partial\theta} = 0 ,$$
$$V = \frac{1}{r}\frac{\partial}{\partial r}(r\hat{A}_r) + \frac{1}{r}\frac{\partial\hat{A}_\theta}{\partial\theta} , \qquad (52)$$
$$W = \hat{A}_z .$$

We assume that V and W have the form

$$V = C_1\frac{\pi}{2b}J_n(\xi r)cos(n\theta) ,$$
$$W = CJ_n(\xi r)cos(n\theta) , \qquad (53)$$

which satisfy (51), provided

$$\begin{bmatrix} \kappa(\frac{\pi}{2b})^2 + \beta\xi^2 - \beta\omega^2/c_o^2 & -\delta\xi^2 \\ p(\frac{\pi}{2b})^2 & \xi^2 - \hat{n}^2\omega^2/c_o^2 \end{bmatrix}\begin{bmatrix} C_1 \\ C \end{bmatrix} = 0 . \qquad (54)$$

For nontrivial solutions, ξ^2 must satisfy

$$[\kappa(\frac{\pi}{2b})^2 + \beta\xi^2 - \beta c_o^{-2}\omega^2]$$
$$(\xi^2 - \hat{n}^2 c_o^{-2}\omega^2) + (\frac{\pi}{2b})^2 p\delta\xi^2 = 0 , \qquad (55)$$

and amplitudes have the ratio

$$C_1/C = R \equiv (\hat{n}^2\omega^2/c_o^2 - \xi^2)/p(\frac{\pi}{2b})^2 . \qquad (56)$$

By substituting (56) and (53) into (52) and solving for the components of vector potential, we have

$$\begin{aligned}\hat{A}_r(r,\theta) =& -\xi^{-1}\frac{\pi}{2b}RCJ_n'(\xi r)cos(n\theta)\ ,\\ \hat{A}_\theta(r,\theta) =& n\xi^{-1}\frac{\pi}{2b}RC\frac{1}{\xi r}J_n(\xi r)sin(n\theta)\ ,\\ \hat{A}_z(r,\theta) =& CJ_n(\xi r)cos(n\theta)\ .\end{aligned} \tag{57}$$

By substituting (57) into (26) and (27) and, in turn, into (4) and (11), respectively, we have

for $r < a$ and $|z| < b$:

$$\begin{aligned}E_r =& i\omega\xi^{-1}\frac{\pi}{2b}RCJ_n'(\xi r)cos(n\theta)\\ & cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ E_\theta =& -i\omega n\xi^{-1}\frac{\pi}{2b}RC\frac{1}{\xi r}J_n(\xi r)sin(n\theta)\\ & cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ E_z =& -i\omega CJ_n(\xi r)cos(n\theta)\\ & sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ B_r =& (-n\xi - n\xi^{-1}\frac{\pi}{2b}R\frac{p\pi}{2b})C\frac{1}{\xi r}J_n(\xi r)sin(n\theta)\\ & sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ B_\theta =& (-\frac{\pi}{2b}\frac{p\pi}{2b}R\xi^{-1} - \xi)CJ_n'(\xi r)cos(n\theta)\\ & sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ B_z =& 0\ ,\end{aligned} \tag{58}$$

For $r < a$ and $z > b$:

$$\begin{aligned}E_r =& i\omega\xi^{-1}\frac{\pi}{2b}RCJ_n'(\xi r)cos(n\theta)\\ & e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,\\ E_\theta =& -i\omega n\xi^{-1}\frac{\pi}{2b}RC\frac{1}{\xi r}J_n(\xi r)sin(n\theta)\\ & e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,\\ E_z =& -i\omega\frac{\pi}{2}(p\hat{n}^2-\beta R)CJ_n(\xi r)cos(n\theta)\\ & e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,\\ B_r =& [\bar{\eta}n\xi^{-1}\frac{\pi}{2b}R - \xi n\frac{\pi}{2}(p\hat{n}^2-\beta R)]\\ & C\frac{1}{\xi r}J_n(\xi r)sin(n\theta)\\ & e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,\\ B_\theta =& [-\bar{\eta}\xi^{-1}\frac{\pi}{2b}R - \xi\frac{\pi}{2}(p\hat{n}^2-\beta R)]\\ & CJ_n'(\xi r)cos(n\theta)\\ & e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,\\ B_z =& 0\ .\end{aligned} \tag{59}$$

For $r > a$, the governing equations for the time harmonic TM modes have exact form as those given by (49) to (53). Therefore, in similar manner, we obtain

$$\begin{aligned}U =& 0\ ,\\ V =& \frac{\pi}{2b}\bar{C}_1K_n(\bar{\xi}r)cos(n\theta)\ ,\\ W =& \bar{C}K_n(\bar{\xi}r)cos(n\theta)\ ,\end{aligned} \tag{60}$$

from which we have

$$\begin{aligned}\hat{A}_r(r,\theta) =& \bar{\xi}^{-1}\frac{\pi}{2b}\bar{R}\bar{C}K_n'(\bar{\xi}r)cos(n\theta)\ ,\\ \hat{A}_\theta(r,\theta) =& -\bar{\xi}^{-1}n\frac{\pi}{2b}\bar{R}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)\ ,\\ \hat{A}_z(r,\theta) =& \bar{C}K_n(\bar{\xi}r)cos(n\theta)\ ,\end{aligned} \tag{61}$$

where

$$\begin{aligned}&\bar{\kappa}(\frac{\pi}{2b})^2 - \bar{\beta}\bar{\xi}^2 - \bar{\beta}c_o^{-2}\omega^2 = 0\ ,\\ &\bar{R} = \frac{\bar{\xi}^2 + c_o^{-2}\omega^2}{p(\frac{\pi}{2b})^2}\ .\end{aligned} \tag{62}$$

The field components are, for $|z| < b$,

$$\begin{aligned}E_r =& -i\omega\bar{\xi}^{-1}\frac{\pi}{2b}\bar{R}\bar{C}K_n'(\bar{\xi}r)cos(n\theta)\\ & cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ E_\theta =& i\omega\bar{\xi}^{-1}n\frac{\pi}{2b}\bar{R}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)\\ & cos\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ E_z =& -i\omega\bar{C}K_n(\bar{\xi}r)cos(n\theta)\\ & sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,\\ B_r =& (-n\bar{\xi} + \bar{\xi}^{-1}n\frac{\pi}{2b}\bar{R}\frac{p\pi}{2b})\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)\end{aligned}$$

$$sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,$$

$$B_\theta = (p\frac{\pi}{2b}\frac{\pi}{2b}\bar{R}\bar{\xi}^{-1}-\bar{\xi})\bar{C}K_n'(\bar{\xi}r)cos(n\theta)$$
$$sin\frac{p\pi}{2}(1-\frac{z}{b})e^{i\omega t}\ ,$$

$$B_z = 0\ ,$$

(63)

and, for $z > b$,

$$E_r = -i\omega\bar{\xi}^{-1}\frac{\pi}{2b}\bar{R}\bar{C}K_n'(\bar{\xi}r)cos(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,$$

$$E_\theta = i\omega\bar{\xi}^{-1}n\frac{\pi}{2b}\bar{R}\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,$$

$$E_z = -i\omega\frac{\pi}{2}(p-\bar{\beta}\bar{R})\bar{C}K_n(\bar{\xi}r)cos(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,$$

$$B_r = [-\bar{\eta}\bar{\xi}^{-1}n\frac{\pi}{2b}\bar{R}-\bar{\xi}n\frac{\pi}{2}(p-\bar{\beta}\bar{R})] \qquad (64)$$
$$\bar{C}\frac{1}{\bar{\xi}r}K_n(\bar{\xi}r)sin(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,$$

$$B_\theta = [-\bar{\eta}\bar{\xi}^{-1}\frac{\pi}{2b}\bar{R}-\bar{\xi}\frac{\pi}{2}(p-\bar{\beta}\bar{R})]$$
$$\bar{C}K_n'(\bar{\xi}r)cos(n\theta)$$
$$e^{-\bar{\eta}(z-b)}e^{i\omega t}\ ,$$

$$B_z = 0\ .$$

Again, according to the uniqueness theorem[10] and for $\hat{H}_z = 0$, we require $\hat{H}_\theta$ and $\hat{E}_z$ be continuous at $r = a$. Substitution of (57) and (61) into (8) and implimentation of the continuity conditions leads to

$$\begin{bmatrix} J_n(\xi a) & K_n(\bar{\xi}a) \\ (-pR(\frac{\pi}{2b})^2\xi^{-1}-\xi)J_n'(\xi a) & (p\bar{R}(\frac{\pi}{2b})^2\bar{\xi}^{-1}-\bar{\xi})K_n'(\bar{\xi}a) \end{bmatrix}\begin{bmatrix} C \\ \bar{C} \end{bmatrix} = 0\ . \qquad (65)$$

The vanishing of the determinant of the coefficient matrix of (65) gives the frequency equation for the TM modes of order n.

VI. Frequency Spectra

In terms of the normalized frequency and wave numbers and the aspect ratio as they are defined by

$$\Omega = \frac{\omega}{\frac{\pi}{2b}c_o}\ ,\quad X = \frac{\xi}{\frac{\pi}{2b}}\ ,$$
$$\bar{X} = \frac{\bar{\xi}}{\frac{\pi}{2b}}\ ,\quad \bar{Y} = \frac{\bar{\eta}}{\frac{\pi}{2b}}\ , \qquad (66)$$
$$\hat{r} = \frac{\pi}{2}\frac{a}{b}\ ,$$

the frequency equation (47) for TE modes can be written as

$$F_1(\Omega, a/b; \hat{n}, n, p) = k\frac{J_n(X\hat{r})}{J_n'(X\hat{r})} + \frac{\bar{X}K_n(\bar{X}\hat{r})}{XK_n'(\bar{X}\hat{r})} = 0\ , \qquad (67)$$

where

$$X^2 = \frac{1}{\alpha}(\beta\Omega^2-\kappa)\ ,$$
$$\bar{X}^2 = \frac{1}{\bar{\alpha}}(\bar{\beta}\Omega^2-\bar{\kappa})\ , \qquad (68)$$
$$\bar{Y}^2 = X^2-\Omega^2\ .$$

From (67) we see that the value of the resonance frequency Ω, for a dielectric disk of given aspect ratio a/b and refractive index $\hat{n}$, depends on the numeral m ($m = 1, 2, 3, \cdots$) of the mth zero root of (67), the nth-order of the Bessel functions ($n = 0, 1, 2, \cdots$), and the pth-order ($p = 0, 1, 2, \cdots$) of the governing equations (9). Therefore, we may denote the corresponding mode to the resonant frequency by TE_{nmp} with n, m, p indicating the spatial variations of the mode in the azimuthal (θ), radial (r), and axial (z) directions, respectively.

In the same manner, the frequency equation (65) for the TM_{nmp} modes can be expressed by

$$F_2(\Omega, a/b; \hat{n}, n, p) = \frac{J_n(X\hat{r})}{J_n'(X\hat{r})} + \frac{\bar{X}(pR+X^2)K_n(\bar{X}\hat{r})}{X(p\bar{R}+\bar{X}^2)K_n'(\bar{X}\hat{r})} = 0\ , \qquad (69)$$

where

$$(\kappa+\beta X^2-\beta\Omega^2)(X^2-\hat{n}^2\Omega^2)+p\delta X^2 = 0\ ,$$

$$\bar{\kappa} - \bar{\beta}\bar{X}^2 - \bar{\beta}\Omega^2 = 0 \ ,$$

$$R = (\hat{n}^2\Omega^2 - X^2)/p \ ,$$

$$\bar{R} = (\Omega^2 + \bar{X}^2)/p \ ,$$

$$p = 1, 2, 3, \cdots (p \neq 0) \ . \tag{70}$$

From the roots computed from (67) and (69), transcendental equations, frequency spectra or $\hat{\Omega}$ vs. a/b curves are plotted in Figs. 2 and 3 for the TE_{nmp} and TM_{nmp} modes, respectively.

We note that

$$\hat{\Omega} \equiv \hat{n}\Omega = \frac{\omega}{\frac{\pi}{2b}c} \ , \quad c = \frac{1}{\sqrt{\epsilon\mu_o}} \ . \tag{71}$$

Hence, $\hat{\Omega}$ is the dimensionless frequency normalized with respect to c, the speed of light in the dielectric disk. The reason for choosing $\hat{\Omega}$ as the coordinate is that the computational results presented as $\hat{\Omega}$ vs. a/b curves in Figs. 2 and 3 are independent of the values of $\hat{n}$, the refractive index, and, therefore, applicable to dielectric disks of any $\hat{n}$.

It is seen in Figs. 2 and 3 that, for each frequency branch, the value of $\hat{\Omega}$ decreases as a/b increases and it approaches to a certain constant value as a/b approaching infinity. This characteristic of the frequency branch of EM waves is very much alike to that of the frequency spectrum for acoustic waves in a circular elastic disk.[11]

We denote $\hat{\Omega}_{nmp}$ as the frequency of the TE_{nmp} mode and we see in Fig.2 $\hat{\Omega}_{nmp} < \hat{\Omega}_{(n+1)mp}$ and $\hat{\Omega}_{nmp} < \hat{\Omega}_{nm(p+1)}$. In Fig. 3, we also see $\hat{\Omega}_{nmp} < \hat{\Omega}_{nm(p+1)}$, but $\hat{\Omega}_{0mp} > \hat{\Omega}_{1mp}$. Yee made a good explanation about this phenomenon by examining the magnitude of the roots of $J_n(X\hat{r}) = 0$ and $J_n'(X\hat{r}) = 0$ $(n = 0, 1)$ which are the approximations of (67) and (69), respectively, by the use of the perfect magnetic face conditions at $r = a$.[2]

Computations of circular frequency f as a function of plate thickness $2b$ are made for circular disks with a constant diameter $2a = 0.325$ in. and $\epsilon_r = 95$. The predicted curves, in solid lines, are compared with Cohn's experimental data[12] which are denoted by dashed lines (f_1 and f_2) and hollowed circles as given in Fig. 4.

We see in Fig. 4. that TE_{010} and TM_{111} are the frequency branches corresponding to the measured f_1 and f_2, respectively. It is also interesting to see that the hollowed circles are very close to the envelope of the higher branches (TE_{110}, TM_{011}, TE_{011}, or TM_{112}), since they are the next higher resonances observed by Cohn.[12]

Predicted f vs. $2b$ for the TE_{011} mode of disks with $2a = 0.16$ in. and $\epsilon_r = 113$ is shown as the solid line in Fig. 5 in which Chow's experimental data (hollowed circles), his prediction (dashed lines),[13] and prediction according to Yee's solution[2,14] (dash and dot line) are slso given for comparison.

It is seen that our prediction is very close to the experimental data and prediction by Chow. We note that the method of solution by Chow appears to give accurate prediction for TE and TM modes except for the TE_{010} mode, which, according to his conclusion, can not exist.[13] It seems to be likely due to the fact that the satisfaction of the continuity conditions at the boundaries of "corner regions", i.e. at $r = a$, $z > b$ and $r = a$, $z < -b$, is more crucial for the fundamental mode ($p = 0$) than for the higher ones.

Comparison of predicted resonance frequencies with Yee's measured values for cylindrical resonators of $SrTiO_3$ for various a/b ratios and for various modes is made as shown in Table I.[2,13]

In summary, by employing the recently derived two-dimensional field equations,[9] closed form solutions are obtained for which continuity conditions at the boundaries of the disk and at the "corner regions" are accommodated.

Frequency equations for the TE_{nmp} and TM_{nmp} modes are obtained in analytical form. Resonance frequencies as a function of the diameter-to-thickness ratio are computed and compared with experimental data by Cohn,[12] Chow,[13] and Yee[2,14] with good agreement.

Acknowledgements

This work was supported by a grant (Contract No. DAAL 03-90-G-0079) from the U. S. Army Research Office.

References

1. R. D. RICHTMYER, "Dielectric resonators," J. Appl. Phys., vol. 10, pp. 391-398, June 1939.
2. H. Y. YEE, "An investigation of microwave dielectric resonators," M. L. Report No. 1065, Microwave Lab., Stanford University, Stanford, Calif., July 1963.
3. M. GASTINE, L. COURTOIS AND J. L.

DORMANN, "Electromagnetic resonances of free dielectric spheres," IEEE Trans. Microwave Theory Tech. (1967 Symposium Issue), vol. MTT-15, pp. 694-700, Dec. 1967.

4. S.D. POISSON, "Mémoire sur l'équilibre et le mouvement des corps élastiques," Mém. Acad. Sci., Paris, Ser. 2, 8 pp. 357-570, 1829.
5. A. L. CAUCHY, "Sur l'équilibre et le mouvement d'une plaque élastique dont l'élasticité n'est pas la même dans tous les sens," Excercices de Mathé matique 4, pp. 1-14, 1828.
6. R. D. MINDLIN, "Influence of Rotary Inertia and Shear on Flexural Motions of Isotropic, Elastic Plates," J. Appl. Mech., 18, pp. 31-38, 1951.
7. R. D. MINDLIN, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., XIX, pp. 51-61, 1961.
8. P. C. Y. LEE, S. SYNGELLAKIS, AND J. P. HOU, "A Two-Dimensional Theory for High-Frequency Vibrations of Piezoelectric Crystal Plates with or without Electrodes," J. Appl. Phys., 61 (4), pp. 1249-1262, 1987.
9. P. C. Y. LEE AND J. S. YANG, "Two-dimensional equations for guided EM waves in dielectric plates surrounded by vacuum," in Proc. of 45th Ann. Symp. on Freq. Control, 1991, pp. 156-165.
10. P. C. Y. LEE, unpublished notes.
11. D. C. GAZIS AND R. D. MINDLIN, "Extensional vibrations and waves in a circular disk and a semi-infinite plate," J. Appl. Mech. 27, pp. 541-547, Sept. 1960.
12. S. B. COHN, "Microwave bandpass filters containing high-Q dielectric resonators," IEEE Trans. Microwave Theory and Techniques, vol. MTT-16, pp. 218-227, April 1968.
13. K. K. CHOW, "On the solution and field pattern of cylindrical dielectric resonators," IEEE Trans. Microwave Theory and Techniques (Correspondence), vol. MTT-14, p. 439, September 1966.
14. H. Y. YEE, "Natural resonant frequencies of microwave dielectric resonators," IEEE Trans. Microwave Theory and Techniques (Correspondence), vol. MTT-13, pp. 256, March 1965.

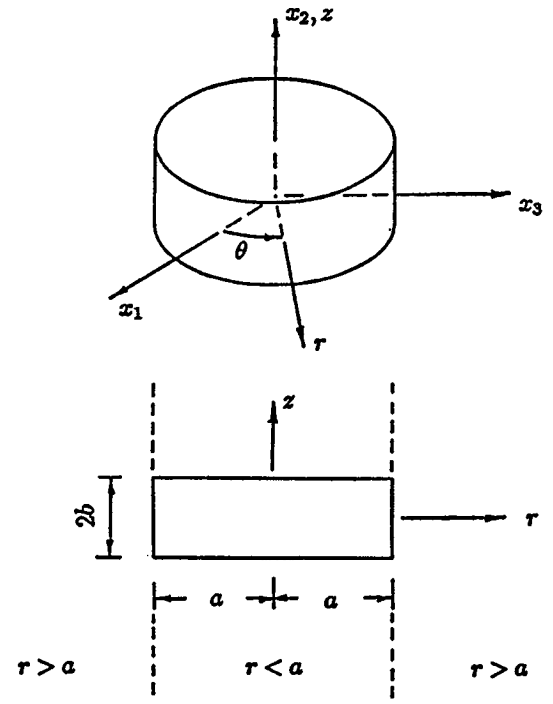


Fig. 1 A Circular dielectric disk embedded in free space.

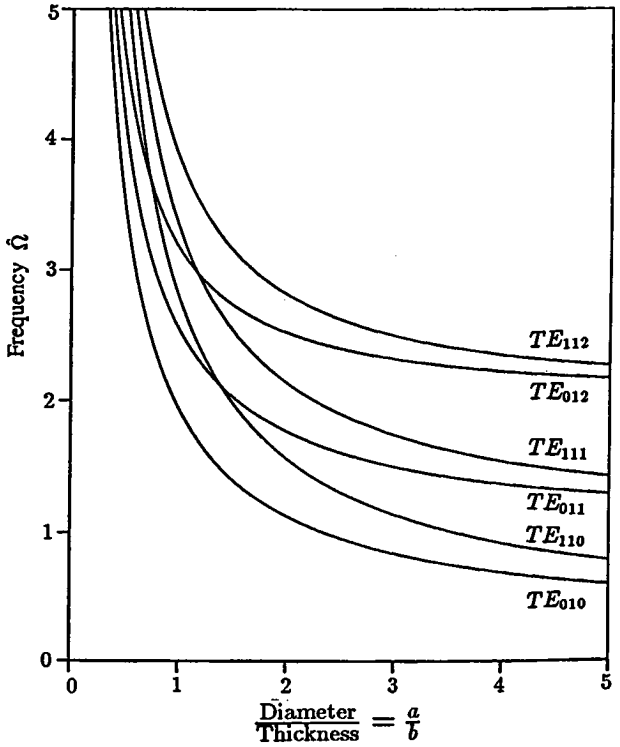


Fig. 2 Frequency spectrum of TE modes in a circular dielectric disk.

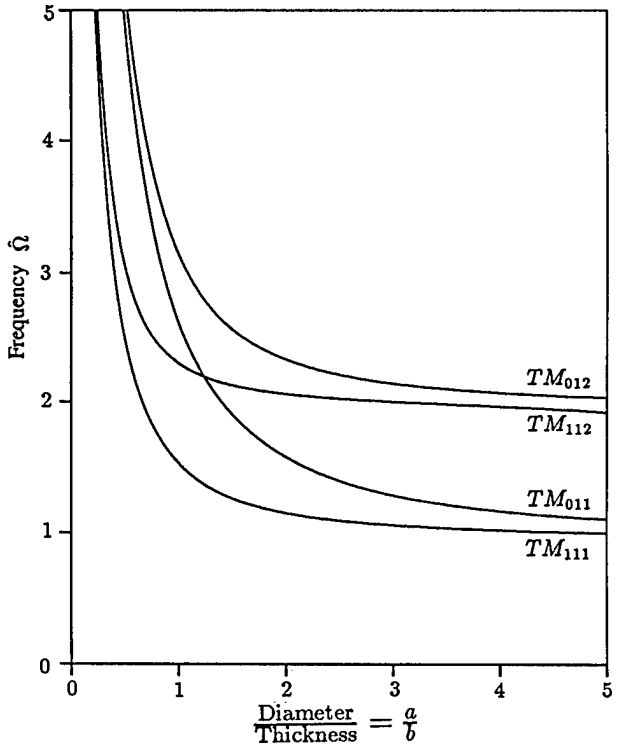


Fig. 3 Frequency spectrum of TM modes in a circular dielectric disk.

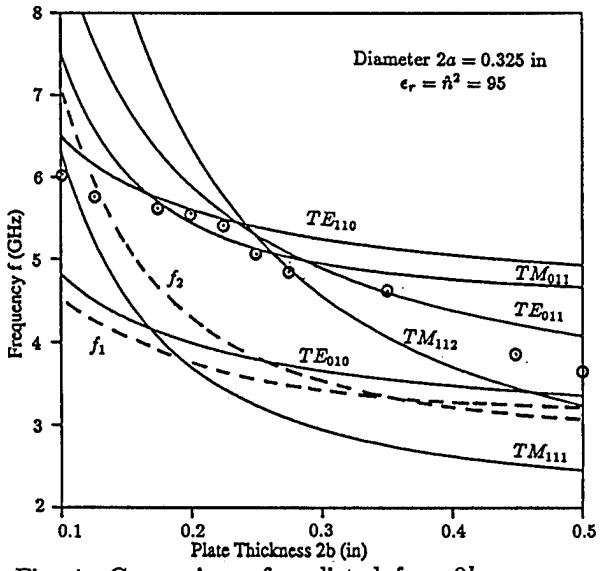


Fig. 4 Comparison of predicted f vs. $2b$ curves, in solid lines, with Cohn's experimental data denoted by dashed lines (f_1 and f_2) and hollowed circles for circular disks with $2a = 0.325$ in. and $\epsilon_r = 95$.

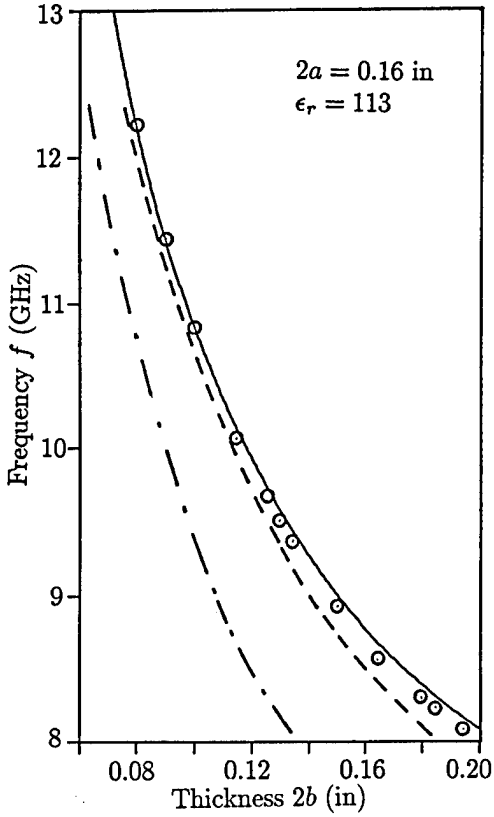


Fig. 5 Comparison of presently predicted f vs. $2b$ for TE_{011} mode with Chow's experimental data (hollowed circles), Chow's prediction (dashed lines), and prediction according to Yee's solution (dash and dot line) for disks with $2a = 0.16$ in. and $\epsilon_r = 113$.

TABLE I. COMPARISON OF PREDICTED FREQUENCIES WITH YEE'S MEASURED VALUES FOR CYLINDRICAL RESONATORS OF $SrTiO_3$ ($\epsilon/\epsilon_o = 279$)

Mode	$\frac{\text{Diameter}}{\text{Thickness}} = \frac{a}{b}$	Radius a (mils)	Resonance Frequencies f (GHz)	
			Measured by Yee	Present Prediction
TE_{010}	0.965	32.8	9.28	10.38
	1.166	32.3	9.71	10.95
	2.301	35.2	10.43	11.87
	3.244	35.2	11.39	12.92
	1.240	15.5	18.50	23.02
	1.000	13.0	21.90	26.46
TE_{110}	0.965	32.8	13.83	14.98
	1.166	32.3	14.24	15.65
	2.301	35.2	14.34	16.49
	3.244	35.2	15.42	17.56
TM_{111}	0.965	32.8	9.30	8.15
	1.166	32.3	10.07	9.04
	2.301	35.2	13.96	12.90
	3.244	35.2	18.15	17.18
	1.240	15.2	17.51	19.40
	1.000	13.0	20.40	20.90
TM_{011}	0.965	32.8	13.66	13.15
	1.175	32.5	14.20	13.68
	2.301	35.2	16.23	15.73
	3.244	35.2	20.40	18.89

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ENERGY TRAPPING RESONATORS AND INTEGRATED FILTERS USING THE AT CUT OF QUARTZ LIKE MATERIALS.

J.Détaint*, J.Schwartzel*, C.Joly*, A.Zarka**,
B.Capelle**, Y.Zheng**, Y.Toudic***, E.Philippot****

*CNET dept PAB/BAG/MCT/CMM. 196, Av.H.Ravera 92220 Bagneux France.
**L.M.C.P. Université Paris VI 4, place Jussieu 75005 Paris France.
***CNET dépt LAB/OCM/MPA route de Tregastel 22301 Lannion France.
****L.P.C.M.S. Université du Languedoc. Place E.Bataillon Montpellier France.

Abstract:

Quartz and its analogues are the only known materials that lead to devices having simultaneously a very high thermal frequency stability in a large temperature range and a very small acoustic dissipation. The square of the coupling coefficient is approximatively doubled from AT quartz to AT berlinite and again doubled from AT berlinite to AT GaPO4 so that it is possible to achieve a large range of applications in the fields of frequency generation or filtering, using these three materials.

In a first part of this communication, the properties of the AT cuts of these three materials are compared on the basis of the one dimensionnal theory of bulk wave propagation in a piezoelectric material (velocity of the shear waves, coupling coefficient, and thermal behaviour). Then a more precise evaluation of the properties of AT cut resonators using the quartz analogues is made in considering the computed properties of energy trapped resonators. They were obtained using a three dimensional model previously reported and based upon a solution of the equations obtained in the theory of the essentially thickness modes (H.F.Tiersten). Using the computed equivalent schemes, the bandwidth achievable at 20 MHz are calculated for 4 poles Jaumann filters having a Chebishev response with 1 dB ripple and a maximal termination impedance of 2 kohm. The obtained results indicate that berlinite and gallium phosphate permit to obtain filters with much larger bandwidth than with quartz. They suggest also that the use of new designs (resonators with elliptical electrodes and/or grooved electrodes, integrated filters etc...) will permit to lower further the equivalent inductance and/or the influence of the stray capacitances. Measured values of the elements of the equivalent scheme of resonators, are compared with computed ones. A very good agreement is observed for berlinite whereas the experimental results obtained for gallium phosphate are more favourable than the computed ones by a factor 1.1 to 1.2, probably due to still underestimated values of the piezoelectric constants of this material.

In a second part, the principle of a tridimensionnal model of the two poles integrated filters is presented. As for the case of the single resonators, this model is founded upon a semi-algebraical solution of the partial derivative equations to the essentially thickness modes proposed by H.F. Tiersten and coworkers. The solution use linear combinations of algebraical solutions of the partial derivative equations that verify the continuity conditions between the different regions of the resonators at a discrete number of points. The electrical equivalent scheme of these four poles structures is derived from the electromechanical analysis under the form of a symmetrical lattice.

We conclude to the important interest of the high coupling quartz analogues for filter and VCXO applications and to the interest to use optimized resonators and filter designs.

1 Introduction:

Since 15 years many works considering the obtainement, the properties and the applications of several quartz like materials having enhanced piezoelectric properties were reported. Berlinite was the first analogue of quartz to be synthesized [1][2][3][4] and evaluated [5]. Berlinite crystals of good quality and sufficient size are now available [6] to permit detailed investigations of devices and the obtainement of industrial prototypes. The obtainement of gallium phosphate crystals having significative dimensions and perfection was reported in the recent years [7] together with the measured constants of this material and the first evaluations of devices [8] [9]. A major advance in the obtainement of large GaPO4 crystals having an improved quality is reported [10] in these proceedings, together with new results of characterization of devices.

To determine precisely the respective potentialities of the quartz analogues in view of an application to I.F. filtering in G.S.M. sets, an investigation was made of the properties of filter type resonators using these materials. In this communication we report some of the results obtained that permit a comparison of the properties of AT resonators and filters using quartz berlinite and GaPO4.

2 Properties of the AT cuts of Quartz, AlPO4 and GaPO4.

In table I the computed or observed basic properties of the piezoelectrically excited slow shear modes existing in the AT cut of quartz, aluminium phosphate and gallium phosphate are presented. The values given for quartz were computed from the constants of James [11] by a one dimensional model. For aluminium and gallium phosphate, they are the results of recent experiments made at CNET. The most remarkable facts to notice in table I are the decrease of the shear velocity from quartz to gallium phosphate and the increase of the dielectric permittivity and mostly of the coupling coefficient. The exact angular positions of the AT cuts in berlinite [12] and GaPO4 [7] [10] are still to be determined as accurately as in quartz, but, since the quantities reported in table I have slow angular variations, they are believed to be accurate to about 5% for berlinite while they represent for GaPO4 the most accurate evaluation presently available. In this latter case, the comparison of computed and experimental values of resonator properties will give further informations on their accuracy (see § 4).

The most important property of quartz like materials is to possess a cut giving a very nearly pure third order thermal behaviour ("true" AT cut). This means that for a certain cut and a certain reference temperature, the 1st and the 2nd order temperature coefficients of the frequency vanish simultaneously. With a good approximation, the variations of the resonance

0-7803-0476-4/92 $3.00 © 1992 IEEE

Table I properties of the AT cut of quartz like materials.

	QUARTZ	AlPO4	GaPO4
V (m/s)	3321.	2942.*	2543.*
$\bar{C}$ (10^9N.m^{-2})	2.921	2.285*	2.310*
Eps_r^S	4.499	4.917*	6.119*
K (%)	8.8	11.*	16.*

frequencies can then be described by the relation:
$\frac{fr(T)-fr(T_{inf})}{fr(T_{inf})} = Tf^{(3)}_{inf}.(T-T_{inf})^3$ where: $Tf^{(3)}_{inf}$ is the third order T.C. referenced to the inflexion temperature T_{inf}.

In table II we compare the third order temperature coefficients and the inflexion temperatures for these three materials. The values given for GaPO4 are still very tentative since they were obtained from measurements made for two near AT cuts with the assumption that as in quartz and berlinite, the 3rd order temperature coefficient has small angular variations in the vicinity of the AT cut. They are to be taken as a first evaluation that will be perfected in the near future. The results in table II indicate a very similar thermal behaviour of quartz and of its analogues but centred at different temperatures. This can be already verified for berlinite (figure 1) by a direct comparison to the variations known for AT quartz.

Table II: 3rd order temperature coefficients of the resonance frequency.

	Quartz	Berlinite	GaPO4
$Tf^{(3)}$(°C^{-3})	$9-10.10^{-11}$	$8-12.10^{-11}$	$4-16.10^{-11}$?*
T_{inf} (°C)	25-28°	62-67°	40-50°? *

*1st estimations from near AT results

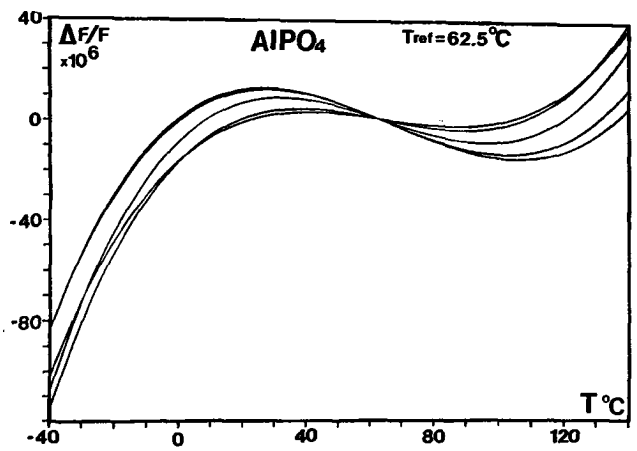


Figure 1: Thermal variations of fr near the AT cut of berlinite.

3 Comparison of energy trapping properties and of energy trapping devices:

Since the properties of real resonators cannot be well described by the one dimensional theory, we have further compared these three materials by an evaluation of the properties of energy trapping resonators made with them. For this purpose a model [12] based upon the theory of H.F.Tiersten [13] was used. This theory that was experimentally verified in many instances, postulate that the thickness variations of the displacements are essentially similar to that of the one dimensional modes whereas the lateral variations are governed by the equation:

$$M_n \frac{\partial^2 \tilde{u}_1^n}{\partial x_1^2} + P_n \frac{\partial^2 \tilde{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 C^*}{4h^2}\tilde{u}_1^n + \rho\omega^2 \tilde{u}_1^n =$$

$$\rho\omega^2(-1)^{(n-1)/2}\frac{e_{26}4V_0e^{j\omega t}}{c^{(1)}n^2\pi^2} \qquad \{1\}$$

In this equation, $\tilde{u}_1(x_1,x_3,t)$ is the lateral dependence of $\hat{u}_1^n(x_1,x_2,x_3,t) = \tilde{u}_1^n \sin(n\pi x_2/2h)$.. $\hat{u}_1$ arises from a transformation of the usual component of the displacement u_1 and of ϕ the potential , made to replace the inhomogeneous boundary conditions $\phi = \pm V_0 e^{j\omega t}/2 at x_2 = \pm h$ by the homogeneous conditions $\hat{\phi} = 0$. The inhomogeneous term of the equation results from this transformation. c* is either $\bar{c}^{(1)}$ for the unelectroded part of the resonator or $\hat{c}^{(1)}$ for the electroded part. $\bar{c}^{(1)}$ Is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix see Table I). $\hat{c}^{(1)} = \bar{c}^{(1)}(1 - 8k^{(1)2}/n^2\pi^2 - 2\hat{R})$ is a constant that includes the electrical effect of the metallization. $4k^{(1)2}/n^2\pi^2$ is approximatively the relative frequency lowering due to the electrical effect of the metallization, $\hat{R}$ is the mass loading. $c^{(1)} = \bar{c}^{(1)}(1-k^{(1)})^2$ Is a "pseudo ordinary elastic constant" which is equal to C_{66} for the Y rotated cuts. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode, this quantity is equal to k_{26} (noted k in Table I) for the Y rotated cuts . M , P are functions of the material constants and of the plate orientation that can be obtained as described in reference [13].

The eigen modes (at V=0) of the resonator are obtained by the simultaneous solution of the equations for the electroded and unelectroded regions taking into account the continuity conditions at the electrode edge. The forced modes are computed as a linear combination of the eigen modes verifying the inhomogeneous equation. At V=0 the homogeneous form of the approximate equation for the two regions reduces to :

$$M_n u^n_{1,11} + P_n u^n_{1,33} + 4\pi^2\rho(f^2 - f^{*2})u_1^n = 0 \qquad \{2\}$$

with : $f^* = f_{ce1} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$ for the electroded region. Or:

$f^* = f_{cl} = \frac{n}{4h}\sqrt{\frac{\bar{c}^{(1)}}{\rho}}$ for the unelectroded region.

Using the coordinate transformations :

$X_1 = \left(\frac{C^*}{M_n}\right)^{1/2}.\gamma.x_1; X_3 = \left(\frac{C^*}{P_n}\right)^{1/2}.\gamma.x_3$; $X_1 = r\cos t; X_3 = r\sin t$

with: $\gamma^4 = \frac{n^2\pi^2}{(2h)^4}$

where: $c^* = \hat{c}^{(1)}$ for the electroded region and $c^* = \bar{c}^{(1)}$ for the unelectroded region. We obtain the equation :

$$u_{1,rr}+\frac{1}{r}u_{1,r}+\frac{1}{r^2}u_{1,\theta\theta}+Au_1=0 \qquad \{3\}$$

$$Where: A=\frac{n\pi}{4}\left(\frac{f^2-f^{'2}}{f^{'2}}\right) \qquad \{4\}$$

that can be separated as $u_1(r,t)=R(r).T(t)$ and solved as described in [12] to obtain the eigen modes from which the forced modes are computed.

The important parameters of equations {1} to {4} for the fundamental mode of the AT cut of the quartz like materials were computed from the published constants corrected for berlinite and GaPO4 by the results given in table I. They are given in table III:

table III: Energy trapping constants for the AT cuts of the quartz like materials.

	M_1	P_1	$\overline{C}$	M_1/P_1
Quartz	11.008	6.922	2.921	1.261
AlPO4	8.718	5.084	2.285	1.309
GaPO4	7.663	3.805	2.310	1.419

Beside the "elastic constant" like variations of the coefficients between quartz and GaPO4, two important observations are to be made about the data presented in table III:
-First, the ratio $\sqrt{\frac{M_n}{P_n}}$ that was shown [14] to represent the "lateral anisotropy" of the plate increases notably from quartz to GaPO4.
-The second fact to be noticed is that the interval (f_{cl}-f_{ce}) in which trapped mode can propagate increase also greatly from quartz to GaPO4, due to the large increase of the electrical frequency lowering $\left(\frac{4k^2}{\pi^2}\right)$. So that at equal geometry, and mass loading, more and more modes are excited as the coupling coefficient increases. This fact leads to diminish the electrode dimensions from quartz to GaPO4, if it is necessary to have a response free from anharmonics or without too many intense anharmonics.

Using the model [12] we have computed the equivalent scheme of 20.0 MHz quartz, AlPO4 and GaPO4 resonators with circular electrodes having in each case a diameter determined so that the response has no symmetrical anharmonics (filter type resonators) with a mass loading of 0.5%. The computed diameter were rounded down to the nearest tenth of mm, and the equivalent schemes of the resonators were computed for these dimensions. The results are presented in table IV. The corresponding computed modes are represented at the same scale in figure 2. The results in table IV and in figure 2 indicate (as previously said) that smaller and smaller electrodes dimensions are required for AlPO4 and then GaPO4 than for quartz. The inductance decreases from quartz to GaPO4, the value of C_0 decreases also due to the reduction of the electrode area, although the dielectric permittivity increases. As expected, the C_0/C_1 capacitance ratio decreases strongly from quartz to GaPO4.

Table IV: Equivalent scheme of resonators designed to have no symmetrical anharmonics.

(max. diameter to have no symmetrical anharmonics with R=0.5%)

material / diameter mm	motional inductance L (mH)	motional capacitance C_1 (fF)	static capacitance C_0 (pF)	C_0/C_1
Quartz 2.5	5.045	12.537	2.367	188
AlPO4 2.	4.123	15.366	1.862	121
GaPO4 1.2	3.761	16.833	0.973	57.8

On table V the bandwidth of 4 poles Jaumann filters having a 1.dB ripple Chebishev response achievable with equal termination resistances of 2000 Ohm and the resonators of table IV are given. As expected, the larger bandwidth is achieved with GaPO4, while berlinite give also an interesting value. The corresponding computed responses are represented in figure 3.

Table V: Achievable bandwidth for 4 pole Jaumann filters.

Max B.W. with suppression of the symmetrical anharm.

	AT Quartz	AT Berlinite	AT GaPO4
B.W. (kHz)	107.6	131.7	145.

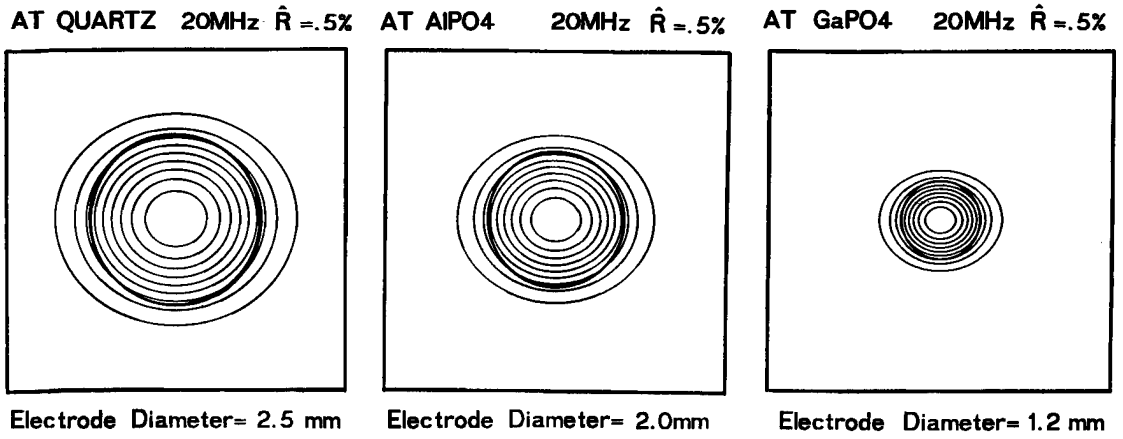


Figure 2: Computed modes of the resonators considered in table III.

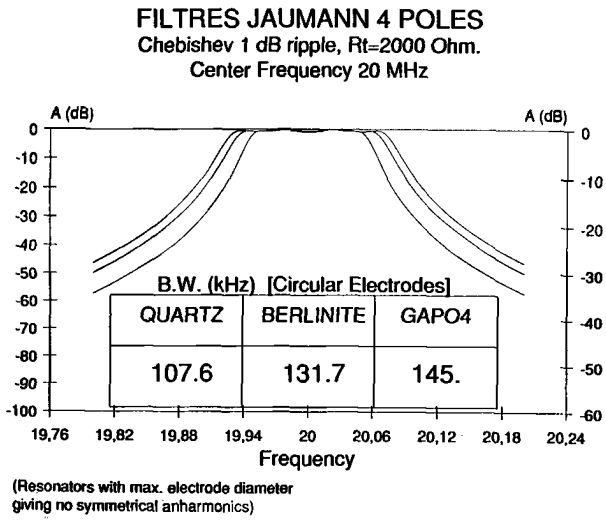


Figure 3: Computed response curve of 4 poles Jaumann filters made using the resonators considered in table IV.

Particularly at V.H.F. frequencies, the small imperfections that are inherent to a practical elaboration process (electrodes misalignment or asymmetry, slight defect of parallelism etc...) can induce the excitation of antisymmetrical modes that are normally unexcited (eigen modes non electrically active by charge compensation on the electrodes with perfect symmetry conditions). These modes can have an as detrimental effect as those of the usual symmetrical anharmonics on a filter response. In figure 4a, we have represented the computed eigen frequencies of the two antisymmetrical modes that exist in the resonators previously considered. It can be easily observed (figure 4b) that if they are electrically active, they can impair greatly the response curve of the filters.

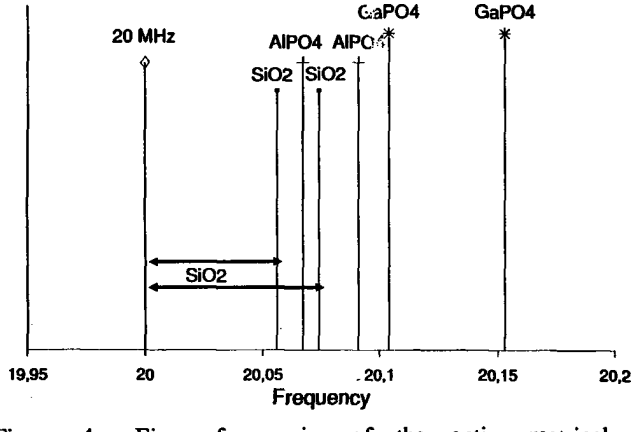


Figure 4a: Eigen frequencies of the antisymmetrical anharmonics.

It was experimentally verified [14] that for small imperfections they appear at frequencies very close to the corresponding eigen frequencies computed for a perfect geometry. So it can be interesting in some practical applications to consider resonators designed so that no antisymmetrical modes can propagate. We have computed for quartz, AlPO4 and GaPO4 20.0 MHz resonators (mass loading= .5%) the electrode

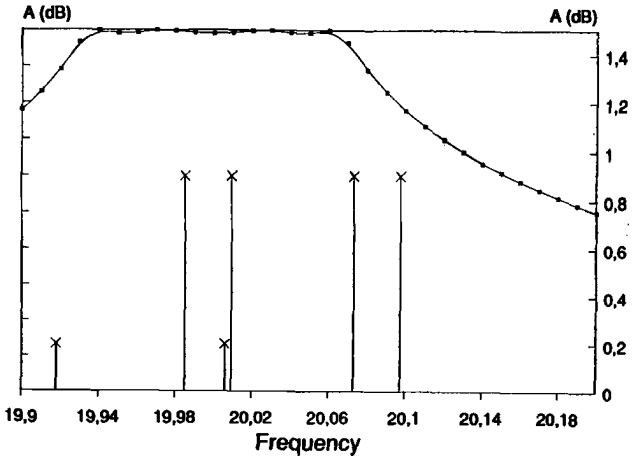


Figure 4b: Position of the anharmonics in the response (berlinite filter).

dimensions below which they do not appear. For these dimensions rounded down to the nearest 0.1mm (or .05mm for GaPO4) the equivalent scheme where computed and are displayed in table VI, where, we can observe that in these conditions, AT berlinite seems to give the lowest inductance (see end of §4). The C_0/C_1 capacitance ratio of GaPO4 is still from far the lowest although somehow degraded. Due to the electrode dimensions, the static capacitance takes smaller and smaller values from quartz to GaPO4. These computed values indicate the possibility of a greater sensitivity to stray capacitances for the high coupling quartz analogues in these conditions. The achievable bandwidth computed with the same specifications as previously considered are given in table VII. They are lower than those of table V, as expected, but still sufficient for many applications particularly if a transposition is made to V.H.F. frequencies where such a "conservative" design can be very useful.

Table VI: Equivalent scheme of 20.0 MHz resonators designed so that the no antisymmetrical nor symmetrical anharmonics can propagate.

(max. diameter to suppress also the antisymmetrical anharmonics)

material / diameter mm	motional inductance L (mH)	motional capacitance C1 (fF)	static capacitance Co (pF)	Co/C1
Quartz 1.6	13.66	4.612	0.969	210
AlPO4 1.3	10.72	5.879	0.786	133
GaPO4 0.75	10.89	5.766	0.380	65.9

Table VII: Achievable bandwidth for 4 poles Jaumann filters using the resonators of table VI.

Max B.W. with suppression of the symmetrical anharm.

	AT Quartz	AT Berlinite	AT GaPO4
B.W. (kHz)	107.6	131.7	145.

4 Comparison of computed and experimental results, discussion:

The good accuracy of the model was previously verified for quartz for many resonator designs [12], [14]. For berlinite, without correction of the constants, the agreement between computed and experiment results was observed to be reasonably good for the frequencies and only fair for the equivalent scheme [12]. A much better agreement is obtained using the values of table I in the computations, the other rotated constants included in M and P being computed from the published data. In Figure 5a the response curve of a 20.523 MHz berlinite resonator having a thickness of 71.4 microns (+-.1) silver electrodes of 1.8mmm diameter and a mass loading of approximatively .5% (+.2%,-1%) is represented. The experimental inductance is 4.56 mH (the mean experimental value for 4 similar resonators is: 4.59 mH); the computed frequency is 20.543 MHz and the computed inductance is 4.73mH. In figure 5b the response curve of a resonator very similar to those of figure 5a is represented. In this case, a condition (very slight unparallelism) exist permitting the excitation of antisymmetrical anharmonics.

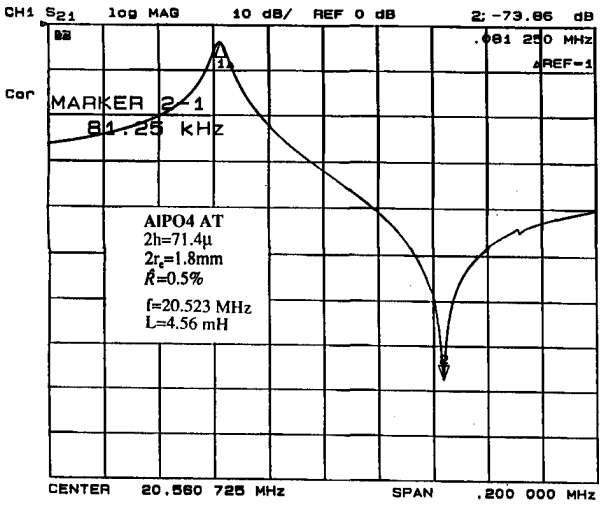


Figure 5a: Response curve of a 20MHz berlinite resonator.

For GaPO4 only resonators having fundamental frequencies in the range 5.0 to 7 MHz were studied. The response curve of a near AT GaPO4 resonator is represented in figure 6. The experimental parameters are: thickness= 208.97 microns, electrode diameter : 3.5mm, mass loading: 0.9%, resonance frequency: 5.9952 MHz, measured inductance: 12.79 mH (the mean experimental inductance measured with 5 similar samples is 13.16 mH). The computed frequency is 6.0035 MHz and the computed inductance is 16.329 mH. In the calculation we have corrected the constants by the measured values of table I (without correction the computed frequency is 5.927158 MHz and the computed inductance is 16.26 mH). The slight discrepancy between the experimental and computed values of the inductance indicates that at least one parameter of table I and III is not accurate enough (we suspect that the coupling coefficient k_{26} may still be somehow larger than the value presently measured and given in table I). This difference indicates also that the computed values above given underestimate some how the actual possibilities of AT GaPO4 resonators. Therefore, to obtain a more accurate evaluation of the possibilities of these resonators, the computed values, previously given of the inductance and bandwidth are to be divided (respectively multiplied) by a factor of about 1.2 .

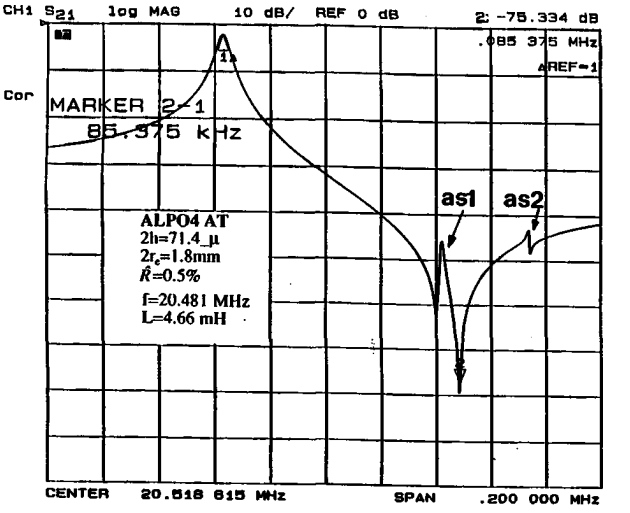


Figure 5b: Response curve of a 20 MHz berlinite resonator with antisymmetrical anharmonics.

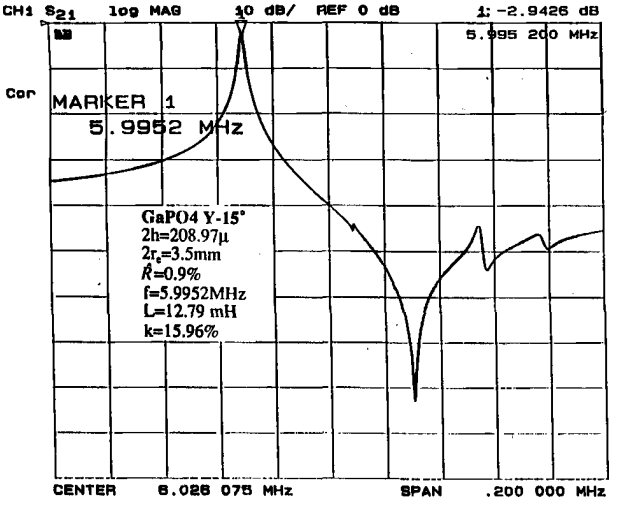


Figure 6: Experimental response of a near AT GaPO4 resonator.

The first observation resulting of the computations (corrected as described above for GaPO4) is the fact that the quartz like materials have very interesting properties for filtering and frequency generation applications (thermal stabilities, achievable bandwidth or shift of oscillators). With the first design considered in § III , the AT resonators using GaPO4 would lead to filters having very important bandwidth and to oscillator with large shifts.

The second observation is to notice that, as previously shown [12] [15] for berlinite, much more interesting designs than those using circular electrodes can be found for the devices that lower the inductance values, allows larger bandwidth for the filters and reduce the influence of stray capacitance in circuits. One of them, the use of electrode geometries respecting the lateral anisotropy is particularly efficient with materials presenting a value of $\sqrt{\frac{M_n}{P_n}}$ differing notably from unity, and, is very simple to apply. The principle of such new designs is recalled in figure 7.

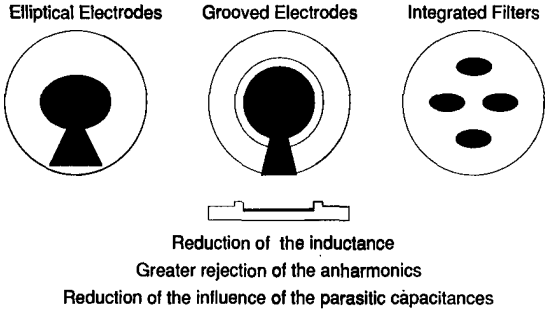


Figure 7: Principle of new designs for resonators and filters.

5 Principle of the modelization of integrated filters:

We consider plane structures with two in plane perpendicular axis of symmetry, having 3 regions with different cut-off frequencies: the electroded region divided in two sets of electrodes characterised by a mass loading $\hat{R}_2$ that are separated by a "coupling" region characterised by a mass loading $\hat{R}_1$ lower than $\hat{R}_2$. The coupling region is supposed to be loaded with an insulating layer so that no electrical lowering is to be considered for this region. The external region surrounding the two previous ones is unelectroded to ensure that an energy trapping behaviour can occur.

Depending on the distance between the electrodes and several other parameters, the resonators can be from strongly to negligibly coupled. This corresponds respectively to the limit cases of the two poles monolithic filters and to the case of acoustically isolated resonators integrated on the same plate. We consider only structures that are symmetrical with respect to the in plane perpendicular axis that lead to the cancellation of the mixed partial derivatives in equation {1} [13]. One fourth of a structure which is rather of the first kind is represented in figure 8, whereas the structure of figure 9 is rather of the second kind.

Several models of monolithic filters were already presented [17], [18], [20], [21], [22]. Our objective here is to obtain an accurate modelization of the cases where either the coupling between resonators is very weak (isolated resonators) or the electrode geometries are quite different from the usual rectangular one. As for the case of the resonators [12] we use the theory of essentially thickness modes [13], so that in each region the lateral dependence of u_1 is governed by equation {1} (§ 3). For each region, specific values of the mass loading and cut-off frequencies are considered in this equation. We have noted with the subscript 1 the mass loading and the cut-off frequency related to the "coupling" region ($\hat{R}_1$, f_{ce1}), with the subscript 2 the quantities related to the electroded regions ($\hat{R}_2$, f_{ce2}), and, with the subscript 3 the quantities related to external unelectroded region (f_{cl}). The cut-off frequencies f_{ce} and f_{cl} are computed using the same relations as in the case of the resonators.

Boundary and continuity conditions : The traction free conditions on the major free surfaces, normal to the thickness are automatically verified as a consequence of the method used to establish the dispersion relations from which equation {1} is obtained. On the surfaces limiting the different regions of the structure we have to specify the continuity of u_1 and of its normal derivative.

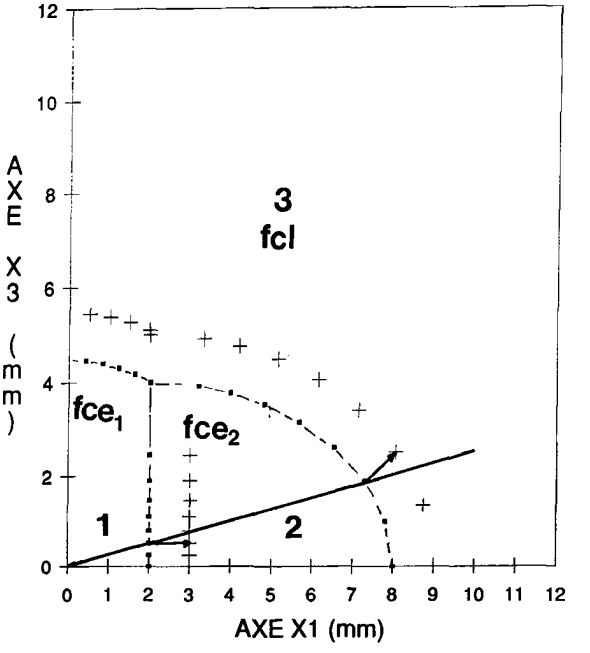


Figure 8: Example of a two poles integrated filter.

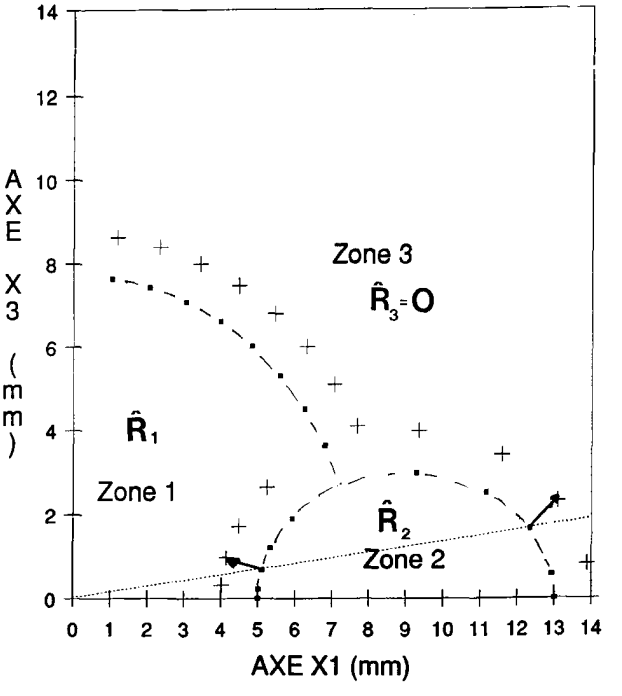


Figure 9: Other example of a two poles integrated filter

In this paper we suppose that the amplitude of the vibration mode near the edges of the plate are negligible so that no boundary conditions at the plate edge are taken into account. It was previously shown that, in this case, their consideration induces a negligible modification of the eigen modes and resonances frequencies.

Eigen mode (short circuit) analysis: The eigen modes of the structure can be obtained by the simultaneous solution of the equations for the different regions taking into account the conditions at the region boundary. For V=0 on all the electrodes, the homogeneous form of the approximate equation reduces, for any region, to equation {2} with $f^*=f_{ce1}$, f_{ce2}, or, f_{cl}.

The same coordinate transformations (simplified elliptical coordinates), as for the case of the resonators [12], are used to reduce equation {2} to an isotropic Helmoltz equation {3}. This partial derivative equation is separated into two ordinary differential equation by setting $u_1(r,t)=R(r).T(t)$:

$$r^2R''+rR'+R(r^2A-\upsilon^2)=0 \qquad \{5\}$$

$$T''+\upsilon^2T=0 \qquad \{6\}$$

υ has to be an integer due to the periodicity of the solutions of the second equation. υ=m

The (symmetrical) solutions bounded at zero and at infinity are respectively for the different regions:

Region I

$$u_1=\sum_{m=0}^{\infty} A_m Be_m(r\sqrt{A})cosmt$$

Region II

$$u_1=\sum_{m=0}^{\infty}\{B'_m J_m(r\sqrt{A})+B''_m Y_m(r\sqrt{A}\}cosmt=$$

$$\sum_{i,m} B_i Bs_m(r\sqrt{A})\cos mt$$

Region III

$$u_1=\sum_{m=0}^{\infty} C_m K_m(r\sqrt{A'})cosmt \qquad \{7\}$$

with: m even; the antisymmetrical solutions with respect to x_3 contains sin(mt) also with m even.

J_m, Y_m, I_m, K_m are respectively the Bessel functions of first and second kind and the modified Bessel function of first and second kind, of order m. Be_m is either I_m or J_m depending on the sign of A (relation {4}of §3). Bs_m is successively J_m and Y_m.

At all points of the boundaries between the regions we have to specify the continuity of u_i and of its normal derivative. As previously discussed, this can be approximated by the expression of these conditions at only a discrete number q of points (figures 8 et 9).

For example at a point $P(x_1^P x_2^P x_3^P)$ between regions I and II, we have:

$$u_1^I(x_1^P x_2^P x_3^P)=u_1^{II}(x_1^P x_2^P x_3^P) \qquad \{8\}$$

$$n_1\frac{\partial u_1^I(P)}{\partial x_1}+n_3\frac{\partial u_1^I(P)}{\partial x_3}=n_1\frac{\partial u''_1(P)}{\partial x_1}+n_3\frac{\partial u_1^{II}(P)}{\partial x_3} \qquad \{9\}$$

If we choose to express these conditions for the points at the region boundaries that correspond to q values of the angle t, we obtain linear relations between the coefficients A_0, A_2, A_{2q-2}, B_0, B_2, B_{2q-2}, C_0, C_2, C_{2q-2}. To obtain a solution we have to truncate the series to q terms. Using a matrix representation of these conditions and some algebra, we eliminate the B coefficients and obtain 2q linear relations between the A and C coefficients.

These relations constitute an homogeneous linear system in the 2q A and C coefficients. The determinant of this system has to vanish to have a non trivial (0) solution in these coefficients. This condition constitutes a frequency equation that is solved numerically for the eigen frequencies. For each eigen frequency ($f_{n\mu}$ where n is the rank of the overtone and μ the rank of anharmonic), the A and C coefficients of the eigen modes are found as the solution of the homogeneous linear system $(A_{2i}^{n\mu}, C_{2i}^{n\mu})$ where i is an index varying in the interval [0,q-1]. The B coefficients are obtained from the relations between them and the A and C ones.

On the whole, the eigen modes for the three region of the structure are given by :

$$\hat{u}_1^n=sin(n\pi x_2/2h)\bar{u}_1^n \qquad \{10\}$$

$$\bar{u}_1^{n\mu}=e^{j\omega t}\sum_{i=0}^{i=q-1} D_{2i}^{n\mu} Be'_{2i}\left(r.\sqrt{A_e^{n\mu}}\right)cos(2it) \qquad \{11\}$$

Where the D coefficients are either A, B, or C and the real coefficients of r in the argument of the relevant Bessel or modified Bessel functions (Be') are now given either by the square root of:

$$A^{n\mu}=[n\tfrac{\pi}{4}(f_{n\mu}^2-f^*2)/f^*2] \qquad \{12\}$$

or $A'^{n\mu}=-A^{n\mu}=\left[n\frac{\pi}{4}(f^*2-f_{n\mu}^2)/f^{*2}\right]$ depending on the sign of A.

The eigen solution must verify the orthogonality relation:

$$\int\int\int_{V_t}\hat{u}_1^{n\mu}.\hat{u}_1^{k\lambda}dV=\delta_{nk}\delta_{\mu\lambda}N^{(n)(\mu)} \qquad \{13\}$$

where V_t is the total volume of the structure and the displacement, is as in the following, defined by piece on the different regions.

Forced vibrations :

The modes forced by the potential $V_{rs}e^{j\omega t}$ applied between the electrodes r and s can be obtained by a linear combination of the short circuit eigen modes. For the potential a supplementary term must be added [19]

$$\hat{u}_1=\sum_n\sum_\mu H_{rs}^{n\mu}\bar{u}_1^{n\mu}\sin(n\pi x_2/2h) \qquad \{14\}$$

$$\hat{\phi}=\sum_n\sum_\mu H_{rs}^{n\mu}\hat{\phi}^{n\mu}+\frac{x_2V_{rs}e^{j\omega t}}{2h} \qquad \{15\}$$

Where:

$$u_1=\hat{u}_1-\frac{e_{26}x_2V_{rs}e^{j\omega t}}{2hc^{(1)}}$$

u_1 being the forced conventional displacement.

Accounting for the fact that the orthogonality of sin $(n\pi x_2/2h)$ has already been expressed in the forms of equation {1} that are relevant for the three regions we substitute in these equations the lateral dependence of $\hat{u}_1$.

Then we multiply these relations by $\bar{u}_1^{k\lambda}$ and integrate respectively on the volumes corresponding to the three regions. The summation of these relations gives (in the case of the Y rotated cuts of class 32 crystals where $c^{(1)}=C_{66}$ and $k^{(1)}=k_{26}$) :

$$\sum_n \sum_\mu H_{rs}^{n\mu} \frac{\pi^2}{4h^2 f_{66}^2}(f^2 - f_{n\mu}^2) \int\int\int_{V_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dV$$

$$= (-1)^{(n-1)/2} \frac{\rho\omega^2 e_{26} 4V_{rs} e^{j\omega t}}{C_{66}^2 n^2 \pi^2} \int\int\int_{V_{ers}} \bar{u}_1^{k\lambda} dV$$

Using the transformed form of the orthogonality relation accounting for the orthogonality of the sinusoidal dependence in x_2:

$$\int\int\int_{V_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dV = 2h \int\int_{S_t} \bar{u}_1^{n\mu} \bar{u}_1^{k\lambda} dS = 2\delta_{nk}\delta_{\mu\lambda} N^{(n)(\mu)}$$

We obtain:

$$H_{rs}^{n\mu} = (-1)^{(n-1)/2} \frac{f^2 h e_{26}}{(f^2 - f_{n\mu}^2) n^2 \pi^2 C_{66}} \frac{4V_{rs} e^{j\omega t} \int\int_{Sers} \bar{u}_1^{n\mu} dS}{2N^{(n)(\mu)}} \qquad \{16\}$$

where: S_{ers} = electroded surface (rs)

S_t = total surface

Electrical response :
Using the constitutive equation for D_2

$$D_2 = e_{26} u_{1,2} - \varepsilon_{22} \phi_{,2}$$

$$D_2 = \frac{\varepsilon_{22}}{2h} V_{rs} e^{j\omega t}(k_{26}^2 + 1) + \sum_n \sum_\mu (-1)^{(n-1)/2} H_{rs}^{n\mu} \bar{u}_1^{n\mu} \frac{e_{26}}{h}$$

The definition of the admittance Y_{rs} between electrodes r and s.

$$Y_{rs} = \frac{\partial I_{rs}}{\partial \Phi_{rs}};$$

and the relation between i and d_2 we obtain: $Y_{rs} = \frac{-1.}{V_{rs} e^{j\omega t}} \int\int_{Sers} D_2^2 dS$

So that the expression of the admittance Y_{rs} is:

$$Y_{rs} = j\omega \left[\frac{S_{ers}.\varepsilon_{22}}{2h}(k_{26}^2 + 1) - \sum_n \sum_\mu \frac{f^2 4k_{26}^2 \varepsilon_{22} h \left(\int\int_{S_{ers}} \bar{u}_1^{n\mu} dS \right)^2}{(f^2 - f_{n\mu}^2) 2n^2\pi^2 N^{(n)(\mu)}} \right]$$

The identification with the admittance of the equivalent scheme (that includes in parallel, one serial arm with the "static" capacitance and as many LC serial (dynamically) arms as there is modes), leads to :

$$C_{Zrs} = \frac{S_{ers}\varepsilon_{22}^S}{2h}(k_{26}^2 + 1) \quad \text{is the capacitance at f = 0} \qquad \{17\}$$

$$C_{rs}^{n\mu} = \left[\frac{8k_{26}^2 \varepsilon_{22}}{n^2\pi^2 2h} \right] \frac{h \left(\int\int_{Sers} \bar{u}_1^{n\mu} dS \right)^2}{N^{(n)(\mu)}} \qquad \{18\}$$

the inductance of the forced mode $(n\mu)$ is:

$$L_{rs}^{n\mu} = \frac{1.}{(C_{rs}^{n\mu} 4\pi^2 f_{n\mu}^2)}$$

As observed for the resonators the parallel capacitance contains C_{Zrs} and an infinite sum of the opposite of the dynamic capacitances of all the modes :

$$C_{0rs} = C_{Zrs} - \sum_n \sum_\mu C_{rs}^{n\mu} \qquad \{19\}$$

Equivalent symmetrical lattice: Applying the Bartlett bissection theorem, it is possible to find a symmetrical lattice equivalent scheme to these structures [21]. This is done considering the symmetrical parts and the antisymmetrical parts of the admittances considered above (admittances for the symmetrical modes with respect to the axis of electrical symmetry of the structure, or for the antisymmetrical modes with respect to this same axis). Then using the Mason transformation we extract from all the arms of the lattice the static capacitance and to put it across the input and the output (figure of reference 21 p.202). Under a condition between the motional parts of the admittances for the symmetrical modes and the antisymmetrical ones, it is then possible to find a ladder equivalent scheme and particularly when no unwanted modes exist, and with several approximations, the conventional ladder equivalent scheme of two poles structures.

Conclusion

The properties of AT resonators made with quartz, berlinite and GaPO4 were compared using calculations made with a three dimensional model that takes in account the energy trapping phenomena existing in real resonators. These computations, reasonably verified by the experiments, have shown that the high coupling quartz like materials lead to very interesting properties in terms of inductance and capacitance ratio. The AT resonators using berlinite and GaPO4 permits to obtain much larger filter bandwidth or oscillator shifts than AT quartz. Berlinite and even more GaPO4 can lead to devices with much smaller dimensions than quartz. This is very favourable for miniaturization. However, in many cases it is probably wise to use new types of designs permitting to lower the inductance and the sensibility to stray capacitances in increasing somehow the electrode area.

For several applications, (for example: VHF filters with suppression of the antisymmetrical anharmonics) the use of berlinite appears to constitute an interesting compromise. For the applications requesting the largest bandwidth or oscillators shifts, the use of gallium phosphate appears to be very promising. The three materials appears to lead to a very favourable and similar 3rd order thermal variations of the resonance frequencies that are centred at different inflexion temperatures.

The principle of a method of modelization of the two poles integrated filters was described. A computer implementation of this model was written and will be applied to optimize acoustically isolated two pole filters.

Acknowledgements: The authors acknowledge the assistance of M. Caumont, LeThuault and Mrs Daniel for the elaboration of the studied samples. They are indebted to M.P.Amstutz for his synthesis programs and to M. Pélissié for helpful contribution.

References:

[1] E.D.Kolb, R.A. Laudise, J. Crystal Growth vol.43 p133 (1978).
[2] B.H.T. Chai, M.L.Shand, E.Bueler, M.Gileo Proc. IEEE Ultrasonics Symp. p.577 (1979).
[3] H.Poignant, L. Le Mareschal, Y.Toudic, Mat.Res.Bull. vol. 14 p.603-612 (1979).
[4] E.Philippot, A.Goiffon, M.Maurin, J.Détaint, J.Schwartzel, B.Capelle, A.Zarka. J.Crystal Growth v.104 p.713-726 (1990).
[5] J.Détaint, A.Zarka, B.Capelle, Y.Toudic, J.Schwartzel, E.Philippot, J.C.Jumas, J.C.Doukhan proc. 40th Annual Frequency Control Symp. p.101,114 (1986).

[6] X.Buisson, A.Zarka, R.Arnaud, B.Capelle, Y.Zheng, J.Schwartzel, J.Détaint, Y.Toudic to appear in Proc. 6th European Frequency and Time Forum
[7] G.F.Engel, P.W.Krempl, J.Stadler Proc. 3rd European Frequency and Time Forum p.50-56 (1989).
[8] P.Krempl, J.Stadler, W.Wallnöfer, W.Ellmeyer, R.Selic. Proc. 5th European Frequency and Time Forum p.143,147 (1991).
[9] F.Huard These Université de Montpellier 27 Sept 1985
[10] E.Philippot, A.Ibanez, A.Goiffon, B.Capelle, A.Zarka, J.Schwartzel, J.Detaint, Proc. 1992 IEEE Frequency Control Symposium (these proceedings).
[11] B. James Proc.42nd Annual Frequency Control Symp. p.146-154 (1988)
[12] J.Detaint, H.Carru, J.Schwartzel, C.Joly, E.Philippot, Proc. 5th.European Frequency and Time forum p.30-40 (1991)
[13] D.S Stevens, H.F.Tiersten J.Acoust.Soc. Amer.vol.79 n°6 p.1455 (1976)
[14] J.Détaint, J.Schwartzel, C.Joly, E.Philippot Proc. 41st Annual Frequency Control Symposium p.314-324 (1987)
[15] J.Détaint, H.Carru, J.Schwartzel, C.Joly, B.Capelle, A.Zarka. Proc. 43rd Annual frequency Control Symposium p.563-573 (1989).
[16] B.Dulmet to appear in Proc. 6th European Frequency and Time Forum (1992).
[17] H.F.Tiersten J.Acoust.Soc. Amer. vol.62 p.1424 (1977).
[18] L.N.Dworsky Proc. 40th Annual Frequency Control Symposium p.161,167 (1986).
[19] R. Holland, E.P.Eernisse Design of Resonant Piezoelectric Devices MIT-Press (1969).
[20] H.Seikimoto Eectron. Communication in Japan vol. 65A,n°12 p10-19 (1982).
[21] R.C.Smythe in Precision Frequency Control E.A Gerber, A.Ballato editors p.202 vol.1 Academic Press (1985).
[22] H. Carru, R.Lefèvre, J.P.Aubry, S. LeChopier. Proc.42nd Annual Frequency Control Symp. p.101 (1988).

NEW PROCESSING TECHNIQUES: SWEEPING OF QUARTZ WAFERS AND A PRACTICAL METHOD FOR PROCESSING QUARTZ RESONATORS UNDER CONTROLLED CONDITIONS

Curt Bowen and Gilbert Houck

Precision Quartz Products
Camp Hill, PA 17011

ABSTRACT: SWEEPING OF QUARTZ WAFERS

There exists in quartz minute amounts of unwanted materials that cause detrimental effects in the processing of wafers (etch channels) and in completed resonator performance (radiation effects).

For some time, electrodiffusion or "sweeping" has been used to remove these impurities from the quartz material. Traditionally, quartz is swept in the bar form prior to being cut into wafers. The apparent cost of performing this operation is high, as the material cost can be as much as 10 times that of the same material unswept. Also losses experienced in subsequent processes (cutting, lapping, x-ray) are being borne with very expensive material. For these reasons and others, it would be advantageous to place the sweeping operation later into the process. That is, experience as much of the blank processing yield loss as possible with unswept quartz, and only sweep those that have successfully completed the blank processes.

Sweeping of blanks has been reported [1] and at least one crystal manufacturer [2] is utilizing this process to reduce etch channels.

Because of the reduced mass, ramping up to the sweeping temperature is much faster with blanks and higher voltages can be applied, which appear to reduce the sweeping cycle time frame substantially.

Etch channel measurements and parameter measurements were made on unswept and swept samples of the same groups.

EXPERIMENTAL

The cultured quartz blanks used in these experiments were obtained from two U.S. and two foreign facilities. Pure Z bars from the U.S. growers were cut into blanks 0.410" square and lapped in various grit sizes to 2.0 MHz. Only blanks were obtained from the foreign sources, one group at 6.6 MHz, 0.405" square and the other group at 12.0 MHz, 0.370" square. Blank groups were kept separate and recorded as A, B, C & D.

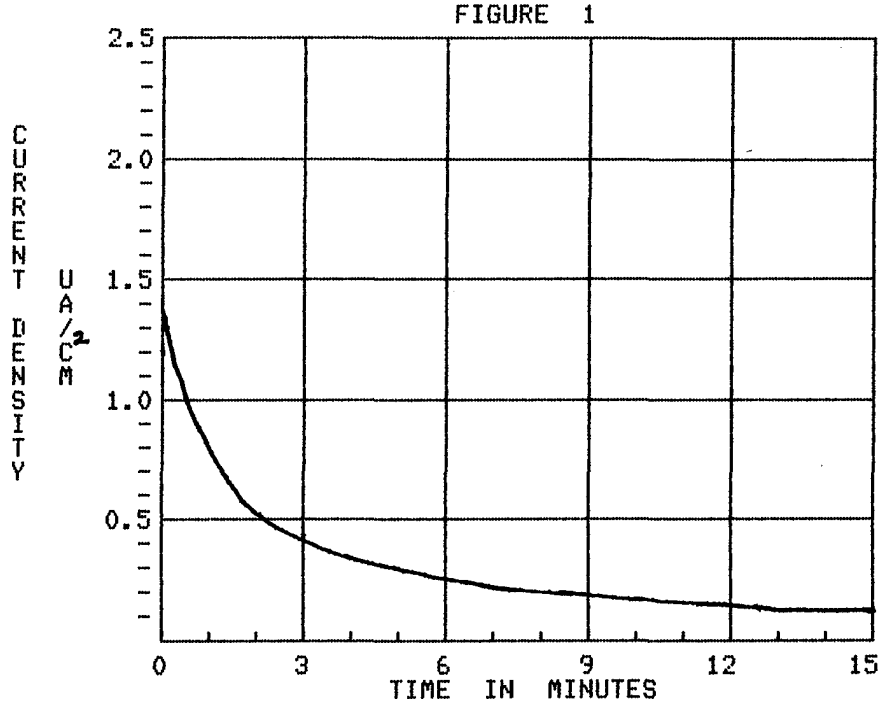
Blanks were then cleaned and etched, groups were separated into subgroups, one to be the unswept control group and the other to be swept. The groups to be swept, separately, were placed between two electrodes in a temperature controlled furnace and brought up to 540 C in 90 minutes. Upon reaching temperature, the voltage was applied and kept constant. Voltage levels used were predetermined to eliminate arcing. Current was recorded at 1 minute intervals for the first 15 minutes and then at 15 minute intervals for a total of 5 hours. Voltage remained applied during cooldown until current reached zero. Current density versus time for the runs are shown in figures 1 thru 4.

Samples from each swept and unswept group were etched in a saturated solution of ammonium bifluoride at 75 C for 2 hours. The blanks were then measured for etch channel density. Results are shown in figure 5.

All blanks were lapped to 15 MHz, processed as regular crystal production units and sealed and tested. Comparison of the test results were made and recorded.

0-7803-0476-4/92 $3.00 © 1992 IEEE

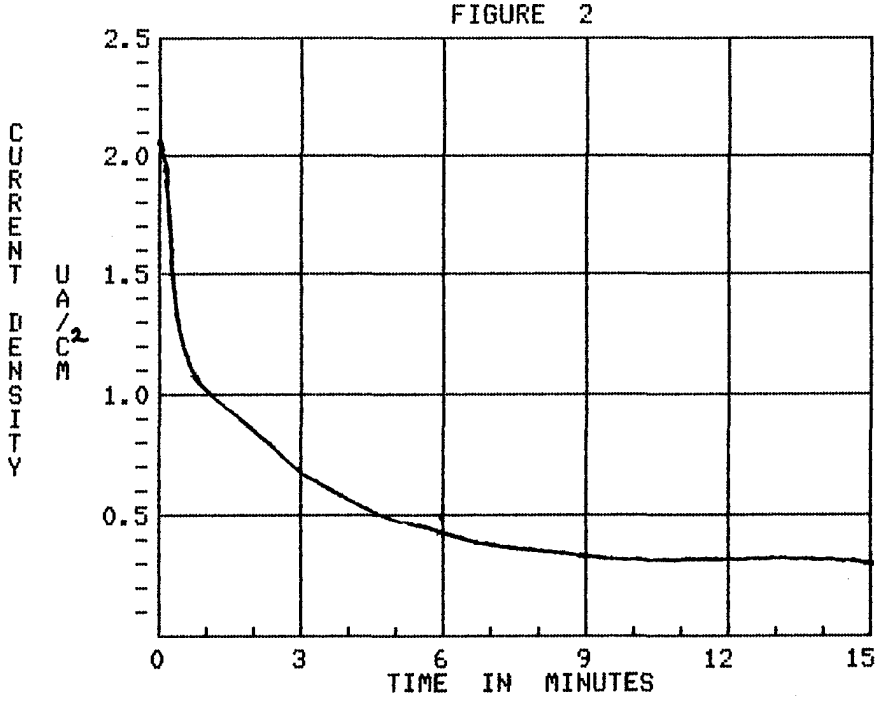
FIGURE 1



Group A Sweep
Voltage 11,324 V/cm

Current Density constant
at .081uA after 188 min.

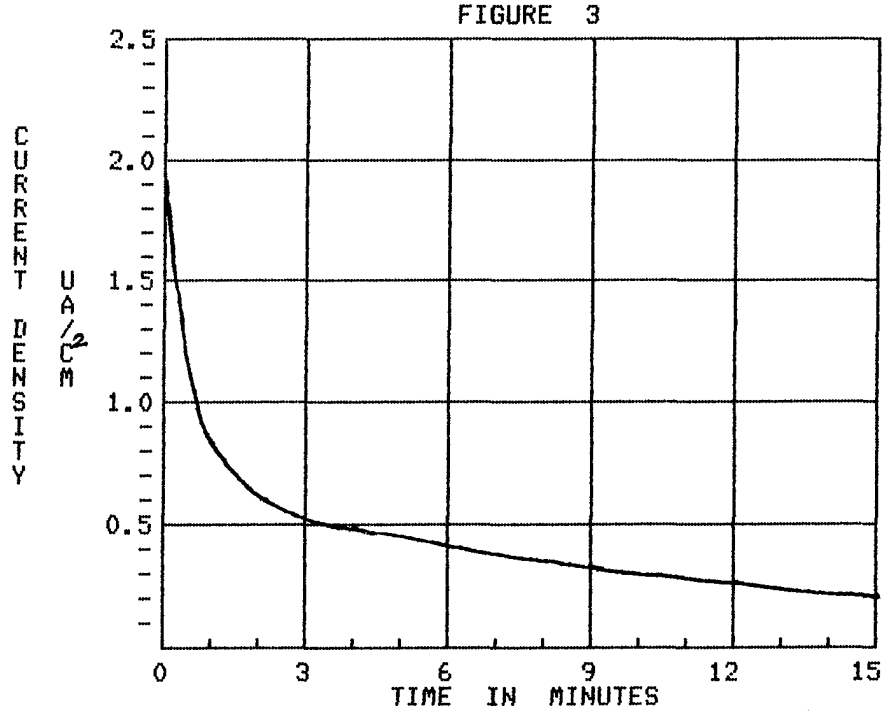
FIGURE 2



Group B Sweep
Voltage 11,265 V/cm

Current density constant
at .096uA after 193 min.

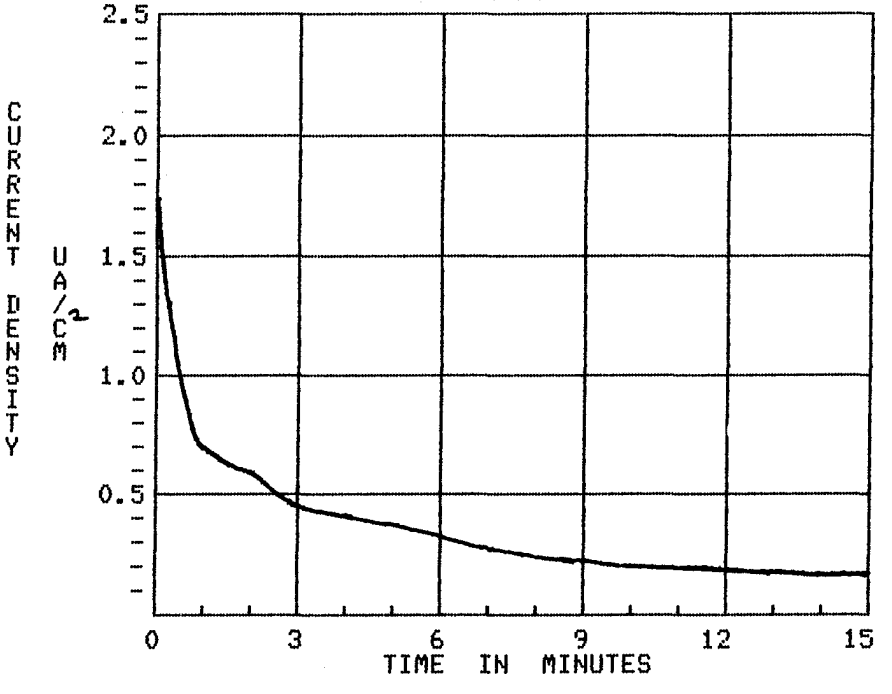
FIGURE 3



Group C Sweep
Voltage 26,428 V/cm

Current density constant
at .121uA after 160 min.

FIGURE 4



Group D Sweep
Voltage 25,045V/cm

Current density constant
at .113uA after 154 min.

Figure 5

ETCH CHANNELS/CM^2

TEST GRP	UNSWEPT	SWEPT
A	40, 50, 60, 100, 200	9, 17, 25 26, 30
B	all greater than 500	150, 200 250, 280, 320
C	9, 20, 22, 53, 60	9, 12, 20, 20, 27
D	28, 40, 50, 60, 100	2, 7, 7, 10, 37

SUMMARY & CONCLUSIONS

Results show that the process of sweeping of quartz in blank form is effective. Exact reduction in etch channel density is not known due to the limited sample size used in this experiment and the wide variation of etch channel density in the original material.

The current density decline with time characteristic is similar to that for bar sweeping except that the time required to reach a steady state current is greatly reduced. This sweeping time reduction is due to the fact that the sweeping voltage per centimeter of thickness is much higher than that typically used for bar sweeping (1000-2000 V/cm for bars and 10,000 to 25,000 V/cm for blanks).

Resistance measurements for resonators fabricated from swept and un-swept wafers showed improvements from 10% average in one group to 30% average in another group.

In any case, it is believed that significant improvement in etch channel density and motional resistance can be achieved. However, it does not appear possible to transform extremely high etch channel material into zero or near zero material.

Future experiments will include further measurements of motional and temperature parameters, and irradiation effects.

ACKNOWLEDGMENTS

The authors thank Chester Kempinski and Robert Warburton of Oak Frequency Control Group for supporting this project with quartz wafers. Also, Darrel Brehm and John Galt for process and test work.

Additionally, we acknowledge supplying of wafers from Walter Kent of Saronix.

REFERENCES

[1] J. J. Martin, "Electrodiffusion or Sweeping of Ions in Quartz," in Proc. 41st Annual Frequency Control Symposium, 1987, pp. 167-174.

[2] Private communication with John Kent and Eugene Hughes of Valpey-Fisher.

A PRACTICAL METHOD FOR PROCESSING QUARTZ RESONATORS UNDER CONTROLLED CONDITIONS

ABSTRACT

The typical process for manufacturing quartz resonators involves a series of operations. These operations include such conditions as high temperature, manipulation of the resonator and its components, high vacuum, and dry atmosphere.

Attempts have been made to construct totally enclosed systems to control the environment in which the crystal is processed. Under vacuum, in dry nitrogen, through ganged transfer systems, and the like, are expensive and difficult to work in.

Because of these problems, and other shortcomings of available equipment, a system was developed to greatly improve the conditions under which resonators are processed, without totally re-tooling the manufacturing line.

The system utilizes a transportable chamber within which partially completed resonator assemblies are kept, unless undergoing another operation such as plating or sealing. The chamber allows access to the crystals while being maintained in a clean, dry, gas environment. The temperature of the chamber can be raised to > 300 degrees centigrade to facilitate the elevated temperature steps in the process, ie. conductive cement curing, pre-plating bake, etc.

INTRODUCTION

The need to improve the conditions under which quartz crystal resonators are manufactured has been continually progressing for many years. Demand for better performance in broader applications requires very clean, and moisture free, manufacturing equipment and environments. Advanced quartz crystals such as very high frequency fundamental mode resonators are a good example.

In many cases, problems related to processing quartz resonators under dirty and humid conditions have been "fixed" by adding cleaning and elevated temperature bake steps to the process. This results in degraded resonator performance and higher cost than if the product was not contaminated during the manufacturing process to begin with.

For these reasons, methods and equipment were developed to maintain the cleanliness of resonators progressing through the manufacturing process. If a clean and dry part can be achieved early in the process, and be maintained in that condition, a better and lower cost crystal will be produced.

ANALYSIS OF SOME CURRENT PROCESSES

Although no two manufacturing processes are exactly the same, many are similar. As related to commonly produced AT, SC, BT, and IT cut quartz crystals, many processes are comprised of five basic steps. They are:

(beginning with a "dirty" quartz wafer of finished mechanical dimensions and characteristics)

1. Preparing (cleaning) the wafer for metalization.
2. Depositing metal(s) on to the wafer using thin film processes.
3. Mounting and electrically connecting the metalized wafer to a hermetically sealable assembly.
4. Adjusting the frequency of the resonator by depositing additional mass on to the metalized wafer.
5. Hermetically sealing the assemble.

The components used in the assembly other than the quartz itself, a base and a cover for example, are equally important. Of course, any part or surface that is ultimately sealed and captive in the enclosure effects the resonators performance.

HIGH TEMPERATURE PROCESSING

A commonly used procedure for raising the temperature of partially completed crystals is the "vacuum bake" oven. This method / equipment is many times used to accomplish the high temperature cure cycle for conductive materials used to secure and electrically connect the quartz wafer to its holder, and / or as a pre-seal bake as a part of a load lock on a dry box or glove box system. It is also used for various intermediate bakes to remove moisture after exposure to atmospheric humidity.

The vacuum bake process can be summarized as follows:

1. Parts to be baked are placed in the interior of a 6 sided cavity (oven).
2. The cavity is evacuated.
3. The walls of the cavity are heated.
4. Parts are baked under vacuum for a period of time.

A shortcoming associated with this method is that the parts to be baked (cleaned) are within a cavity, surrounded by an insulator (vacuum), being heated in a system where the walls of the cavity are at the highest temperature in the system.

Many times, the cleanest (hottest) part of the system is the cavity itself, and not the parts being processed in it. As the surfaces of the oven become contaminated through use and exposure to atmosphere, the contamination is redistributed to everything in the system including the parts (resonators) being processed.

Static and forced air ovens are also widely used for various bake steps, or for storage of parts. Similar problems exist with these types of ovens.

ALTERNATIVE PROCESS

In order to improve the conditions under which resonators (or hybrid oscillators with open crystals) are processed, two aspects need to be addressed:

1. The environment in which the parts are processed and exposed to throughout the manufacturing process.
2. The quality (cleanliness) of the system(s) used to elevate parts to high temperature.

A transportable chamber system was developed to facilitate the manufacturing process. The purpose of the chamber is dual, first, to serve as a clean and dry medium in which to keep parts (crystals) during or in between processes, and 2nd, to be the medium in which parts are raised to high temperatures for various process steps.

Because these two functions are combined, handling of the resonators is reduced. Loading and un-loading of parts into stationary vacuum or forced air ovens and in an out of nitrogen filled storage boxes for example is eliminated.

THE TRANSPORTABLE CHAMBER

A 8 inch square by 3 inch high stainless steel enclosure with a sliding glass cover was developed. Provisions for introducing and exiting nitrogen gas are provided. This arrangement allows the gas to flow into and through the chamber, thus entraining and replacing its atmospheric contents (after being open) with a clean and dry gas.

The chamber also incorporates a stainless steel thermocouple probe to monitor the internal temperature of the chamber.

ELEVATED TEMPERATURE

In order to avoid contamination associated with heated wall type ovens, the temperature of the contents of the chamber are elevated by

infrared radiation. The glass cover of the chamber is mostly transparent to the wavelength if infrared light provided be the "light source" portion of the system. The transportable chamber is inserted into the light source chassis which properly positions the chamber for even distribution of IR light and also connects the nitrogen gas source and the thermocouple.

Nitrogen gas is controlled by a flow meter. The thermocouple is connected to a temperature controller that drives a zero crossing triac circuit to regulate the AC power to the IR bulbs. This allows accurate and noise free control of the light (heat) source and provides the ability to program multi-segment temperature ramp and soak profiles.

Because of the nitrogen atmosphere within the chamber, temperatures in excess of 300 degrees C can be accomplished.

PROCESS UTILIZATION

Referring to the basic process steps commonly used to manufacture AT, BT, and doubly rotated quartz resonators described previously, the "Controlled Environment Processing System" (reference figure 1) can be utilized as follows:

1. Pre Baseplate Blank Bake - load etched and cleaned wafers into the chamber. Insert the chamber into the light source chassis, ramp the temperature to 300 C over 30 minutes and soak for 1 hour. Reduce the temperature to 30 C over 30 minutes.

2. Baseplate Mask Load - Move the chamber to a "clean station" for loading of the baseplate mask. The loaded masks can be held (stored) in the chamber or baked again. Nitrogen flow should be maintained on the chamber to prevent room atmosphere from re-entering the chamber.

4. Mount and Paste - Unload the baseplate masks and mount the plated wafers into holders. Apply the conductive paste and put the resonator assembly directly into the chamber. Maintain nitrogen flow into the chamber.

5. Conductive Paste Cure - Insert the chamber into the light source chassis and program the appropriate cure profile. Times can be from minutes to hours and temperature can be from 25 to 300 C.

6. Frequency Calibration (plating) - Move the chamber with "cured" resonators to the plating equipment, maintain nitrogen flow. Insert a crystal (or group of crystals) into the plating equipment and plate to frequency. Upon completion of plating return the crystals to the transportable chamber.

7. Pre-seal Bake - With a group of plated crystals in the transportable chamber, insert and program the temperature to 90% of the temperature used for curing of the conductive paste. Maintain temperature for 1/2 to 2 hours. Include cleaned covers in the chamber to be baked and used at the sealing operation.

8. Seal - For a sealing system enclosed in a nitrogen "dry box", move the chamber into the loadlock section (vacuum bake) of the system. Do not vacuum bake (reference previous discussion on vacuum baking) but use the loadlock only to gain access to the drybox. Once inside the drybox, unload and seal the resonators.

RESULTS

Frequency Calibration - A benefit of processing crystals under clean and dry conditions is that frequency shifts, both before and after sealing, are minimized or eliminated. Some manufacturers use constantly changing "offsets" at the frequency calibration operation. This is necessary to adjust for variations in processing conditions that effect the resonators frequency, such as, humidity level. As the humidity level changes, or exposure time changes, the offsets must also change. The process described maintains a dry atmosphere and thereby eliminates this frequency shift problem.

Reductions of 10 - 20 to one for undesired frequency changes during and after the plating operation are routinely observed using the Controlled Environment Processing System. That

Figure 1 : Technical Data Sheet

CONTROLLED ENVIRONMENT PROCESSING SYSTEM
for QUARTZ CRYSTALS and HYBRID OSCILLATOR PRODUCTS

Features : Model HC-643-N2

* Standard and custom chambers for lot sizes 30 to 300 pieces
* Compatible with all crystal and oscillator packages
* Stainless steel construction
* Portable, move product between operations in one fixture
* Contents continuously bakeable to 300° C
* Inert gas (N_2) purge feature
* Directly transferable into glovebox equipment
* Contents storeable under N_2 flow at room temperature
* Operate with N_2 at any location
* Contents continuously protected from the outside environment
* Warmup from 25 to 200° C in less than 5 minutes
* Optimum results when used with "clean station" type process
  - Only Class 100,000 or better manufacturing area
  - Class 50 work stations where resonators are exposed
* Compatible with additional, low cost, N_2 chambers
* **Patent Pending**

Typical Process Utilization:

* Pre-baseplate high temperature (300° C) blank bake
* Baseplate mask bake (175° C)
* High or low temperature conductive cement cure
  - Ramp and soak cure cycle
  - Continuous N_2 flow/purge with constant temperature control
* Hybrid circuit, epoxy component attach
* Nitrogen atmosphere (oxygen free), fluxless solder reflow
* Pre-final frequency adjust bake (250° C)
* Pre-seal high temperature bake
* 150° C, N_2 staging - overnight, weekend

Specifications - Model HC-643-N2

	Chamber
Outside Dimensions:	9.0 X 9.0 X 3.5 inches
Internal Dimensions:	8.0 X 8.0 X 2.5 inches
Access:	Sliding glass lid, opens to full internal dimensions of the chamber
Temperature Feedback:	SS probe with Type J thermocouple
Nitrogen Inlet Port:	Quick connector
	Light Section
Overall Dimensions:	21" wide 9" high X 12" deep
Nitrogen:	Quick connectors
Nitrogen Gas Requirement:	20 to 100 PSI, regulated
Chamber Pressure:	Flow Control, 0 to 20 SCFH
Electrical Requirements:	115 to 130 VAC @ 12 amps maximum
High Temperature Safety:	Independent setpoint, probe fault detector, requires reset when tripped
Temperature Control:	Proportional, with Ramp and Soak

Precision Quartz Products, Inc. **(717) 761-1070**
320 North 24th Street, Camp Hill, PA 17011

is, because the resonators are clean and dry prior to the plating step, the frequency remains stable through the pre-seal bake and sealing operation.

This results in the ability to produce resonators to very tight frequency calibration specifications at high yield.

Ageing - Frequency change after the assembly is sealed (ageing) is related to the cleanliness of the device. By using the process described,ageing performance has been improved by a least 10 to 1 without changing the basic processing equipment (which includes oil pumped vacuum systems and other less than ideal conditions).

10 MHZ range resonators with consistent ageing performance in the low 10 to the minus 10th and 11th range have been mass produced.

Other Performance Parameters - There is evidence that parameters such as resistance and Q, curve fit / smoothness, hystersis and others are related and improved by the processes described. To date, comprehensive testing and documentation is not complete for these parameters.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

SWEEPING INVESTIGATIONS ON GROWN-IN Al-Li^+ AND Al-OH^- CENTERS IN NATURAL CRYSTALLINE QUARTZ

Harish Bahadur
National Physical Laboratory,
Hillside Road,
New Delhi-110012

ABSTRACT

Good quality natural quartz is used as a starting material in hydrothermal synthesis of high quality cultured quartz. It is therefore important to investigate the role of impurities present in natural quartz. Such information would be useful in evaluating the growth controlling role of these impurities. In the present work, we have used near infrared absorption in the region of 3100-3700 cm^{-1} to study a variety of optically clear samples of natural quartz crystals of Brazil and Arkansas origin. The crystals were hydrogen-rich and hydrogen-lean with approximately same or varying aluminum concentration. The alkali -H concentration in hydrogen-rich samples was much higher than that in hydrogen-lean samples.

The crystals were investigated in their as received (unswept or Li-swept) and H-swept conditions. The very narrow variation in the strength of grown in Al- Li^+ centers in all samples suggests that Al-OH^- centers in natural quartz become stable only after a certain maximum Al-Li^+ concentration. This concentration appears to be independent of the origin of the growth and reflects a characteristic feature of the growth mechanism of natural quartz crystals.

INTRODUCTION

Alpha quartz plays a key role in a number of precision frequency control and ultrasonic devices. In fact, in the present day technology for various kinds of precision equipment the name "quartz" has become the symbol of quality. Most of the high stability oscillators use quartz crystals as a frequency standard and sometimes there are stringent sensitivity and stability requirements for many such applications e.g. in aerospace devices, filters, accelerometers, microbalance etc. Therefore, in recent years a significant attention has been directed towards fundamental studies of quartz crystals, and in particular, towards nature and behavior of point defects that are present in the quartz material.

Good quality natural quartz is used as a starting material in hydrothermal synthesis of high quality synthetic quartz; the later are also called as 'cultured quartz'. At present most of the industrial demands on quartz crystals are met by the different grades of commercially available cultured quartz except for those where a radiation resistant quartz is required for various aerospace applications. The radiation induced instability in quartz oscillators are known to be due to the presence of various impurities in quartz material and post irradiation effects in modifying these impurity-related defects. Because of the fact that natural quartz is used in growing cultured quartz it is important to investigate the role of impurities present in the starting material. Such information would be useful in evaluating the growth controlling role of these impurities. Beyond this, it is, however, expected that through greater understanding of the relationship between structural characteristics of specific defects and the performance of a quartz device, it will be possible to make a superior quartz for a given application.

A variety of interesting point defects are found to occur in quartz crystal [1-9]. Aluminum is the most pervasive impurity in quartz and is present as Al^{3+} at substitutional sites of Si^{4+}. The charge compensation at Al^{3+} is provided by one of the monovalent impurities such as Li^+, Na^+, K^+ or H^+ or an electron hole. Figure 1 illustrates the three primary types of defect centers arising from charge-compensated aluminum in the quartz crystal lattice. Usually, these centers are designated as Al-M^+ where M^+ is either Li^+, Na^+ or K^+ and Al-OH^- in case of an interstitial proton compensating the electron excess defect at Al; the proton forms an OH^- molecule with adjacent oxygen. The Al-hole centers are designated as Al-h^+ or $[Al_{Si}^+]^o$ Weil [6] has proposed a different scheme of designation of these centres. According to him, Al-Na^+ center is denoted as $[AlO_4/Na^+]^o$, Al-Li^+ as $[AlO_4/Li^+]^o$, Al-OH^- as $[AlO_4/H^+]^o$ and the Al-hole centers as $[AlO_4]^o$. In addition to the Al-associated centers there are other types of point defects in quartz [6]. They include e.g. oxygen vacancies, Ge-associated centers and numerous OH^- related defects.

A variety of techniques are used to characterize the point defects in quartz. For example, the Al-Na^+ centers have a strong acoustic loss peak at 53K and a weaker peak at 135K; Al-hole, oxygen vacancies and some of the Ge-related paramagnetic defects are detected by the ESR techniques. Martin [10] has presented the evidence that acoustic loss peaks at 23K, 100K, and 135K are associated with the presence of Al-hole center. Al-OH^- centers are observed by infrared absorption [11-23]. The absorption bands at 3367 and 3306

0-7803-0476-4/92 $3.00 © 1992 IEEE

cm^{-1} respectively correspond to long bond and short bond Al-OH$^-$ centers. There is however, no spectroscopic signal that has been yet found for direct monitoring of Al-Li$^+$ centers in quartz.

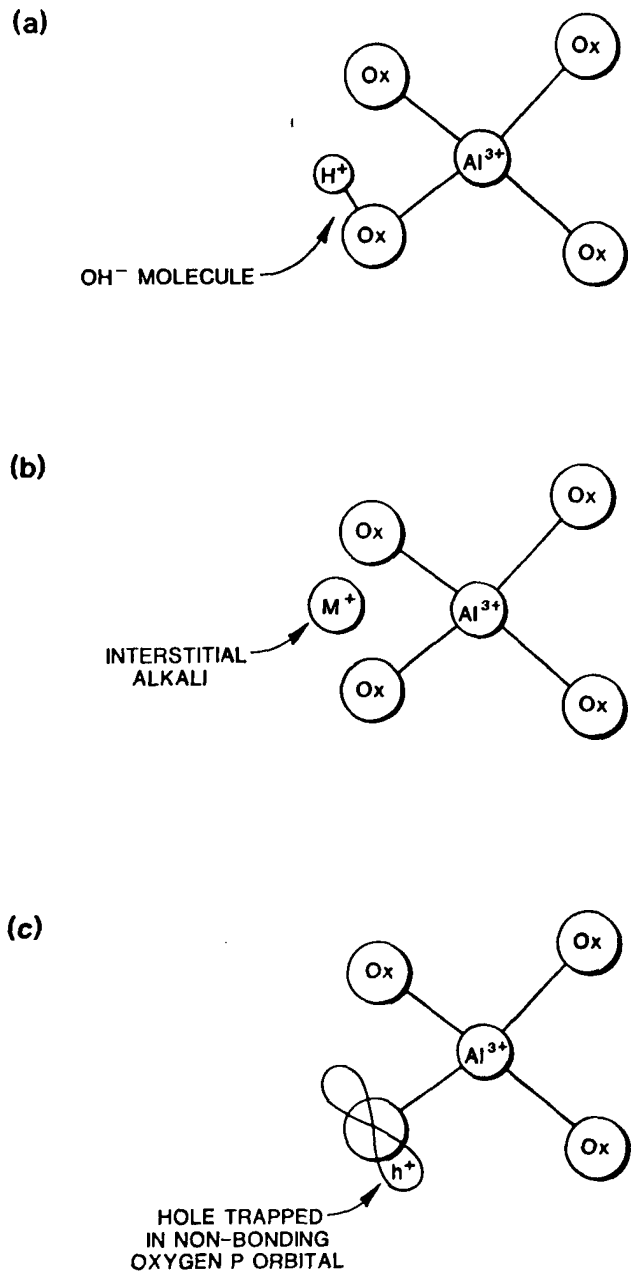


Figure 1: Schemsatic representation of (a) Al-OH$^-$ center, (b) the Al-M$^+$ center where M$^+$ is either Li$^+$ or Na$^+$, and (c) the $[Al_{e^-}]^o$ center [after Halliburton et.al ref.3].

Previous studies [10,11,24-26] have shown that in the as-grown quartz Li$^+$ is the majority of alkali ions at Al^{3+} sites in cultured material while in the natural crystals both Li$^+$ and protons compensate the excess $3+$ electronic charge of aluminum; very little of Al^{3+} is charge compensated by Na. With the above background, if a natural quartz is swept with hydrogen and the sweeping is complete, the combined strength of both the short and long - bond Al-OH$^-$ bands would represent the total aluminium concentrations. Thus, the difference of combined strength of Al-OH$^-$ bands between the H-swept and as received (or Li-swept) condition would exhibit the total concentration of Al-Li$^+$ centers present in quartz.

In the present work we report our investigations on grown-in Al-Li$^+$ and Al-OH$^-$ centers in optically clear samples of natural quartz from Brazil and Arkansas. In particular, we have monitored OH$^-$ related infrared absorption bands in the range 3100-3700 cm^{-1}.

EXPERIMENTAL PROCEDURE

Two natural stones, one from Arkansas mines and the other mined in Brazil were used in the present investigation. Samples of matching thickness were cut such that the major faces of the plates were perpendicular to c-axis. The c-axis was thus always parallel to the direction of the unpolarized beam of light. In this case the electric vector $\vec{E}$ of the incident light was always perpendicular to the c-axis. All spectra were recorded for E$\perp$c. After the crystals were cut and polished, they were mounted in an optical Dewar equipped with three CaF_2 windows and one aluminum window. The temperature was controllable to $\pm$ 2K. The cryostat tail could be rotated by 90° to position the CaF_2 window in the path of the light beam. Near infrared absorption was measured with a Beckman 4240 double beam spectrophotometer in which the ratio recording of sample and reference signals was done electronically. The spectrophotometer resolution was 5 cm^{-1}.

Samples were swept by Na, Li and hydrogen. Prior to sweeping the samples were given an optical finish. Our sweeping process began by vapor depositing gold electrodes on the two sample faces. If Li$^+$ or Na$^+$ was to swept, the appropriate salt was first evaporated on one side, then a gold electrode was deposited over the salt. The electrodiffusion was carried out at a temperature of 490^oC. Alkali sweeps i.e. taking alkali into the crystal, were done in a vaccum while an H_2 atmosphere was used for sweeping protons into the crystal. The applied electric field was approximately 30V/cm for alkali sweeps and 1500V/cm for hydrogen. A digital multimeter and a laboratory computer were used to monitor the sweeping current as a function of time.

Near the completion of sweeping, we believe only one species of ion was being transported and the sweeping current reached a steady state; this usually took about 20-24 hours for hydrogen and about two hours for alkali sweep. Once the sweeping was complete the furnace temperature was slowly programmed down. During the cool down process the electric field remained still on to prevent back diffusion of alkali ions or protons. The data were collected from a few samples per crystal. Samples were measured in unswept, Li-swept, Na-swept and H-swept conditions for their low temperature infrared spectra.

The impurity contents in natural quartz can not be typified to their origins becuase the growth conditions may vary for different stones. Generally speaking, Brazilian crystals are regarded as good quality natural quartz and are used in hydrothermal synthesis of high purity cultured quartz. In the present study we have found that the relative strength of the Li-H lines in the Brazilian crystal was much higher than that of the Arkansas one. Using the H-swept spectra of the two crystals we have found the aluminum (Al_l+Al_s) (Brazilian)/(Al_l+Al_s)(Arkansas) varying between 1.05 and 1.46. This shows that some of the samples from the two kinds of crystals used had about similar aluminum concentration and others varying. Based on the relative strength of the Li-affected OH lines represented as (Li-H) of these crystals we have used 'high-H' and 'low-H' notations respectively for the 'Brazilian' and 'Arkansas' crystal. The aluminum concentration estimated from the spectra was 'several hundred' ppm in both high-H and low-H samples. This is a rather high impurity level but would be useful for observing Al-related defects and developing defect characterization methods. Two samples investigated from high-H quartz were designated as H_1 and H_2 and two from low-H one as L_1 and L_2.

RESULTS AND DISCUSSION

Figure 2 shows the IR scans of the two crystal of Brazil and Arkansas origin in their as-received conditions. Figures 3 and 4 respectively show the spectral change upon Na- and H- sweeping of the samples from the Brazilian (High-H) crystal. Out of the numerous bands only the bands at 3306 and 3367 cm^{-1} have been assigned a model. The absorption band at lower frequency represents short-bond Al-OH^- centers and the one at higher wave number results due to absorption by long-bond Al-OH^- centers. Other infrared bonds in the region shown have not yet been assigned specific atomic models. It has been observed [11,12] that in the as-received or unswept condition the majority of absorption bands are associated with the presence of a Li ion and a proton together. Na- or H-sweeping was found to affect the spectra significantly. The results on increase of Al-OH^- bands on our samples under H-sweep are in accordance with the previous work [3, 5,11,14, 27].We have labled the short-bond Al-OH^- centers as Al_s and the long-bond Al-OH^- as Al_l in the present work.

Figure 5 represents a bar diagram showing the production of Al-OH^- center in unswept and H-swept conditions for samples H_1, H_2, L_1 and L_2. Presuming that H-sweeping has completely removed the Li-H defects, the sum of Al_l and Al_s in the H-swept condition would represent the total aluminum concentration. Therefore, subtracting $[Al_l + Al_s]$ bars in unswept condition from those in H-swept condition would result in the concentration of Al-Li^+ centers. In the case of Na- swept crystals subtracting these bars from the H-swept versions of the crystals would result in a sum of Al-Na^+ and Al-Li^+ centers. This would be so because often some Li ions are left over in the sample during the Na-sweep as can be seen by some common infrared spectral lines between the as-received (or Li-swept) and the Na-swept condition shown in Figure 3 for a Brazilian crystal. In the case of Na-swept condition of the crystal the left over Li-H related lines are very weak compared to their strength in the as-received or Li-swept condition.

We shall now estimate the relative impurity concentrations in each sample. For this purpose, we refer to the use of bar diagram of Figure 5 for all the samples H_1, H_2 L_1 and L_2. The length of each vertical bar represents the peak absorption coefficient of the Al-OH^- lines. At present we are not considering the case of other spectral bands and therefore their line strength have not been plotted in Figure 5.

To obtain an estimate of the Al-concentration ratio we compare the combined Al-OH^- bars for the H-swept condition, assuming that the sweeping has completely replaced all alkalis with H. Thus, Al concentration ratio for samples H_1 to L_1 comes out to be 27.5/18.8 = 1.46 and for samples H_2 to L_2 the ratio was 1.05. The Li ratio in samples H_1 to L_1 was 9.3 and that in the samples H_2 to L_2 was 4.23.

From the above, it can thus seen that in the high-H stone the aluminum concentration within the same stone varied by about 30%. The aluminum distribution in the low-H stone was almost uniform (variation $\sim$ 5%). The Li concentration in the samples from high-H stone was always much high compared to the samples from the low-H stone. The Li-concentration within the same stone in high-H quartz was found to vary by about 20% while in the low-H quartz it was almost the same. One must be reminded here that these impurities are all randomly distributed within the quartz lattice.

Table (I) indicates the defects concentration ratio. Sample L_1 is used as a comparison reference. The Al-related data are derived in three ways, as shown in Table (I); from Al_l, Al_s and their sun $[s+l]$. The $[s+l]$ ratio always fall between Al_s and Al_l ones. This fact may be used to estimate the error ranges. Thus, the percentage error for concentration ratio in the samples H_1 and L_1 for Al was 3.5%; for Al-OH^- it was of the order of 16.5%; for Al-Li^+ as 0%, and for Al-Na^+ of the order of 4%.

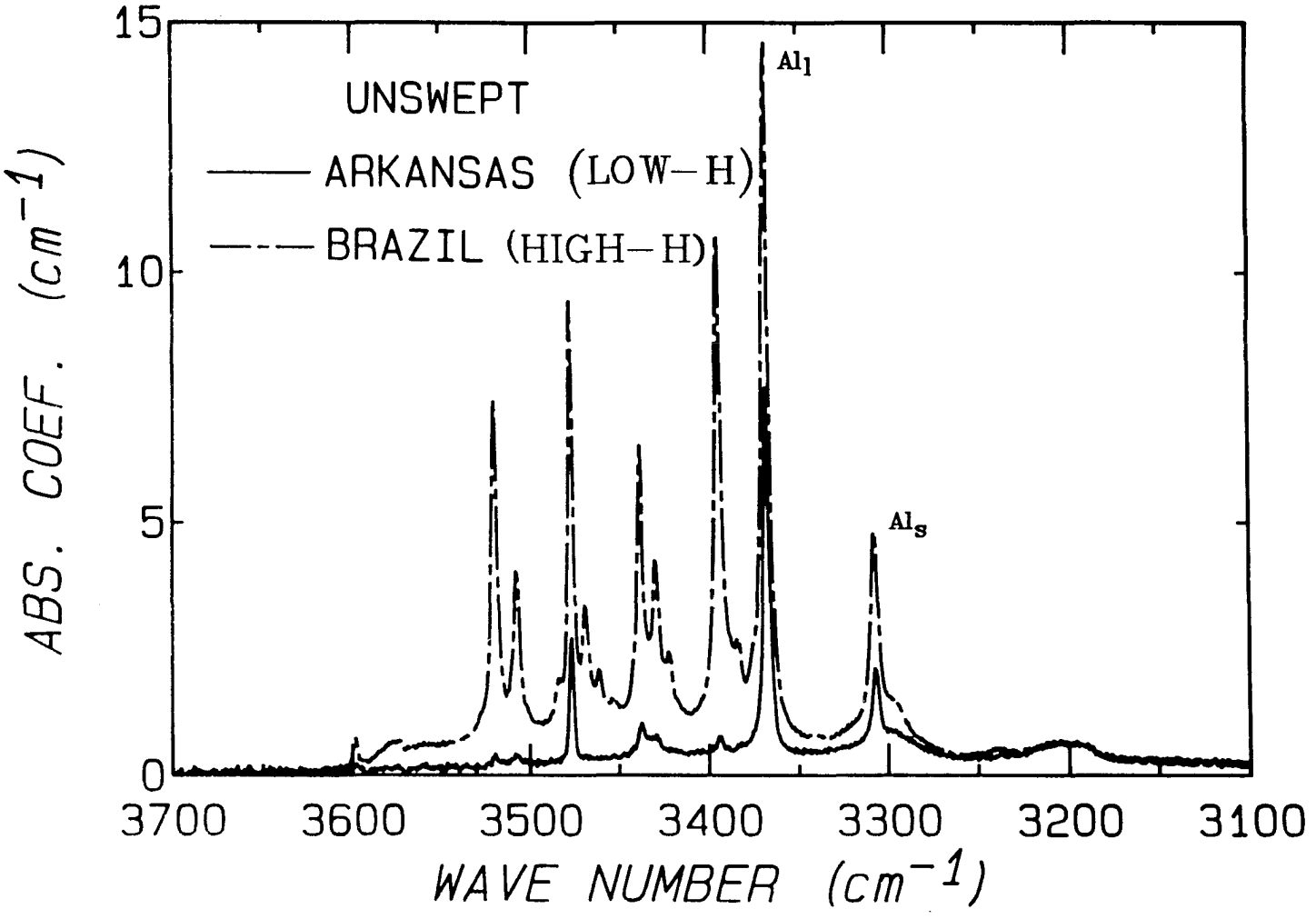


Figure 2: Infrared absorption spectra of natural Brazilian (High-H) and Arkansas (Low-H) quartz crystals in their as-received conditions. The bands at 3306 wavenumbers represents the short-bond Al-OH centers and at 3367 wavenumbers represents the long-bond Al-OH centers. These bands are respectively denoted as Al_s and Al_l.

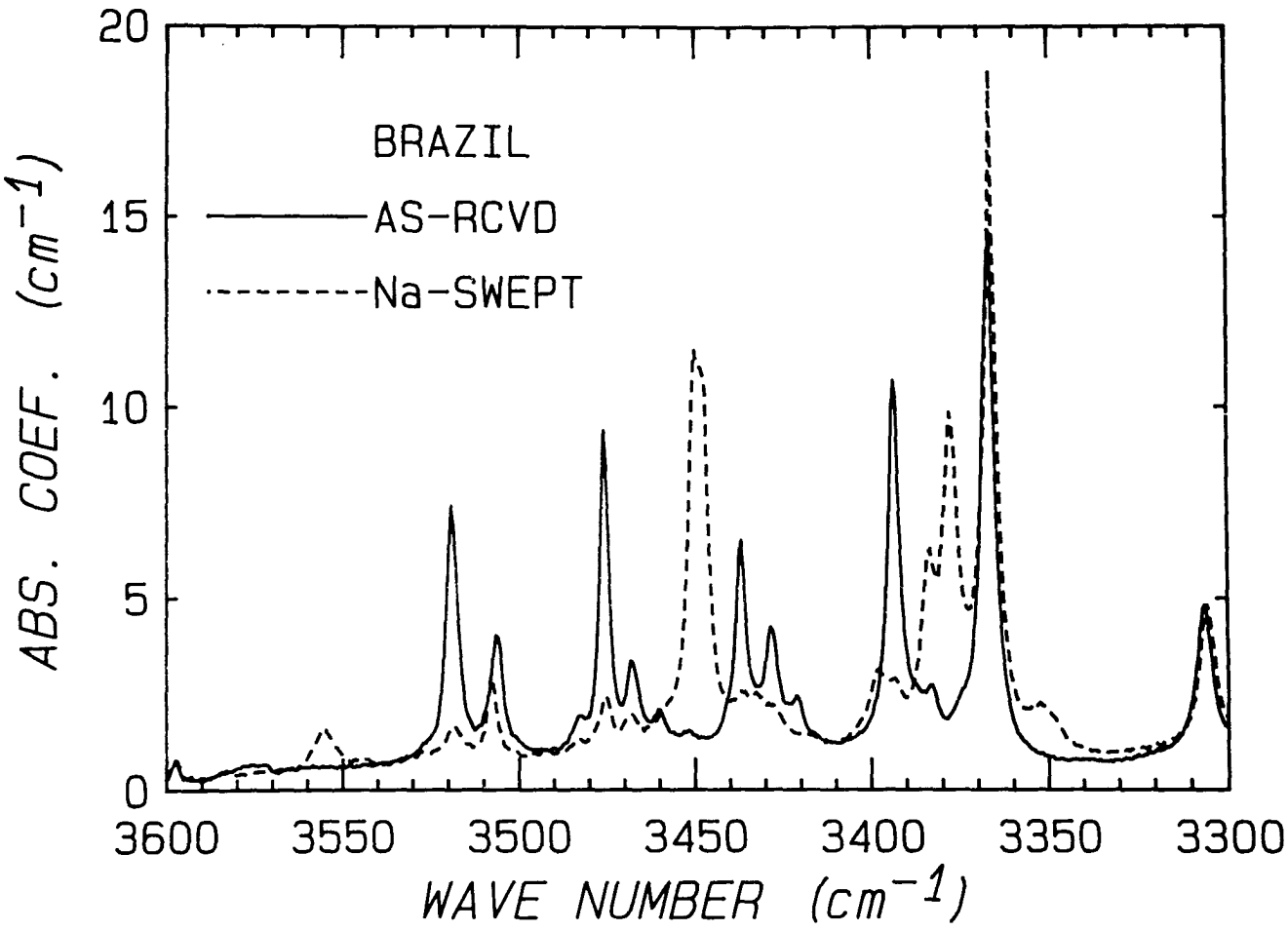


Figure 3: Infrared spectra of the Brazilian crystal in its as-received condition and after Na-sweeping. It may be observed from the two spectra that there are some small strength bands in the Na-swept version which are also observed in relatively higher strength in the as-received spectra. This shows that there are some left over Li-H bands in the Na-swept condition.

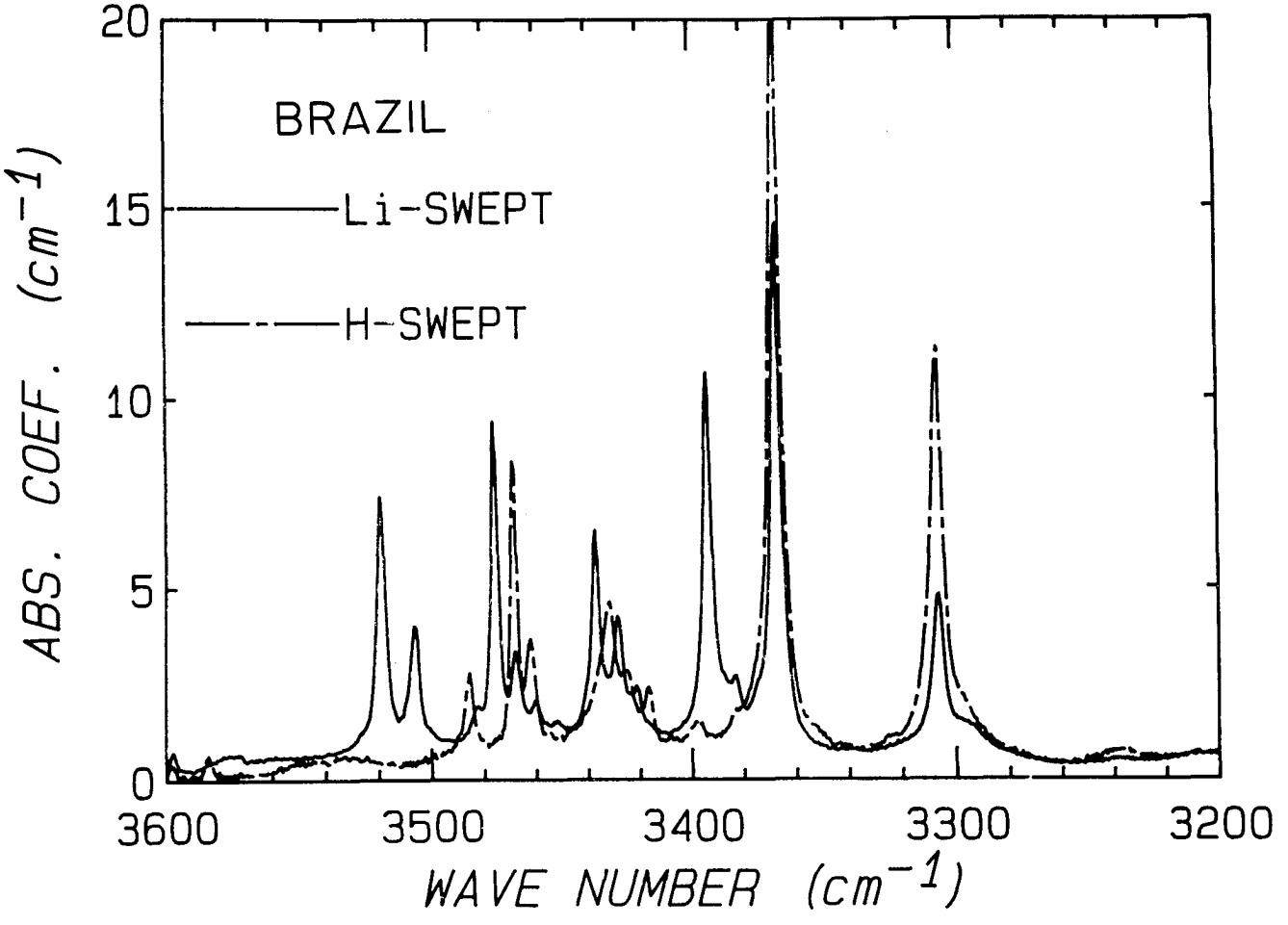


Figure 4: Infrared spectra of the same Brazilian crystal (Figure 3) in its Li-swept and H-swept conditions. Hydrogen sweeping increases the strength of the Al-OH centers existing prior to sweeping.

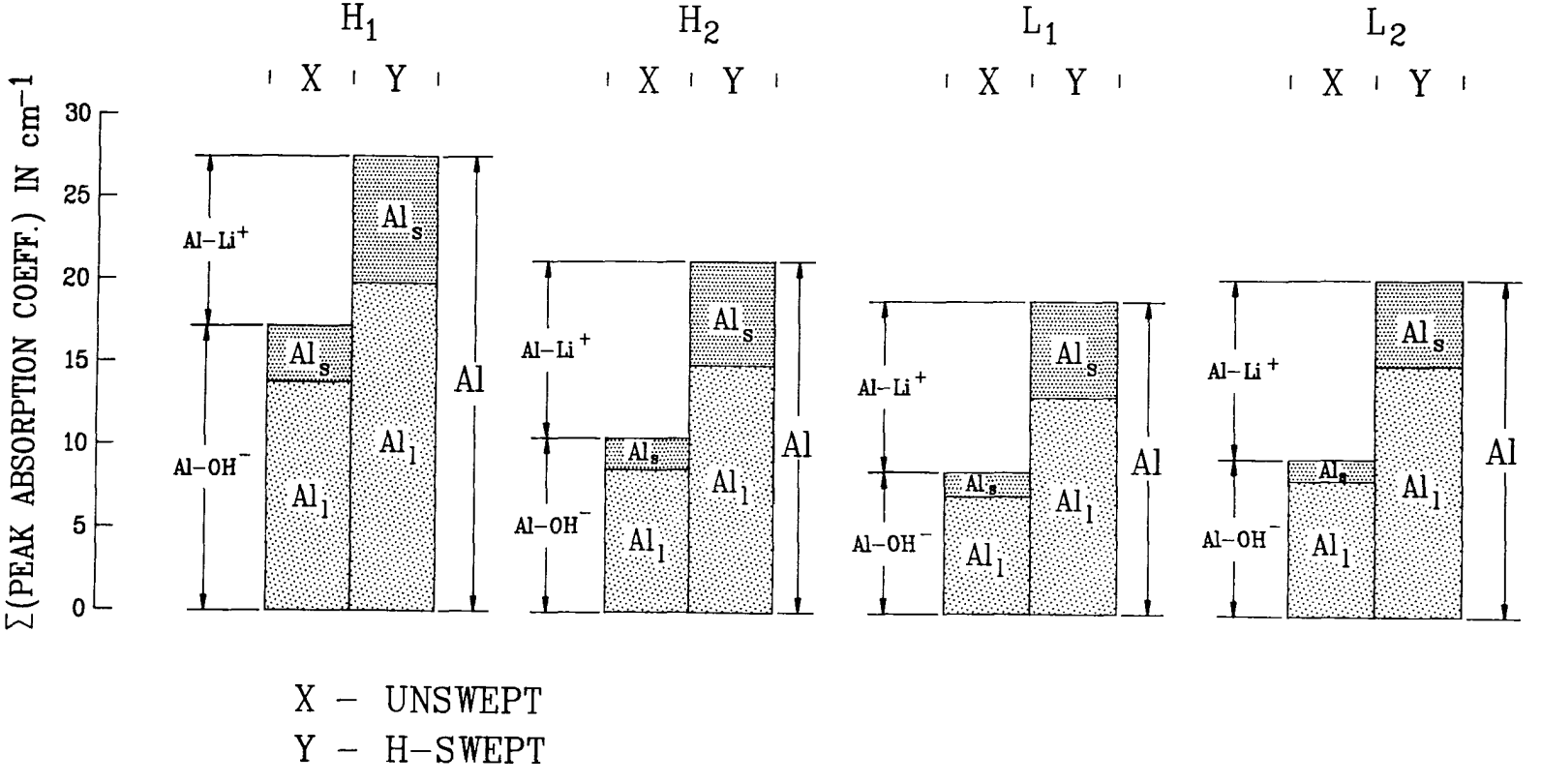


Figure 5: Bar diagrams showing the production of Al-OH^- centers in samples of unswept and H-swept natural quartz crystals of high-H and low-H concentrations. It may be noted that the Al-Li^+ centers in all samples occur in equal strength.

ESTIMATED DEFECT CONCENTRATION RATIOS

Ion	Defects	Band	Ratio	Samples: peak abs. coef. [1/cm]				Concentration Ratios			
				H_1	H_2	L_1	L_2	H_1/L_1	H_2/L_1	L_1/L_1	L_2/L_1
	Grown-in										
Al	X	[s+l]		27.5	21.4	18.8	20.2	1.46	1.14	1.00	1.07
		[Al_s]		7.7	6.5	5.8	5.5	1.33	1.12	1.00	0.95
		[Al_l]		19.8	14.9	13.0	14.7	1.52	1.15	1.00	1.13
H	Al-OH^-	[s+l]		17.2	10.5	8.5	9.7	2.02	1.23	1.00	1.14
		[Al_s]		3.4	1.8	1.4	1.4	2.43	1.29	1.00	1.00
		[Al_l]		13.8	8.7	7.1	8.3	1.94	1.23	1.00	1.17
		[s+l]	Al-OH/Al	0.63	0.49	0.45	0.48	1.4	1.08	1.00	1.06
		[Al_s]		0.44	0.28	0.24	0.25	1.83	1.17	1.00	1.05
		[Al_l]		0.70	0.58	0.55	0.56	1.28	1.05	1.00	1.01
Li	Al-Li^+	[s+l]		10.3	10.9	10.3	10.5	1.00	1.06	1.00	1.02
		[Al_s]		4.3	6.2	4.4	6.4	0.98	1.41	1.00	1.45
		[Al_l]		6.0	4.7	5.9	4.1	1.02	0.8	1.00	0.7
		[s+l]	Al-Li/Al	0.37	0.51	0.55	0.52	0.68	0.93	1.00	0.94
		[Al_s]		0.56	0.95	0.76	1.16	0.74	1.25	1.00	1.53
		[Al_l]		0.30	0.32	0.45	0.28	0.67	0.71	1.00	0.62
	Li-H	all	S_0	36.5	16.94	3.9	4.0	9.36	4.34	1.00	1.02
	Y	all	S_1	5.3	1.77	0.5	0.95	10.6	3.54	1.00	1.9
			S_1/S_0	0.15	0.10	0.13	0.24	1.13	0.8	1.00	1.86

X ..... Al-OH, Al-Li and very little Al-Na.

Y ..... Li-H left in after Na-sweeping.

Interestingly, it may be noted from Table I that the grown-in Al-Li^+ ratios appear to be nearby constant; varying between 1.00 and 1.06. The grown-in Al-Li^+ line strength in samples H_1, H_2, L_1 and L_2 correspond to 10.3, 10.8, 10.3 and 10.5 wave numbers. We have calculated this difference from the spectra published in literature by Brown and Kahan [13] and Kats [14]. It comes out to be of the same order. Table II depicts these results.

From the Tables I and II and Figure 5 it seems that the grown-in Al-Li^+ ratios in natural quartz appear to be nearly constant; varying over a very narrow range between 1.00 and 1.06. In other words, the differences in the line strength which indicate Al-Li^+ content are about equally strong in all natural quartz crystals. It can be speculated that the grown-in Al-OH^- in natural quartz would become stable only if the total aluminum concentration exceeds the Al-Li^+ concentration value.

TABLE -(II)

Strength of Al-Li^+ centers in different samples of natural quartz crystals

Samples used in the present work	Difference in $[Al\text{-}OH]_{H\text{-swept}} - [Al\text{-}OH]_{unswept}$
H_1	10.3 wavenumbers
H_2	10.8 wavenumbers
L_1	10.3 wavenumbers
L_2	10.5 wavenumbers
Samples from literature	
[Ref. 13]: Fig.9 [Brown and Kahan] thickness = 3 mm.	10.4 wavenumbers
[Ref. 14]: [Kats] Fig. 4.14	11.0 wavenumbers

It will be instructive from the present studies to simulate these observations on cultured quartz crystals. There are however two difficulties at present in doing so. Firstly, the conventionally grown quartz crystals do not show the presence of Al-OH^- bands in their as-grown or unswept conditions. Secondly, there is no direct spectroscopic signal that canbe used for measuring the Al-Li^+ centers. Nevertheless, it might prove fruitful to grow cultured quartz with aluminum as high as comparable with natural quartz. In such a case during the growth some of the aluminum may start getting compensated with hydrogen after a certain maximum number of Al-Li^+ centers are stabilized. Further investigations in this direction are necessary to establish the similarity between natural and cultured quartz.

CONCLUSION

The present study shows that Al-OH^- centers in natural quartz crystals become stable only after a certain maximum Al-Li concentration. This concentration is independent of the origin of growth and reflects a characteristics feature of the growth of natural quartz crystals.

ACKNOWLEDGMENTS

The author expresses his appreciation to Professor E.S.R. Gopal, Director, National Physical Laboratory and Dr. T.K. Saksena, Head, Ultrasonic Section of NPL for their encouragement in the work of present paper. He is thankful to Professor Joel J. Martin of Oklahoma State University for supporting this work and Dr. J.C. King of Sandia National Laboratories, Albuquerque, NM (now retired) for his encouragement in work on natural quartz crystals. N. Balani's sincerity in preparing the manuscript is worthy of appreciation.

REFERENCES

1. L.E. Halliburton, J.J. Martin and D.R. Koehler, "Properties of Piezoelectric Materials" in Precision Frequency Control, E.A. Gerber and A. Ballato, eds., Orlando, FL: Academic Press, Vol.1, pp.1-46 (1985).

2. L.E. Halliburton, "Defect Models and Radiation Damage Mechanisms in Alpha Quartz", Proc. Int. Conf. on Insulators, Univ. Utah, Salt Lake City, UT, Aug. 1984.

3. 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).

4. J.A. Weil, "The Aluminum Centers in Alpha Quartz", Radiat. eff. Vol.26, pp. 261-265 (1975).

5. L.E. Halliburton, M.E. Markes and J.J. Martin, "Point Defects in Synthetic Quartz: A Survey of Spectroscopic Results", in Proc. 34th Annual Frequency Control Symposium, EIA, pp.1-8,(1980).

6. J.A. Weil, "A Review of Electron Spin Resonance Spectroscopy and Its Application to the Study of Paramagnetic Defects in Crystalline Quartz", Phys. Chem. Minerals Vol.10, pp.149-165 (1984).

7. S.W.S. McKeever, "Thermoluminescence in Quartz and Silica", Radiation Protection Dosimetry, Vol.8, pp. 81-98 (1984).

8. L.E. Halliburton, J.J. Martin and W.A. Sibley, "Study of Defects Produced by the Growth, Post Treatment and Fabrication of Quartz", Oklahoma State University, RADC-TR-81-276. October 1981. Copies available from Rome Air Development Command, Hanscom, AFB, MA 01731.

9. L.E. Halliburton, J.J. Martin and W.A. Sibley, "Study of Defects Produced by the Growth, Post Treatment and Fabrication of Quartz", Oklahoma State University, RADC-TR-83-309, Jan.1984. Copies available from Rome Air Development Command, Hanscom, AFB, MA 01731.

10. J.J. Martin, "Aluminum-Related Acoustic Loss in AT-cut Quartz Crystals", J. Appl. Phys. 56, pp.2536-2540 (1984).

11. Harish Bahadur, "Infrared Characterization of Natural and Cultured Quartz: The Effect of Electrodiffusion and Irradiation", J. Appl. Phys. 66, pp. 4973-4982 (1989).

12. Harish Bahadur, "Sweeping and Irradiation Effects on Hydroxyl Defects in Crystalline Quartz", Proc. 45th Annual Frequency Control Symposium, pp. 37-57 (1991). Doc. # 91CH2965-2, IEEE-445 Hoes Lane, Piscataway, NJ 08854, USA.

13. R.N. Brown and A. Kahan, "Optical Absorption of Irradiated Quartz in the Near I.R.", J. Phys. Chem. Solids, 36, pp.467-476 (1975).

14. A. Kats, "Hydrogen in Alpha Quartz", Philips Research Reports. 17, pp.133-195; 201-279 (1962).

15. H.G. Lipson, F. Euler and A.F. Armington, "Low Temperature Infrared Absorption of Impurities in High Grade Quartz", Proc. 32nd Annual Frequency Control Symposium,NTIS, pp. 11-23 (1978).

16. F. Euler, H.G. Lipson and P.A. Ligor, "Radiation Effects in Quartz Oscillators, Resonators and Materials", Proc. 34th Annual Frequency Control Symposium, NTIS, pp. 72-80 (1980).

17. H.G. Lipson, A. Kahan, R.N. Brown and F. Euler, "High Temperature Resonance, Loss and Infrared Characterization of Quartz", Proc. 35th Annual Frequency Control Symposium, NTIS, pp. 329-334 (1981).

18. F. Euler, H.G. Lipson, A. Kahan and A.F. Armington, "Characterization of Alkali Impurities in Quartz", Proc. 36th Annual Frequency Control Symposium, NTIS, pp. 115-123 (1982).

19. H.G. Lipson, A. Kahan and J.O'Connor, "Aluminum and Hydrogen Defect Centers in Vacuum Swept Quartz", Proc. 37th Annual frequency Control Symposium, NTIS, pp. 169-173 (1983).

20. H.G. Lipson and A. Kahan, "Effects of Vacuum Sweeping and Radiation on Defect Distribution in quartz", IEEE Trans. Nucl. Sci.,NS-31, pp. 1223-1229 (1984).

21. H.G. Lipson and A. Kahan, "Infrared Characterization of Aluminum and Hydrogen Defect Centers in Irradiated Quartz", J. Appl. Phys. 58, pp. 963-970 (1985).

22. J.J. Martin, "Electrodiffusion (Sweeping of Ions in Quartz - A Review", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, UFFC-35, pp. 288-296 (1988).

23. W.A. Sibley, J.J. Martin, M.C. Wintersgill and J.D. Brown, "The Effect of Radiation on the OH Absorption Bands in Quartz Crystls", J. Appl. Phys. 50, pp. 5449-5452 (1979).

24. J.J. Martin, Ho B. Hwang, H. Bahadur and G.A. Berman, "Room Temperature Acoustic Loss Peaks in Quartz", J. Appl. Phys. 65, pp. 4666-4671 (1989).

25. E.R. Green, J. Toulouse, J. Wacks and A.S. Nowick, "Effect of Irradiation and Annealing on the Electrical Conductivity of Quartz Crystals", Proc. 38th Annual Frequency Control Symposium, NTIS, Doc. #AD-A217381, pp. 32-37 (1984).

26. J. Toulouse and A.S Nowick, "Dielectric Loss of Quartz Crystals Electrodiffused with Either Na^+ or Li^+ ", J. Phys. Chem. Solids, 46, pp. 1285-1292.

27. D.M. Dodd and D.B. Fraser, "The 3000-3900 cm^{-1} Absorption Bands and Anelasticity in Crystalline Alpha Quartz", J. Phys. Chem. Solids, 26, pp. 673-686 (1965).

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

AGEING OF CRYSTAL ELEMENTS MANUFACTURED OF SWEPT SYNTHETIC QUARTZ

L. N. Romanov, S. V. Kolodieva

All-Union Research Institute for the Synthesis of Materials,
Russia, 601600, Alexandrov, Vladimir region, Institutskaya, 1,
VNIISIMS

Abstract:

Piezoquartz properties improvement, attained under conditions of electric field and high temperature (so called sweeping) is not irreversible because of essential contribution in the sweeping process of the polarisation effects with great (months, years) relaxation times. Depolarisation, rate of which exponentially increases vs temperature, causes the long-term ageing of resonators made of the swept quartz.

Introduction

According to many research works, sweeping of piezoquartz lowers etch channel density in this material, increases Q-value of the crystal units made of the swept material, decreases series resonance resistance, raises frequency stability with time and under irradiation.

Too marked effect of the resonator characteristics improvement with swept crystal units and relative simplicity of sweeping process realization have result in working out of the typical industrial piezoquartz sweeping process. Sweeping regims of different swept quartz producers are slightly differed. As a rule the lumbered culture quartz Y-bars without seed and with machined and oriented X- and Z-surfaces are subjected to sweeping. The sweeping is carried out in air atmosphere at the temperature 500±20°C and the electric field intensity 1,5÷3,0 kV/cm. Sweeping process duration varies from hours to few tens of days.

It should be noted that many producers and users of the swept piezoquartz are influenced by simplified understanding of physics of processes, occuring in quartz effected by electric field at high temperatures. Usually the characteristic stability of the resonator manufactured from quartz that was subjected to high temperature electrical effect is believed to be a result of the stable effect, consisting in irreversible removal of alkaline ions from crystal volume. Such simplificated understanding is explained by a lack of systematizied data on characteristic changes of crystal units manufactured from swept quartz after the lapse long time enough, by scant information about properties of piezoquartz subjected to the sweeping procedure and by the absence of the estimation criterion of the removed impurity ions quantity.

However at last time the tendency to the more detailed study of the quartz sweeping process became appreciable. It is possible that this tendency is connected with the enhanced requirements to the quality of piezoquartz material used in new high frequency and microwave communication technique. There are works (see for example [1,2]) devoted to such important problems as the type of electrode, possibility of physical-chemical reactions suppressing the diffusion of hydrogen in a crystal, and other phenomena on intersurface quartz-electrode-gas boundaries. A very interesting data on swept quartz material inhomogeneity that depends on quality of metalic film-electrode are given and possible variants of more homogeneous sweeping realization are discussed in [3]. D. W. Hart et al. [4] analysing the process of the electrodiffusion in quartz have approached a negation of dominating role of point defects removal from the sweeping volume. They state that the defects modification, but not their removing from a crystal, is the reason of quartz properties improvement after sweeping.

0-7803-0476-4/92 $3.00 © 1992 IEEE

However after this statement authors [4] declare that electrical cleaning in air or hydrogen medium removes alkaline ions from quartz and replaces them on hydrogen ones. The fact that this phenomenon inheres in quartz has been proved long ago [5] and it is beyond doubt. However it can't be realized by the use of the characterizied above industrial sweeping process. Phenomena taking place during process of the lumbered culture quartz Y-bars sweeping is complicated enough. That is why the characteristic improvement of the resonator manufactured from swept Y-bars is reversible at the least partially.

In this paper experimental data on ageing of crystal units produced from swept quartz are presented and a physical interpretation of derived results is proposed.

Experimental methods

Y-bars 15X15 mm (5 samples) were sawed out synthetic quartz, grown on horizontal seeds with top surface screened. As it is known such growing method ensure high quality of quartz. After sawing Z and X surfaces were subjected to mechanical lapping, finishing and X-ray orientation accurate to ±10'. Z-surfaces of bars were coated by the Pt layer using method of cathode sputtering.

The sweeping of bars was carried out in the standard regime: 530°C, 1.5 kV/cm; duration of the sweeping process was controlled by time dependence of current [6], high temperature stage of the sweeping process was stoped if the current during one hour doesn't change more than ±3%.

Crystal units of plane-covex geometry 14 mm in diameter for precision thermostatic resonators on frequency 5 MHz (the fifth oscillation mode) were produced of two swept lumberd cultured quartz Y-bars. Two other bars were used for thermodepolarisation research in the short circuit regime. Before crystal units making from the face side of each bar were sawed by one plate of 12 mm along Y-axis; from them specimens for IR-tests were made. One bar was not subjected to the sweeping. Crystal units for control resonators, the same as from swept bars, were made from it.

IR-tests were realized on the spectrophotometer "Specord M85" in the scanning regime for the homogeneity control of swept quartz material along Z-axis.

Q-value of resonators was determined by means of damping factor measurement.

Resonator frequency was measured accurate to $\pm 5 \cdot 10^{-3}$ Hz.

Capacitance bridge 1620-A (General Radio Co) and laboratory special cell with the coaxial electrodes were used in order to receive the temperature-frequency dependence of imaginary part of the complex dielectric permeability.

Results and Discussion

As can be seen from the table where typical results of the sweeping influence on the main resonator characteristics such as Q-value and frequency are presented, the sweeping appreciably increases resonator Q-value and decreases relative frequency time-change (long-term ageing) and radiation stimulated frequency drift. In the case of wafer sweeping this effect keeps for a long time, however characteristics of resonators manufactured from the swept lumbered quartz bars essentially vary in due course: for the 2.5 years Q-value is reduced about 10%, $\Delta f/f$ is rised more than 5 times. Radiation stability is reduced too.

Observed ageing of crystal units produced from swept bars is the more remarkable the lower Q-value before sweeping. Moreover, as can be seen from the table sweeping conditions efficiently effect on the resonator characteristic stability level.

The efficiency of Y-bar sweeping process was controlled in our experiments by time dependence of current. As already noted high temperature stage of sweeping process stopped as soon as the absence of current value changes exceeding the measurement error ±3% was attained during 1 hour.

Ageing of crystal units manifestation in spite of observance above formulated criterion is evidence of the fact that constancy of electric current value does not mean the completeness of driving out of the sweeping volume alkanile ions that are responsible for the time and radiation stimulated resonator characteristic instability.

Macroscopic volume polarization of quartz appearing under the conditions of electric field and high temperature is at the bottom of sweeping current decrease. Volume polarization is forming due to the

point defect modification, consisting at first in the replacement of an alkanile ions by hydrogen ones and second by creation of the space charge that arises from alkaline ions electrodiffusion from anode to cathode.

Fig. 1 demonstrates the experimental results confirming the presence of macroscopic ("high voltage polarization" according to Ioffe [5]) in swept Y-bar. The curve under the abscissae axis is obtained in short circuit regime at electric field cutted off, i.e. we deal with depolarization current. If the current observed during sweeping had been the current of charge transfer then no current could be observed in the absence of electric field. Polarization and depolarization greatly and in complicated manner depend on temperature. Unusual behaviour of quartz resistance vs temperature is shown in Fig. 2. Dependence of R/RO vs temperature demonstrates presence of extremum at the temperature correlating with the most intense polarization process. For each temperarure RO and R correspond to initial and final sweeping current. The constancy of current magnitude in quartz at continious effect of electric field of high intensity at high temperatures means the completion of the relaxation polarization process and consequently formation of the stable thermoelectret state in quartz. The possibility of a formation in quartz of several polarization types with different relaxation times which define the thermoelectret life-time is proved by dielectric spectoscopy methods, coulometry and by depolarizarion analysis. The life-time of thermoelectret state formed during sweeping which is conducted at conventional V-T-t conditions is $\geq 10^8$ s at room temperature (about 3 years). During this time period quartz will slowly relax to the initial (non-polarized) state, i.e. it will return to the physical properties inherent in the nonpolarized sample. This relaxation process reveals just in a very slow return of crystal unit Q- and resonance frequency values observed to the initial ones, i. e. a phenomenon known as a long-term crystal ageing occurs.

Recuring to proposed by Vig et al. method of sweeping completeness control it should be said that this method although does not mean the completeness of the

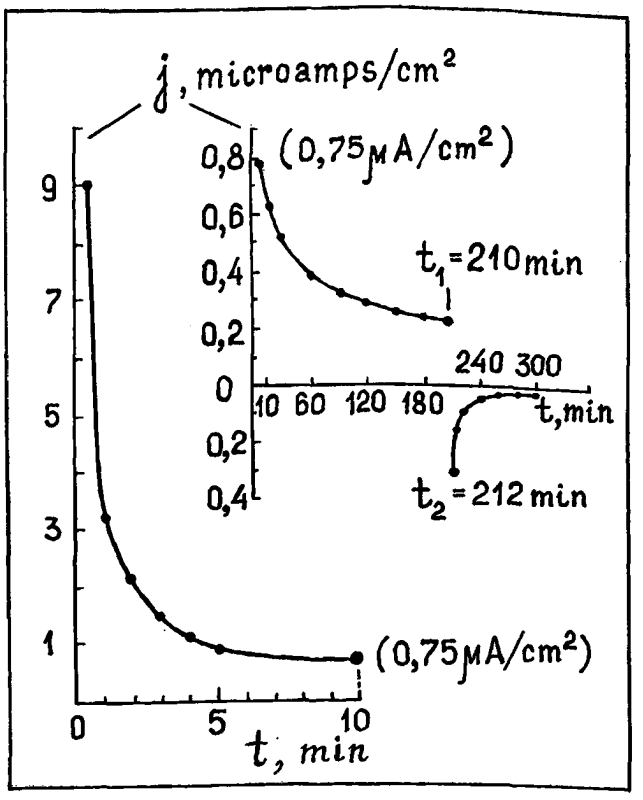


Fig. 1. Sweeping current density vs time (1200 V/cm; T=530°C)

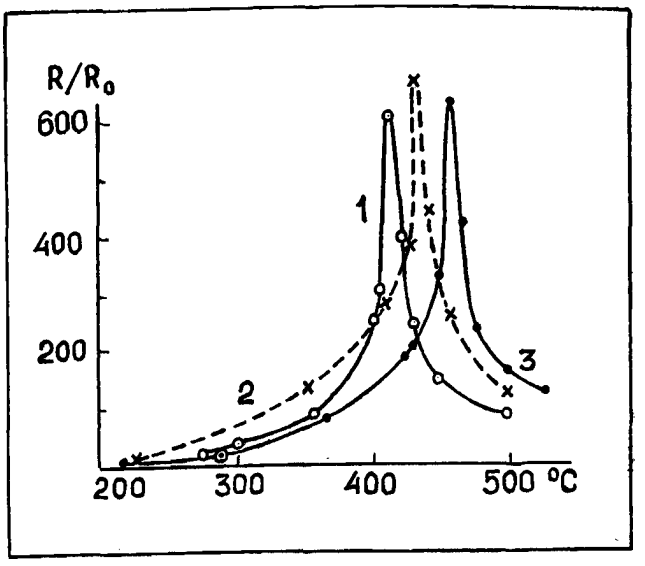


Fig. 2. Resistance vs temperature E(V/cm): 1 - 1320; 2 - 380; 3 - 40

process of alkaline ion removal from the crystal nevertheless it can be used for sweeping process control. The constancy of current magnitude is a reliable signal for ceasing high temperature stage of sweeping and for the following cooling. As a rule current becomes unchangeable 2-4 hours after electric field switching on. Therefore further Y-bar maintenance under the field (during tens hours and even tens days) seems to be unjustified.

Taking into account the statement of the authors [3] concerning the influence of the quality of electrode covering on the sweeping homogeneity, it should be stressed that even in case of uniform Pt-layer deposited by cathode sputtering the sweeping inhomogeneity of quartz bar

Comparative data on time and radiation stimulated characteristic change of resonators manufactured from as-grown and swept quartz

Characterization of resonator	Q-value, $x10^6$			Fractional frequency deviation f/f, $x10^{-8}$		
	at once after manufact. (after irradiat.)	after 6 months	after 2.5 years	at once after manufact. (after irradiat.)	after 6 months	after 2.5 years
Resonator of unswept lumbered cultured quartz bar						
unradiated	2.25	2.20	2.27	0	+30.0	+137.0
irradiated (γ=10^5 R)	1.80	2.20	2.28	+68.5	+23.0	+6.0
Resonator of swept lumbered cultured quartz bar						
unradiated	2.38	2.32	2.04	+1.2	+3.8	+6.9
irradiated (γ=10^5 R)	2.40	2.35	2.35	+1.0	+1.8	+9.0
Resonator of swept quartz wafer						
unradiated	2.51	2.55	2.50	+0.5	+0.8	+1.5
irradiated (γ=10^5 R)	2.50	2.48	2.58	+0.6	+0.5	+1.2

Notes: Quartz crystals from that bars and wafers are lumbered have been grown using horizontal seeds with screened top side.

in the direction of Z-axis is observed (Fig. 3), although such covering is known to be inexhaustible source of hydrogen ions.

The sweeping inhomogeneity is inevitable since it is caused by the volume polarization occuring at any blocking electrodes. In conclusion it should be fixed attention on that for ensuring long-time stability of precision resonators it is necessary to make sweeping of lumbered cultured quartz bars under conditions of electrical transfer excluding relaxation polarization processes, especially ones which result in thermoelectret state of quartz with large . The choice of such conditions requires consideration of many factors such as electrode types, the manner and efficiency of the electrical contacts of the electrode with the quartz surface, ensuring conditions of exclusion any chemical reactions suppressing H entering the material at the interface "quartz-electrode", correlations between V-T-t etc. We doubt whether these conditions could been carried out by lumbered quartz bars using. The sweeping of quartz wafers seems to be more perspective and the hopeful results have been already obtained (see the table). In this case of course there are many problems too which we attempt to solve.

Another variant of avoiding the uncontrolled instability of resonator characteristics with crystal units made out of swept bars consists in quantitative estimation of thermoelectret life-time and accordingly of the duration of the period whithin which resonator characteristics will be stable. These estimations can be obtained by dielectric relaxation technique (Fig. 4) for different temperatures. Information obtained from such estimation is of a great significance when elaborating resonator technology especially when choicing the method and regime (temperature, time) of electrode covering.

Conclusion

1. The resonator characteristic improvement with crystal elements made of swept lumbered cultural quartz bars is explained by macroscopic polarization formed in quartz under conditions of electric field and high temperature at typical industrial regimes of sweeping and only partially by alkaline ion removal from

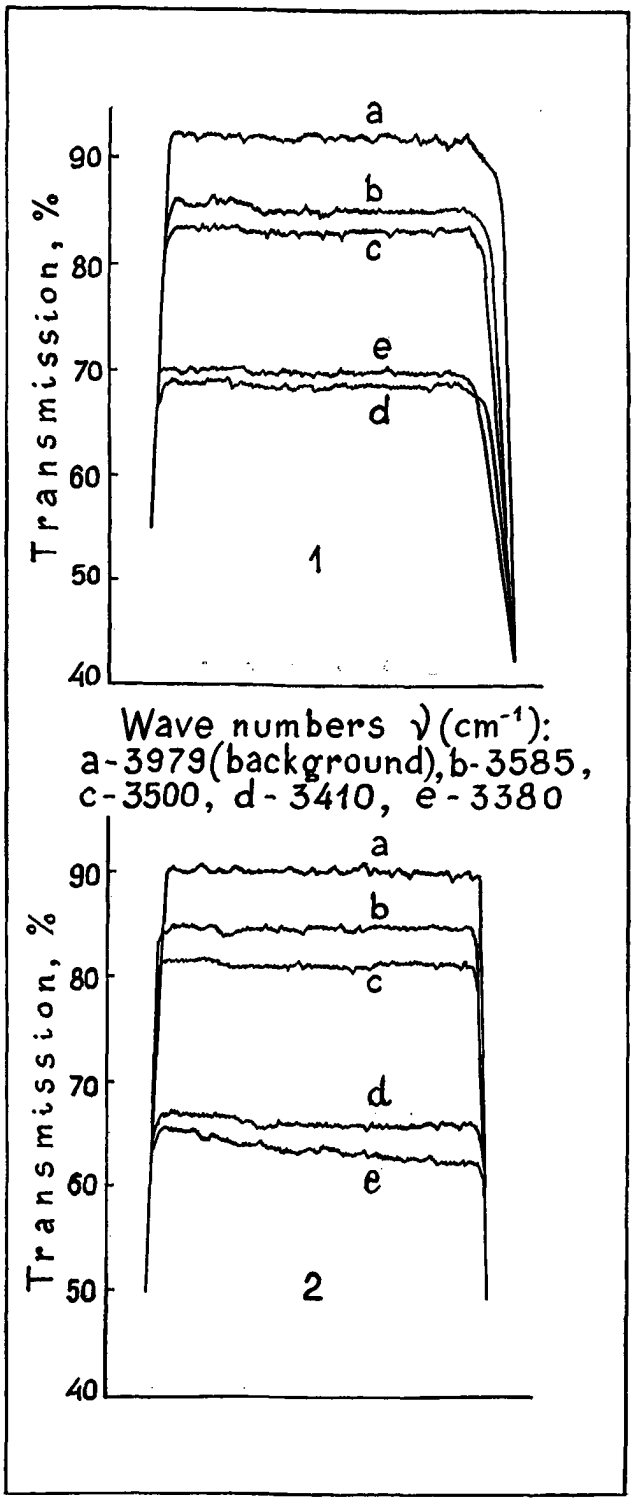


Fig. 3. Variation of the infrared absorption along the scanning line in parallel with Z-axis of lumbered quartz Y-bar before (1) and after (2) sweeping

quartz bars.

2. Swept quartz is a thermoelectret with a very large ($\approx 10^7 \div 10^9$ s) lifetime at room temperature. Depolarization speed which determines thermoelectret life-time and consequently the instability level of resonator characteristics depend on temperature in complicated manner. It should be taken into account when elaborating the resonator technology and especially when choicing the method and regime (temperature, time) of electrode covering.

3. The method of sweeping control by dependencies of current vs time proposed by Vig et al. is more sensitive than other known methods and may be successfully applied for sweeping process control. The absence of current changes after its intensive fall under conditions of electric field is a signal for ceasing high temperature stage of sweeping.

Aknowledgements

We would like to take this opportunity to thank Drs. Juergen H. Staudte and Ya. L. Vorokhovsky for the interest to this work and for useful discussions.

References

1. J. G. Gualtiery, L. Calderon and R. T. Lareau, "Update on Possible Electrode Mechanisms in the Sweeping of Alpha Quartz". Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

2. J. G. Gualtiery, R. T. Lareau and D. W. Eckart, "Electrode Effects in the Sweeping Process of Alfa Quartz". IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 37, no. 5, pp 393-403, 1990.

3. J. G. Gualtiery, J. Kosinski and R. A. Murray, "Nonuniformities in the Air-Sweeping of Quartz". Proceedings of the 44th Annual Symposium on Frequency Control, 1990.

4. D. W. Hart, J. Frank, D. Smith and J. J. Martin, "A Study of the Electrodiffusion Process in Quartz". Proceedings of the 44th Annual Symposium on Frequency Control, 1990.

5. А. Ф. Иоффе, "Механические и электрические свойства кристаллов", Избранные труды., том 1, Ленинград: изд. Наука, 1974, 328 с.

6. J. G. Gualtiery and J. R. Vig, "Sweeping and Irradiation Studies in Quartz". Proceedings of the 38th Annual Frequency Control Symposium, 1984.

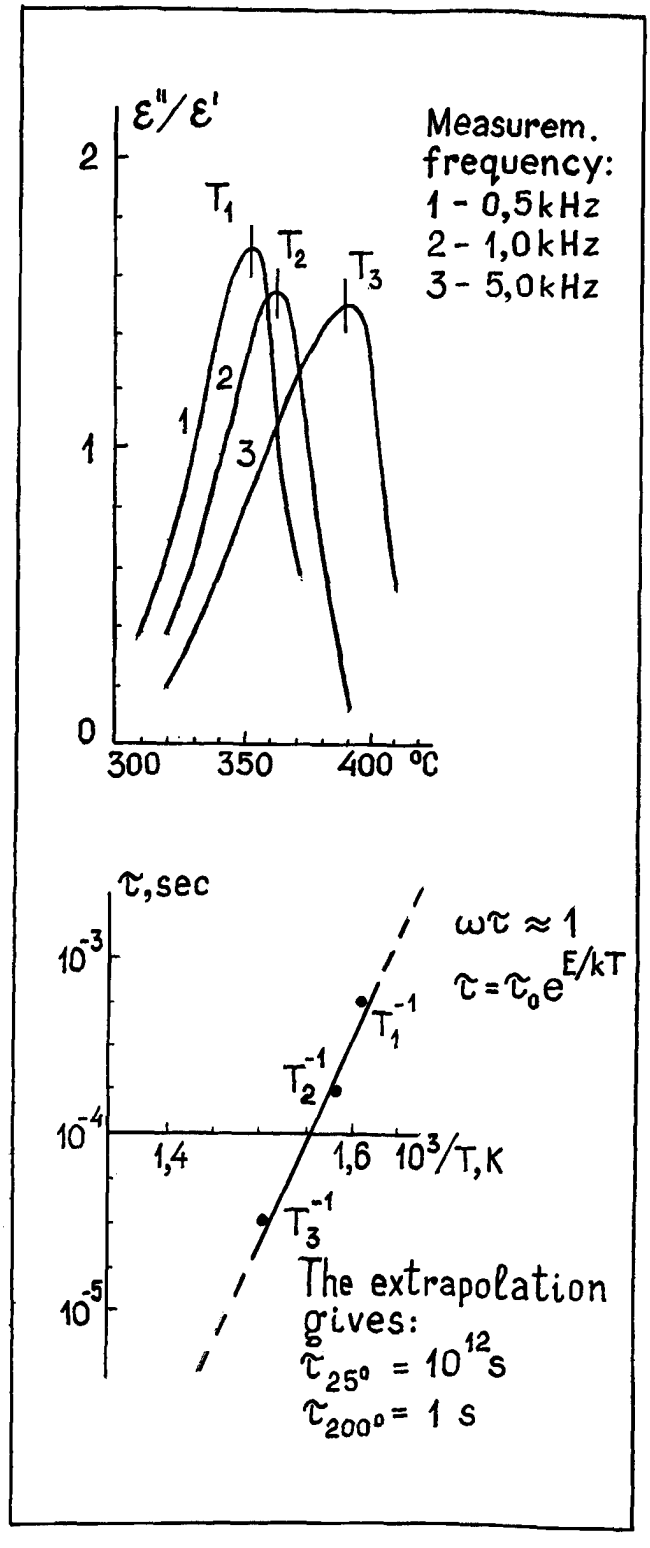


Fig. 4. Relaxation time τ of swept quartz temperature (back to nonpolarizied state)

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

RADIATION RESISTANCE OF SYNTHETIC QUARTZ CRYSTALS

I.V.Kabanovich

Scientific Research Institute "Fonon",
Krasnobogatyrskaya Str. 44, Moscow 105023, Russia

The method for estimating radiation resistance of synthetic quartz crystals grown in the same cycle, on the basis of analysis of changes caused by radiation of isotope Co^{60} in Q-factor and frequency of trial resonators made of that cycle crystals, is proposed. The method is reliable to select quartz crystals appropriate for manufacture of radiation-resistant resonators; it is also effective and convenient to optimize the conditions of high quality quartz crystal growth and develop the construction of precision resonators.

Introduction

The development of nuclear-power engineering and space-system engineering raises the actuality of creation of equipment with high operational reliability in conditions of strong ionizing and impulse radiation. Particular attention is given to the stability of parameters of quartz resonators, filters and pickups in equipment to be exploited under the influence of radiation.

The main requirements to the construction of resonators, which are necessary to achieve high radiation resistance, include the creation of high vacuum, use of materials with a low oxidation potential for electrode covering, excitation of crystal element in electrodes with a gap, etc. But experience shows that much attention should be also payed to selecting quartz of high crystal perfection for manufacture of crystal elements.

It is known that the radiation resistance of quartz is reduced significantly by including into the crystal lattice ions of aluminum and univalent ions compensating electric charge. Such quartz after ionizing radiation turns dark colour. That property is made use of for a rough estimating radiation resistance of quartz. After prior irradiation the most transparent crystals or their parts are selected. However, that way is not sufficient for reliable estimating radiation resistance of quartz, since crystal can contain defects which reduce the radiation resistance without causing colouring in the visible-light range. It is known also that small absorption of infrared light at wavelength of 3500 (or 3585) cm^{-1}, that is high Q-factor, is a necessary but insufficient condition of high radiation resistance of quartz.

The radiation resistance can be estimated by optical method with greater reliability if the absorption spectrum in the range of wavelength 3700 cm^{-1} to 3200 cm^{-1} is measured at 77K [1]. In addition to that method, the determination of the aluminum content by electron-spin resonance was proposed [2]. However, in our opinion, the complexity of necessary equipment hampers wide practical application of the method. Besides, it is not clear whether it allows to take into

0-7803-0476-4/92 $3.00 © 1992 IEEE

account the influence of non-homogeneousness of distribution of impurities on the radiation resistance. Finally, radiation itself induces complicated processes of creation and redistribution of new defects and that makes it still more difficult to prognosticate the radiation resistance of quartz devices before their producing [3].

Method for estimating radiation resistance

We propose a method for estimating radiation resistance of quartz crystals grown in the same cycle, which in our opinion is more acceptable to producers of quartz and resonators. The method is grounded on the analysis of changes caused by radiation in parameters of trial resonators made of a few crystals of the same cycle to be estimated.

Preliminaryly, we measure the linear defect density of every crystal, counting pits formed by hydrothermal etch on the pinacoid surface, and sort crystals into groups with close values of that parameter within each one. Treating each group separately, we select from it the crystal with the greatest values of the growth rate and linear defect density, thus securing the quality of the rest to be not less than that of the selected crystal.

As resonator's construction influences considerably its parameters, a systematic use of our method requires to standardize the construction and manufacture technology of the trial resonator. The most prospective is a vacuum resonator with AT-cut piezoelement with fifth harmonic frequency of $5\cdot10^6$c.p.s., for example, the resonator of which the IEC suggested measuring directly the Q-factor to estimate the Q-factor of quartz. Producer can choose for own systematic use another trial resonator with a similar piezoelement. For example, one can use the resonator with crystal element excited in a gap, which has the merit of lack of electrode covering and hence lack of the influence of processes in treated surface and electrode film on measurements. To spare time and means for test one can use a dismountable construction of resonator and change from test to test trial crystal elements excited in a gap. However, in that case one should secure rather perfect dismountable construction to minimize changes occuring during test in parameters due to non-perfect fastening of crystal element.

We measure fractional changes caused by radiation of isotope Co^{60} in Q-factor ($\delta Q=\Delta Q/Q$) and frequency ($\delta f=\Delta f/f$) of trial resonators.

It is known that at ionizing irradiation the dynamic resistance of AT-cut crystal element, as a rule, increases, reachs its maximum and at further irradiation decreases again. In practice, it is more convenient to observe changes in the Q-factor which have the reciprocal behaviour in comparison with that of the dynamical resistance: the Q-factor of resonators with crystal element of good quality synthetic quartz under the influence of ionizing radiation decreases firstly, reachs its minimum after radiation dose of 10^3-10^5 roentgen and then increases approaching initial value or even surpassing it after radiation dose of 10^6-10^7 r.

The frequency of resonators with synthetic quartz crystal element at ionizing radiation, as a rule, increases and reachs saturation. As one should catch reliably very small changes ($\delta f \sim 10^{-7}$) caused by radiation, one should be sure that deviations during test which connected with imperfection of trial resonator construction are less than they. The best in this respect is precision resonator with metallized piezoelement.

In the case of quartz of high Q-factor, the typical behaviour of fractional

changes in the Q-factor and frequency for precision resonators with plano-convex AT-cut piezoelement of the diameter of 12.5 mm, the fifth harmonic frequency of 5×10^6 c.p.s., is shown in figures 1 and 2, respectively.

The higher the quality and radiation resistance of quartz, the less are fractional changes and radiation doses necessary to reach the minimum of Q-factor and saturation of frequency. This property requires to fix the optimal radiation doses so that for as wide range of comparatively high radiation resistance as possible (in the case of quartz of high Q-factor) they will allow to measure values of the Q-factor close to the minimum and values of the frequency close to the saturation. We propose as optimal the radiation doses of isotope Co^{60} equal to 10^5 r and 10^7 r to determine δQ and δf, respectively.

Analysing test results for resonators of quartz crystals grown at various conditions and having various quality, we have established that only taking into consideration both δQ and δf one can get reliable estimation of radiation resistance. In our opinion, frequency changes first of all result from inclusions of ions of aluminum into crystal lattice and Q-factor changes first of all reflect non-homogeneousness of impurity distribution and also depend on the nature of defects.

The effectiveness of the method, its sensitivity to various factors influencing the radiation resistance, is well illustrated by examples of Table 1, at which test results for resonators with piezoelements made of swept as well as unswept quartz crystals grown at various conditions and containing various density of linear defects are presented.

Comparing test results for resonators of quartz crystals of cycle C-1920, grown in a solution of Na_2CO_3 at the rate of 0.12 mm/day, we can see that the increase

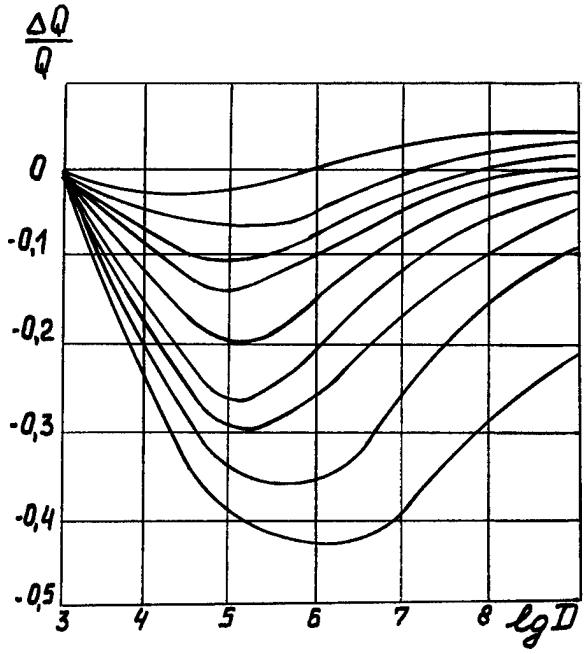


Fig.1. Radiation dependence of the fractional change of Q-factor for resonators made of quartz of various quality.

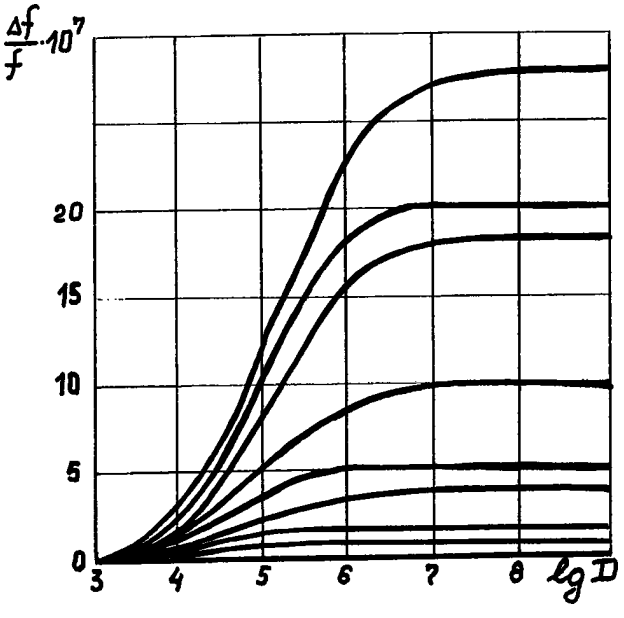


Fig.2. Radiation dependence of the fractional change of frequency for resonators made of quartz of various quality.

Table 1. Results of tests of resonators with piezoelements with fifth harmonic frequency of 5×10^6 c.p.s., made of quartz crystals grown at different conditions; here Q is the Q-factor of resonators before irradiation, δQ_1, δf_1 are fractional changes after the radiation dose of 10^5 r, δQ_2 and δf_2 are fractional changes after the radiation dose of 10^7 r. The quantities of importance are Q, δQ_1 and δf_2.

Cycle No.	Growth rate mm/day	Linear defect density pc/cm^2	Addition to solution of Na_2CO_3	Sweeping	Q-factor $\times10^6$	Q-factor changes $\times100$		frequency changes $\times10^7$	
						δQ_1	δQ_2	δf_1	δf_2
C-1920	0.12	25	without	unswept	2.40	-20	-8.0	1.0	5.0
C-1920	0.12	25	without	swept	2.48	-2.0	-0.5	0.1	0.4
C-1920	0.12	400	without	unswept	2.25	-29	-19	2.5	9.0
C-1920	0.12	400	without	swept	2.45	-2.7	-0.9	0.1	0.6
C-1261	0.12	25	without	unswept	2.25	-31	-3.6	1.9	18.0
C-1261	0.12	25	without	swept	2.40	-25	-1.0	1.2	15.0
C-2225	0.20	50	without	unswept	1.91	-52	-51	6.0	28.0
C-2225	0.20	50	without	swept	2.10	-43	-38	3.0	20.0
C-2116	0.12	50	$LiNO_3$	unswept	2.37	-14	1.2	-0.08	-0.17
C-2116	0.12	50	$LiNO_3$	swept	2.50	2.0	4.0	-0.03	-0.03
C-2116	0.12	300	$LiNO_3$	unswept	2.25	-20	-0.4	-0.2	-0.3
C-2116	0.12	300	$LiNO_3$	swept	2.50	2.0	0.4	-0.02	-0.02

in the linear defect density from 25 cm^{-2} to 400 cm^{-2} entails considerable increases in values of δQ and δf. The rate of Q-factor restoration after further irradiation is greater at the less linear defect density. We also can see the effectiveness of sweeping, which reduces tenfold values of δQ and δf.

Test results in the case of crystals of cycle C-1261, grown in a similar solution and at the same rate as that of crystals of cycle C-1920, even at the low linear defect density of 25 cm^{-2} show much greater values of δQ and δf. Sweeping of crystals of cycle C-1261 gives an insignificant improving effect, which once more emphasizes the difference between crystals of cycles C-1920 and C-1261. This difference is explained by the fact that due to high content of aluminum in charge crystals of cycle C-1261 contain aluminum ten times as much as crystals of cycle C-1920 do. Quartz of cycle C-1261 after the radiation dose of 10^7r had turned dark colour.

Crystals of cycle C-2225 were grown at a similar solution and with use of the same charge as in the case of cycle C-1920, but at the comparatively great rate of 0.2 mm/day. Though that quartz does not turn colour at irradiation its radiation resistance is low. After first irradiation the Q-factor decreases by 52%

and remains low after further irradiation. Sweeping yields no significant increase in the radiation resistance. The cause of the low quality in this case is a non-homogeneous distribution of impurities in the grown layer. Impurities were included as zonary layers along growth front, which was seen at investigating optical homogeneousness and by means of preferential etch. It is the zonary layers that hamper sweeping in this case. Indeed, a bar of that cycle's crystal, which contained a zonary layer boundary (for example, the boundary between seed and grown layer) did not yield to sweeping, while a bar without zonary layer boundaries was effectively swept.

Analysing test results for resonators of quartz crystals of cycle C-2116 differing from cycle C-1920 by addition of solt $LiNO_3$ to the solution, we can again note effectiveness of the proposed method and underline the positive effect of using solts of lithium to grow high-quality quartz. The values of δQ and δf in this case are less, that is the radiation resistance is greater, in comparison with that of quartz of cycle C-1920. For quartz crystals, grown at low rate in a solution with lithium additions, sweeping is very effective. The characteristic feature of swept quartz of cycle C-2116 is the increase of the Q-factor after already the radiation dose of 10^5 r, which grows at further irradiation.

Applications

The proposed method is not only effective but also convenient to develop technology of quartz crystal growth. It allows to get promptly estimations of contributions of various technological factors to changes in crystal perfection and thus to conduct purposeful search for the optimum growth conditions to achieve the utmost quality of crystals in the Q-factor and radiation resistance.

We optimized crystal growth conditions making use of large autoclaves operating at middle pressures (600-800atm.). The most careful measures taken to raise quality (low rate of growth, seeds of high quality quartz, mechanical and chemical treatment of seed surface, charge with small quantities of impurities and others) allowed to achieve $\delta Q<15\%$ and $\delta f<2\cdot10^{-8}$.

We should like to note that searching for the optimal growth conditions we have failed to grow crystals such that resonators of them would have $\delta Q<10\%$ without quartz sweeping, even though the Q-factor surpassed $3\cdot10^6$. Thus, quartz sweeping remains a necessary operation to raise the radiation resistance.

At the same time, we have estableshed that though after sweeping of quartz with high density of linear defects one can achieve a considerable effect of increase in its Q-factor and radiation resistance with respect to continuous radiation, but its radiation resistance to impulse radiation is low in comparison with that of swept quartz with low density of linear defects. We have no satisfactory explanation of the effect. The highest radiation resistance to impulse influences can be achieved with swept quartz whose Q-factor before sweeping was greater than $2.8\cdot10^6$ and linear defect density less than 50 cm^{-2}.

Taking into attention the close connection between values of fractional changes caused by radiation in the frequency of resonator and values of corrections to the cut angle of crystal element with the purpose of securing a given temperature-dependence of frequency, the proposed method can be implemented to produce resonators with high requairements to the identity of temperature-dependence of frequency.

Conclusions

The proposed method for estimating radiation resistance on the basis of analyses of fractional changes caused by fixed doses of ionizing radiation of isotope Co^{60} in the Q-factor and frequency of trial resonators is effective for practical use at testing quality of quartz crystals grown in the same cycle.

The radiation resistance of quartz is considerably decreased by linear defects in crystals. Quartz with low linear defect density is radiation- resistant not only at continuous radiation but also at impulse radiation.

The sweeping is necessary operation to raise the radiation resistance of quartz grown at any conditions, but is the most effective in the case of homogeneous crystals with low linear defect density.

The proposed method is effective and convenient to develop both quartz growth technology and construction of precision resonators to be exploited at the conditions of high radiation level.

References

[1] L.E.Halliburton, M.E.Markes, J.J. Martin, "Point defects in synthetic quartz: a survey of spectroscopic results," in Proceedings of the 34th Annual Symposium on Frequency Control, USA, 1980,pp.1-7.

[2] M.E.Markes, L.E.Halliburton, "Defects in synthetic quartz: radiation induced mobility of interstitial ions," J.Appl.Phys., vol. 50, pp.8172-8186, December 1979.

[3] H.G.Lipson, A.Kahan,"Distribution of aluminum and hydroxide defect centers in irradiated quartz", in Proceedings of the 38th Annual Symposium on Frequency Control USA, 1984, pp.I0-I5.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

The development and production of high quality quartz material for precision crystals

A. P. Pogrebnyak, S. N. Abdrafikov.

The plant "Kristall", Uyzhnouralsk, Russia

To produce the synthetic quartz crystals for the high-quality resonators we put into practice the selection of the seeds, we used the synthetic concentrate of the charge, additions of the salts of the lithium. In this case we have the serial production of the material with Q-value more than 2,4 millions and with density of etch channel less than $100cm^{-1}$

The additional methods of testing of the quality of the quartz are described, for example, hydrothermal etching of the crystals.

Now the synthetic quartz crystals finded wide application for produce crystal units and filters, but problem of cultivate of the technological parameters of growth of the quartz crystal is actual. The propose of this problem is raizing the homogeneity and the perfection of the crystals. In connection with the expansion of frequency's range and the conductions of the exploiting of the resonators more and more high demands are maked to synthetic quartz crystals.

Quartz is the most element in the precision crystals. The technology of hydrotermal synthesis of quartz which fit for production of precision crystal units was working out and put into practice at our plant. In that quartz are quaranted low content of impurity and theirs even distribution along thickness of growth layer, small density of inclusions and etch channels, high Q-value. The production of quartz with aforesaid characterictics are reached by following ways: 1. Use of synthetic quartz crystals as charge for growth.

2. Selection of the seed's material.

3. Add of the lithium's salts to mother liquor.

4. Control of technological parameters during the growth process.

Let's consider these methods more carefully.

1. The initial charge.

Usually the natural vein quartz are used as initial charge for hydrothermal synthesis of quartz. The sum of admixtures in this quartz total 0,02 ppm. The fundamental impureties in the vein quartz are water, aluminium, sodium, iron. After the recrystalization of vein quartz we have material with sum of admixtures total 0,002 ppm and use it as charge for production of the crystals with low content of impurety elements.

2. Selection of seed's material.

The seed is named any peice of crystal which distine for grow it to more large dimensions. In our manufacture the plate from synthetic quartz is used as seed. To the seeds are make the following

0-7803-0476-4/92 $3.00 © 1992 IEEE

demands:

- the plate must be monocrystal, it shouldn't have cracks, bounds of the blocks.
- it shouldn't have the alien inclusions
- seed should be homogeneity, must have low quantity of linear defects.
- it should be growth in such conditions as the crystals will be growth.

Making of stock of quality seed-plates we began by growth of natural quartz pieces. In this material we've testing the absence of twins, alien inclusions, quantity of linear defects. The growth material are cut up for produce the seed plate and initial natural plate are use for growth reiterate. From seeds which were produce by this way, are growth the crystals. In crystals we've testing homogeneity and density of etch channel. To examinate the quality of all crystal we developed indestructive method of estimate - the hydrothermal etching of crystals in the autoclav.

The defects as etch pits are reveal by the method of hydrothermal etching. Figure 1 shows the external view of etch pits. The quality of crystal are estimate in quantity of etch pits on the surface.



Figure 1. The external view of basic surfase of crystal with etch pits, 50x magnification.

The etch pits are the three-cornered pits which are continue in inside of the crystal at the depth of 0,5 mm.

If density of etch pits is small we have the linear interdependency between density etch pits on the basic surface of crystal and density etch channels inside of the crystal. The density etch channels was determined by calculation after etching AT-cut plate in accordance with Standard IEC, Pub. 758. Figure 2 shows relationship this. All grown crystals are divide into groups depending on density etch pits. If the crystal have density less than 400 cm^{-1} we cut up theirs for seeds for commodity production of crystals. All commodity crystals are test by hydrothermal etching method and are divide into groups. It guarantee that density etch channels inside of the crystal less than 100 cm^{-1}.

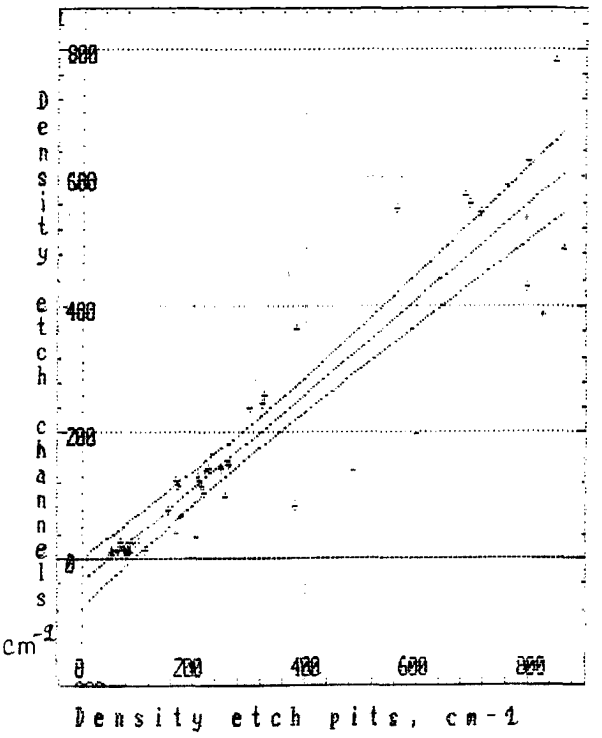


Figure 2. The relationship between density etch pits on the surfase of crystal and density etch channels inside of crystal.

3. The addition of lithium salts to mother liquor.

The pozitive influence of addition of the lithium salts to quality of the quartz crystals, in particular to Q-value was discribed row authors [1-2]. At our plant we apply the additions of lithium salts with purpose to produce quartz material with Q -value more than 2,4 millions. We have selected the technological parameters and the concentrations of the solutions of lithium nitrate. It was found that the addition of lithium change the growth rates of the pyramids basic <c>, negative trigonal prism <-x>, dipyramid <s> in comparison with growth rates in the solusions without litium salts. In our conditions the growth rate of dipyramid <s> is reduce to zero. Figure 3 shows the changes of growth rates of the pyramids <c> and <-x>. The growth rate of basic <c> pyramids is decrease and rate of <-x> pyramid is increase when concentration of lithium nitrate is increase. Absense of the dipyramid <s> in the crystals, which were grown with addition of the lithium nitrate, is well observe at Y-cut sample after gamma-rays irradiated, as shows figure 4.

4. Control of the technological growth parameters.

The structure of crystal is very sensitive to the change of conditions which it was grown. The ensuring of the homogeneity of the crystal are achieve by constant control and adjusting of technological growth parameters if it is necessary. The automatized system of centralize control and adjusting was devised. There is constant control of five technological parameters including the temperature of the crystaline zone, the temperature of dissolve zone. Simultaneously the calculation of gradient of temperature and the exchange of this parameters from the tasks conditions is carry out. The mistake of measuring channel is ± 0,3 %.

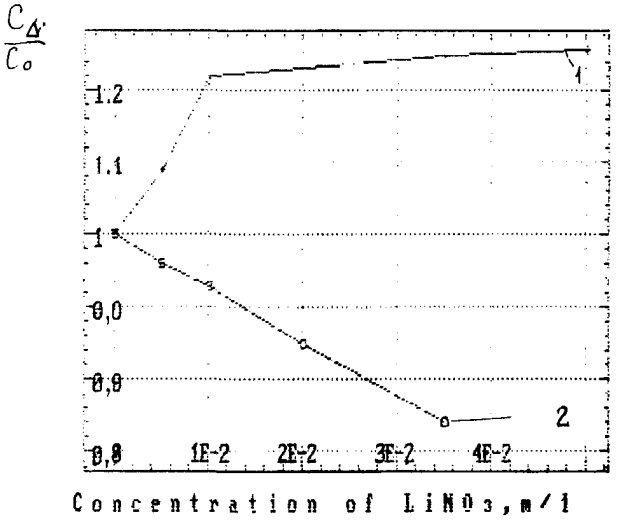


Figure 3. The exchange of growth rate after addition of lithium nitrate.
1- for <-x> pyramid,
2- for <c> pyramid.

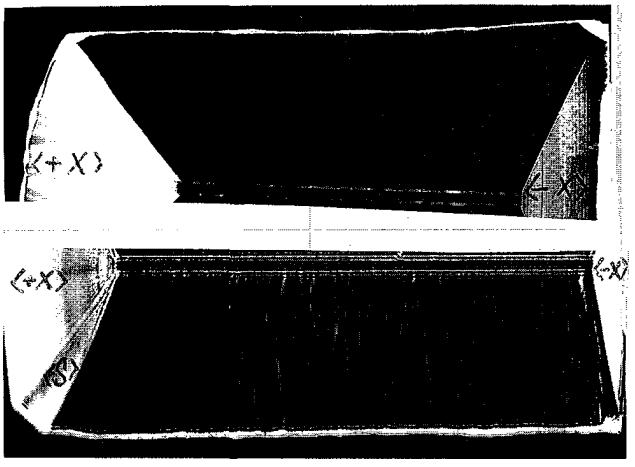


Figure 4. The rotografs of quartz samples Y-cut after gamma-irradiation.
1-with addition lithium nitrate,
2-without addition lithium nitrate.

The algoritms and programs for put into commision of the autoclave for adjusting of temperature rate and for prognosis of growth rate were working out and put into practice.

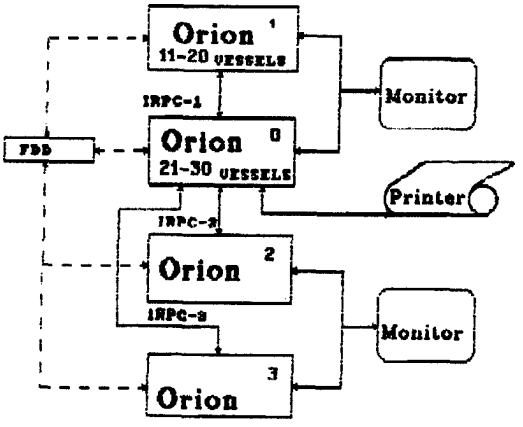


Figure 5. The structure scheme of control and adjusting complex on the basis of microprocessor "Orion".

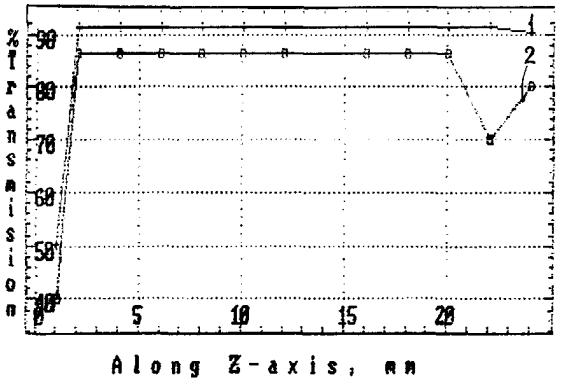


Figure 6. The typical diagram of infrared absorption of our quartz along Z-axis. 1- wave 3800 cm-1, 2- wave 3500 cm-1.

Figure 5 shows the structure scheme of control and adjusting complex.

The control of homogeneity of crystal are carry out by infrared spectrometry. Figure 6 shows the typical diagram of infrared absorption of sample Y-cut.

Synthetic quartz was make by above-stated ways is fit to produce precision crystal units. High quality of this material difine the stability of task properties of crystals and repetition of characterictics, for example, as in internally heated quartz resonators.

References

[1] D. Chakraborty and P. Saha Effects of solvent and growth conditions on mechanical Q ofquartz single crystal Indian J. Phys. 48, 439-449, 1974.

[2] N. C. Lias, E. E. Kolb and R. A. Laudise The growth of high acoustic Q quartz at high growth rates. - J. Crystal Growth 19, 1-6, 1973.

[3] B. Sawyer Q capability indications from infrared absorption measurement for Na_2CO_3 process cultured quartz. - IEE Trans. Sonics Ultrasonics SU-19, 41-44, 1972.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

NUMERICAL CALCULATION OF TEMPERATURE FIELD FOR THE CASE OF SIMPLE CONVECTION MODEL IN AUTOCLAVE

W. Hofman
Institute of Electronic Materials Technology /Z-18/
01-919 Warszawa, POLAND

ABSTRACT

A numerical study of temperature field for the case of 8.5 litre autoclave, filled with water, was performed using the iteration method. The natural convection in liquid was assumed to be steady, axisymmetric and laminar. Two thermal steady states of the vessel were considered, with weak cooling and with intense heat exchange between the surroundings and the baffle region. For both states almost identical local temperatures, under the plunger (350°C) and under the bottom of the vessel (390°C) were achieved. The obtained results show several common features for both thermal states. Firstly, the velocity of water flow is few times higher in the upper than in the lower zone. Secondary, the strong concentration of isotherms exists in the interzone region. Thirdly, the radial and axial temperature gradients become greater when additional cooling is applied to the vessel.

INTRODUCTION

It is known commonly, that practically during hydrothermal growth of quartz crystals only three parameters of the process are taken into account namely,the temperature in two points under the bottom and under the cover of the autoclave and the pressure of solution. In production run of quartz crystals in Poland during more than ten years it was found, that in spite of maintaining those three parameters at the same exact levels in many processes there were certain runs in which the crystals obtained had different sizes, up to +/-10%, from those standards bars. Repeated and exact analysis of such processes induced us to set forth a thesis that a main cause of above mentioned irregularities are local, long-term changes of temperature distribution, close to the external surface of the high pressure vessel, coming into enviroment of crystal growth. Those disturbances by changing the temperature field in the autoclave can have an effect on growth rate of crystals.

EKSPERIMENTAL

The object of the experiment was 8 litre vertical, steel autoclave of Autoclave Engineers Inc. The longitudinal half-section of that vessel is shown in Fig.1. The outside dimensions of the vessel are 60" by 10" and inside 53" by 3.5". The aspect ratio of internal chamber equals .067.

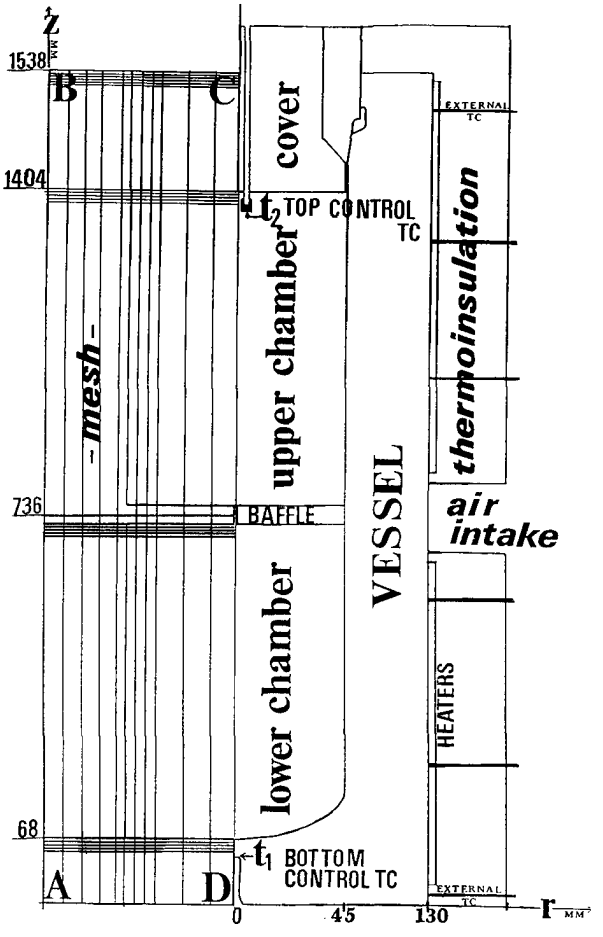


Fig.1 Schematic diagram of the autoclave (right side), model adopted for calculations (left side).

0-7803-0476-4/92 $3.00 © 1992 IEEE

The chamber is devided into two halfs by means of a steel baffle. 12mm thick baffle with 3.5 percent opening was used. Each chamber of the autoclave had independent heating section which was controlled by its control thermocouple, t_1 for bottom and t_2 for top zone. There was an air ventilating duck, around the body of the vessel, between the heating sections. By means of that drug the central part of the vessel, at the level where the baffle was located, could be cool directly. The set of external thermocouples were installed in proper way along the t_1DABCt_2 contour. 20 cm thick thermoinsulating layer was put on the heaters. The base and the top of the autoclave were also covered with thermoinsulating material. The vessel was leveled exactly and filled with water up to 78% of its free volume. After chamber was sealed, the blower and heaters were switched on. When the temperatures, 390 $\pm 1^{o}$C at t_1 point and 350 $\pm 1^{o}$C at t_2 point were reached, the water pressure was 19900 to 20000 psi. For that state the external temperatures were measured by means of installed thermocouples. In the second part of experiment, the blower was switched off and the system was naturally air cooled. By small reduction of electric power in heating system the temperatures: 390 $\pm 1^{o}$C and 350 $\pm 1^{o}$C at points t_1 and t_2 were achieved once again (the change of electric power was less than 10%). When new thermal steady state had been achieved the pressure of water increased about 200 psi. In order to obtain the previous pressure a little water was let out and external temperatures were measured again. As proper, long thermowell was not available the temperature distribution along the centerline of the system was measured at normal conditions namely, at temperature of water equals 90°C at point t_1 and 50°C at point t_2, one time with forced cooling and second time with natural air flow. It should be mentioned, that external temperature changes recorded at these new conditions were similar to those recorded at the hight pressure. The smoothed profiles of temperature measured along AB and DC edges are shown in Fig.2. The curves for the first thermal state were plotted with dashed line and was marked as I_{AB} and I_{DC} respectively. It is seen ,that for both AB profiles there are deep minima at the level of air channel. Moreover there are local maxima, lower and upper, connected with heaters work. For the state with forced cooling the minimum is positivly deeper and the first maximum higher than for the state with natural cooling. In the case of inside profiles, which were shifted to high

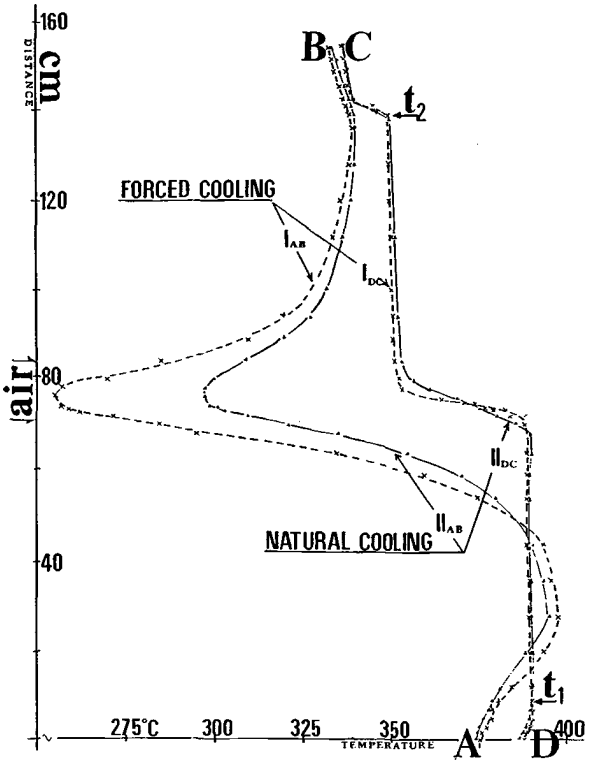


Fig.2 Axial temperature profile for AB line and CD centerline

temperature region in Fig.2 (curves: I_{DC} and II_{DC}) there is sharp temperature gradient near the baffle. That gradient is smaller for the second thermal state. It is characteristic that there is another sharp temperature gradient in the lower part of the plunger. Similar results were reported by Johnson et al [1].

MODEL OF CONVECTION

To determine the temperature field of the autoclave, nine differential equation have to be solved simultaneously, namely the equation of energy for walls and the baffle and the equations of motion, energy and continuity for fluid in each chamber of the reactor. As there are no experimental data for boundary conditions for the equations of motion, the solution was possible when few simplified assumptions were adopted for the model.

The list of assumptions is as follows:

(1) The real vessel shown in Fig.1 is replaced by heavy-walled cylindrical tube with flat thick bases (E shape profile at the left side in Fig.1).

(2) The temperature distribution at the outer surface of the vessel is axisymmetric.

(3) There is one closed loop of laminar fluid convection in each chamber of the vessel, as it is shown in Fig.3a, and flow along the central axis is opposite to flow near walls.

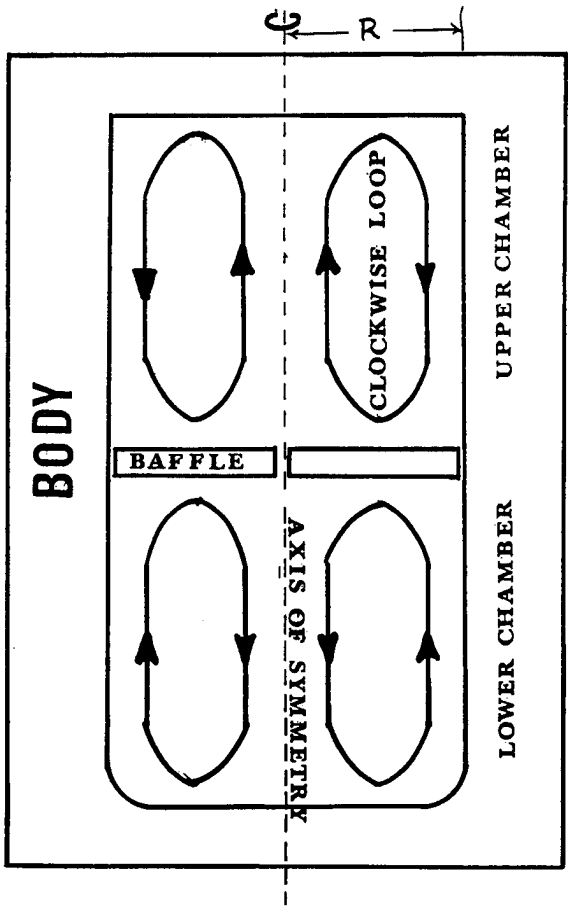


Fig.3a Flow convection loops in autoclave

(4) The radial changes of axial velocity ϑ_z was desribed by a simple staircase function shown in Fig.3b [2]. In the case of clockwise circulation in the region called INTERNAL the fluid streams move with the velocity ϑ_z above a circular section. The area of that section equals 45% of internal area of the vessel. In thin TRANSIENT region the velocity drops to zero. The area of annular cross section of transient region equals 8% of internal area of the vessel. Next to the transient region there is EXTERNAL region where flow is reversed. The value of flow velocity and the area of cross section of external region is the same as for internal region. In the last region called PERIPHEDRAL the velocity drops to null again. In the case of counter-clockwise circulation the sequence of stairs should be reversed.

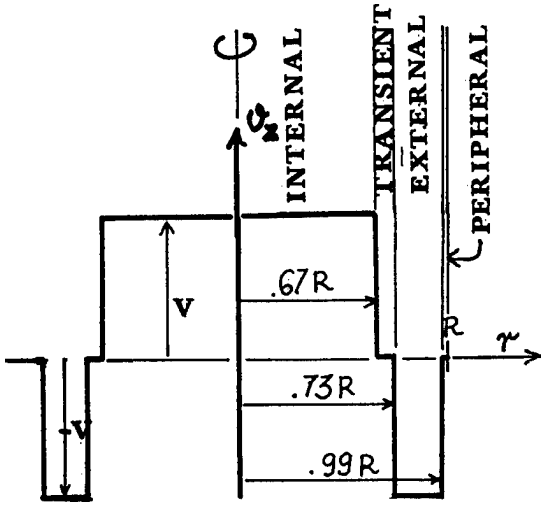


Fig.3b Radial changes of axial velocity in the upper chamber

(5) A linear dependence between the velocity ϑ_z and the current difference of average temperatures of INTERNAL and EXTERNAL regions was assumed,

$$|\vartheta_z| = p \; |\langle T_{int} \rangle - \langle T_{ext} \rangle| \qquad (1)$$

where p is constant in calculations and the average temperatures of INTERNAL and EXTERNAL regions are defined as,

$$\langle T_{int} \rangle = (1/V_{int}) \int_{Vint} T(r,z)dV \qquad (2)$$

$$\langle T_{int} \rangle = (1/V_{int}) \int_{Vext} T(r,z)dV \qquad (3)$$

where V_{int} and V_{ext} are the volumes of Internal and External regions and the temperature T(r,z) is a function of radial (r) and axial (z) coordinate because the system is axisymmetric.

(6) As the percent baffle opening is small the heat exchange by convection between autoclave chambers was neglected.

The adoption of above assumptions reduces the number of differential equations from nine to three .The solution that has to be found, should satisfy the set of three equations of energy,

$$\Delta T(r,z) + \left(\gamma c \vartheta_z / \lambda_{wat}\right) \frac{\partial T(r,z)}{\partial z} = 0 \quad (4)$$

where the second (convection) term disapperes for walls and baffle domains and: γ, c, λ_{wat}, are: the desity, the heat capacity and the thermal conductivity of fluid.

For the investigated problem the following boundary conditions were used,

$$-\lambda_{ves} \frac{\partial T(r,z)}{\partial n}\Big|_{S_{int}} = (T_{ves}(r,z) - T_{wat}(r,z)) \quad (5)$$

at the internal walls (also, at the surfaces of the baffle), where T_{ves} and T_{wat} are the temperatures of vessel internal wall and water, close to the internal wall, α is the heat transfer coefficient between the wall and water, λ_{ves} is the thermal conductivity of vessel wall and a derivative is taken perpendilary (n vector) to the vessel wall,

$$T(r,z)\Big|_{S_{out}} = T_{out}(r,z) \quad (6)$$

for outer boundaries where T_{out} is the profile of temperature measured along ABCD contour,

$$\frac{\partial T}{\partial r}\Big|_{r=0} = 0 \quad (7)$$

at the axis of symmetry.

Below there are property values taken for calculations:

density of fluid $\gamma=7.80\text{x}10^{-4}$ [g/mm^3]

heat capacity of fluid c=4.18 [J/g deg]

thermal conduc.of fluid $\lambda_{wat}=5.80\text{x}10^{-4}$ [W/mm deg]

thermal conduc.of vess. $\lambda_{ves}=2.66\text{x}10^{-2}$ [W/mm deg]

heat transfer coefficient $\alpha=3.00\text{x}10^{-3}$ [W/mm^2 deg]

The formulated problem was solved by the difference method using a rectangular mesh with a very fine and constant vertical step and with a coarse, and variable horizontal step. To obtain the unique and convergent solution the velocity ϑ_z had to satisfy the following condition,

$$\vartheta_z < \frac{2\,\lambda_{wat}}{\gamma c d_z} \quad (8)$$

where d_z is a vertical step of the mesh.

The set of difference equations for almost 40000 nodal points was solved by the iteration method. The calculations were stopped when the maximum temperature difference between two succeeding iteration cycles was less then 10^{-2}deg.

NUMERICAL RESULTS

The distributions of selected izotherms for both investigated states and for the case of weak convection ($\vartheta_z \cong 0$) are shown in Fig.4. The locations of bottom, the baffle and the lower surface of cover are shown by short horizontal segments. The locations of ventilation duck and the control thermocouples: t_1,t_2 are marked with letter symbols.

Analysing the distribution of isotherms a characterystic concentration of lines can be seen along the internal vessel walls and along baffle surfaces. To make the estimation easer, the values of normal component of temperature gradient were given for the regions with the highest temperature changes. When corresponding values are compared, it is seen, that for the state with forced cooling the gradient components are distinctly greater especially for the region around the baffle. For instance the axial and radial gradients shown in Fig.4a are one and a half time greater then those in Fig 4b In the table I, the average values of integral temperatures are presented for both states.

TABLE I

	<Twat> [^{o}C]	<DT> [^{o}C]	<DTt> [^{o}C]	<DTb> [^{o}C]
state with forced cool.	369.1	39.7	4.2	1.2
state with natural cool.	370.3	37.7	3.8	0.9
difference	-1.2	2.0	0.4	0.3

where:

<Twat> is the average water temperature,

<DT> is the average difference of temperature between the upper and the lower chamber

<DTt> is the average difference of temperature between internal and external streams in the top chamber,

<DTb> is the average difference of temperature between internal and external streams in the bottom chamber.

In the case of average water temperatures the calculated results corresponds to experimental date. The transition from one state to another caused the increase of water pressure to 200 psi. According to pressure-temperature date for water in hydrothermal conditions, the increase of average temperature for $1.2^{o}C$ should cause the increase of pressure of about 250 psi [3]. In the case of average temperature difference between the chambers and between the streams in the same chamber, the higher values were obtained for the state with forced cooling. Those results confirm an assumption that the temperature disturbances at the outer surface of the vessel can lead to the significant internal variations of the temperature.

It is known from practice, that growth rate of quartz crystals depends highly on changes of the temperature difference between t_1 and t_2 points. It is expected, that the same regularity is true for the difference of the average integral temperatures . It is probable that <DT> directly influence on supersaturation in crystallization zone so it can change crystal growth rate.

Finally, the changes in isotherm distributions were calculated for forced cooled state for p=.3 and p =.4 [mm/s deg] what gave ϑ_z =.5m/h and 1m/h in the upper chamber, respectively (Fig.5) .As it can be seen when ϑ_z velocity becomes higher isotherms are shifted up in the top chamber and the axial temperature gradient becomes smaller in the interzone region. Certainly, the radial temperature differences <DTt> and <DTb> increase and they are equal $4.6^{o}C$ and $1.3^{o}C$ for p =.3 and $6.9^{o}C$ and $2.3^{o}C$ for p =.4 [mm/s deg].

CONCLUSIONS

There is high probability, that during hydrothermal processes there are local temperature changes at the outer surface of autoclave due to failure of heating elements and thermoinsulation or simply due to variations of air temperature around the vessel.

Those disturbances, can change the average temperatures of solution and in this way they can alter the effective suppersaturation in the region of crystal growth and the intensities of convection in both chambers of the reactor.

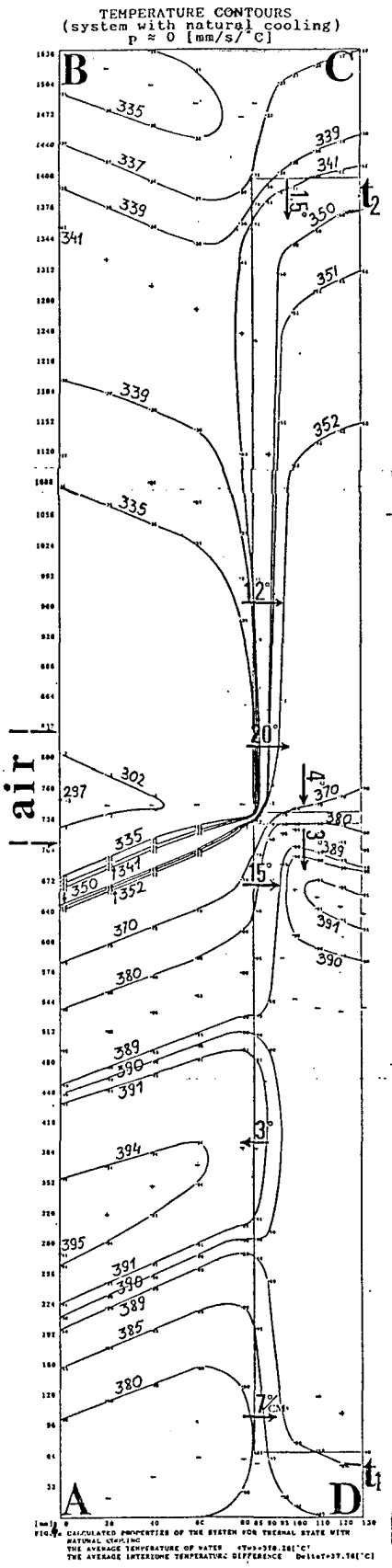
It seems useful to instal additional control thermocouples, especially in the cooling region , as additional indicators for process estimation.

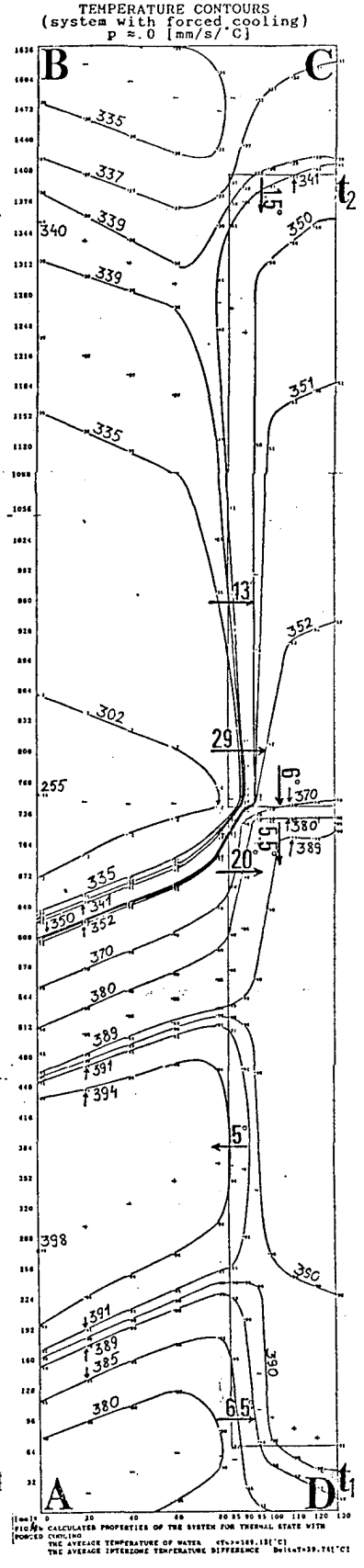
REFERENCES

[1] G.R.Johnson, R.A.Irvine, J.Foise, "Eksperimental Determination of the Relationship Among Baffle, Temperature Difference and Power for the Hydrothermal Growth of Quartz," Proc.of 43rd Ann.Symp. on Freq.Cont., 1989, pp.447-458.

[2] W.Łobodzinski, K.Eckersdorfowa, T.Gozdecki, R.Łojek, L.Michalski, "Opracowanie Projektu Koncepcyjnego Układu Ogrzewania Autoklawu do Hodowli Kwarcu, "Praca ITR-u, 1971, §5, pp.16-17.

[3] G.C.Kennedy "Pressure-Volume-Temperature Relations in Water at Elevated Temperatures and Pressures," Am.J.Sci., vol.248, pp.540-546, 1950.


TEMPERATURE CONTOURS
(system with natural cooling)
p ≈ 0 [mm/s/°C]
B
C
A
D
air
335
337
339
341
350
351
352
339
335
302
297
370
380
389
390
391
394
395
385
12°
20°
15°
3°
7°
1.5°
4°
3°
t2
t1


TEMPERATURE CONTOURS
(system with forced cooling)
p ≈.0 [mm/s/°C]
B
C
A
D
335
337
339
340
341
350
351
352
302
255
370
380
389
391
394
398
385
13
29
20°
5°
6.5
1.5°
6°
5.5°
t2
t1

TEMPERATURE CONTOURS IN AUTOCLAVE FOR DIFFERENT VALUES OF p COEFFICIENT

p=.030[mm/s/°C] p=.040[mm/s/°C]

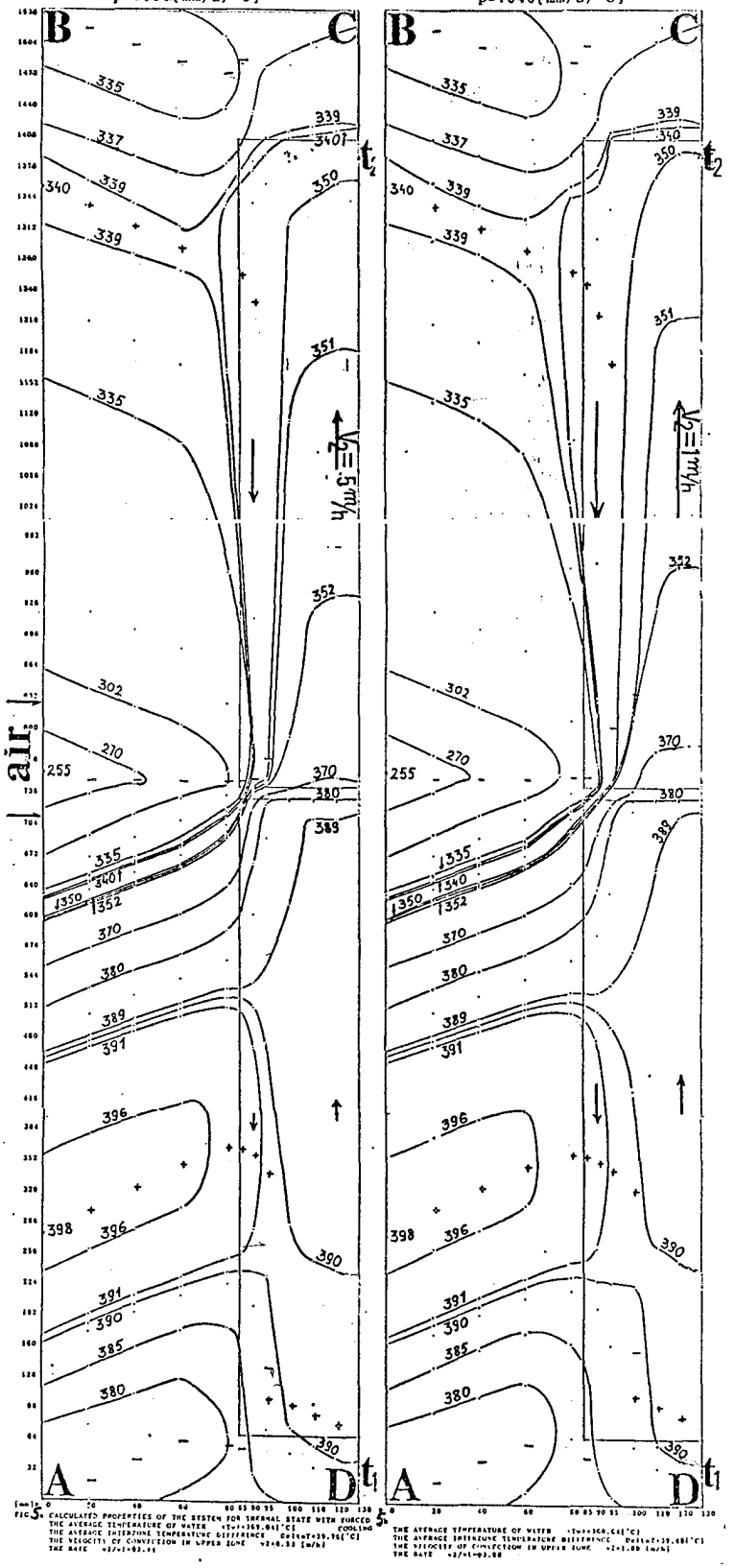


FIG 5. CALCULATED PROPERTIES OF THE SYSTEM FOR THERMAL STATE WITH FORCED COOLING

THE AVERAGE TEMPERATURE OF WATER t_{av}=363.84[°C]

THE AVERAGE INTERZONE TEMPERATURE DIFFERENCE Δt_{av}=39.76[°C]

THE VELOCITY OF CONVECTION IN UPPER ZONE v_2=0.53 [m/h]

THE RATE v_2/v_1=83.4%

THE AVERAGE TEMPERATURE OF WATER t_{av}=360.64[°C]

THE AVERAGE INTERZONE TEMPERATURE DIFFERENCE Δt_{av}=39.68[°C]

THE VELOCITY OF CONVECTION IN UPPER ZONE v_2=1.00 [m/h]

THE RATE v_2/v_1=93.8%

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DISREGARDED SOURCES OF RANDOM ERRORS IN THE NONLINEAR CONSTANTS OF QUARTZ

Carl K. Hruska

Piezoelectricity Research Laboratory
Department of Mathematics and Statistics, York University
North York, Ontario M3J 1P3, Canada

Standard errors of the third-order nonlinear material constants of quartz obtained by the least-squares fit reflect only the random errors from a single source. The source is the experimental data set used to determine the nonlinear constants. This paper points towards other possible sources of errors, namely the linear material constants and the crystal specimen orientation. To minimize their effect, the most important factor is the accuracy of the linear elastic constants. If these constants have standard errors of 1%, they produce nontrivial errors in the nonlinear constants, about half the size of the errors due to the experimental errors. It appears that the errors caused by the other linear constants and by the orientation angles are of substantially lesser importance.

Introduction

The third-order nonlinear material constants of quartz are computed from experimental data using overdetermined linear systems and the least-squares method. This method provides the estimates of the nonlinear constants and their standard errors [1].

The standard errors of the nonlinear constants obtained by the least-squares process reflect only the random errors in the experimental data. The effect of errors in other quantities participating in the calculation are not taken into account. Prominent among these quantities are the linear material constants and the crystallographic orientation angles of the quartz crystal specimens used to provide the experimental data.

This paper addresses - for the first time - the question of the effect on the estimates of the nonlinear constants exerted by random errors in the linear material constants and in the specimen orientation angles.

The work is done for the nonlinear electromechanical constants obtainable by the transit-time method employing the dc field interactions. This method can yield three (out of the total number of four) complete sets of the nonlinear constants which are the electroelastic, electrostrictive and third-order dielectric constants. This work will thus be centered on the errors of these constants.

The numerical values of all quantities needed here are stated for right-hand quartz and the frame of reference according to the IEEE 1978 Standard [2].

Determination of the nonlinear constants

The third-order material nonlinearities in quartz are studied by means of various experiments. Past experience indicates that the theoretical relationship of m independent observations of an experimental quantity L to the third-order nonlinear constants can be recorded using a system of linear equations

$$L_i = \sum_{j=1}^{n} M_{ij} x_j + E_i \,, \qquad (1)$$

where $i = 1,2,\ldots,m$.

In system (1) L_i represent individual observations of the experimental quantity L. The coefficients M_{ij} as well as the absolute terms E_i are functions of a number of parameters describing the linear material properties of quartz, the experimental crystal specimens used and other conditions of the conducted experiment. The quantities x_j are the sought nonlinear constants.

It is accepted that the observations of L_i are not accurate because of experimental errors. To minimize the effect of these errors, the number m of observations L_i is made as large as possible. It is always larger than the number n of the unknown nonlinear constants to be determined from the linear system. The errors in L_i then make system (1) overdetermined.

The values of the unknown nonlinear constants x_j are sought from system (1) using a data reduction process called the least-squares fit. It provides the estimates of the values of the nonlinear constants x_j and their standard errors. They are referred to as the solution of system (1).

For the solution of the above least-squares process to be valid, a number of assumptions should be satisfied. The theoretical definitions of quantities M_{ij} and E_i in system (1) should be

0-7803-0476-4/92 $3.00 © 1992 IEEE

correct. The values of these quantities should be determined with an absolute accuracy. Observations L_i should be free of systematic errors. The only uncertainty allowed in system (1) are the random experimental errors in observations L_i which are the sole reason for the nonzero standard errors of the estimates of the nonlinear constants. Failure to comply with these assumptions may reduce the least-squares process to a mere application of the least-squares algorithm [1]. It may no longer produce the estimates of x_j and their standard errors with the meaning attributed to them by the theory of probability. Unfortunately, under the practical conditions, minor infractions of these assumptions are unavoidable. They have always been silently tolerated.

Among the parameters defining coefficients M_{ij} and absolute terms E_i are the linear material constants of quartz. They are the six elastic moduli, c_{11}, c_{12}, c_{13}, c_{14}, c_{33}, c_{44}, the two piezoelectric moduli, e_{11} and e_{14}, and the two dielectric constants, ε_{11} and ε_{33}.

The orientation of the crystal specimens used in the experiment is generally given by the IEEE orientation symbol $(xzlwt)\psi_k/\phi_k/\theta_k$ [2] with three orientation angles ψ_k, ϕ_k, and θ_k describing each specimen; k = 1,2,...,s, where s is the number of crystal specimens used to determine the experimental values L_i. The angles ψ_k, ϕ_k, and θ_k also belong to the parameters defining M_{ij} and E_i in system (1).

In order to calculate M_{ij} and E_i in system (1), the 'best' available values of the above parameters are used. These values will be denoted

$$c_{11}^B,\ c_{12}^B,\ \ldots,\ c_{44}^B,\ e_{11}^B,\ e_{14}^B,\ \varepsilon_{11}^B,\ \varepsilon_{33}^B, \quad (2)$$

$$\psi_1^B,\ \phi_1^B,\ \theta_1^B,\ \ldots,\ \psi_s^B,\ \phi_s^B,\ \text{and}\ \theta_s^B\ . \quad (3)$$

Accordingly, the solution of system (1), i.e., the estimates of nonlinear constants x_j and their standard errors, will be denoted x_j^B and $\vartheta^B(x_j)$, respectively (j = 1,2, ..., n).

The best available values of the above parameters need not be necessarily their true values. As such, they may produce coefficients M_{ij} and absolute terms E_i which are not quite accurate. This is one of the typical infractions of the assumptions of the least-squares process which must also be overlooked in this paper. This applies here as well as in the next section where the parameter values are randomized.

Problem definition and solution method used

Returning to the fact that the best available values (2) and (3) are not necessarily the true values of the parameters defining M_{ij} and E_i. It can be asked how the solution of system (1) would be affected by any possible errors in the parameters. In order to provide the answer, a simulation process has been used in this paper.

Suppose that the 'best' value of a parameter p is called p^B and that the accuracy of this value is characterized by a standard error $\vartheta^B(p)$. In order to asses the effect of $\vartheta^B(p)$ on the solution of system (1), the parameter value will be randomized using the following equation

$$p^R = p^B + N(0,1).\vartheta^B(p)\ . \quad (4)$$

In Eq.(4) N(0,1) is the standard normal variate whose value has been obtained at random; p^R is a randomized value of parameter p. Suppose that the randomized value p^R replaces p^B in the calculation of coefficients M_{ij} and absolute terms E_i and that the solution of system (1) is then obtained. If this process is repeated many times, with N(0,1) always selected at random, a spectrum of different solutions of system (1) is obtained. This spectrum illustrates the effect of standard error $\vartheta^B(p)$ on the estimates of the nonlinear constants x_j.

The above randomization process was applied to all 10 linear constants, starting with

$$c_{11}^R = c_{11}^B + N(0,1).\vartheta^B(c_{11})$$

and ending with

$$\varepsilon_{33}^R = \varepsilon_{33}^B + N(0,1).\vartheta^B(\varepsilon_{33})\ .$$

Being selected by a (quasi-) random computer generator, the values of N(0,1) were generally different in each case. All randomized values of the linear constants were then used simultaneously to produce randomized coefficients M_{ij} and absolute terms E_i and then the solution of system (1). To isolate the effect of the variations in the linear constants, all other quantities in system (1) were naturally kept constant. This entire process was repeated 100 times. A set of 100 least-squares solutions of system (1) was thus obtained to assess the comibend effect of standard errors $\vartheta^B(c_{11})$, $\vartheta^B(c_{12})$, ..., and $\vartheta^B(\varepsilon_{33})$ on the estimates of the nonlinear constants x_j.

In some cases, when the effect of the standard errors of only some linear constants was of interest, e.g., of the six elastic constants alone, or of an individual constant, the randomization was applied selectively. The remaining linear constants were then kept constant and equal to their original values (2).

Later, the same randomization was also applied to all orientation angles, starting with

$$\psi_1^R = \psi_1^B + N(0,1).\vartheta^B(\psi_1)$$

and ending with

$$\theta_s^R = \theta_s^B + N(0,1).\vartheta^B(\theta_s)\ .$$

Again, a set of 100 solutions was obtained to assess the combined effect of standard errors $\vartheta^B(\psi_1)$, $\vartheta^B(\phi_1)$, ..., and $\vartheta^B(\theta_s)$ on the estimates of x_j.

To serve as the set of the best available values (2) of the linear constants of quartz, the values published by Bechmann [3] were selected. The reason for this selection was that it is probably the most widely used complete set. The choice does not suggest that the values provided by other authors are in any way inferior. Moreover, the choice of any particular set is unlikely to affect our results in any substantial way.

The set of linear constants [3] dates back to the times when it was not common to provide the values of material constants with their standard errors. As none are available, the analysis made in this paper is done assuming that the relative standard errors in the linear constants are generally equal to r%.

There is no way to say how the standard errors of the individual linear constants of any given type may differ in magnitude from one another. As a result, the following uniform definitions of the standard errors have been adopted for the standard errors of the elastic, piezoelectric and dielectric constants, respectively,

$$\vartheta^B(c_{11}) = \vartheta^B(c_{12}) = \ldots = \vartheta^B(c_{44}) = r.(|c^B_{11}| + |c^B_{12}| + \ldots + |c^B_{44}|)/600 , \quad (5)$$

$$\vartheta^B(e_{11}) = \vartheta^B(e_{14}) = r.(|e^B_{11}| + |e^B_{14}|)/200 , \quad (6)$$

$$\vartheta^B(\varepsilon_{11}) = \vartheta^B(\varepsilon_{33}) = r.(|c^B_{11}| + |\varepsilon^B_{33}|)/200 . \quad (7)$$

These definitions have been used wherever the relative standard error of any given type of the linear constants was said to be r%.

The set of the best available values (3) of the orientation angles must naturally coincide with the nominal orientation angles given in the description of the experiment providing observations L_i. Data regarding the errors of such nominal values are usually hard to obtain and/or difficult to interpret in simple statistical terms. The analysis made in this paper is done using a simplifying assumption that the standard error of all nominal values of the orientation angles is the same and generally equal to δ angular minutes. Consequently,

$$\vartheta^B(\psi_1) = \vartheta^B(\phi_1) = \ldots = \vartheta^B(\theta_s) = \delta . \quad (8)$$

As indicated above, each randomization process was repeated 100 times, producing 100 solutions of system (1). This means that a spectrum of 100 estimates has always been obtained for each nonlinear constant x_j. The mean value and the standard deviation was obtained for each spectrum and denoted x^R_j and $\vartheta^R(x_j)$, respectively; $j = 1,2, \ldots,n$.

The standard deviations $\vartheta^R(x_j)$ measure the spread in the estimates of the nonlinear constants x_j. For the sake of simplicity, they have been converted to a standardized percentage form

$$\vartheta^R_p(x_j) = 100.\vartheta^R(x_j)/|x^R_j| , \quad (9)$$

where $j = 1,2, \ldots, n$. They are referred to in this paper as the percentage errors of the estimates of the nonlinear constant x_j due to standard errors (5)-(7) of the linear constants or standard errors (8) of the orientation angles.

In order to evaluate the importance of errors assocciated with the linear constants and orientation angles, errors (9) are compared in this paper with similarly defined quantities

$$\vartheta^B_p(x_j) = 100.\vartheta^B(x_j)/|x^B_j| , \quad (10)$$

$j = 1,2, \ldots, n$, which are the percentage errors in the estimates of x_j due to random experimental errors in observations L_i. In making any such comparisons it should be noted that for all calculations made in this work the differences between x^R_j and x^B_j were found negligible.

In general, the errors of 31 nonlinear electromechanical constants of quartz can be investigated using Eqs.(9) and (10). They include constants of 4 types:

$$f_{111}, f_{113}, f_{114}, f_{122}, f_{124}, f_{134}, f_{144}, f_{315}, \quad (11)$$

which are the eight independent electroelastic constants,

$$l_{11}, l_{12}, l_{13}, l_{14}, l_{31}, l_{33}, l_{41}, l_{44}, \quad (12)$$

which are the eight independent electrostrictive constants,

$$\kappa_{111}, \quad (13)$$

which is the (only) independent third-order dielectric constant, and

$$c_{111}, c_{112}, c_{113}, c_{114}, c_{123}, c_{124}, c_{133}, c_{134}, c_{144}, c_{155}, c_{222}, c_{333}, c_{344}, c_{444}, \quad (14)$$

which are the fourteen independent third-order elastic constants. The electrostrictive constants (12) are elements of the total electrostrictive tensor according to Nelson [4] which includes the vacuum electric-stress tensor.

Stating the values of $\vartheta^R_p(x_j)$ for all 31 individual nonlinear constants (11)-(14), i.e., for $\vartheta^R_p(f_{111})$, $\vartheta^R_p(f_{113})$, etc., would make the results too numerous and difficult to evaluate. Moreover, while the values of $\vartheta^R(x_j)$ for the same type of nonlinear constants x_j are of similar magnitude, the same is not always true for the corresponding values of x^R_j. Some of the percentage errors $\vartheta^R_p(x_j)$ may thus become misleadingly large or small. For these reasons it is practical to reduce the number of results by summarizing them

as follows

$$\vartheta_p^R(f_{***}) = 100.[\vartheta^R(f_{111}) + \vartheta^R(f_{113}) + \ldots + \vartheta^R(f_{315})] / [|f_{111}^R| + |f_{113}^R| + \ldots + |f_{315}^R|] , \quad (15)$$

$$\vartheta_p^R(l_{**}) = 100.[\vartheta^R(l_{11}) + \vartheta^R(l_{12}) + \ldots + \vartheta^R(l_{44})] / [|l_{11}^R| + |l_{12}^R| + \ldots + |l_{44}^R|] , \quad (16)$$

$$\vartheta_p^R(\kappa_{***}) = 100.\vartheta^R(\kappa_{111})/|\kappa_{111}^R| \quad (17)$$

$$\vartheta_p^R(c_{***}) = 100.[\vartheta^R(c_{111}) + \vartheta^R(c_{112}) + \ldots + \vartheta^R(c_{444})] / [|c_{111}^R| + |c_{112}^R| + \ldots + |c_{444}^R|] , \quad (18)$$

and refer to them as the characteristic percentage errors of the electroelastic, electrostrictive, third-order dielectric and third-order elastic constants, respectively.

The same type of summarization can be applied to the percentage errors (10) to arrive at the characteristic percentage errors $\vartheta_p^B(f_{***})$, $\vartheta_p^B(l_{**})$, $\vartheta_p^B(\kappa_{***})$, and $\vartheta_p^B(c_{***})$ which are representing the effect of the errors in observations L_i. It is in terms of the characteristic percentage errors that the results will be stated and compared.

Results

The described procedure was applied to an overdetermined linear system of 24 equations formulated by Kittinger, Tichy and Friedel [5] for their transit-time observations of the dc field interactions with quartz. As shown later by Hruska [6], the maximum potential of the linear system is to estimate the complete set of the electroelastic constants (11), electrostrictive constants (12) and the third-order dielectric constant (13). The system is not suitable to determine the third-order elastic constants (14), even though 12 of them are also present in the system. In keeping with this, the present study shows only the effect of the errors of the linear constants and of the orientation angles on the nonlinear constants (11)-(13).

The experimental data in [5] were based on crystal specimens of only two different orientations. As further details were not readily available, the calculations made in this paper proceeded as if only two physically different specimens were actually used. As a result, only six different values of orientation angles, ψ_1^B, ϕ_1^B, θ_1^B, ψ_2^B, ϕ_2^B, and θ_2^B, have been treated as the possible sources of errors in the nonlinear constants.

As stated earlier, in order to isolate the effect of the errors in the linear constants and, later, *in the orientation angles, all other* quantities substituted into the linear system had to be treated as fixed parameters. This also applied to the values of the third-order elastic constants (13) taken from Thurston, McSkimin and Andreatch [7] and used in the calculation. The fact that the third-order elastic constants themselves also depend on the linear constants and orientation angles of another set of experimental crystal specimens has been disregarded.

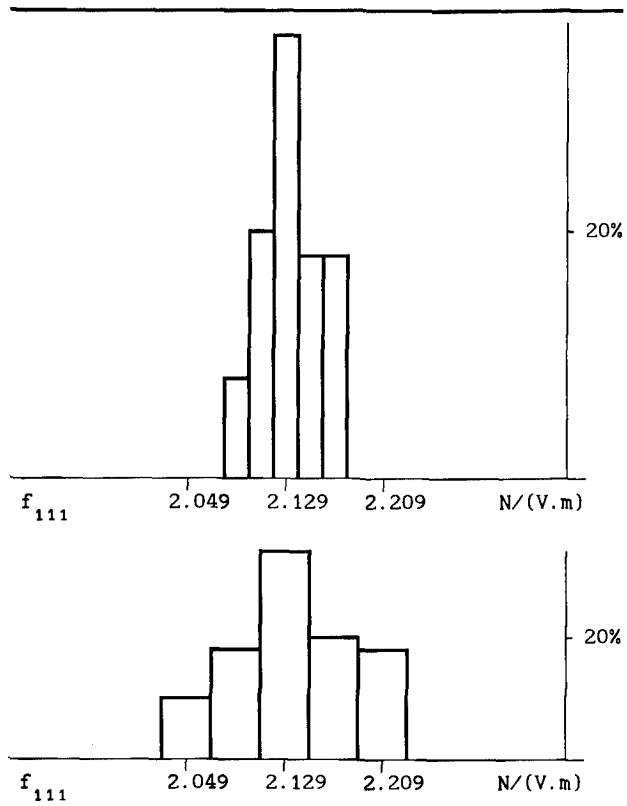


Fig.1. Histogram of the spectrum of values of electroelastic constant f_{111} of quartz obtained by randomization of the linear material constants. The upper and lower histogram correspond to the standard errors of 1% and 2% in the linear constants, respectively.

Under the above circumstances, the following results have been obtained:

*1) If the standard errors of the linear constants are 1%, i.e., if r ≈ 1 in Eqs.(5)-(7), then the characteristic percentage errors (15)-(17) of the electroelastic, electrostrictive and third-order dielectric constants are about 2%, 30%, and 30%, respectively.

*2) For the standard errors of the linear constants up to 4%, i.e., for r = const ≤ 4 in Eqs.(5)-(7), the characteristic precentage errors (15)-(17) are proportional to r.

*3) For the standard errors of equal relative size, i.e., r = const in Eqs.(5)-(7), the standard errors of the elastic constants have about a ten times larger effect on the calculated

nonlinear constants than those of the piezoelectric or dielectric constants. Among the elastic constants alone, the standard errors of c_{13} and c_{14} cause the largest variations in the nonlinear constants.

Tab.I. Comparison of the third-order nonlinear electromechanical constants of quartz calculated from system (1) using different sets of values of the linear constants.

	(a)	(b)	(c)
f_{111}	2.13	2.14	2.15
f_{113}	-0.53	-0.54	-0.53
f_{114}	0.22	0.24	0.24
f_{122}	-1.12	-1.12	-1.12
f_{124}	0.75	0.76	0.76
f_{134}	1.68	1.64	1.64
f_{144}	0.09	0.09	0.09
f_{315}	-0.88	-0.89	-0.89
l_{11}	6.55	4.86	4.36
l_{12}	-10.09	-8.37	-7.98
l_{13}	15.02	14.24	14.13
l_{14}	-6.59	-5.91	-5.79
l_{31}	2.63	2.33	2.26
l_{33}	-3.98	-3.35	-3.29
l_{41}	-0.03	-1.40	-1.51
l_{44}	-1.44	-0.75	-0.67
$\kappa_{111}.10^{20}$	4.37	3.72	3.54

The linear constans used are: (a) by Bechmann [3], (b) by Bechmann with the elastic constants taken from McSkimin, Andreatch and Thurston [8], (c) by James [9]. The electroelastic constants f_{ijk} are in N/(V.m), the electrostrictive constants l_{ij} are dimensionless, the third-order dielectric constant κ_{111} is in F/V.

*4)If the standard errors of the orientations angles are 10', i.e., if $\delta = 10$ in Eq.(8), then the characteristic percentage errors (14-(17) of the electroelastic, electrostrictive and third-order dielectric constants are about 0.02%, 0.5%, and 0.5%, respectively.

*5)For the standard errors of the orientation angles not exceeding 20', i.e., for $\delta = \text{const} \leq 20$ in Eq.(7), the characteristic percentage errors (15)-(17) are proportional to δ.

The above errors can be compared with the characteristic percentage errors of the nonlinear constants associated with the random errors in observations L_i. Their values for the electroelastic, electrostrictive and third-order dielectric constants are about 4%, 60% and 60%, respectively [6].

As folows from Eqs.(15)-(17) the characteristic percentage errors (15)-(17) are obtained using standard errors $\vartheta^R(x_j)$ each of which is in its turn calculated from a spectrum of 100 generally different estimates of x_j obtained by the randomization process. Fig.1. presents an example of such a spectrum summarized in a histogram.

In the context of this paper it is not without interest to show how the solution of system (1) changes if the source [3] of the best linear constants (2) used in this paper is replaced by another set of linear constants. Tab.I. shows estimates x_j^B of the nonlinear constants obtained using the linear constants by Bechmann [3] in column (a). The values in column (b) are obtained when the elastic constants from [3] are replaced by those according to McSkimin, Andreatch and Thurston [8]. The column marked (c) is computed for a complete set of independently obtained linear constants published recently by James [9]. The differences among the columns are typically of the order of 1% for the electroelastic constants and of 10 % or more for the electrostrictive and third-order dielectric constants.

Conclusion

The principal, long-recognized source of errors in the third-order nonlinear constants of quartz are the experimental errors, i.e., the errors in the observations which need to be made in order to provide the fundamentally new data from which the nonlinearities are calculated.

This paper has identified and analyzed two additional sources of errors which are the errors in the linear material constants of quartz and in the crystallographic orientation of the crystal specimens used for the experiments. Of the two, the errors in the linear constants, specifically in the linear elastic constants, appear to be more important.

If the standard errors in the linear constants are about 2%, then their effect on the nonlinear constants seems comparable with the effect of the experimental errors. Smaller standard errors of the linear constants will produce a proportionally smaller effect. But it may still be large enough to make an appreciable contribution to the overall errors in the nonlinear constants.

This work has employed a number of definitions introduced to make the simulation process more manageable and to serve as simple tools of evaluation of the results. They may not be necessarily the best tools for the purpose. If they are found insufficient, more suitable ones will have to be found at a later stage.

This is a pilot study of a problem which has not been dealt with before. It has been made for one particular experiment [5]. Its results apply primarily to this experiment. It is believed that they apply qualitatively also to other experiments and experimental data sets. However, in order to be sure, this needs to be tested.

Acknowledgement

The author wishes to acknowledge the assistance of Peggy Ng who provided a vital component of the computer program speeding up the simulation process. Pauline Hruska helped with the final stages of preparation of this paper.

References

[1] N.R. Draper and H. Smith, Applied Regression Analysis. London: J.Wiley, 1966, ch.2, pp. 59 and 61.

[2] IEEE Standard on Piezoelectricity, IEEE Std 176-1978. New York: IEEE, 1978, ch.3, pp. 18-19 and 25-26. Available from The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017.

[3] R. Bechmann, "Elastic and piezoelectric constants of alpha-quartz," Phys.Rev., vol.110, pp. 1060-1061, 1958.

[4] D.F. Nelson, "Theory of nonlinear electroacoustics of dielectric, piezoelectric, and pyroelectric crystals," J.Acoust.Soc.Am., vol.63, pp.1738-1748, 1978.

[5] E. Kittinger, J. Tichy, W. Friedel, "Nonlinear piezoelectricity and electrostriction of alpha quartz," J.Appl.Phys., vol.60, pp.1465-1471, 1986.

[6] C.K. Hruska, "The contribution of the transit-time and the resonator method towards determination of the nonlinearities in quartz," in the Proceedings of the 5th European Frequency and Time Forum, Communautes Scientifiques de Besancon et de Neuchatel, Besancon, France, March 1991, pp. 137-142. Available from E.N.S.M.M., Universite de Besancon, France.

[7] R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr, "Third-order elastic coefficients of quartz," J.Appl.Phys., vol.37, pp. 267-275, 1966.

[8] H.J. McSkimin, P. Andreatch, Jr, and R.N. Thurston, "Elastic moduli of quartz versus hydrostatic pressure at 25° and -195.8°C," J. Acoust.Soc.Amer., vol.34, pp. 1271-1274, 1962.

[9] B.J. James, "A new measurement of the basic elastic and dielectric constants of quartz," in the Proceedings of the 42nd Annual Frequency Control Symposium, Baltimore, Md, USA, June 1988, pp. 146-154. Available from IEEE, Catalogue No. 88CH2588-2.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

LATTICE DISTORTION AND PARAMETER VARIATION IN QUARTZ AND BERLINITE CRYSTALS; METHODS FOR IMPROVING THE QUALITY.

A. ZARKA*, B.CAPELLE*, Y. ZHENG*, J. SCHWARTZEL**,
J. DETAINT**, X. BUISSON***, Y. TOUDIC****.
* L.M.C.P. 4 PLACE JUSSIEU. 75252. PARIS CEDEX. FRANCE
**C.N.E.T./PAB/ BAG. 132 AV. H. RAVERA. 92220. BAGNEUX. FRANCE.
***S.I.C.N. 4 RUE DU RADAR. 74005. ANNECY.CEDEX. FRANCE
****C.N.E.T./LAB/OCM. ROUTE DE TREGASTEL.2230. LANNION. FRANCE.

ABSTRACT

X-ray transmission topography and plane-wave imaging technique using synchrotron radiation have been employed to reveal lattice defects and to measure minute distortions in plates of synthetic quartz and berlinite crystals. Furthermore, for the berlinite, studies of the optimal conditions of growth have been performed in order to obtain larger crystals. This was done using an extension of different techniques previously applied to quartz crystals.To control the quality of the obtained crystals, X-ray topographic studies were done and typical features of the samples are presented.

INTRODUCTION

Lattice distortions and structural defects in synthetic quartz have been studied by various authors using X-ray diffraction techniques (1-6). In the present study, X-ray transmission topography (7) and plane-wave imaging technique using Synchrotron radiation have been employed to reveal lattice defects and to measure minute distortions in Y-cut plates of various synthetic quartz samples. The aim was to obtain information concerning the cell dimension variations associated with lattice growth defects observed in these samples.

The second part of the paper concerns the berlinite.The berlinite $AlPO_4$ (8) is a material analogous to the quartz. Its structure is obtained by an alternated substitution of the silicon by phosphorus and aluminium atoms. Its piezoelectric properties are more important than those of quartz. Berlinite is also one of the rare materials with crystalline orientations which gives very few fluctuations of the resonance frequencies in relation to the temperature variations. The researches on the growth of this material and the realisation of piezoelectric devices have been initiated at the end of the seventies in different laboratories (C.N.E.T., Bell Laboratories, Allied Signal, G.E.C.). The great majority of the results concerning the devices using berlinite have been obtained in the begining of the eighties at the laboratoire de Bagneux (C.N.E.T., France)(9).The high quality of material obtained permits the construction of devices with very high surtensions.

QUARTZ CRYSTALS

SPECIMEN GROWTH AND SAMPLE PREPARATION

The three quartz crystals under investigation have been choseen from the standard production of various companies. Specimen Q1, Q2, and Q3 were respectively grown at the Toyo company (Japan), University of Pekin (China) and S.I.C.N. (France). From the initial bars, plates oriented parallel to $(01\bar{1}0)$ planes (commonly named Y-cut plates) were cut (Fig.1). After mechanical and chemical polishing, these samples, about 1mm thick, were examined by X-ray transmission spherical wave topography and by plane-wave reflection topography.

X-RAY TRANSMISSION TOPOGRAPHY

X-ray topographs of each of the three samples are presented in Fig. 2 (Q1), Fig.3 (Q2) and Fig.4 (Q3).

0-7803-0476-4/92 $3.00 © 1992 IEEE

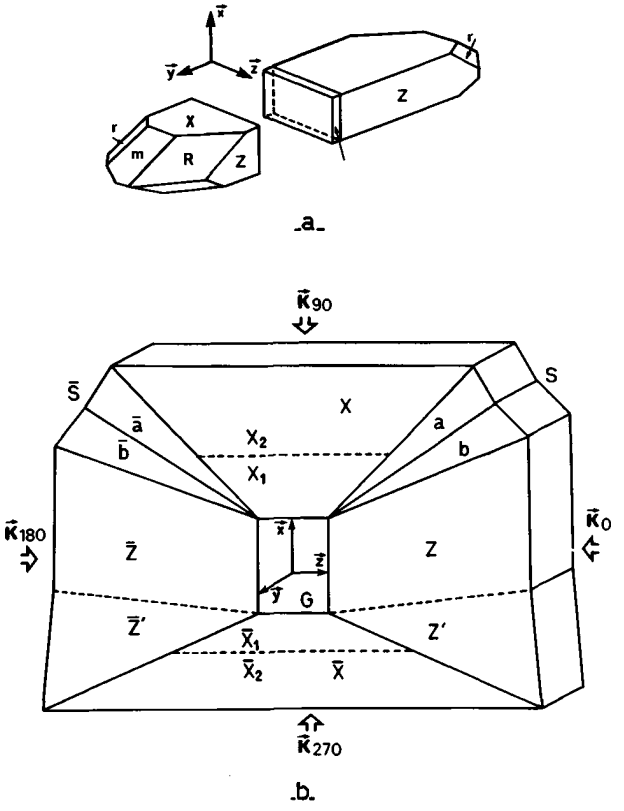


Fig 1 a/ Schematic drawing of a synthetic quartz crystal. The X, Y, Z orthogonal axis used in piezoelectricity The Y-cut plates are cut perpendicular to the Y axis.
b/ Drawing of the various growth sectors or zones observed in the Y plates : X, Z, G (seed);. We have labelled K_0, K_{180}, K_{90} and K_{270} the projections of the diffraction vectors following the four azimuths used for the plane-wave topography method.

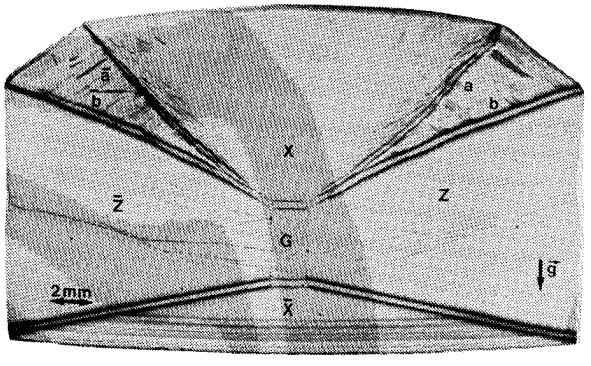


Fig. 2 X-ray transmission topograph of the sample Q1. $10\bar{1}0$ reflection. MoKα_1.The seed G was cut in zone Z and presents very few dislocations in the different growth zones.

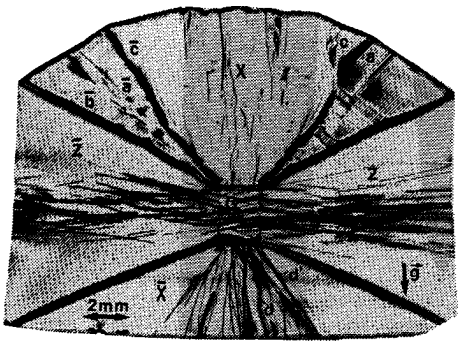


Fig. 3 X-ray transmission topograph of the sample Q2. $10\bar{1}0$ reflection. MoKα_1. Note the important density of the dislocations in zone Z due to the dislocations present in the seed G. In zone X we can note the bundles of the dislocations created on the seed. Two particular dislocations d and d' create rotations of the lattice in this sector X. In c we can remark an unusual sector created at the end of the growth.

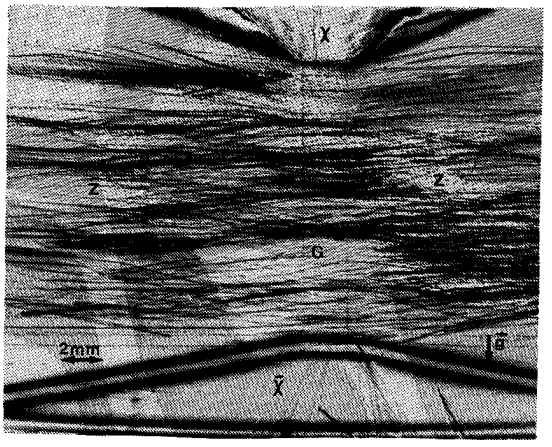


Fig.4 X-ray transmission topograph of the sample Q3 $10\bar{1}0$ reflection. MoKα_1. This central part shows that the seed G contains many dislocations which have been developped in the grown crystal.

Some comparative remarks can be deduced from these topographs:

Crystal growth : These three samples show classical growth sectors (10). However, a more detailed observation shows two Z' and Z' sectors quite different from the Z usual zones.

As in the case of the two growth sectors a and b composing the S zones these two Z' faces are of vicinal type and are initiated on the seed at the corners of the X face. The development of these two "faces" gives on the external final morphology a reentrant angulation more or less accentuated on the two Z sides of the crystal.

Apart from the classical a and b faces composing the S zones, the Q2 sample (Fig. 3) presents a supplementary c vicinal face. Moreover the X sector is composed of a great number of small growth zones corresponding to different vicinal faces which contribute to the very irregular form presented by this X "face".

Dislocations: From the observations of the topographs (Fig. 2,3 and 4) it appears that all the seeds of the studied samples were cut in a Z zone of synthetic quartz. Following the sample, these three seeds contain few (Q1), several (Q2) or many (Q3) dislocations. This initial density of the dislocations is directly related to the final amount of these linear defects in the three samples.

PLANE-WAVE REFLECTION TOPOGRAPHY

Method

Plane-wave reflection topography was used for measuring local variations in lattice spacing and orientation in these Y-cut plates. For the experiments the synchrotron radiation was employed as a source. On the first axis of a double axis spectrometer a multiple reflection monochromator is adjusted so as to release an extended wave of narrow angular and spectral divergence. The principal characteristics of this beam are (11,12): λ= 1.2378 A, $\Delta\lambda/\lambda$ 7.10^{-6}. Intensity = 2.10^7 photons/cm^2/ sec.(1.72 GeV, 100MA) and area 1.5x 1.5 cm^2. On the second axis, the sample, adjusted for the $04\bar{4}0$ reflection is mounted. The Bragg angle is 35°6 and for this case, the calculated half-width of the rocking curve is 1.3 seconds of arc for a perfect crystal.

Four reflection profiles corresponding to four azimuth positions labelled 0, 180, 90, 270 have been recorded for all the three samples.

Then, on different regularly spaced positions on each profile about ten topographs were obtained. An analysis of these series of topographs gives the $\Delta d/d$ values corresponding to the variations of the lattice parameter d (spacing of the $(01\bar{1}0)$ planes) and the $\Delta\Phi$ values corresponding to the rotations between each considered zone and a reference zone (Z sector in our case) by the following relations (13):

$$\Delta\Phi_X = 1/2\ (\Delta\theta_0 - \Delta\theta_{180}) \quad (1a)$$

$$\Delta d/d = -1/2\ \text{cotg}\ \theta_B\ (\Delta\theta_0 + \Delta\theta_{180}) \quad (1b)$$

with : $\Delta\theta_0$: Angular difference between the position on the 0 profile where the considered zone gives the maximum intensity and the position where the reference zone has its intensity maximum.

$\Delta\theta_{180}$: Angular difference between the maximum intensity position for the same zone considered as for $\Delta\theta_0$ but obtained from the 180 profile.

θ_B : Bragg's angle.

We have also calculated the following values:

$$\Delta\Phi_Z = 1/2\ (\Delta\theta_{90} - \Delta\theta_{270}) \quad (2a)$$

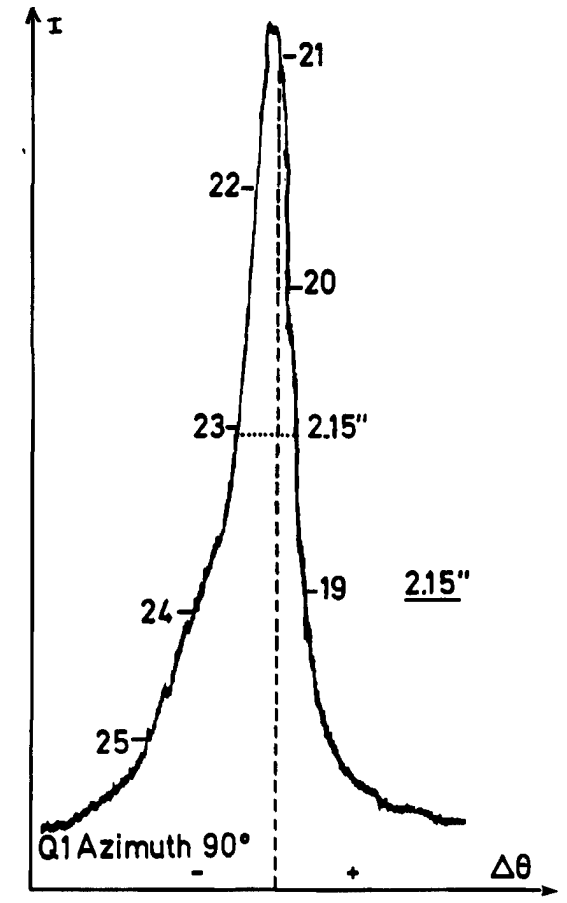
$$\Delta d/d = -1/2\ \text{cotg}\ \theta_B\ (\Delta\theta_{90} + \Delta\theta_{270}) \quad (2b)$$

Where $\Delta\theta_{90}$ and $\Delta\theta_{270}$ are obtained from the 90 and 270 rocking curves and topographs. In fact, the results concerning the $\Delta d/d$ should not depend on the considered azimuth coupled pairs but should allow for a comparison between the values obtained from the expression (1b) and (2b).

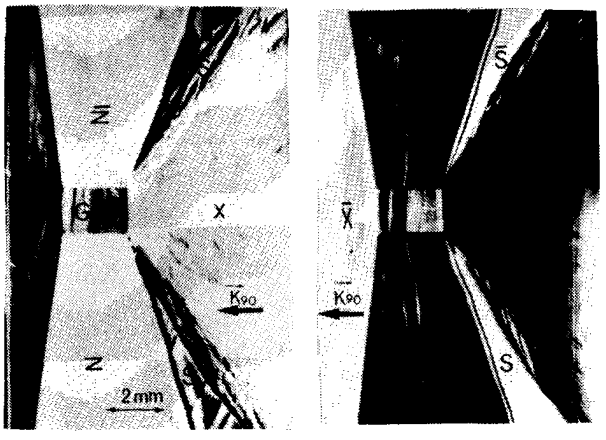
The $\Delta\Phi_X$ and $\Delta\Phi_Z$ notations and values are different because they refere to the calculated rotations around the X and Z axis. The 0-180 and 90-270 diffraction vector projections lie parallel to the Z and X axis respectively as is shown in Fig.1.

Reflection profiles

The profiles obtained for the three samples present different features for their half-width and for their maximum intensity. The experimental half-width is generally about 3" (the theoretical value for the perfect crystal is 1.3"). This is due to the fact that the crystal is not perfect and composed of various growth sectors. So the enlargement and the splitting presented by the profiles depend on the quality of the crystal



a



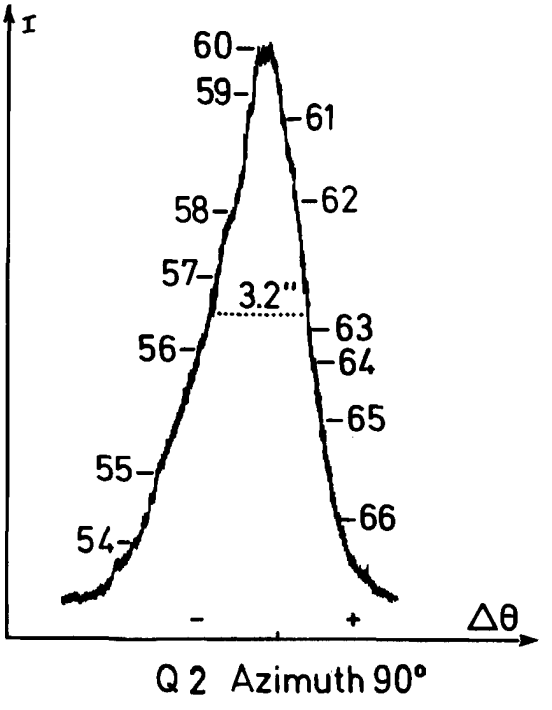
b c

Fig.5 a) Reflection profile of the sample Q1 obtained for the azimuth 90. b) Plane-wave reflection topograph corresponding to the position 24 on the reflection profile. The X and Z zone are out reflection in contrary to the S zones. c) Plane-wave topograph corresponding to the position 21 on the profile. Note that only the X, Z and Z' zones and some part of the seed are in reflection.

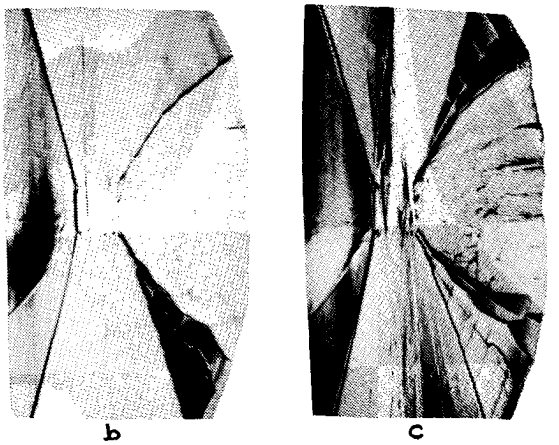
and the different azimuths used.

Plane-wave topographs

In Figs 5-6-7 are represented one profile and two typical topographs obtained for each of the studied samples.



a



b c

Fig.6 a) Reflection profile of the sample Q2 obtained for the azimuth 90. Note that this profile is relatively large (3.2"). b) Plane-wave topograph. Position 55 on the profile. The contrasts are complicated and do not present the symmetry about the X axis. Note the particular local contrast created by the dislocations d and d' in the $\bar{X}$ growth sector. c) Plane-wave topograph. Position 57 on the profile. The contrasts are very complicated. We can note that the $\bar{S}$ zone is out of contrast, in contrary to the symmetrical S sector.

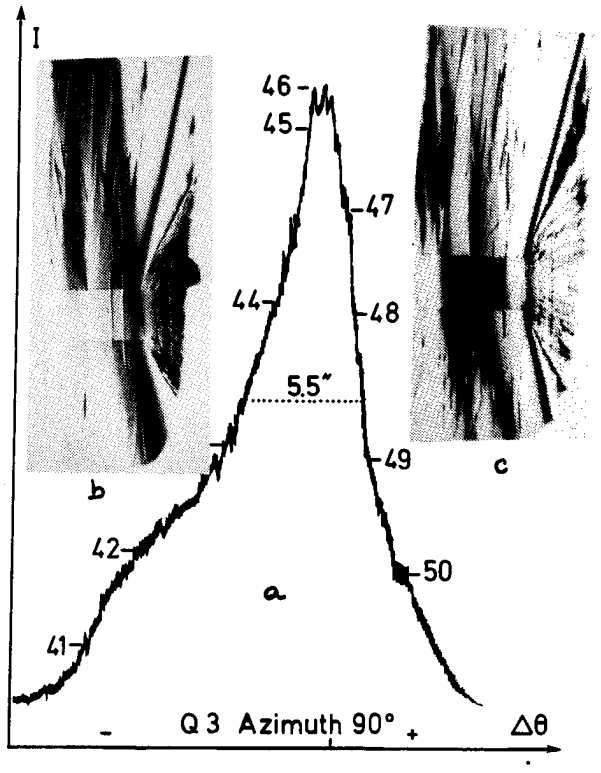


Fig.7 a) Reflection profile of the sample Q3 obtained for the azimuth 90. This broad profile is due to the great inhomogeneity of the crystal. b) Plane-wave topograph. Position 44 on the profile. We can note the great density of the dislocations which create great deformations in the Z sectors.The X sector presents a good contrast while the S zones are out of contrast. Plane-wave topograph. Position 47 on the profile.The deformations due to the dislocations in the Z zones are very complicated.(c).

Q1Sample

In Fig. 5 are shown two topographs of this sample. We can note the symmetry about the X axis of the contrast obtained for two equivalent sectors (Z and $\bar{Z}$ for example) and the quite perfect homogeneity of this contrast in the X, Z and $\bar{Z}$ zones.

These two features will not be verified for the two other Q2 and Q3 samples. The quantitative results relative to this slice are given in the drawing of the Fig. 8 and are in good agreement with those obtained by Yoshimura et al.(6) for similar crystals grown also at the Toyo company.

Q2 Sample (Fig. 6)

This plate shows two interesting features:

- a very clear asymmetry of the contrast with respect to the X axis concerning essentially the Z and $\bar{Z}$ zones.

- the lattice rotations due to the dislocations, particularly those situated between the Z and Z′zones and the bundle of dislocations in the $\bar{X}$ sector. The values obtained for the deformations of the (0110) planes in this plate are given in Fig. 9.

Q3 Sample

A profile and two topographs of this slice are given in Fig.7. The numerous dislocations give important deformations particularly in Z, $\bar{Z}$, Z′, and $\bar{Z}'$zones and are at the origin of the heterogeneous contrast of these topographs. The complexity of the deformations is also revealed by the obtained four profiles which show very different half-width values (respectively 3.2", 2.5", 5.53 and 12" for the 0, 180, 90 and 270 azimuth positions) depending essentially on the lattice rotations created by the high density of the dislocations. The quantitative results relative to this slice are given in Fig. 10.

Lattice parameter variations

The distribution of the lattice parameter of the $(01\bar{1}0)$ planes in the various zones with respect to the reference Z growth sector is quite similar over these three samples (Fig. 8, 9 and 10). Following the zones the $\Delta d/d$ is about 2.10^{-6} to $3.\ 10^{-5}$. However, some general features can be noted.

1) There is a good agreement between the obtained value for the seed G and the Z growth zone (reference). This result agrees with the quite good growth continuity of the samples from the seeds belonging to the Z zone.

2) The S (a, b or c growth sectors), $\bar{S}$, $\bar{X}$ growth zones show important lattice parameter variations with respect to the Z reference zone. These growth sectors correspond to "fast " faces and show important local parameter variations at the level of growth bands due to incorporated impurities.

Orientation variations of the lattice planes

One of the advantages of the technique of the plane-wave topography is related to the four azimuth settings used for this study. The rotations of the lattice planes can be analysed with respect to two perpendicular X and Z axis, and a precise analysis concerning the local deformations can be done. Concerning this point, the three samples show great differences due to localized individual defects (Figs. 8, 9 and 10).

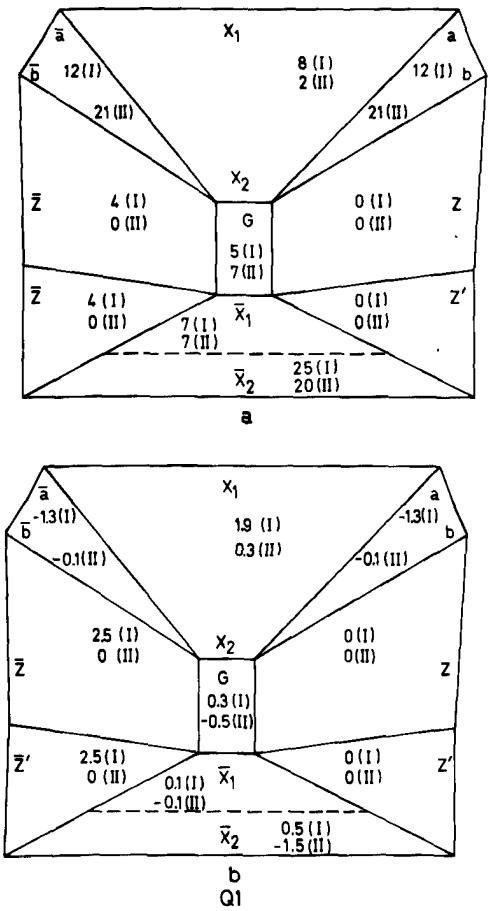
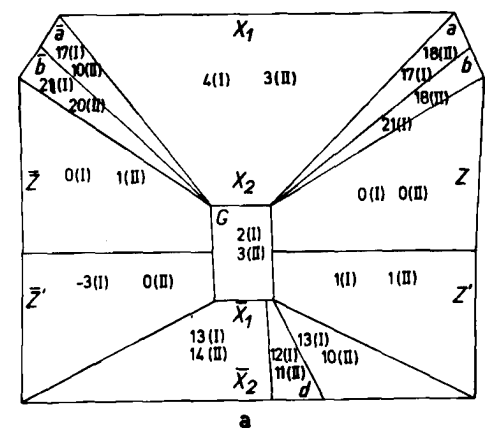


Fig.8. Sample Q1. a) Results of the calculations obtained in the various zones concerning the variations of the parameter $(\Delta d/d)$ (x 10^6) from the plane-wave reflection topographs. The comparisons of the results obtained from the relations (1b) and (2b) respectively noted (I) and (II) are in good agreement. b) Results of the desorientations $\Delta\Phi$ (in second of arc) obtained respectively from the (1a) and (2a) relations and corresponding to the rotations about the X (I) and Z (II) axis.



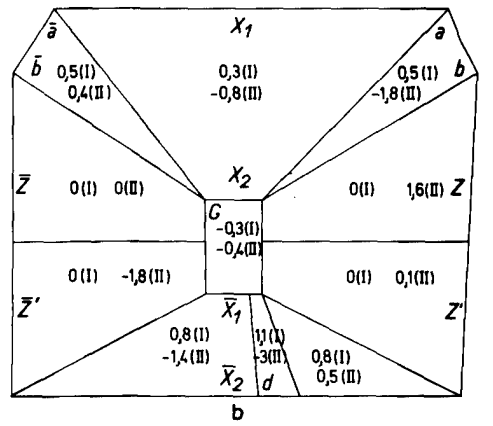


Fig. 9. Sample Q2 a) Results of the calculations obtained for the $(\Delta d/d)$ (x 10^6). b) Results of the rotations $\Delta\Phi$ (second of arc) about the X axis (I) and Z axis (II). We can note important rotations of the lattice with respect to the Z reference sector. The d and d' dislocations (in the X zone) induce rotations about the X axis.

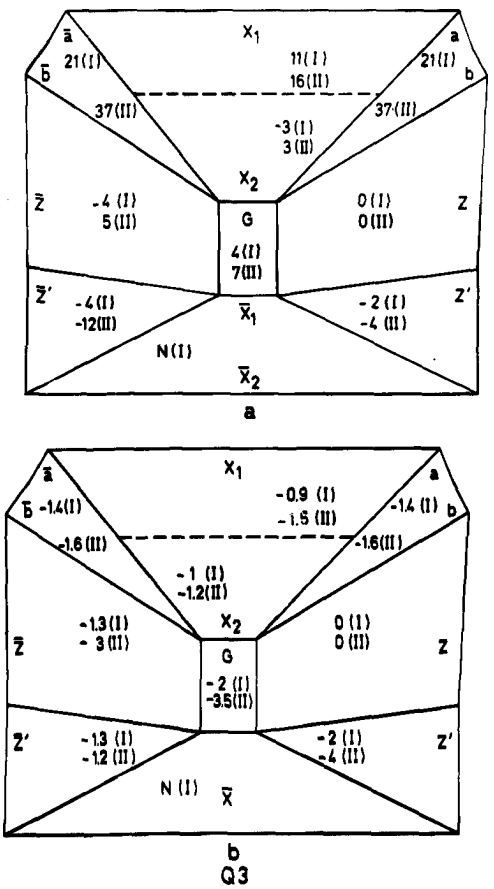


Fig. 10.Sample Q3 a) Results of the calculations obtained for the ($\Delta d/d$) (10^6). b) Calculations of the rotations $\Delta\Phi$ (second of arc) of the various zones about the X axis (I) or the Z axis (II). The reference zone is the Z sector.

Measurement precisions

Measurement couples 0-180 and 90-270 allow the control of the result validity obtained for the parameter variations $\Delta d/d$. Apart from some results the agreement is generally satisfactory. The important sources of errors are due to several factors:

- The precision on the maxima of intensity is variable following the sample and is related to the homogeneity more or less effective in the various growth zones.
- The stability of the setting during the exposure of the film and correlatively the quality of the mounting which can change from an experiment to another one.

All these factors contribute to an imprecision for the obtained results which we have estimated to $5.\ 10^{-6}$ for $\Delta d/d$ and about 0.2" for $\Delta\Phi$ in the X, $\bar{Z}$, Z' and $\bar{Z}'$ sectors. For the sectors where the lattice parameter variatiations are important ($\bar{X}$, S, $\bar{S}$) and which correspond to positions located on the sides of the profiles, the lattice parameter variations were estimated at $8.\ 10^{-6}$ and the rotations at 0.5".

BERLINITE CRYSTALS

PROPERTIES OF THE CRYSTALS

Presently, the most interesting are applications which give piezoelectric features better than those given by the quartz crystals. These characteristics are necessary in order to obtain filters with bandwidth twice greater than those of quartz or oscillators with important frequency bandshift with a high stability of the frequency. Another type of application is the possibility to obtain high frequency resonators by several techniques of chemical etching. Contrary to quartz, the method is very simple and permits a significantly lower price for these devices.

The use of berlinite for acoustical waves devices and also in thermostated oscillators leads to important advantages which results in electromechanical coupling coefficients superior to those of quartz. The expected advantages have permitted the construction of resonators with relative shift between resonance and the antiresonance frequency which are twice those of quartz crystals. Numerous resonators with frequency between some MHz to more than 75 MHz have been made at the C.N.E.T. and other french industries. These resonators have, at the same frequency, lower dimensions than those of quartz. Mathematical models have been done in such a way that their properties do not change when the resonators are made. These models have also demonstrated that different geometries could bring very important advantages (14).

GROWTH OF CRYSTALS

In the past ten years, many efforts were made to develop hydrothermal methods that can produce large berlinite crystals of high perfection needed for piezoelectric applications.

Growth cycles were first performed in H_3PO_4 acid at 240°C (15) followed by systematic investigation of the solubility of berlinite which demonstrated that solubility in sulfuric acid is much higher than in all other known solvants (16). It is well known that hydrothermal growth method is based on thermal variation of the solubility of a material in a solvant to obtain a good single crystal. In case of berlinite, sulfuric acid at high temperature (about 220°C) and under high pressure is used. Under such conditions, the solubility is retrograde (i.e. decreases with the temperature). The technique used in industry comprises vertical autoclaves in which a spatial vertical gradient is established (Fig. 11). In the upper part of the autoclave there is a basket with nutriant materials which are small synthetic crystals. These crystals were synthetized from commercial product in a preliminary growth. In the lower part (separated from the higher part by a diaphragm) single crystals seeds are placed. The autoclave is then filled with a known quantity of sulfuric acid solution of aluminium phosphate. The temperature gradient and the correspondant solubility variation creates convection current.

In the lower part, crystallization takes place on the seeds and the solution concentration decreases. Therefore, the solution moves upwards, dissolves nutriant and becomes more concentrated. Consequently, the solution get down in the crystallization zone where excess of concentration leads to a growth on the seeds. This is more favorable for the nucleation.

Presently the S.I.C.N. Company produces about 7500 slices/year of a good quality crystals which permits the realisation of prototypes or little series of piezoelectric devices.

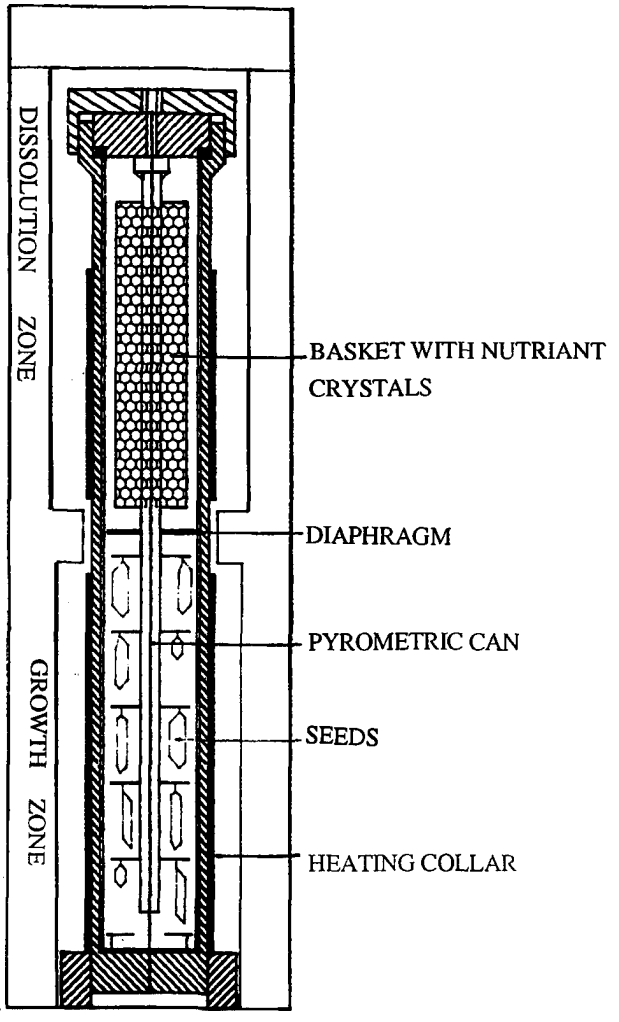


Fig. 11 Technique of growth.

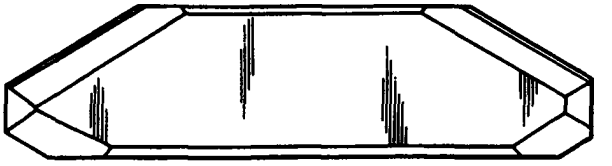


Fig. 12 Morphology of the crystals.

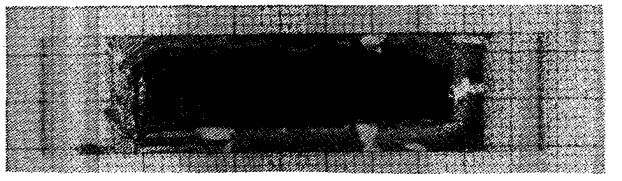


Fig. 13 Mounting used to stick the seed crystals.

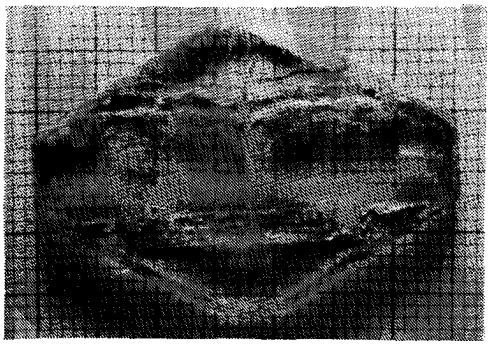


Fig. 14 Obtained crystal from the previous seed.

METHODS FOR GROWING LARGER CRYSTALS

The observation of the crystal morphology (Fig.12), their size and the growth parameters have shown that the growth velocity along the Y direction is very low. Therefore, the studies of the optimal conditions to extend the size of the sample have been performed using two particular techniques:

- The first technique (17) is used to elongate the crystal. Its consists in joining together several plates of berlinite having the same orientation. Its remains always a slight desorientation between the different parts of this seed and the as-grown crystal is not perfect. In figure 13 is presented a photograph of two crystals which were joined together to form a single seed with the mounting which was employed. In figure 14 is shown the obtained crystal and it may be observed that the quality is not perfect. To avoid or to reduce this type of imperfection one can growth the crystal in several steps as it was done for quartz crystals.
- This corresponds to the second technique which is the so-called "crossed-growth" (18). In the first stage, the seed is choosen and cut in a manner that the different desoriented parts or the zones where the dislocations are numerous are avoided. The growth is performed with such seed and the density of the defects(dislocations) are less numerous. If this not the case, new cycle is done in the same manner. This process may be repeated until the final crystal is of a good quality (Fig.15).

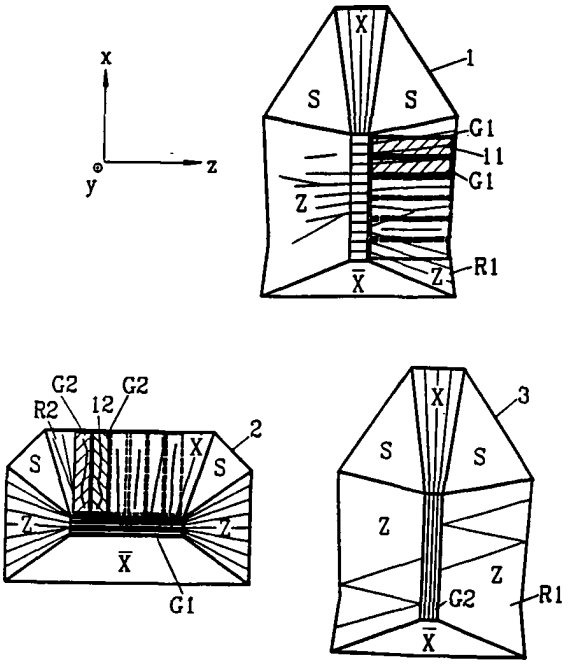


Fig. 15 "Crossed-growth technique.

METHODS OF STUDY

In order to examine the different crystals obtained by described techniques for growth, experimental methods have been employed. Among them, observations at the optical microcope and X-ray topography. This second technique was used with laboratory equipments and with the Synchrotron radiation at L.U.R.E. (Orsay. France). In this last case we have used the Laue setting (19) which is the simplest. It consists of setting the sample perpendicularly to the white incident beam and recording the diffracted beam on a film. Due to the characteristics of the beam, all the diffracted spots are topographs and from their analysis the features of the defects present in the crystal can be determined.

RESULTS

In figure 16 is shown an optical photograph of a Z-cut slice. It may be observed that the quality is enough good. However, it remains an imperfect zone which corresponds to the joining zones of the seed. This may be also observed in the following figure 17 which is a topograph of the same sample. Numerous dislocations which were created at the

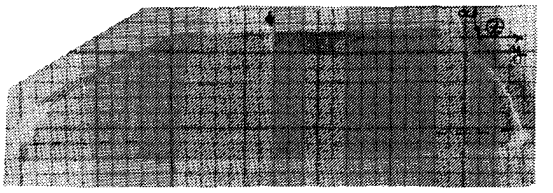


Fig. 16 Optical photograph of a Z-cut slice.

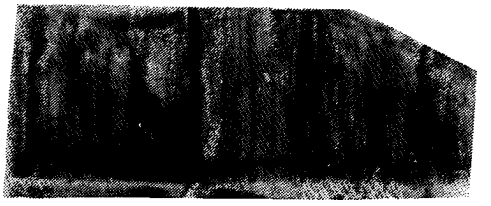


Fig. 17 Topograph of the crystal presented in Fig.16.

level of the seed and above all those which are in the imperfect zones are visible.

Another example is presented in figure 18. This topograph was obtained from a Z-cut slice (X seed). This crystal presents a good quality and it may be observed the different growth zones. Nevertheless, these different parts of the crystal present numerous dislocations which were created at the start of the growth on the seed, or during the development of the crystal.

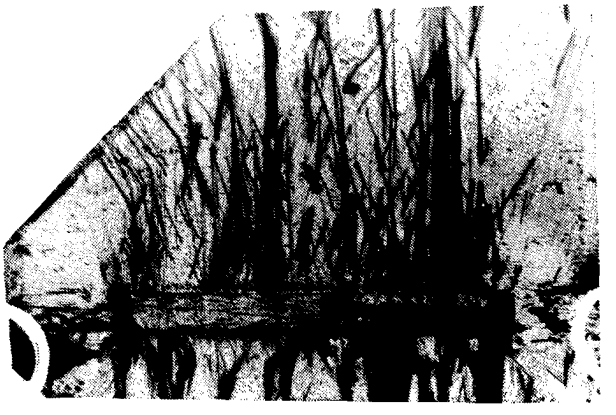


Fig. 18 Topograph of a Z-cut slice.

The topograph presented in figure 19 corresponds to another crystal which was grown in the same conditions as the previous samples. But in this case the different parts of this slice are enough desoriented and this can be observed by the different contrasts obtained. This is due to the misorientations of the lattice introduced at the level of the joining zones.



Fig 19 Topograph of a crystal which presents different desoriented zones.

The last example is given in the figure 20 which represents a topograph of a Z-cut crystal and only the seed and the X zone are visible. As the previous example the main features which are observed are the following. The start of the growth is not absolutly perfect and growth bands parallel to the seed are visible at the begining of the growth. It remains also several desoriented zones which are due to the sticked zones of the seed where bundles of dislocations are located. These defects are less numerous than in other cases but they are still present and mark the different joined zones of the seed. Interference fringes due to progressive variation of the parameters between the seed and the growth zone may be also observed. These fringes are due to wave fields recreation of the X-rays at the boundary between the two slightly desoriented zones. They are visible all along the Y direction. The study of this sample by plane-waves X-ray topography (20) has shown (Fig. 21) that a desorientation of about one minute of arc exists between the lower part of this slice (seed) and the upper part (as-grown crystal).

Therefore, the topograph obtained from these different crystals have shown that quite all the samples have a good

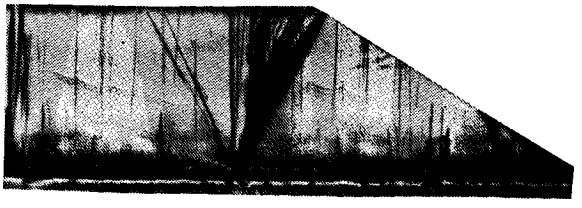


Fig. 20 Topograph of a Z-cut crystal.

quality in spite of the defects and the slight desorientations they present.

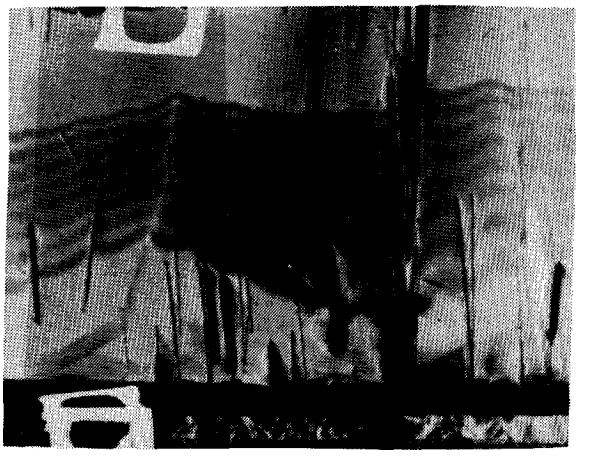


Fig. 21 Plane-wave topograph of the same sample as in

CONCLUSION

The analysis of the X-ray spherical wave topographs of various Y-cut plates of synthetic quartz crystals has shown that the coherence of the lattice between the seed and the grown crystal, following the (0110) planes, is directly related to the density of the dislocations present in the seed. For the three studied samples, the local deformations are essentially related to the density of the dislocations. The plane-wave topography has shown that the local deformations can be estimated with a good precision and related to the growth defects..

Due to their piezoelectric properties, berlinite crystals of large size are needed for industrial purpose. In this work, we have shown that it is possible, by using different techniques of growth to obtain large crystals of good quality. We have used two techniques of growth. The first one starts by joining several crystals of the same orientation to make a single seed. If this is not sufficient, the second technique which is the "crossed growth" can be used. The obtained X-ray topographs from different samples have confirmed the good quality of the crystals.

REFERENCES

(1) W. H. SPENCER & K. HARUTA. J. Appl. Phys. 1966. 37. 549.

(2) A.R. LANG & V.F. MIUSCOV. J. Appl. Phys. 1967. 38. 2477.

(3) Y. EPELBOIN & V.F. PATEL. J. Appl. Phys. 1982 53 271.

(4) S. HOMMA & M. IWATA. J.Cryst. Growth. 1973. 19 125.

(5) J. YOSHIMURA & KOHRA. J.Cryst. Growth. 1976. 33. 311.

(6) J. YOSHIMURA, T. MIYAZAKI, T. WADA, K. KOHRA, M. HOSAKA, T. OGAWA & S. TAKI. J.Cryst. Growth. 1979. 46. 691.

(7) A. R. LANG. Acta Cryst. 1959. 15. 249.

(8) J.H.SHERMAN. Proc. 40th Frequency Control Symposium. 1986. pp. 91-95.

(9) J. DETAINT, M. FELDMAN, J.HENAFF, M. POIGNANT, Y. TOUDIC. Proc. 33th Frequency Control Symposium. 1979. pp. 70-74.

(10) A. ZARKA, LIU LIN & X. BUISSON. J. Cryst. Growth 1981. 54. 394.

(11) M. SAUVAGE & J.F. PETROFF in " Synchrotron Radiation Rechearch " Plenum Press". 1980. Chap. 18.. 607.

(12) J. F. PETROFF, M.SAUVAGE, P. RIGLET and M. HASHIZUME. Phil. Mag.(1980). A42. 319.

(13). S. KIKUTA, K. KOHRA & Y. SUGITA. Jap. J. Appl. Phys. 1966. 5. 1047.

(14).J..DETAINT,A.ZARKA,B.CAPELLE,Y.TOUDIC,J .SCHWARTZEL,E.PHILIPPOT,J.C.JUMAS, A.GOIFFON,J.C.DOUKHAN. Proc. 40th Annual Frequency Control Symposium .1986. pp. 101-104..

(15) E.PHILIPPOT, A.GOIFFON, J.C.JUMAS, C.AVINENS, J.DETAINT, J.SCHWARTZEL, A.ZARKA. Proc. 42st Annual Frequency Control Symposium 1988. pp. 138-145.

(16) B.H.T. CHAI, E. BUEHLER, J.J. FLYM. U.S. Patent. 1982 N° 4. 324773.

(17)J..DETAINT,J.SCHWARTZEL.Y.TOUDIC, E.PHILIPPOT,B.CAPELLE, A.ZARKA, A. GOIFFON, R. ARNAUD. Brevet d'invention. n° 233 072 D. 13606 MD. 24. 05. 1991. France.

(18) A.ZARKA,J..DETAINT,J.SCHWARTZEL, Y.TOUDIC,B.CAPELLE,Y.ZHENG E.PHILIPPOT, X. BUISSONR. ARNAUD. Brevet d'invention. n° 233 191 D. 13719 DL. 14. 11. 1991. France.

(19) T. TUOMI, K. NAUKHARINEN and P. RABE. Phys; Stat. Sol. A. 25. 1974. p. 93.

(20) A. ZARKA, LIU LIN and M. SAUVAGE . J. Cryst. Growth. 62. 1983. pp. 409-423.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PIEZOELECTRIC RESONATORS FROM $La_3Ga_5SiO_{14}$ (LANGASITE) - SINGLE CRYSTALS

V.B.Grouzinenko, V.V.Bezdelkin

Scientific-Research Institute "Fonon"
Moscow, Russia

Abstract

The possibility of application of langasite (LGS) single crystals in different BAW resonator types for TCXO's, VCXO's, wide-band filters and other devices is shown. The tables presenting a number of fundamental physical properties of LGS-crystals and resonators manufactured of them are given. The dependencies of frequency-temperature characteristics, frequency coefficients, motional parameters and resonance spacing vs piezoelectric element orientation and size for different modes of vibration are analyzed. It is shown that LGS-resonators have considerably smaller motional inductance and capacitance ratio values and size compared to these parameters of quartz crystal units. Possible parameter values of LGS-resonators operating in flexure, longitudinal, face-shear, and thickness-shear modes are discussed.

Introduction

Investigations of physical properties of $La_3Ga_5SiO_{14}$ (LGS) single crystals made by a number of authors, showed that these crystals have 5 times lower sound frequencies attenuation values compared to those of quartz and approximately 3 times higher values of electromechanical coupling coefficient.

Resonators with piezoelectric elements fabricated of $La_3Ga_5SiO_{14}$ possess greater Q-value, lower motional inductance L_1 and motional resistance R_1-values, greater resonance spacing compared to analogous parameters of quartz crystal units.

This makes them promising for application in VCXO and TCXO.

In Table 1 the values of several physical properties of LGS- and those of quartz crystals are compared.

The analysis of this table allows to make the conclusion that according to the well-known Mason's classification, LGS-crystals can be considered as belonging to the group of piezoelectric with a moderately strong piezoelectric coupling. The motional inductance of resonators fabricated of LGS-crystals will be 2-5 times lower then that of quartz crystal units analogous in design, size and orientation, owing to greater piezoelectric constants of LGS-crystals.

Comparable in their values of elastic moduli of the shear mode and 2.2 times greater density of LGS-material compared to quartz makes it possible to manufacture langasite resonators operating in face shear and thickness shear modes at frequencies 1.2 - 1.5 times lower than those of quartz, the size of resonators being the same.

Table 2 lists attenuation values of hypersound (at frequency of 900 MHz) for different wave types in the quartz- and LGS-crystals [1-3].

Table 1

Property		Material	
		LGS	Quartz
Hardness (Moos)		6.6	7
Density, g/cm^3		5.8	2.65
Thermal expansion coefficient, $10^{-6}\,°C^{-1}$	α_{11}	5.1	7.1
	α_{33}	3.6	13.25
Phase transition temperature, °C		1470 ($t_{melt.}$)	573
Piezoelectric constants, 10^{-12} K/N	d_{11}	-6.2	2.31
	d_{14}	5.4	0.67
Elastic compliances, ($x10^{-12}$), M^2/N	s_{11}	8.75	12.77
	s_{33}	5.31	9.6
	s_{12}	-4.02	-1.79
	s_{13}	-1.88	-1.22
	s_{44}	21.99	20.04
	s_{66}	25.54	29.12

0-7803-0476-4/92 $3.00 © 1992 IEEE

Table 2

Attenuation value (dB/cm)		Material
Shear mode	Longitudinal mode	
1.9	0.5	LGS
4.0	2.8	Quartz

From Table 2 it follows that with flexure and longitudinal modes one can obtain considerable advantage in quality factor Q when using LGS-crystals.

Experimental

Thickness-shear mode LGS-resonators

The first order frequency-temperature coefficients(FTC) vs crystal orientation are shown in Fig. 1 for LGS-crystal units operating in thickness-shear mode (curve 1). It is evident that for angles of cut -48°, -12.5°, 0° and 45° the first order FTC has a nearly zero value. Curve 2 in Fig. 1 illustrates the second order FTC vs angle of cut, this coefficient takes the zero value for angles of cut 0° and 36° [5].

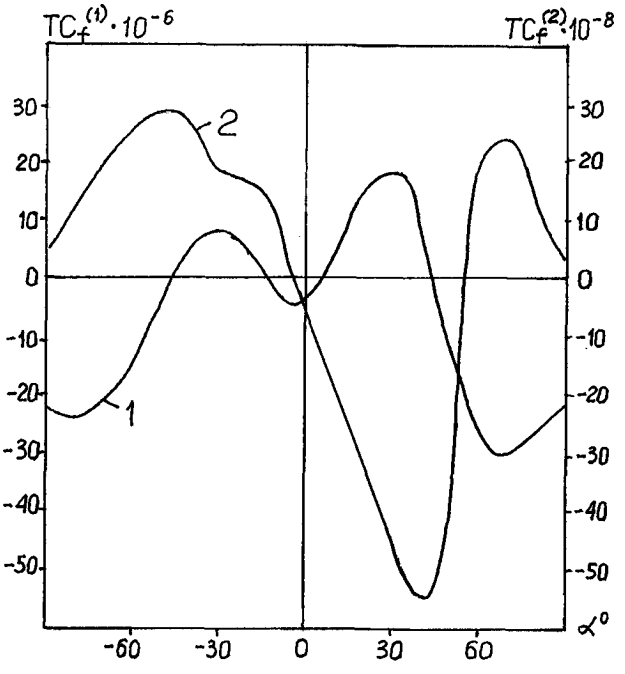


Fig.1: Dependence of the first and the second order frequency-temperature characteristics (curve 1, $TC_f^{(1)}$),(curve 2, $TC_f^{(2)}$) respectively, vs the angle of cut α^0

In accordance with the well-known relation, the electromechanical coupling coefficient

$$k = d'_{26}\sqrt{\frac{c'_{66}}{\varepsilon'_{22}}}$$

depends on the piezoelectric element orientation (Fig.2) and has its maximum value (0.14 ... 0.17) for cut angles ranging from -30° up to +15°. Based on this it is considered to be the most effective to use single-rotated crystal cuts with α-value near to zero and -12°.

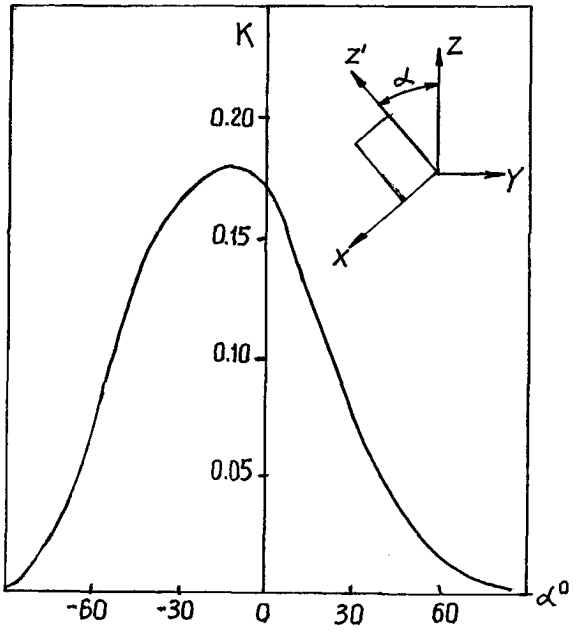


Fig. 2: Dependence of the electromechanical coupling coefficient B_s on the angle of cut °

During the experiments we have studied LGS-crystal resonators operating at the fundamental frequency from 5 to 36 MHz with crystal elements of plane-disk configuration having the thickness value from 46 up to 276 μm. The piezoelectric elements had vacuum-evaporated Ag-electrodes with their diameter depending on the frequency and ranging from 0.7 to 6.2 mm. Piezoelectric elements were mounted in crystal enclosures by using electroconductive compound. LGS-resonators were fabricated in enclosures analogous to those of the EB- and DX-types of IEC 122-3.

Typical frequency-temperature characteristics of LGS-resonators operating in thickness-shear mode are given in Fig. 3. These characteristics have a quadratic parabola shape, the steepness of its branches and the turnover point temperature of which being mainly defined by the angle of cut. The frequency

instability of such LGS-resonators operating over the wide temperature range corresponds to that of BT-cut quartz crystal units.

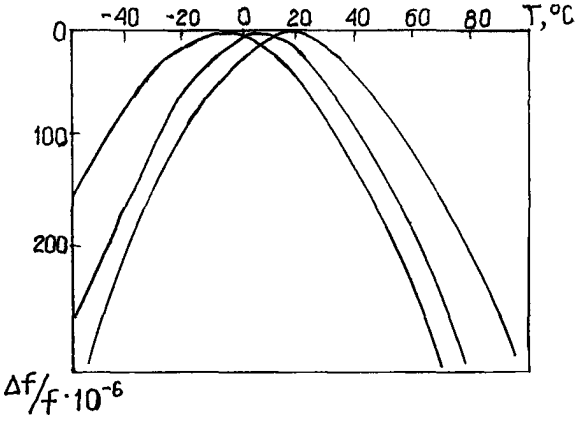


Fig. 3: Typical frequency-temperature characteristics of thickness-shear LGS-resonators

The turnover temperature may be approximately determined from the curves in Fig. 4, characterizing its dependence on the cut angle $\alpha°$ for LGS-resonators operating at the fundamental frequency.

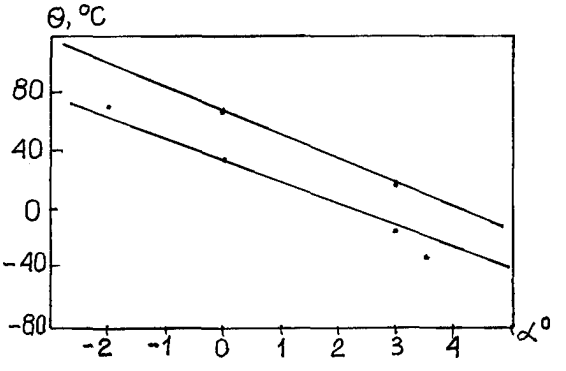


Fig. 4: Dependence of the turnover point temperature Θ on the angle of cut $\alpha°$

The dependences of frequency coefficient and resonance spacing vs angle of cut are given in Fig. 5.

For our experiments we have manufactured 36 MHz LGS-resonators having the thickness of the piezoelectric element active zone ~ 45 mm and the electrode diameter of 0.5 mm, R_1 ranging from 5 to 30 Ohm, the quality factor ~ 10 000, and the resonance spacing 0.5 ... 0.6%.

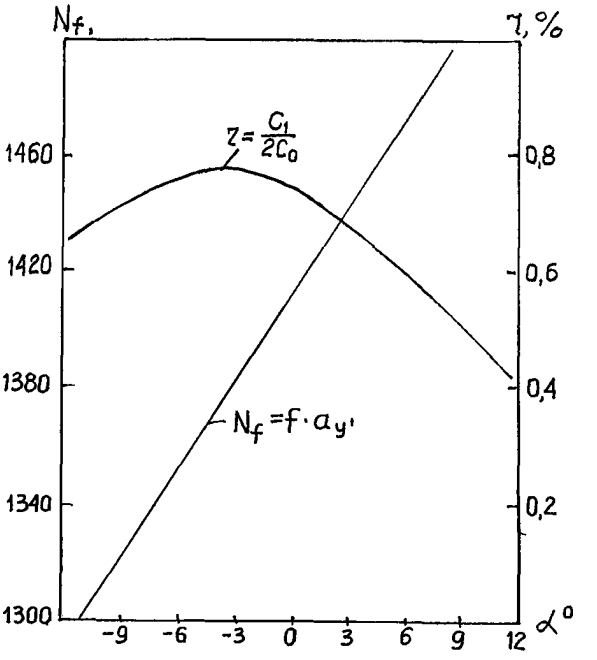


Fig. 5: Dependence of the frequency coefficient N_f and the resonance spacing on the angle of cut $\alpha°$

Longitudinal mode LGS-resonators

Crystal elements for this resonator type were fabricated with their thickness along the X-axis and the length along the Y-axis. The chemical nickel plating method was used for metallization the crystal plate being processed by the 7 μm abrasive; the nickel electrodes had the thickness of 0.5 ... 0.7 μm. Their wire leads (ϕ 0.1 mm) were soldered at nodal points on the plate surfaces. The resonators were fabricated operating over the frequency band of 128 ÷ 520 kHz. The LGS-resonators with zero FTC were obtained when using angles of cut -6° ... +4°, the resonance spacing attained being 0.8% ... 1.1%.

Typical frequency-temperature characteristics are given in Fig. 6. The steepness coefficient of the parabola for LGS-resonators equals to a value over the range (5 ÷ 8)x 10^{-8} x $°C^{-2}$ and slightly changes depending on the cut angle variations. Increasing crystal element thickness results in decreasing the steepness of the frequency-temperature characteristic. Compared to the longitudinal mode quartz crystal units LGS-resonators have approximately 1.2 ... 1.5 times lesser frequency-temperature stabi-

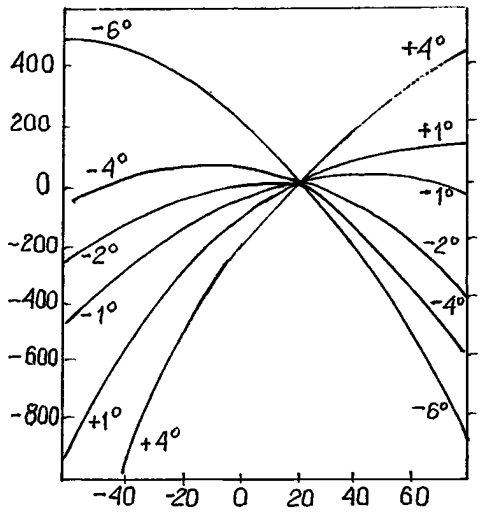


Fig. 6: Typical frequency-temperature characteristic of longitudinal mode LGS-microminiature resonators (Freq. 350 kHz; enclosure volume 0.25 cm^3; $a_{z'}/a_{y'}$ = 0.28; angle of cut xys/$\pm\beta°$; $\beta°$ = -6° ÷ 4°)

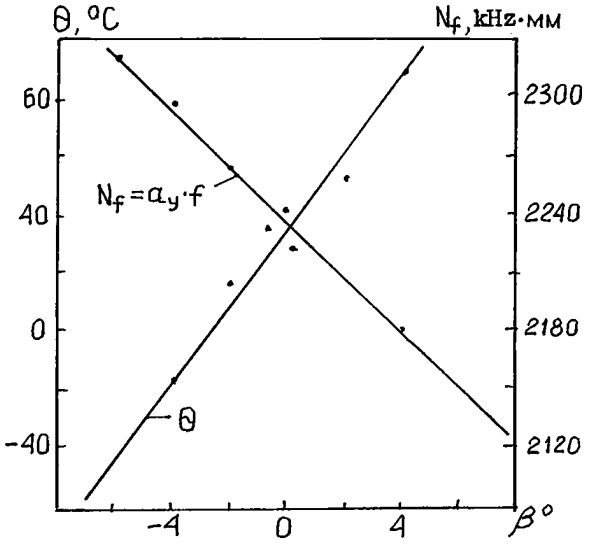


Fig. 7: Dependence of the frequency-temperature characteristic $N_f = a_y \cdot f$ and the turnover temperature θ on the angle of cut $\beta°$ of longitudinal mode LGS-resonators

lity, they have, however, 3 ... 6 times greater resonance spacing, lower motional inductance L_1 and resistance R_1-values and 30% lower value of frequency coefficient, which is extremely important for miniaturization of BAW LGS-resonators operating in LF-range. By increasing the relative crystal element width, the frequency coefficient value, the FTC parabola steepness, the motional inductance and resistance decrease simultaneously the turnover point on FTC-curve increases. With the crystal element width (a_z) to the element length (a_y) ratio values close to 0.24, the LGS-resonator may exhibit anomalous values of motional parameters and constants due to the greater coupling of the fundamental longitudinal mode to the unwanted flexural and shear modes in the "length-width" plane. In the case of LGS-resonators the influence of technological process errors on the turnover point of FTC-curve is lesser than for the quartz units and in general the parameters of the LGS-resonators have better reproducibility.

In Fig. 7 dependence of the inflexion point of the FTC and of frequency coefficient on the angle of cut are represented for LGS-resonators operating in longitudinal mode.

Analogous experimental investigation was made for LGS-resonators operating in flexural and face-shear modes. The lapping of crystal elements, their plating and soldering of wire leads were effected by using the same methods as those used for longitudinal mode LGS-resonators. In Table 3 averaged results of our studying fundamental parameters of laboratory sample LGS-resonators are given. In Table 4 possible frequency ranges and expected values of fundamental LGS-resonator parameters are represented.

The data of Table 4 point at the evident propabilities for the producing LGS-crystal units possessing the frequency-temperature characteristics close to those of the quartz crystal units of BT-, CT-, X-cuts etc.

The experimental results have shown that the force-sensitivity of LGS-resonators operating in thickness-shear mode has an order of magnitude lesser value compared to that of AT-cut quartz crystal units, the zero force-sensitivity being observed in the direction close to that of the crystallographic X-axis [4].

Low values of the force-sensitivity enable to design LGS-resonators withstanding severe mechanical and temperature effects and with decreased hysteresis of FTC's [6,7].

Table 3

Experimental results

Mode of vibration	Frequency, kHz	Quality factor, Q	Resonance spacing, %
Thickness-shear fundamental	13500 20000 36000 5000 5350	11500 27000 7000 4800 11000	0.45 0.63 0.6 0.72 0.43
Longitudinal	130 250 260 224 406 520	46000 23000 17000 20000 22000 23000	0.8 0.9 0.83 0.85 1.00 0.6
Face shear $a_z/a_x = 1$ $a_z/a_x = 0.39$ $a_z/a_x = 0.23$ $a_z/a_x = 0.167$	 200 400 500 650	 35000 25000 15000 10000	 0.25 0.27 0.22 0.25
Flexural (in the z'y' - plane) $a_{z'}/a_{y'} = 0.4$	150	18000	0.35

Table 4

Expected parameters of LGS-resonators

Parameter	Mode of vibration			
	Flexural	Longitudinal	Face shear	Thickness shear
Frequency range, kHz	10 ÷ 1000	250 ÷ 1000	400 ÷ 2000	1000 ÷ 150000
Resonance spacing, %	0.01 ÷ 0.4	0.5 ÷ 0.9	0.2 ÷ 0.4	0.7 ÷ 1.1
Motional inductance, H	5 ÷ 100	0.3 ÷ 6	0.1 ÷ 10	$(0.01 \div 6)\times10^{-3}$
Steepness coefficient of FTC, $10^{-8}\,°C^{-2}$	4,5 ÷ 10	5 ÷ 8	4.5 ÷ 8	4 ÷ 6
Motional resistance, Ohm	100 ÷ 2000	25 ÷ 500	5 ÷ 200	1 ÷ 150

Investigations of LGS-resonators operating in thickness shear and longitudinal modes revealed their higher compared to quartz crystal units stability against drive level variations. The 50 times increase of the drive level (from 0.05 up to 2.5 mWt) resulted in only insignificant changes of the nominal frequency, motional resistance and spectral characteristics. It should be mentioned, however, that these experiments were carried out by using microminiature resonators and it would be incorrect to apply the above results to precision resonators.

For carrying out more detailed experiments aimed at the improvement of parameters and designing LGS-resonators for high stability oscillator applications as a result, it is necessary to realise the fabrication of high-precision LGS-resonators taking into account the present state-of-the-art.

Conclusion

The experiments made evidence of perspectives of using LGS single crystals for manufacturing piezoelectric BAW resonators operating over a wide frequency range intended for the application in instruments where VCXO's and wide band filters, especially low-frequency ones, are used.

As far as the fundamental parameters LGS-resonators are concerned, they correspond to those of certain type of quartz crystal units with crystal elements of X-, CT-, BT- and other cuts. Besides that LGS-resonators are superior to quartz crystal units in their parameters such as a wide resonance spacing, higher environmental effects, lower motional resistance and inductance, etc. It should be also mentioned that technological processes of manufacturing LGS-resonators are superior to those of quartz crystal units bearing in mind their correspondance to the ecological standards.

References

[1] Kaminsky A.A., Silvestrova I.M. et al. Investigation of Technology of $(La_{1-x}Nd_x)_8Ga_5SiO_{14}$ crystals. Spectral Laser and Electron Chemical Properties. Phys. Stat. Sol., 1983, v.80, p.607-620

[2] Kaminsky A.A., Mil B.V. et al. "Acoustical and electrical properties of new nonorganic material $La_3Ga_5SiO_{14}$. The Theses of Reports at the X All-Union Acoustic Conference, Moscow, 1983, p.20-23.

[3] Silvestrova I.M., Pissarevsky Y.V. et al. "Temperature dependencies of elastic properties of $La_3Ga_5SiO_{14}$ single crystal, Ph.T.T., 1986, No.9, p.2875-2877.

[4] Roudenkov A.P., Jakounin Y.P. et al. "Force sensitivity of lithium tantalate and gallium silicate of lanthanum", Electronic Engineering, issue 3(76), 1989.

[5] Bezdelkin V.V., Grouzinenko V.B. et al. "Piezoelectric vibrator operating in shear modes", Pat. U.S.S.R. No.4886192 of the 26th November 1991.

[6] Hammond D.L., Adams C.A. et al. "Hysteresis effects in quartz resonators", in Proc. 22nd Annu. Symp. Freq. Contr., 1968, pp.55-66.

[7] Kusters J.A. and Vig J.R. "Hysteresis in quartz resonators - A review", IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, v.38, No.3, May 1991.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

APPLICATION OF LANGASITE CRYSTALS IN MONOLITHIC FILTERS OPERATING IN SHEAR MODES

S.A.Ssakharov, I.M.Larionov, A.V.Medvedev

Scientific-Research Institute "Fonon"
Moscow, Russia

Abstract

The basic information on piezoelectric material $La_3Ga_5SiO_{14}$ called "langasite" is given. The techniques of growing up langasite crystals and the influence of different methods of charge preparing on piezoelectric properties of the grown material are described. The difference in properties of crystals grown along crystallophysic Z- and Y-axes is explained. The data on anisotropy of elastic wave velocities, on piezoelectric constants and their temperature dependencies vs orientation with respect to crystallophysic axes are presented for piezoelectric substrates operating in thickness shear mode. It is shown that experimental results presented coincide well with the theoretical data to the permissible degree of accuracy. In the course of investigation monolithic filters have been designed operating in shear mode over the frequency range from 5 MHz to 18.5 MHz. Filters operating at 5 MHz, 10.7 MHz and 18.5 MHz have been manufactured as a result. The frequency-temperature characteristics and ageing of filters operating over the temperature range from -60°C to +85°C were investigated. The comparison of monolithic quartz crystal- and langasite crystal filters was made. In conclusion certain technological features of manufacturing langasite elements and those of monolithic filters using them are given.

Introduction

In the Proceedings of the 45th ASFC our paper was published which described investigations of "strong" piezoelectric materials with a particular attention to their application for monolithic filters and resonators operating in thickness shear mode. The present paper is a continuation of our investigation and deals with one of those materials, namely with langasite (LGS) crystal, which has found its application in mid-band monolithic filters and resonators for VCXO's.

The first published mention of LGS-crystals is dated 1982 [1], where a feature of these crystals modified by Nd was discovered to generate stimulated laser radiation. Suppositions were made already then concerning its other possibilities for a practical application. The first investigation results of electromechanical and piezoelectric properties of LGS-crystals had been published in 1983 [2,3]. Subsequent research works were aimed at more thorough investigation of piezoelectric properties of these crystals.

Growing up LGS-crystals

LGS-crystal synthesis is made from oxides via forming an intermediate $La_3Ga_5SiO_{14}$ phase (perovskite structure), which terminates at temperatures ranging from 1400°C up to 1450°C. The run time is 3-4 h. In accordance with the thermal analysis results obtained, LGS-crystal melts congruently at T_{melt} = 1470°C and exhibits no phase transitions up to the room temperature. This makes it possible (as distinct from quartz- and berlinite crystals) to grow up single crystals of this composition $La_3Ga_5SiO_{14}$ from the melt by using (which is extremely important) the Tchokhralsky method for the growing.

At present LGS-crystals are mass-produced by two growers, using different methods of charge synthesis: the thermal one [4] and the method of selfdistributing high temperature synthesis. The thermal analysis consists of the following: the initial materials for the growth process are prepared by heating mixtures of pertinent high purity oxides at temperatures 1150°C - 1200°C during 4-5 h. A multiphase mixture consisting of $La_3Ga_5SiO_{14}$ + $LaGaO_3$ + Ga_2O_3 + SiO_2 is obtained as a result.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The complete LGS-crystal synthesis occurs during preliminary mixture melting in the growth container (crucible). At this synthesis stage special attention is paid to the control of Ga_2O_3 volatile components. After charge melting and its complete synthesis, the process of growing up crystals is continued.

The condition of successful self-distributing high-temperature synthesis (SHS) is self-deliberate attaining during the synthesis process of the temperature insignificantly exceeding the melting temperature of the material in question. The temperature obtained is gouverned by the relationship of free metal giving off heat during oxidation and the quantity of oxygen-inert substance, which is,however, required for substance synthesis of specified composition. The modern theoretical physical chemistry methods enable to calculate the temperature resulting from chemical reaction up to saturation under adiabatic conditions by using the laws of thermodynamics. The calculation of adiabatic temperature is made under condition of maintaining the complete enthalpy of the system under constant pressure of 1 atu.

The value of complete system enthalpy depends on the quantity of metal gallium in the initial charge. The synthesis process itself consists of the following: metal gallium in a liquid state is mixed with oxides, the mixing temperature being maintained at the expense of self-heating the mixture. High mixing intensity (frequency 50 Hz, amplitude 3-4 mm) enabled obtaining homogenous mixtures with a developed metal gallium surface, which is an obligatory condition for mixture oxidation process during burning regimes.

The mixture oxidation is effected in a quartz reactor of flowing type, in which after local mixture heating by means of a resistive heater in oxide containing medium, the oxidation (burning) reaction is distrubuted within the total mixture volume. The burning temperature is controlled by a pirometer and a Pt - Rd thermocouple.

The crystals are grown up after the complete charge synthesis.

LGS-single crystals are grown up by the Tchokhralsky method in the growth crucibles with induction heating equipped with an instrument for diameter control of the crystal being grown by its weighting results.

In Fig. 1 a heating assembly for LGS-crystal growth is shown.

For the growth process platinum crucibles are used. Preliminary synthesis occurs in the air atmosphere (oxygen containing medium) and crystals grow in the gas flow mixture, containing 3% of O_2 and 97% of N_2. During the growth process in a pure nitrogen atmosphere a noticeable evaporation of gallium oxide is observed.

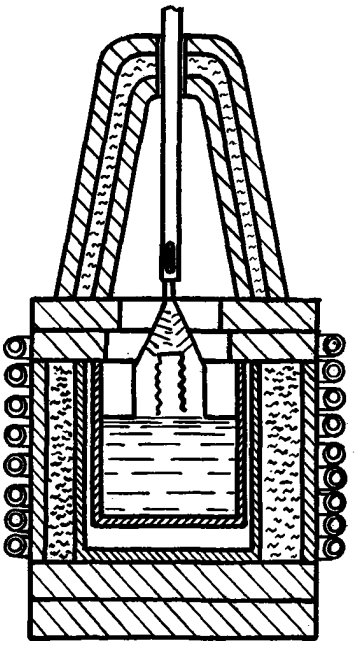


Fig. 1: Heating assembly for LGS-crystals growing

With oxigen content increase, platinum appears in the melt. The growth is effected by using seeds oriented along the Z-axis. The grown up crystals have hexagonal cross-section with rounded angles. The perpendicular to the hexagon face coincides with the direction of the X-axis.

Grown up crystals have up to 60 mm in diameter and weight up to 3 kg. They are shown in Fig. 2.

At present investigations are made to grow up LGS-crystal along the Y-axis in order to increase the yield while crystal cutting and to use a group type technology for manufacturing langasite piezoelectric elements. Table 1 gives principal piezoelectric characteristics of LGS-crystals grown up from the charge synthesized by both the above methods.

LGS-crystal structure

LGS-crystal structure belongs to the structure type of $Ca_3Ga_5GeO_{14}$ and has a spatial group P321.

The LGS cell parameters in accordance with the data given in [5] are:

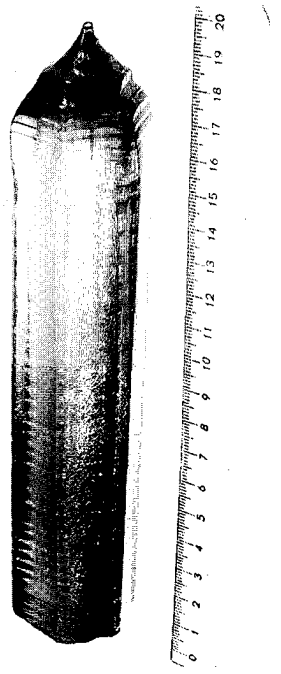


Fig. 2: "As grown" langasite crystal

During the investigation a piezoelectric tester has been developed in order to determine the sign of piezoelectrically active X-axis, which eliminates the possibility of incorrect crystal orientation in the case of a left-handed modification.

Piezoelectric, dielectric and elastic properties of LGS-crystals

Electromechanical properties of LGS-crystals has been investigated by using complex methods, basically the resonance ones [7] . By applying HF electric field to samples of various orientation different modes of vibration were excited, every one of them corresponding to definite piezoelectric moduli and constants.

Piezoelectric moduli d_{11} and d_{14} have been obtained by investigating compression-elongation vibrations along X-cut bars (the length of the bars made angles 0°, -30°, +30° and +45° with respect to Y-axis in the ZY-plane) and Y-cut bars (the length made +45° with respect to X-axis in XZ-plane). Piezoelectric coefficients d_{11} and d_{14} were determined from the investigation of compression-elongation vibrations along the thickness of X-cut plates.

Table 1

Comparison of LGS-crystal characteristics grown up by different methods

Synthesis technique	k_s, %	N kHz mm	FTC $10^{-6}{}^{\circ}C^{-1}$	Q	$\rho \cdot 10^{-15}(\frac{Ohm}{cm})$		$v \cdot 10^{-3}(\frac{m}{s})$	
					11	33	L	S
SHS	15	1380	1.6	15000	1.21	0.81	5.78	2.34
Thermal analysis	13	1405	2.0	15000	1.25	0.83	5.70	2.30

a = 8.162 Å, c = 5.087 Å and in accordance with the data given in [6] they are: a = 8.1662 Å, c = 5.074 Å, the crystal density measured by X-ray method equals to 5.751 g/cm^3. In conformity with the symmetry class, LGS-crystal can have two enantiomorphous modifications: the left-handed and the right-handed ones. (Crystals investigated in the process of this work belonged to the right-handed modification). The enantiomorphism sign is determined by means of a polarizing microscope based on changes of orthoscope pattern while turning polarizer and by using the Eary's figures.

Dielectric constants have been obtained from the measurement of X- and Y-cut plates capacitance at 1 kHz by using an automatic capacitance bridge; the electric field intensity being 100 - 200 V/cm^2.

In Table 2 the measuring results of LGS-crystal constants and their first order temperature coefficients are listed.

Acoustic and piezoelectric properties of LGS-crystals

The LGS-crystal properties described herein are considered in the standrad

Table 2

Crystallographic constants of LGS-crystal

	Constant value	First order temperature coefficient
c^E_{11}	19.02 10^{10} Pa	-47 10^{-6} °C^{-1}
c^E_{12}	10.63 10^{10} Pa	-100 10^{-6} °C^{-1}
c^E_{13}	9.19 10^{10} Pa	-130 10^{-6} °C^{-1}
c^E_{14}	1.47 10^{10} Pa	-370 10^{-6} °C^{-1}
c^E_{44}	5.34 10^{10} Pa	-30 10^{-6} °C^{-1}
c^E_{66}	4.235 10^{10} Pa	36 10^{-6} °C^{-1}
e_{14}	-0.45 C·m^{-2}	-
e_{22}	0.077 C·m^{-2}	-
ε^S_{11}	1.68 10^{-10} F·m^{-1}	153 10^{-6} °C^{-1}
ε^S_{33}	4.36 10^{-10} F·m^{-1}	-760 10^{-6} °C^{-1}
α_{11}	-	5.15 10^{-6} °C^{-1}
α_{33}	-	3.65 10^{-6} °C^{-1}

crystallographic axial system. The positive direction of the Y-axis goes out of the minor rhombohedral face. The positive angle is counted counterclockwise in accordance with Fig. 3. All LGS-crystal properties were calculated for the angle range from -90° to +90°.

In Fig. 4 cross-sectional views by means of a representative surface cut by YZ plane are shown for the piezoelectric coefficient e_{26}.

Phase velocities and frequency constant

It is well known that the solution of Green-Christoffel equation for rotated Y-cuts of trigonal system class 32, gives only one piezoelectrically active mode propagating along the X-axis with the velocity v_1 determined by the relation:

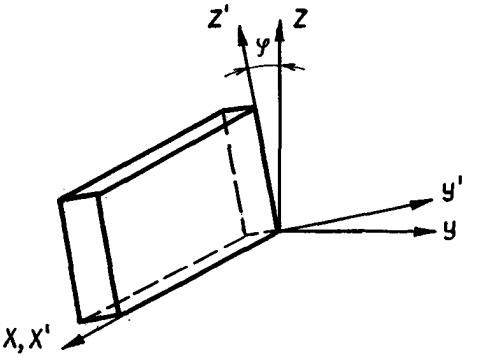


Fig. 3: Standard LGS-crystal plate orientation

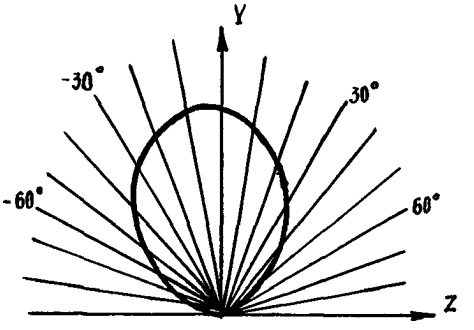


Fig. 4: Cross-sectional views of piezoelectric coefficient e_{26}

$$v_1 = \sqrt{\frac{I}{\rho_s}\left(c'_{66} + \frac{e'^2_{26}}{\varepsilon'_{22}}\right)}$$

where: $c'_{66} = c_{66}\cos^2(\varphi) + c_{44}\sin^2(\varphi) + c_{14}\sin(2\varphi)$

$$e'_{26} = -e_{11}\cos^2(\varphi) - e_{14}\sin(\varphi)\cos(\varphi)$$

$$\varepsilon'_{22} = \varepsilon_{11}\cos^2(\varphi) + \varepsilon_{33}\sin^2(\varphi)$$

and two active modes in the YZ-plane.

In Fig. 5 cross-section of the velocity surface by the YZ-plane is shown, the velocity of piezoelectrically active mode being calculated with taking and without taking into account piezoelectric

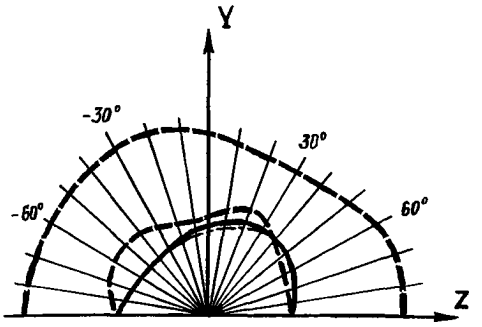


Fig. 5: Cross-section of the velocity surface by the YZ-plane

effect. From the figure it can be seen that piezoelectrically active mode is a slow transverse mode within the angle range from -90° to +37° and it becomes a rapid one within the range from 37° to 90°.

One of the constants, characterizing the thickness shear mode of an infinite plate is the frequency constant defined by the expression:

$$N = t \cdot f$$

where t - plate thickness (mm)
f - plate frequency (kHz)

In the upper part of the Fig. 6 the dependency of frequency constant vs angle of rotation φ is shown.

In accordance with the above expression the frequency constant of zero Y-cut should be equal to 1376 kHz mm.

We investigated frequency constants along the Z-axis of the grown crystals.

For this purpose from the upper, middle and the lower parts of a grown boule rectangular Y-cut plates with the accuracy up to 5' have been cut. After mechanical lapping and polishing up to the thickness equal to 0.5 mm the frequency and plate thickness were measured. With the accuracy of frequency constant measurement of ± 3 kHz mm, the spread in values along the crystal growth axis did not exceed ± 5 kHz mm and it did not exceed ± 10 kHz mm from crystal to crystal.

We have investigated 20 boules. The averaged frequency constant was equal to 1380 kHz mm.

It is necessary to note, that calculation of frequency constant without taking into account piezoelectric effect gives N=1359 kHz mm which differs considerably from the experimental value obtained.

Based on this we have made a conclusion that piezoelectric effect should not be neglected even while making approximate calculations.

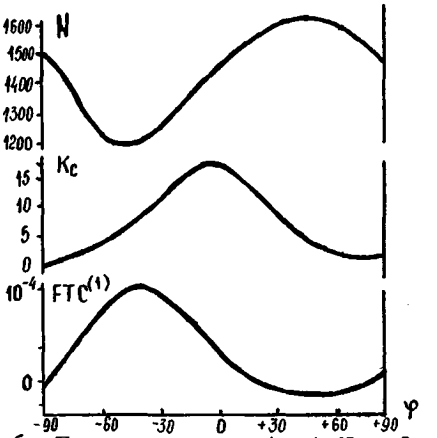


Fig. 6: Frequency constant N, electromechanical coupling coefficient k_s and the first order frequency temperature coefficient $FTC^{(1)}$ vs the rotation angle

Electromechanical coupling coefficient

One of the most important crystal constants defining its application field is electromechanical coupling coefficient k_s.

For thickness shear vibrations the electromechanical coupling coefficient is described by the expression:

$$k_c^2 = \frac{e_{ij}^2}{c_{jj}^D \, \varepsilon_{ii}^S}$$

where

$$c_{jj}^D = c_{jj}^E + \frac{e_{ij}^2}{\varepsilon_{ii}^S}$$

As it has been shown earlier, the neglection of piezoelectric effect can cause incorrect results, therefore, we cannot ignore its influence.

After simple transformations, we obtain the following expression for the electromechanical coupling coefficient:

$$k_c^2 = \frac{e_{ij}^2}{c_{jj}^E \, \varepsilon_{ii}^S + e_{ij}^2}$$

or for the rotated Y-cuts we obtain the following:

$$k_c^2 = \frac{e_{26}'^2}{c_{66}' \varepsilon_{22}' + e_{26}'^2}$$

The calculation results are given in the middle part of Fig. 6.

The analysis shows, that electromechanical coupling coefficient obtains its maximum value (17%) at the angle $\varphi = -10°$ and tends to zero at $\varphi = \pm 90°$.

Frequency vs temperature characteristic

In Fig. 7 temperature dependencies of elastic, dielectric constants and of thermal linear expansion coefficient of LGS-crystal are given.

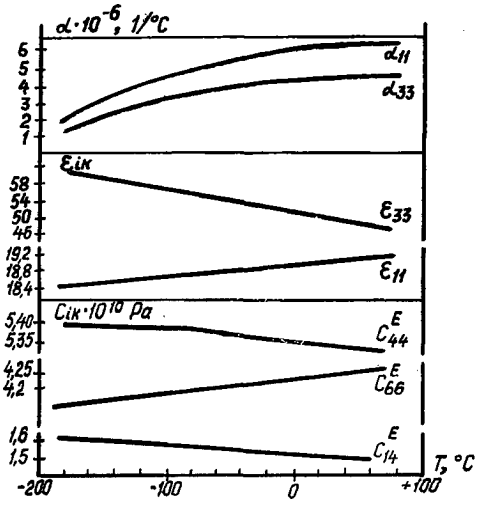


Fig. 7: Dependence of elastic, dielectric constants and of linear thermal expansion coefficient of LGS-crystal vs temperature

From Fig. 7 it follows that the elastic moduli and the dielectric permeability coefficients have different signs of the first order temperature coefficients, which supposes the possibility of temperature-compensated cuts existence.

The existence of the first order frequency vs temperature coefficient $T_f^{(1)}$ equal to zero at the room temperature opens wide application possibilities for this crystal in various types of equipment operating over a wide temperature range.

By using equation given in [8] we have calculated the first order temperature coefficient for langasite crystal.

The calculation results are given in the bottom part of Fig. 6. From Fig. 6 analysis it follows that $FTC^{(1)}$ equals to zero when $\varphi = -84°$ and $\varphi = +6°$.

We have investigated different cuts of the type yxl/$\pm\varphi$ over the angle range φ from $-14°$ to $+14°$.

In Fig. 8 the frequency-temperature dependence is given for the angle φ range from $-4°$ to $+4°$ over a wide temperature range.

The analysis of the data received shows that the most optimum angle of cut for application over a wide temperature range from $-60°$ to $+85°C$ is considered the angle $\varphi = +1°30'$ [9].

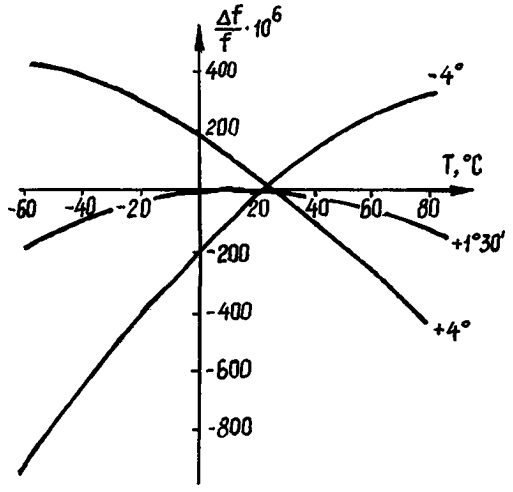


Fig. 8: Frequency vs temperature dependence for cut angle range φ from $-4°$ to $+4°$

Monolithic filter design features using LGS-crystal

In the course of the present investigation an engineering design method of monolithic filters using LGS-crystals has been developed. When designing monolithic LGS-filters it is necessary to take into account the influence of piezoelectric effect which is usually neglected in quartz crystal monolithic filter design.

Investigations of the dependence of acoustic coupling coefficient on propagation direction in XZ-plane (Fig. 9) showed, that the maximum value this coefficient has in the direction of the X-axis, which dominates the geometrical shape of yxl/+1°30' crystal element.

The dimensions of the crystal element are obtained under condition of unwanted response suppression and the minimum resonator motional resistance.

The most optimal dimensions can be considered the following ones:

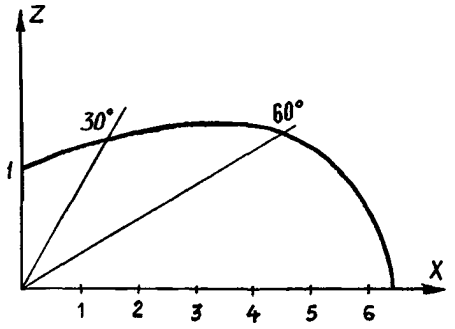


Fig. 9: LGS-element acoustic coupling coefficient vs orientation in YZ-plane. The acoustic coupling coefficient value in the Z-direction is taken as a unity

t_x – 30 t in the X-axis direction

t_z – 20 t in the Z-axis direction,

where

t - is the crystal element thickness

At present investigations are aimed at the development of monolithic filters using strip type crystal elements.

Electrode dimensions were determined from the energy trapping criterion and the condition of unwanted response suppression. These investigations have shown that the most optimum electrode dimensions correspond to the following relations:

$$l_x = (8 - 10)\, t$$

$$l_z = (6 - 8)\, t$$

where

t-is crystal element thickness, mm

l_x-is electrode dimension in the X-axis direction

l_z-is electrode dimension in the Z-axis direction

While determining the frequency lowering it is necessary to take into account the unwanted response existence, appearing at a frequency corresponding to half-mass lowering, which is very important for extremely wide-band filters. This kind of spurious vibrations can have a considerable influence on forming filter characteristic slope on the high frequency side and in certain cases on the pass band itself.

The value of frequency lowering can be represented as follows:

$$\Delta = \Delta_1 + \Delta_2$$

where

$$\Delta_1 = f_c \left(\frac{4}{\pi^2} k_{26}^2 + \frac{c_{11}}{c_{66}} \frac{\rho_e t_e}{\rho_s t_s} \right),$$

$\Delta_2 = 0.002 f_c$, additional coefficient for designing filters with relatively wide pass band exceeding 0.6%.

k_{26}^2 - electromechanical coupling coefficient;

f_c - centre frequency of a filter;

ρ_e, ρ_s - electrode and substrate material density, respectively;

t_e, t_s - evaporated electrode thickness and substrate thickness, respectively.

While further investigation the coefficients in Beaver's equation have been defined more exactly

$$k_{i\lambda} = \alpha \left(\frac{2l_x + d}{t} \right)^{-2} \exp \left\{ -\beta \sqrt{\frac{\Delta}{f_c}}\, d/t \right\}$$

where

$k_{i\lambda}$ is the acoustic coupling coefficient for $i\lambda$ piezoelectric element of monolithic filter

$\alpha = 3.05$

$\beta = 2.0$

l_x - is electrode dimension in the X-axis direction

d - interelectrode distance

and also the coefficients for the static capacitance C_o calculation

$$C_o = 0.435 \frac{S}{2t} \quad [pF]$$

and of motional inductance L_1 of a particular resonator L_g

$$L_g = 4.5 \frac{2t}{S} \quad [mH]$$

where

S-is the electrode area.

By using this method monolithic filters have been designed operating over the frequency range from 5 to 18.5 MHz with relative pass bands ranging from 0.4 to 0.85%.

Practical results

In the framework of our investigation monolithic LGS-crystal filters have been developed which are mass-produced at present. Their parameters are given in Table 3 compared to MCF analogues.

Table 3

Comparative characteristics of quartz crystal filters and LGS-filters

Manufacturer	F_o	Bandwidth		Ripple in pass-band	Stop band attenuation		Insertion attenuation	Operating temperature range	Outline dimension
		dB	kHz	dB	dB	kHz	dB	°C	
Fonon LGS-478	5.34	3.0	±22	1.5	80	35	2.0	-60...+85	1.0x23x53.0
QK F 1243	5.35	3.0	±7.5	0.5	75	65	2.0	-20...+85	17.5x26.5x60.5
Fonon LGS-500-1	10.7	3.0	±25	2.0	80	65	2.0	-60...+85	5.0x14.0x19.0
QK FTF No.107.0	10.7	3.0	±22	1.0	60	60	2.0	-40...+85	19.0x26.7x36.2
Fonon LGS-500	18.5	3.0	±50	2.0	80	125	2.0	-60...+85	5.0x14.0x19.0
QK FTF No.117	18.05	3.0	60	2.0	60	160	6.0	-40...+85	11.0x26.7x60.7

In Fig. 10 the realized characteristic of a band-pass monolithic LGS-filter is given.

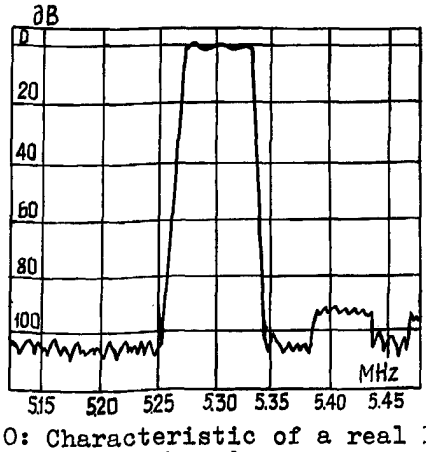


Fig. 10: Characteristic of a real LGS-filter of the 12th order

In Fig. 11 the temperature vs centre frequency characteristic of the same LGS-filter is presented.

In Fig. 12 the ageing characteristic of this LGS-filter is presented after 5000 h of the test at +85°C.

These filters have been also tested for their radiation resistance. Positive results have been obtained.

In Fig. 13 a pass-band LGS-filter is shown compared to MCF with analogous performance characteristics.

Evident superiority of LGS-filter can be seen in size and weight.

It should be also noted that LGS-filter manufacturing technology is more simple compared to that of filters with discrete elements.

In Fig. 14 the design of 12th order monolithic LGS-filter is shown.

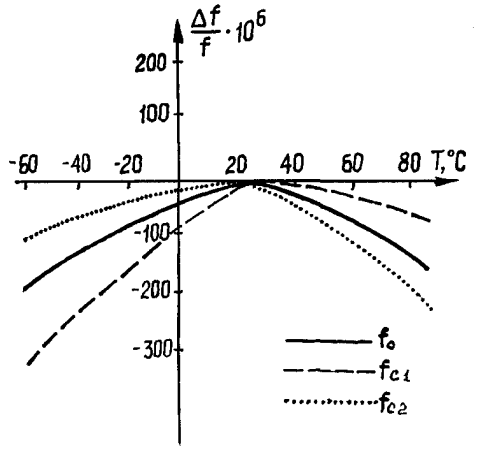


Fig. 11: Dependence of LGS-filter centre frequency and cut-off frequencies f_{c1}, f_{c2} on temperature

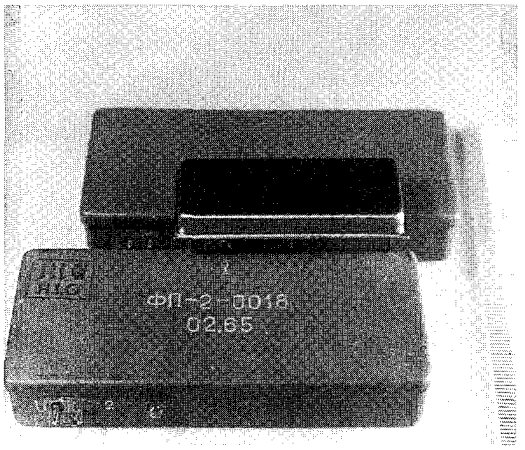


Fig. 13: LGS, 5 MHz-filter and its quartz crystal filter analogue

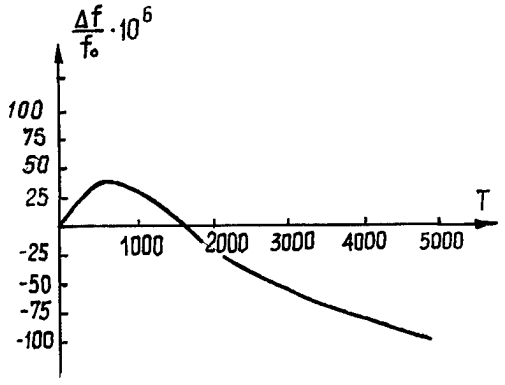


Fig. 12: Dependence of centre frequency vs time of 18.5 MHz LGS-filter at +85°C

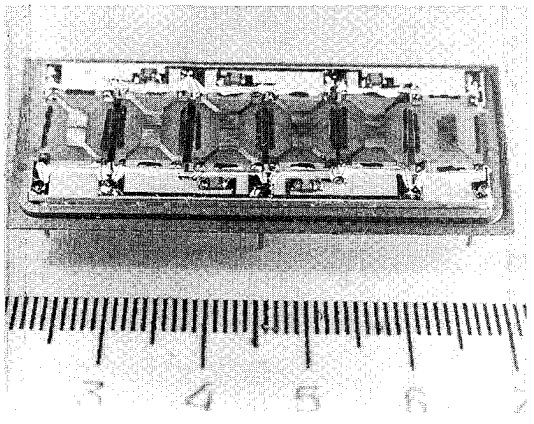


Fig. 14: Design of 12th order, LGS, 5 MHz-filter

In the course of the present investigation LGS-piezoelectric elements have been also developed and manufactured operating at 4 MHz and 10.7 MHz.

These elements have a strip configuration, their quality factor Q exceeds 15000 and the motional resistance 100 Ohm.

While using these piezoelectric elements in VCXO's, the pulling sensitivity obtained was up to 0.5%.

Technological features of langasite filter manufacturing

There is a number of special features in the process of LGS-crystal mechanical treatment, which make this technological process different from the quartz crystal treatment process. LGS-crystals are considerably softer than the quartz ones. Their Moos hardness equals to 5.5, which gouverns their mechanical treatment. While cutting LGS-crystals the best results are obtained when multi-blade cutters and machines with internal cutting edge are used with the diamond grain size of 50/40 μm. The use of diamond cutting wheels having smaller grain sizes causes LGS-crystal cracking as a result of its overheating.

In the process of X-ray orientation the difficulties arise connected with a correct definition of the positive X-axis direction and this causes incertaintly in the choice of the specified cut angle direction. A piezoelectric tester is used, therefore, for determination of the positive X-axis direction.

As far as chemical etch process of LGS-crystals is concerned, it is necessary to note their solubility practically in all nonorganic acids and in the majority of alkali.

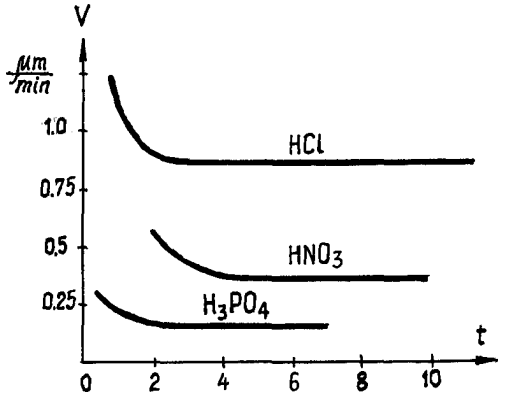


Fig. 15: Etch rate of LGS-plates in different acids vs the layer depth

Fig. 15 shows the dependence of LGS-crystal etching rate by different acids on etch layer depth. An alkali etchant was also selected for etching LGS-crystals at the room temperature without impairing crystal element roughness.

Other technological operations of LGS-monolithic filter manufacturing do not differ significantly from those used in traditional quartz crystal technology.

Conclusion

The use of LGS-crystal as a piezoelectric substrate makes it possible to enlarge substantially the application area of monolithic filters. They can replace quartz crystal filters with discrete elements having added inductance coils over the frequency range from 5 MHz to 18.5 MHz with a relative pass bandwidth of 0.3% to 0.8%. LGS-filters are, moreover, superior to them in size and weight, reliability and they are easier in production.

At present investigations are aimed at the enlargement of LGS-filter frequency range both towards the lower frequencies up to 2 MHz and towards the higher frequencies up to 25 MHz (it is supposed up to 50 MHz).

Investigations of other crystals are also made, namely of gallium germanate group, in particular, of calcium gallium germanate ($Ca_3Ga_4GeO_{14}$) and strontium gallium germanate ($Sr_3Ga_5GeO_{14}$). The authors expect that these crystals will also find their application in piezoelectric technique.

References

[1] Kaminsky A.A., Sarkissov S.E., et al. Generation of stimulated Nd-ion radiation in trigonal acentric $La_3Ga_5SiO_{14}$-crystal. Reports of the U.S.S.R. Academy of Science, 1982, v.264, No.1, p. 93-95.

[2] Kaminsky A.A., Silvestrova I.M., at al. Investigation of trigonal $(La_{1-x}Nb_x)_3Ga_5SiO_{14}$-crystals.

[3] Kaminsky A.A., Mil B.V., Silvestrova I.M. et al. Nonlinearity-active material $(La_{1-x}Nb_x)_3\ Ga_5SiO_{14}$. Izv. of the U.S.S.R. Academy of Science, Physics Series, 1983, v.47, No.10, p.1903-1909.

[4] Andreev A.I., Doubovik M.F. New piezoelectric "langasite" - $La_3Ga_5SiO_{14}$ - the material with a zero frequency-temperature coefficient of elastic vibrations. Letters to Journal of Theoretical Physics, 1984, v.10, issue 8, p.487-491.

[5] Mil B.V., Batachin A.B. et al. Modified gallates with the structure $Ca_3Ga_5GeO_{14}$. Report of the U.S.S.R. Academy of Science, 1982, No.26, p.1385-1389.

[6] Stassevich V.A. Single crystals technology, Moscow, Radio and Communication, 1990, p.273.

[7] Bronnikova E.G., Larionov I.M., et al. Monolithic filters and resonators using piezoelectric material gallium silikate of lantanum, Electronic Engineering, Series "Radiodetails and Components", issue 2 (63), 1986.

[8] Zelenka J. Temperature dependence of resonant frequency in accurate AT- and BT- quartz resonators and its dependence on excitation level. Electronics & Telecommunications, v.3, No.3, 1969.

[9] Ssakharov S.A., Larionov I.M. et al. "Monolithic crystal filter", Patent application No.4886181/22 of 28.11.90.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

DILITHIUM TETRABORATE ($Li_2B_4O_7$) FABRICATION TECHNOLOGY

J. G. GUALTIERI, J. A. KOSINSKI, and W. D. WILBER
US Army Electronics Technology and Devices Laboratory,
Fort Monmouth, NJ 07703-5601

Y. LU
Rutgers University, Department of Electrical and Computer Engineering, Piscataway, NJ 08855-0909

S. T. LIN, M. MURRAY, and W. RUDERMAN
INRAD Inc., 181 Legrand Avenue, Northvale, NJ 07647-2498

Abstract: Dilithium tetraborate (DLTB)* is a promising piezoelectric material for frequency control and signal processing applications. Since the first announcement of this material along with its favorable properties [1], considerable progress has been made in the growth of single crystals and oriented thin films for bulk and surface acoustic wave devices.

The growth and fabrication technology of DLTB crystals and devices of the type reported recently at this symposium [2] are discussed. Also reviewed are recent worldwide efforts to grow large bulk crystals and thin films.

Special emphasis is given to: (1) the Czochralski technique including the effort to eliminate a number of crystal defects arising from the use of this method and the activity to grow large diameter crystal boules; and (2) the sol-gel method for thin film preparation including the role of acid and water on crystallization and side products. Other lesser-used techniques for bulk and thin film production of DLTB, such as, Bridgman growth, RF sputtering, vapor deposition, and laser ablation deposition are discussed.

Introduction

The search for exceptional piezoelectric materials for bulk and surface acoustic wave devices is a continuous exercise. $Li_2B_4O_7$ is a relatively new material that combines the properties of having high piezocoupling and crystal cuts (orientations) for which the temperature coefficients of frequency and time delay are zero. None of the piezoelectric materials currently in use or under development can match this exceptional combination of properties.

As a consequence of its 4mm (C_{4v}) symmetry, DLTB is piezoelectric, pyroelectric, optically uniaxial (negative), not enantiamorphic, subject to growth twinning (c-axis reversals) and nonferroelectric (poling not required). DLTB is a congruently melting (m.p. = 917°C) compound in the Li_2O-B_2O_3 system. Crystals are transparent and colorless. DLTB has: lattice spacings a = b = 9.477 Å and c = 10.280 Å, a Mohs hardness of 6 (quartz is 7), a low density (2.451 g/cm^3), acoustic velocities slightly greater than $LiNbO_3$ and $LiTaO_3$, a SAW reflectivity per stripe greater than 5 times $LiNbO_3$, $LiTaO_3$, or quartz - suggesting miniaturization, solubilities described as very soluble in acids, slightly soluble in bases, and insoluble in organic solvents. $Li_2B_4O_7$ is etched by hot water.

Advances in DLTB fabrication technology have been occurring, to a large extent, without U.S. participation. In many cases, foreign technology advances have been reported in scientific journals that are difficult to obtain. The aim of this paper is to describe our efforts to grow bulk crystals and thin films of DLTB and to review, cite, and list additional references found in the foreign literature. We hope this disclosure will both serve as motivation for increased U. S. participation in this field and as a reference source for researchers interested in joining the worldwide effort to provide the frequency

* The laser community is using the acronyms LBO and LTB to designate lithium triborate, a promising frequency doubler crystal. To avoid confusion, we suggest the use of nomenclature for $Li_2B_4O_7$ that indicates the proportions of the constituents denoted by Greek numerical prefixes [3], namely, dilithium tetraborate [4]. We also have adopted the acronym DLTB in accordance with this choice of terminology.

0-7803-0476-4/92 $3.00 © 1992 IEEE

control community with this important material.

Experimental Procedures / Results

Melt-Growth Techniques

Czochralski Growth: Crystal pulling is the most important melt-growth technique and is due originally to Czochralski. Advantages of this method are that the growing boule is not in contact with the crucible and the boule may be viewed during the growth run. Disadvantages are the possible contamination of the molten material by the crucible material and the difficulty in obtaining reproducible crystal shapes.

In view of the low melting point of DLTB, resistance heating has proved simpler than RF heating. The furnace design is shown schematically in Fig. 1. The heating element was Kanthal wire that was first coiled and then wound on an alundum core. The winding was spaced to give the desired temperature gradient (approximately 10°C/cm) in the crystal growth zone. The seed was attached to a 2 mm diameter pulling rod by means of platinum wire. The pulling rate was 0.2 mm/h and the rotation rate was 3 RPM with periodic

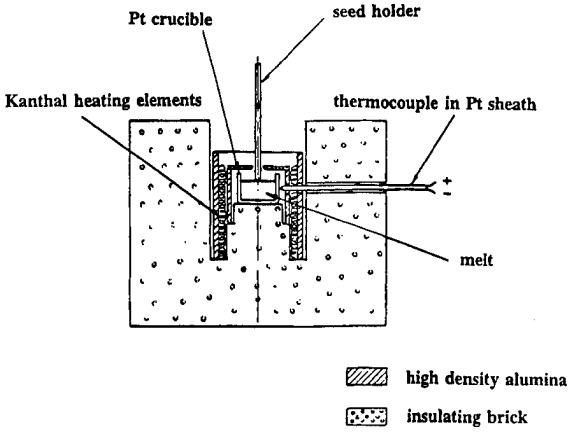


Fig. 1. Furnace design for Czochralski growth of $Li_2B_4O_7$.

reversals of rotation. The important elements of DLTB growth are the heat shields, the purity of starting material, and the temperature control system. Heat shields and after heaters are required because DLTB develops cracks if large thermal gradients are present. The control system for our Czochralski crystal growth station is shown in Fig. 2.

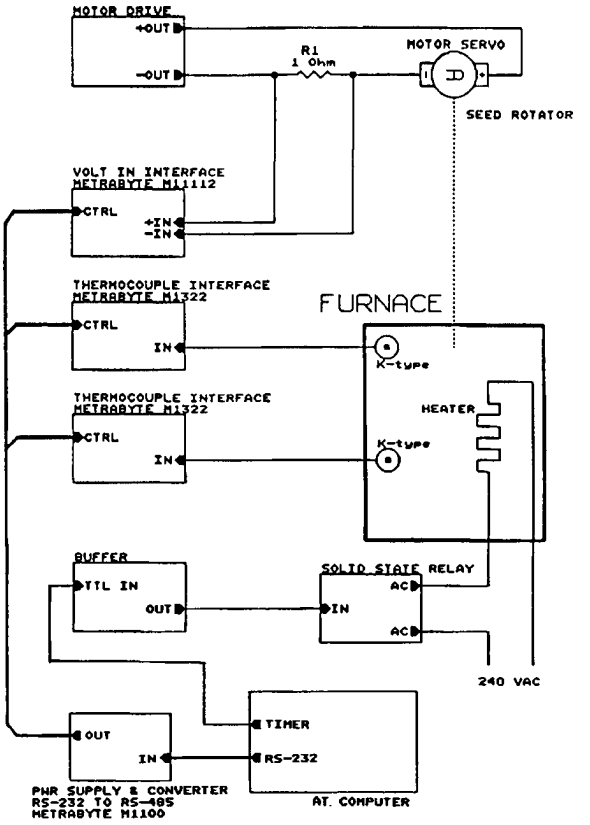


Fig. 2. Czochralski growth station control system.

Temperature is measured at one or more points inside the furnace. The measurement is done using K-type thermocouples, each connected to a Metrabyte signal conditioning module. This module sends the temperature information, in digital format, via an RS-485 bus, to the computer. This communication is more robust than RS-232 and is less affected by RFI/EMI. The actual temperature is compared against the set point programmed for the present time. Based on a proportional-integral-differential (PID) algorithm, the computer calculates the power required for the heater.

The furnace is also provided with a DC servomotor to rotate the seed. The servomotor is driven at constant speed by a motor driver. The resistor R_1 (1 ohm) is used to measure the torque of the motor. This torque should be relatively constant during the normal mode of operation, but may greatly increase when the crystal is about to freeze. This measurement and data communication is accomplished by another Metrabyte signal conditioning module.

The parameters of the system - actual temperature, set point, torque, power level

- are written every second to a file. That file later can be analyzed and changes made to the control program, temperature profile, PID parameters, etc., in order to optimize the crystal quality and size. This file also provides information about subtle effects caused by changes between day and night and the effect of many people starting their work at the same time, i.e. power drop.

Our use of high purity starting material (99.999%) may be the reason why growth defects were not encountered even though automatic diameter control (ADC) was not used. Some bubbles were observed near the seed but these disappeared after about 1.5 cm of growth. Other practical features are the use of platinum crucibles, oxygen or air ambients, and usually either <100> or <001> growth directions. However, <110> growth directions also have been used to obtain 45° x-cut samples for SAW devices [5-7]. Most of our growth runs were along <100>. Experimental boules (see Fig. 3) were typically 4 cm in diameter. No difficulties are expected in scaling to produce 7 cm diameter boules. The boule in Fig. 3 was suddenly removed from the melt to reveal the shape of the solid-liquid (S-L) interface.

Typical analysis of crystals grown with high purity starting material and high purity platinum crucibles (in PPM) was: Al 1.6, Ca 1, Fe 1, K 1, Na 0.3, and Si < 5.

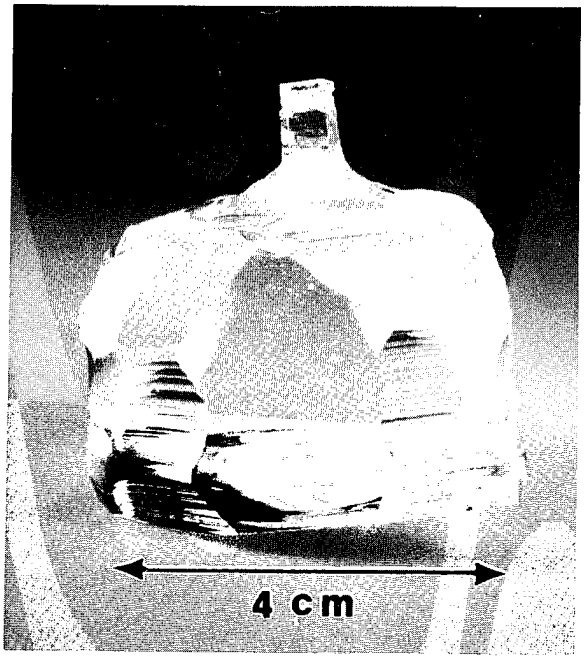


Fig. 3. An x-grown boule showing growth facets, a flat S-L interface, and small bubble inclusions that typically appeared only beneath the seed area.

The crystal boules were colorless and exhibited excellent transmission from 200 nm to 1700 nm. The transmission of a 40 mm thick polished Z-cut $Li_2B_4O_7$ crystal is shown in Fig. 4. X-ray rocking curves were made using a double crystal x-ray spectrometer. Cu K-alpha radiation was used to produce the typical rocking curve shown in Fig. 5.

Recent Contributions: In the past, many investigators have prepared DLTB by mixing lithium carbonate with boric acid and melting the mixture in platinum crucibles [8-12]. Czochralski growth was attempted by Garrett, et al. [13] from nickel crucibles because he found that Pt-crucibles were slowly oxidized above 900°C. To obtain good quality crystals, he used a counter rotating seed and crucible. Whatmore, et al. [14] and Shorrocks, et al. [15] grew large single crystals having good optical quality. They used heat shielding to achieve thermal stability.

Robertson and Young [16] used a borate rich starting material to compensate for the loss of B_2O_3 from the melt during growth. This was effective in eliminating the production of Li-rich core defects. They developed a model employing growth induced fluid flow at the S-L interface to explain the origin of the core defects.

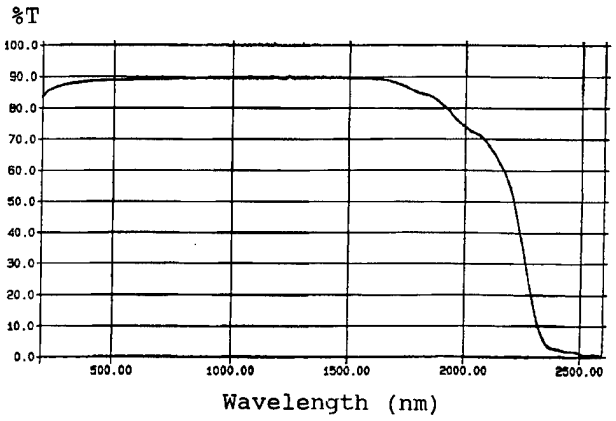


Fig. 4. Z-axis transmission spectrum of a 40 mm thick DLTB crystal.

Fukuta, et al. [17], used 99.99% $Li_2B_4O_7$ powder or purified polycrystal as starting material, a rotation rate of 10-40 rpm, a linear growth rate of 0.5 to 5 mm/h, with growth along <100> and <001>. Crystal cracking and void incorporation were suppressed by using low temperature gradients, a thermal reflector, slower growth rates, and higher purity starting materials. They found that voids correlated with growth rate changes and could be

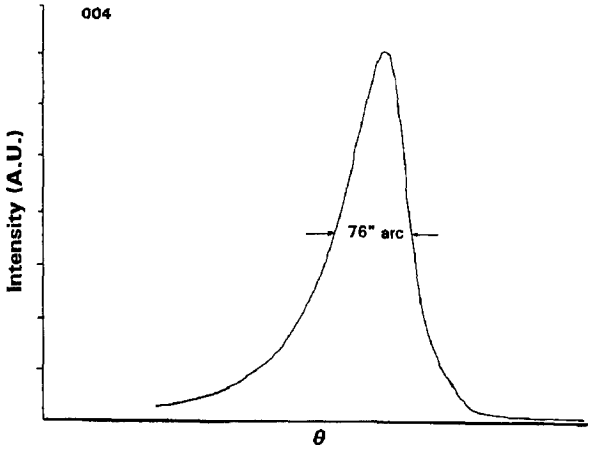


Fig. 5. X-ray rocking curve for the 004 reflection of DLTB, grown by the Czochralski method.

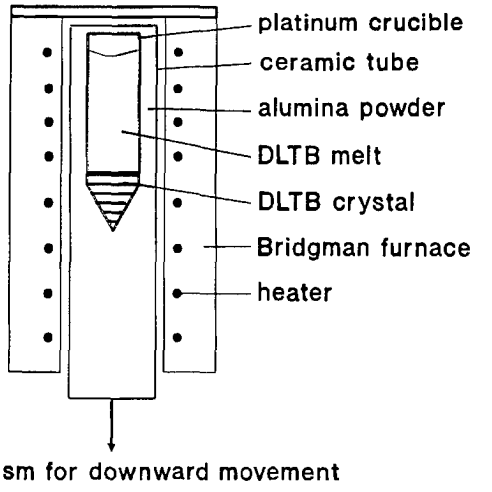


Fig. 6. Schematic diagram of Bridgman growth of DLTB [20].

attributed to constitutional supercooling and the evaporation of boron rich $Li_2B_4O_7$. They found that a-axis crystals were more difficult to grow than c-axis crystals.

Adachi, et al. [18], also found that DLTB is very sensitive to temperature gradients and is subject to cracking. They concentrated on reducing radial and vertical temperature gradients. They found that starting materials must be pure and they added B_2O_3 to the melt to reduce cracking and coring. They also used ADC to improve boule shape and quality. These investigators pointed out that seed rotation rates influenced the shape of the S-L interface and that the best crystals were grown with a flat interface. Flat S-L interfaces can be obtained by optimizing the pulling/rotation rates. Later, they reported that the formation of narrow central core defects in the growing boule depended on the shape of the S-L interface [19].

Bridgman Growth: In this method, DLTB is completely melted in a Pt-crucible in an upper heating chamber. The crucible is then lowered through a negative temperature gradient and crystallization progresses upward from a seeded tip (see Fig. 6). Crystal size is easily scaled and vertical and radial temperature gradients may be designed to be quite low to insure crack-free growth. In this method, the natural convection in the melt is minimized and power and temperature stabilization is more easily maintained. The crystal assumes the shape of the crucible so ADC is unnecessary. This technique has a number of advantages and may be more appropriate for large-scale production of DLTB.

Growth of high-quality, core-free, and sub-boundary free DLTB crystals was reported by Fan Shi-ji, et al. [20] They used <001>, <100>, and <110> orientations and found no defect dependence on growth rate below 0.3 mm/h if convex or plane S-L interfaces were employed. Coring was found only when the S-L interface was concave. They were able to grow good quality crystals with only 99.9% pure raw material, fabricated in advance using pressure to create blocks of DLTB. However, they found that recrystallized fragments of DLTB gave better results than either stoichiometric composition or Li-rich composition raw material. Since there was no forced convection of the melt during growth and only a small temperature gradient, their growth rate was only 0.3 mm/h. This is half the growth rate obtainable with the Czochralski method for core-free, good quality crystals. However, in Bridgman growth, cracks resulting from thermal stress (because of the anisotropic thermal expansion) are avoided mainly by using these small temperature gradients and stable temperature fields.

Very large crystals can be produced by the Bridgman / Stockbarger method. Axial temperature gradients can be reduced to very low values by appropriate design of heaters and radiation baffles. Heat transport is influenced by: 1) the geometry which minimizes the degree of natural convection in the melt and 2) radiation through the transparent melt.

Thin Film Techniques

Sol-gel Growth: This low-processing temperature technique has proved useful for fabrication of lithium niobate thin films [21]. Sol-gel growth of DLTB has been carried out by Yamashita, et al. [22-24]. In their method, the coating solution preparation is accomplished in a dry nitrogen atmosphere. Solutions are prepared by adding Li-metal to methanol at temperatures of 0° to 2°C to form lithium methoxide, $LiOCH_3$, solution. Boron tri-n-butoxide, $B(OC_4H_9)_3$, is then added to this solution at a rate of 10 mL/min at 20°C. The resulting solution is stirred at 40°C until transparent, then diluted to a concentration of 0.4M $LiOCH_3$ and 0.8M $B(OC_4H_9)_3$ by methanol. These procedures were conducted in a dry nitrogen atmosphere. Then water and varying amounts of acetic or hydrochloric acid are added.

Films are fabricated by dip coating. Substrates are dipped into the coating solution and withdrawn at fixed speeds. This process also is carried out in a dry nitrogen atmosphere. The coatings are air-dried at 200°C for 10 minutes. Film thickness is increased by repeating the dipping procedure from 8 to 15 times. Films are then heat treated for 30 minutes in air at 600°C and sometimes for 60 minutes in dry nitrogen at 800°C.

Yamashita, et al. [22] found that crystallization starts at 250°C in thick (3.5 μm) films. Thin films were found to be resistant to crystallization up to 550°C. They established that as water is added to the coating solution, the crystallization of a single $Li_2B_4O_7$ phase predominates. This occurs when the mole ratio (R_w) of water to the total alkoxy groups, i.e., the sum of $-OCH_3$ from $LiOCH_3$ and $-OC_4H_9$ from $B(OC_4H_9)_3$ in solution becomes 1.6. Thus, the water content of the solution greatly influences the crystallization behavior of the film. Increasing the water content decreases the crystallization temperature and suppresses the formation of phases other than $Li_2B_4O_7$.

Later it was found that by preparing films with R_W = 2.4 and R_A = 2.3, where R_A is the mole ratio of acetic acid, the films become highly oriented along the (122) plane on Si (100) substrates with heat treating to 800°C [23]. Similarly, by adding HCl instead of acetic acid with R_W = 2.4 and R_H = 1.3, the degree of (122) orientation on Si (111) became 0.98 [24].

The addition of HCl not only aids in crystallization, but also aids in suppressing random orientations. However, the addition of a large amount of HCl was found to promote the growth of α-quartz at the interface between the $Li_2B_4O_7$ film and the Si substrate at temperatures of 800°C even in a dry nitrogen atmosphere. Lowering the HCl content to R_H = 0.3 and the temperature to 600°C eliminates this problem [24].

It seems that the addition of HCl suppresses crystallization up to 600°C, where preferential crystal growth occurs along (122) on further heating to 800°C. The preference for the (122) orientation was attributed not to favorable lattice matching conditions but to the relatively small number of oxygen atoms in the (122) plane compared to other possible planes. This helps to minimize the lattice distortion energy between DLTB and the Si substrate [24].

Sol-gel processing has advantages of high purity, low temperature, and precise control for the fabrication of fibers and thin film and bulk monolithic structures.

RF Sputtering: RF sputtering is a well-known and widely used technique for preparing thin films. Sputtering targets can be pressed from DLTB powders and sintered at about 700°C. Uno [25] used a mixture of argon and oxygen as ionization gas at a pressure of 0.8 Pa. Heated fused quartz and Si (100) substrates were used to obtain c-axis oriented films that were shown to be piezoelectric. The sputtered DLTB films decomposed at high RF power and could only be grown reproducibly using RF powers of less than 300 W.

Vapor Deposition: DLTB was found to decompose when heated in a Pt-crucible above its melting point in a vacuum [25]. The decomposition was effectively suppressed using a 50 Pa overpressure of argon. c-axis oriented films were produced on fused quartz and silicon (100) substrates that were heated to 480°C. It was found that the deposition rate depends on gas pressure, crucible temperature, distance between the crucible and the substrate, and melt surface area. Uno used a 13 mm diameter crucible and was able to achieve deposition rates of only 0.05 to 0.15 μm/h, too low for practical film production. With the crucible temperature optimized to 1000°C, he concluded that a deposition rate of 0.5 μm/h could be achieved if a 45 mm diameter crucible were used.

Laser Ablation Deposition: We report here our preliminary results on DLTB films prepared by the excimer laser ablation method.

In laser ablation, short high energy pulses of light are focused onto a target in order to generate a plasma of the target

material. A thin film of target material is grown by collecting the laser-generated plasma on an appropriate substrate (see Fig. 7) . This method is well suited for growth of refractory material films because it is the only method capable of transferring complex molecular structures from ceramic targets to substrates at a high rate. Recently, there have been many experimental and theoretical investigations of laser ablation, especially for production of high temperature superconducting films such as $YBa_2Cu_3O_7$ [26].

Targets of DLTB about 64 % relative density were prepared by isostatic pressing (up to 3400 Atm) and sintering at 700°C for 12 h. The film deposition was accomplished by excimer laser ablation of a rotating DLTB target with the incident beam (KrF exciter laser, 248 nm wavelength, 15 ns pulse width) at a 42° angle with respect to the target surface in a vacuum chamber under 100 mTorr oxygen pressure. The beam energy density was varied, from 3 to 5 J/cm^2, by changing the laser spot size on the target surface by adjusting the lens to target distance.

Si (100) substrates were heated to up to 800°C. After deposition, the oxygen pressure was increased to 250 Torr while the substrate cooled to room temperature in about 30 minutes. Work is underway to analyze the films using x-ray diffraction and Auger analysis to evaluate stoichiometry. Results indicate that films prepared at elevated temperatures are lithium rich, probably due to a loss of B_2O_3 during deposition on the heated substrate. Films prepared at room temperature and elevated temperature exhibited a piezoelectric response under the application of stress along the normal direction to the plane of the film. Similar piezoelectric behavior was reported for films prepared by RF sputtering and vacuum deposition [25].

Discussion / Conclusion

Both the Czochralski and Bridgman methods can be used to grow good quality DLTB bulk crystals for SAW and BAW device applications. Good Czochralski technique depends on a number of factors including: vertical and radial temperature gradients near the growth surface, the starting material purity, and growth and seed rotation rates optimized to obtain a flat S-L interface. The elimination of core defects depends critically on the adjustment of growth and rotation rates and control of the boule diameter. The Bridgman method has a number of advantages related to temperature stabilization and melt convection which may make it more appropriate for future large scale production of DLTB.

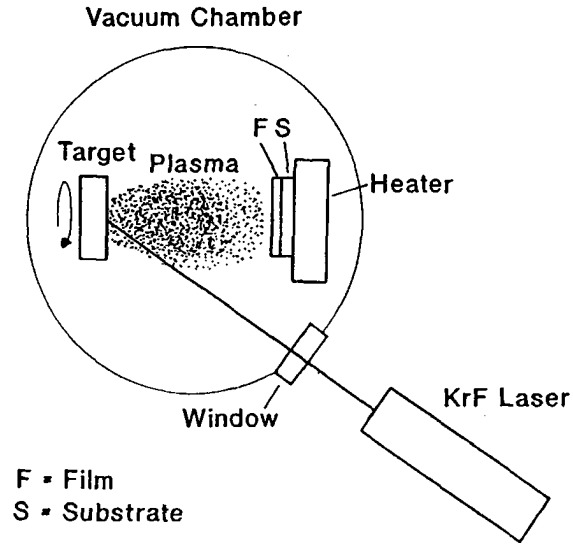


Fig. 7. Diagram of the laser ablation thin-film deposition system.

DLTB films can be prepared by a number of methods. The sol-gel technique, useful for fabrication of ceramic materials, provides excellent highly ordered crystalline films of DLTB using a simple dip coating procedure. Preferential crystal growth along the (122) plane was attributed to the smaller number of oxygen atoms contained in this plane, possibly minimizing the interfacial distortion required for its nucleation.

While DLTB films can be deposited using RF sputtering and vapor deposition techniques, these methods have decomposition and low growth rate problems to overcome before they can challenge the sol-gel method.

The laser ablation deposition method shows great promise for high rate growth of complex molecular structure refractory materials on substrates. The required experimental setup for laser ablation is very simple. With the exception of an appropriate target, all the items in a laser ablation system are within the means of modern industrial and university laboratories.

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 Letters, vol. 17, no. 1, pp. 11-12, 8 Jan. 1981.

[2] A. Ballato, J. Kosinski, S. Mallikarjun, and T. Lukaszek, "A Comparison of Predicted and Measured Properties of Doubly- Rotated Lithium Tetraborate Resonators," Proceedings of the 45th Annual Symposium on Frequency Control, May 1991, pp. 217-221.

[3] International Union of Pure and Applied Chemistry Inorganic Rules as published in J. Amer. Chem. Soc., vol. 82, pp. 5525-5544, 1960.

[4] M. Natarajan, R. Faggiani, and I. D. Brown, "Dilithium Tetraborate, $Li_2B_4O_7$," Crystal Structure Communications, vol. 8, pp. 367-370, 1979.

[5] H. Abe, H. Saitou, M. Ohmura, T. Yamada, and K. Miwa, "Lithium Tetraborate ($Li_2B_4O_7$) SAW Resonators," IEEE Ultrasonics Symposium Proceedings, 1987, pp. 91-94.

[6] S. Matsumura, T. Omi, N. Yamaji, and Y. Ebata, "A 45° X Cut $Li_2B_4O_7$ Single Crystal Substrate for SAW Resonators," IEEE Ultrasonics Symposium Proceedings, 1987, pp. 247-250.

[7] S. Furusawa, S. Tange, Y. Ishibashi, and K. Miwa, "Raman Scattering Study of Lithium Diborate ($Li_2B_4O_7$) Single Crystal," Journal of the Physical Society of Japan, vol. 59, no. 5, pp. 1825-1830, May 1990.

[8] G. S. Smith and G. E. Rindone, "High-Temperature Energy Relations in the Alkali Borates: Binary Alkali Borate Compounds and Their Glasses," Journal of the American Ceramic Society, vol. 44, no. 2, pp. 72-78, Feb. 1961.

[9] J. Krogh-Moe, "The Crystal Structure of Lithium Diborate, $Li_2O{\cdot}2B_2O_3$," Acta Crystallographica, vol. 15, pp. 190-193, 1962.

[10] J. Krogh-Moe, "Refinement of the Crystal Structure of Lithium Diborate, $Li_2O{\cdot}2B_2O_3$," Acta Crystallographica, vol. B24, pt. 2, pp. 179-181, 1968.

[11] L. A. DeWerd, T. H. Kim, and T. G. Stoebe, "Thermoluminescent Behavior of Crystalline and Glassy Lithium Borate," Materials Research Bulletin, vol. 11, pp. 1413-1418, Nov. 1976.

[12] S. R. Nagel, L. W. Herron, and C. G. Bergeron, "Crystal Growth of $Li_2B_4O_7$," Journal of the American Ceramic Society, vol. 60, no. 3-4, pp. 172-173, March-April 1977.

[13] J. D. Garrett, M. N. Iyer, and J. E. Greedan, "The Czochralski Growth of $LiBO_2$ and $Li_2B_4O_7$," Journal of Crystal Growth, vol. 41, pp. 225-227, 1977.

[14] 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 Letters, vol. 17, no. 1, pp. 11-12, 8 Jan. 1981.

[15] 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 Devices," IEEE Ultrasonics Symposium Proceedings, 1981, pp. 337- 340.

[16] D. S. Robertson and I. M. Young, "The Growth and Growth Mechanism of Lithium Tetraborate," Journal of Materials Science, vol. 17, pp. 1729-1738, 1982.

[17] K. Fukuta, J. Ushizawa, H. Suzuki, Y. Ebata, and S. Matsumura, "Growth and Properties of $Li_2B_4O_7$ Single Crystal for SAW Device Applications," Japanese Journal of Applied Physics, vol. 22, Supplement 22-2, pp. 140-142, 1983.

[18] M. Adachi, T. Shiosaki, and A. Kawabata, "Crystal Growth of Lithium Tetraborate ($Li_2B_4O_7$)," Japanese Journal of Applied Physics, vol. 24, Supplement 24-3, pp. 72-75, 1985.

[19] T. Shiosaki, M. Adachi, and A. Kawabata, "Growth and Properties of Piezoelectric Lithium Tetraborate Crystal for BAW and SAW Devices," IEEE International Symposium on the Application of Ferroelectrics (ISAF) Proceedings, Lehigh University, Bethlehem, PA, June 1986, pp. 455-464.

[20] Fan Shi-ji, Shen Guan-shun, Wang Wen, Li Jin-long, and Le Xiu-hong, "Bridgman Growth of $Li_2B_4O_7$ Crystals," Journal of Crystal Growth, vol. 99, no. 1-4, pt. 2, pp. 811- 814, Jan. 1990.

[21] S. Hirano and K. Kato, "Preparation of Crystalline $LiNbO_3$ Films with Preferred Orientation by Hydrolysis of Metal Alkoxides," Adv. Ceram. Mater., vol. 3, pp. 503-506, 1988.

[22] H. Yamashita, T. Yoko, and S. Sakka, "Preparation of $Li_2B_4O_7$ Thin Films by Sol-Gel Method," Journal of Materials Science Letters, vol. 9, no. 7, pp. 796-800, July 1990.

[23] H. Yamashita, T. Yoko, and S. Sakka, "Preparation of Highly Oriented $Li_2B_4O_7$ Films by Sol-Gel Method," Nippon Seramikkusu Kyokai Gakujutsu Runbushi, vol. 98, no. 8, pp. 913-916, 1990.

[24] H. Yamashita, T. Yoko, and S. Sakka, "Preparation of $Li_2B_4O_7$ Films With Preferential Orientation by Sol-Gel Method," Journal of the American Ceramic Society,

vol. 74, no. 7, pp. 1668-1674, July 1991.

[25] T. Uno, "Fabrication of $Li_2B_4O_7$ Piezoelectric Thin Films," Japanese Journal of Applied Physics, vol. 27, Supplement 27-1, pp. 120-122, 1988.

[26] D. B. Chrisey and A. Inam, Pulsed Laser Deposition of High T_c Superconducting Thin Films for Electronic Device Applications," (and references therein) MRS Bulletin, Vol. XVII, No. 2, pp. 37-43, Feb. 1992.

Additional References for $Li_2B_4O_7$ Fabrication Technology

[27] B. S. R. Sastry and F. A. Hummel, "Studies in Lithium Oxide Systems: III, Liquid Immiscibility in the System Li_2O-B_2O_3-SiO_2," Journal of the American Ceramic Society, vol. 42, no. 2, pp. 81-88, Feb. 1959.

[28] B. S. R. Sastry and F. A. Hummel, "Studies in Lithium Oxide Systems: V, Li_2O-Li_2O-B_2O_3," Journal of the American Ceramic Society, vol. 42, no. 5, p. 218, 1959.

[29] J. Liebertz, "Crystal Growth From Melts of High Viscosity," Progress in Crystal Growth and Characterization, vol. 6, no. 4, pp. 361-369, 1983.

[30] M. Adachi, S. Yamamichi, M. Ohira, T. Shiosaki, and A. Kawabata, "Crystal Growth of $Li_2B_4O_7$ and its Leaky SAW Properties," Japanese Journal of Applied Physics, vol. 28, Supplement 28-2, pp. 111-113, 1989.

[31] T. Shiosaki, "Recent Developments in Piezoelectric Materials," Ferroelectrics, vol. 91, pp. 39-51, 1989.

[32] R. Komatsu, T. Suetsugu, S. Uda, and M. Ono, "The Growth and Characterization of Lithium Tetraborate Single Crystal," Ferroelectrics, vol. 91, pp. 103-107, 1989.

[33] T. Shiosaki and A. Kawabata, "Future Trends in Piezoelectric Materials and Applications," Ferroelectrics, vol. 95, pp. 9-14, 1989.

[34] T. P. Balakireva, V. V. Lebol'd, V. A. Nefodov, M. V. Provotorov, and A. A. Maier, "Czochralski Growth of Lithium Tetraborate Single Crystals," Inorganic Materials, vol. 25, no. 3, March 1989, pp. 462-464. Translated from Izvestiya Akademii Nauk SSSR, Neorganischeskie Materialy, vol. 25, no. 3, pp. 524-526, Mar. 1989.

[35] V. V. Zaretskii and Ya. V. Burak, "New Incommensurate $Li_2B_4O_7$ Crystal," Soviet Physics - Solid State, vol. 31, no. 6, June 1989, pp. 960-963. Translated from Fizika Tverdogo Tela, vol. 31, no. 6, pp. 80-84, June 1989.

[36] Ya. V. Burak, B. N. Kopko, I. T. Lyseiko, A. O. Matkovskii, R. R. Slipetskii, and U. A. Ulmanis, "Color Centers in $Li_2B_4O_7$ Single Crystals," Inorganic Materials, vol. 25, no. 7, July 1989, pp. 1038-1039. Translated from Izvestiya Akademii Nauk SSSR, Neorganicheskie Materialy, vol. 25, no. 7, pp. 1226- 1228, July 1989.

[37] Ya. V. Burak, Ya. O. Dovgii, and I. V. Kityk, "Energy Band Structure and Characteristics of Chemical Bonds in $Li_2B_4O_7$ Crystals," Soviet Physics - Solid State, vol. 31, no. 9, September 1989, pp. 1634-1636. Translated from Fizika Tverdogo Tela, vol. 31, no. 9, pp. 275-278, Sep. 1989.

[38] S. F. Radaev, L. A. Muradyan, L. F. Malakhova, Ya. V. Burak, and V. I. Simonov, "Atomic Structure and Electron Density of Lithium Tetraborate $Li_2B_4O_7$," Soviet Physics - Crystallography, vol. 34, no. 6, November-December 1989, pp. 842- 846. Translated from Kristallografiya, vol. 34, no. 6, pp. 1400-1407, Nov. - Dec. 1989.

[39] K. Ya. Borman and Ya. V. Burak, "Characteristics of the Thermal Expansion of $Li_2B_4O_7$," Inorganic Materials, vol. 26, no. 2, February 1990, pp. 372-373. Translated from Izvestiya Akademii Nauk SSSR, Neorganische Materialy, vol. 26, no. 2, pp. 440-442, Feb. 1990.

[40] Yang Ping, Jiang Shusheng, Wang Luchun, Zhao Jiyong, Jiang Xiaoming, Jiang Jianhua, Chao Zhiyu, Tian Yulian, Xian Dingchang, Yang Wenjuan, and Yang Yao, "Synchrotron Radiation Topographic Study on the Dislocations in Lithium Tetraborate Crystal," Chinese Physics Letters, vol. 8, no. 2, pp. 78-81, 1991.

[41] T. Lukasiewicz and A. Majchrowski, "Czochralski Growth of Lithium Tetraborate Single Crystals," Materials Letters, vol. 11, no. 9, pp. 281-283, July 1991.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

Piezoelectric, Elastic, and Ferroelectric Properties of $KTiOPO_4$ and Its Isomorphs

David K.T. Chu, John D. Bierlein
Du Pont Experimental Station, Central Research & Development, P. O. Box 80356, Wilmington, Delaware 19980-0356

Robert G. Hunsperger
Department of Electrical Engineering, University of Delaware, Newark, DE 19716

Abstract

Potassium titanyl phosphate ($KTiOPO_4$, KTP) has been shown to possess some promising bulk properties such as relatively large electro-mechanical coupling coefficients and moderate material parameter temperature dependence in frequency control applications. In this article, results of the bulk elastic, piezoelectric and ferroelectric properties of KTP and its isomorphs will be discussed. Influences on piezoelectric and dielectric properties caused by ionic dopants such as barium and gallium will also be presented.

Introduction

The KTP crystal structure belongs to the orthorhombic crystal class mm2, it has five independent nonzero piezoelectric tensor elements (e_{ij}) and nine independent nonzero elastic tensor elements (c_{ij}) [1]. Most of the elastic, piezoelectric and dielectric constants of KTP can be determined by measuring the thickness vibrations of crystal plates with various crystallographic orientations [2]. A combination of bar and plate samples with specific orientations enables one to completely determine the piezoelectric-elastic-dielectric (P-E-D) matrix of KTP [3]. Curie temperatures and dielectric properties of $KTiOPO_4$ (KTP), $RbTiOPO_4$, $KTiOAsO_4$, $RbTiOAsO_4$, $CsTiOAsO_4$, Ba:KTP, and Ga:KTP were measured with small-signal relative dielectric permittivity analysis, acoustic resonance analysis, and optical second harmonic generation. Thermodynamic study on the order of ferroelectric phase transition in KTP has been conducted using optical second harmonic generation technique. Knowing the P-E-D matrix of KTP will allow one to study the various type of surface waves which could be excited through interdigitated transducers. A fundamental understanding of ferroelectric properties in KTP would assist the analysis on doping effects.

Experimental and Results

The orientation of the different crystal cuts used to completely determine the P-E-D matrix of KTP (or any other crystals belong to mm2 point group) is shown in Fig.1. The determined P-E-D matrix of flux grown KTP is shown in table I. Thermal dependence of the following properties of KTP and its isomorphs were obtained: expansion coefficients (α_{ij}), elastic constants (τ_c) (see tables II and III). Both $RbTiOPO_4$ (RTP) and $RbTiOAsO_4$ (RTA) exhibit large negative thermal expansion along the z axis (c axis).

Four z-cut samples grown by different techniques were used to determine the effects caused by various crystal growth techniques. The various growth techniques are: 1) low temperature hydrothermal (LTHT) 2) High temperature hydrothermal (HTHT) 3) flux by Ferroxcube (FF) 4) flux by Chinese Beijing (CF). These four

0-7803-0476-4/92 $3.00 © 1992 IEEE

different techniques have resulted crystals possessing different physical properties such as different ionic conductivities, optical damage threshold and dielectric losses [4]. The magnitude of ionic conductivity (undoped) scatters from 10^{-6} $(ohm\text{-}cm)^{-1}$ for Chinese Beijing flux grown KTP to 10^{-9} $(ohm\text{-}cm)^{-1}$ for low temperature hydrothermal grown KTP[4]. The measured electromechanical coupling coefficients of these four different samples are shown in table IV. The major electromechanical coupling coefficients and elastic constants of KTP's isomorphs such as $RbTiOPO_4$ (RTP), $RbTiOAsO_4$ (RTA), and $KTiOAsO_4$ (KTA) were also measured (see table V).

Three computer simulation programs were written to assist in the investigations of the coupling coefficient as a function of crystal orientation, anharmonicity of piezoelectric resonances and determination of four bulk constants. Figures 2 and 3 show the coupling coefficients of KTP vary with respect to crystal orientations.

Curie temperature and dielectric response of $KTiOPO_4$, $RbTiOPO_4$, $RbTiOAsO_4$, $KTiOAsO_4$, $CsTiOAsO_4$, Ba:KTP, and Ga:KTP were measured with electrical impedance analysis, acoustic resonance, and optical second harmonic generation (see tables VI, VII). All the isomorphs and doped KTP exhibit lower Curie temperatures than KTP, ranging from 637°C for $CsTiOAsO_4$ to 957°C for hydrothermal grown KTP. All samples follow the Curie Weiss law. Doped KTP (Ba, Ga) and all KTP family members except CTA show a large dielectric relaxation in the temperature range 500°C - 800°C at frequencies below 100 KHz. For barium doped KTP, it is found that Curie temperature decreases with increasing doping concentration (see Table VIII). It is also noticed that even small doping concentration (< 1%) can significantly modify the dielectric (see Fig.3). Second harmonic generation study in a KTP wedged cube, with edges parallel to the three crystal axes, showed that KTP possesses a second order phase transition (see Fig.5). Possible modes for surface waves generation such as surface acoustic waves and Bleustein-Gulyaev waves have been identified and will be discussed in elsewhere.

Conclusion

KTP family exhibits a combination of relatively large bulk electromechanical coefficients, medium temperature dependence coefficients, and single acoustic wave mode generation from principle cuts which makes it attractive for may piezoelectric applications. The ability to fabricate acoustic and optical waveguides in KTP substrate by ion-exchange will lead to acousto-optic applications. Ferroelectrically, KTP exhibits a second order phase transition. A small amount of external dopants such as Ga and Ba have remarkable influences on both real and imaginary part of the KTP dielectric permittivity.

References

[1] J. F. Nye, *Physical Properties of Crystals*, Oxford University Press, Ely House, London, 1957

[2] R. W. Whatmore, J. Cryst. Growth, **40**, 530 (1980)

[3] M. Onoe, H. F. Tiersten, and A. H. Meitzler, J. Acoustic. Soc. Amer. **35**, 36 (1967)

[4] P. A. Morris, A. Ferretti, and J. D. Bierlein, J. Cryst. Growth. **109**, 367 (1991)

Table I. The measured P-E-D matrix of flux grown KTP crystal

Elastic Tensor (N/m^2)

1.7e11	0.3333e11	0.4e11	0	0	0	0	0	0.5428
0.3333e11	1.74e11	0.35e11	0	0	0	0	0	0.707
0.4e11	0.35e11	1.495e11	0	0	0	0	0	1.636
0	0	0	0.5919e11	0	0	0	0.403	0
0	0	0	0	0.5454e11	0	0.332	0	0
0	0	0	0	0	0.4313e11	0	0	0
0	0	0	0	0.332	0	$11.44\epsilon_0$	0	0
0	0	0	0.403	0	0	0	$11.47\epsilon_0$	0
0.5428	0.707	1.636	0	0	0	0	0	$15.47\epsilon_0$

Piezoelectric Tensor (C/m^2) Dielectric Tensor

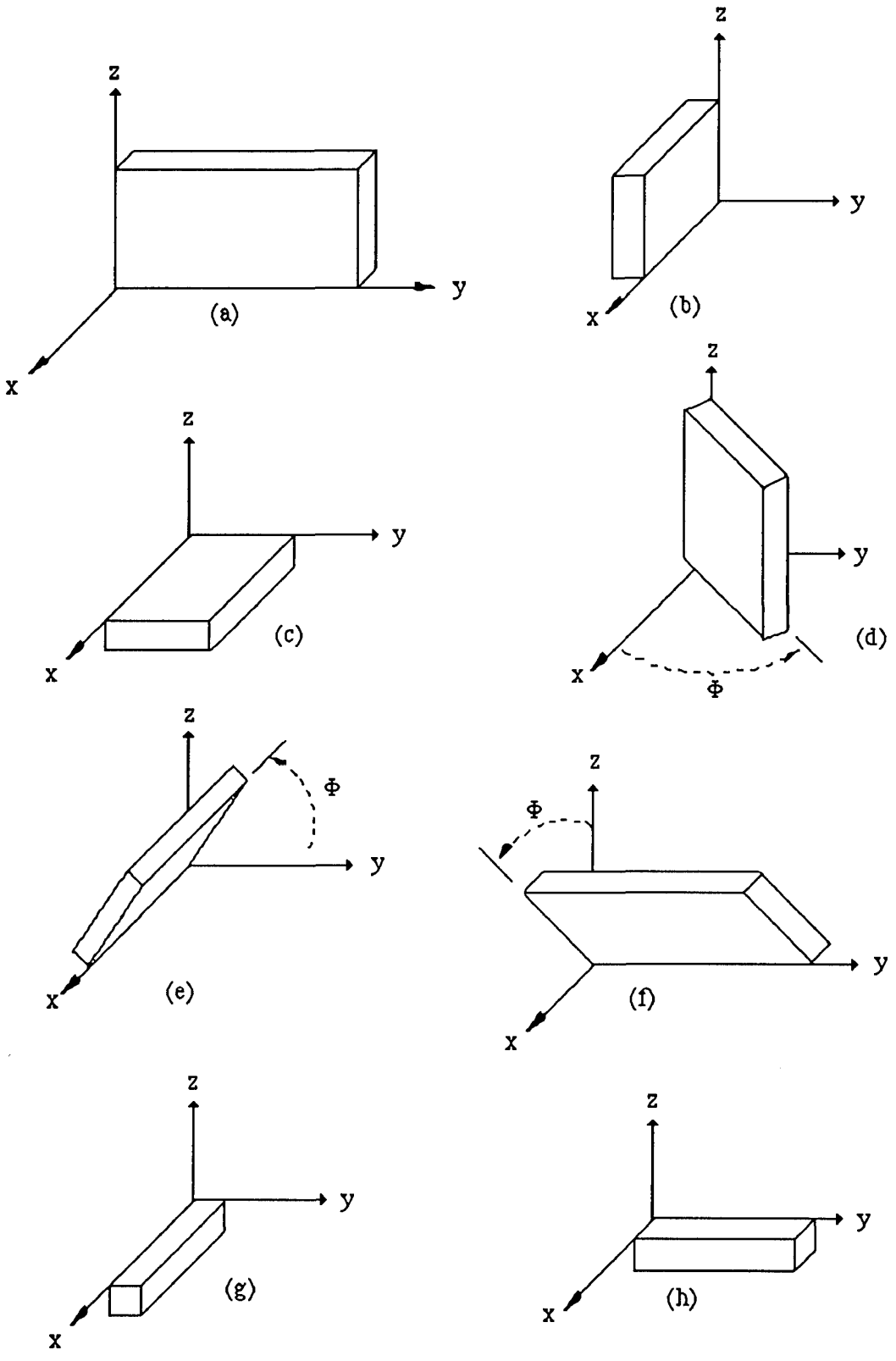


Fig.1 The set of crystal cuts to completely determine the piezoelectric, elastic, and dielectric matrix of KTP crystal

Table II. Thermal expansion coefficients of KTP and its isomorphs (20°C - 100°C)

Crystal\α_{ij} (ppm/°C)	α_{11}	α_{22}	α_{33}
KTP	6.8	9.6	-1.3
KTA	7.0	11.8	0.3
RTP	10.8	13.3	-5.9
RTA	15.1	8.5	-7.0

Table III. Comparison of temperature dependence coefficients of major elastic constants between KTP and its isomorphs.

Crystal	τ_{c33} (ppm/°C)	τ_{c44} (ppm/°C)	τ_{c55} (ppm/°C)	τ_{c66} (ppm/°C)
KTP	-220	-134	-143	-100
RTA	-190	-154	-100	NA
RTP	-118	-166	-122	NA

Table IV. The electromechanical coupling coefficients for KTP crystals grown by different growth techniques

	LTHT	HTHT	FF	CF
Ionic Conductivity	Low ($\approx 10^{-9}\ \Omega^{-1}cm^{-1}$)	Medium	High	Highest ($\approx 10^{-6}\Omega^{-1}cm^{-1}$)
k_t	38%	36%	35%	33%
k_{15}	20%	NA	NA	14%
k_{24}	21%	NA	NA	16%

Table V. Some major piezoelectric coupling coefficients and elastic constants of KTP and its isomorphs (RTP, RTA, KTA)

	k_t (%)	k_{24} (%)	k_{15} (%)	c^E_{33} (N/m^2)	c^E_{44} (N/m^2)	c^E_{55} (N/m^2)
KTP	34	16	14	1.5×10^{11}	5.5×10^{10}	4.3×10^{10}
RTP	31	7	8	1.4×10^{11}	5.6×10^{10}	7.5×10^{10}
RTA	38	5	9	1.2×10^{11}	4.4×10^{10}	4.6×10^{10}
KTA	34	NA	NA	1.2×10^{11}	NA	NA

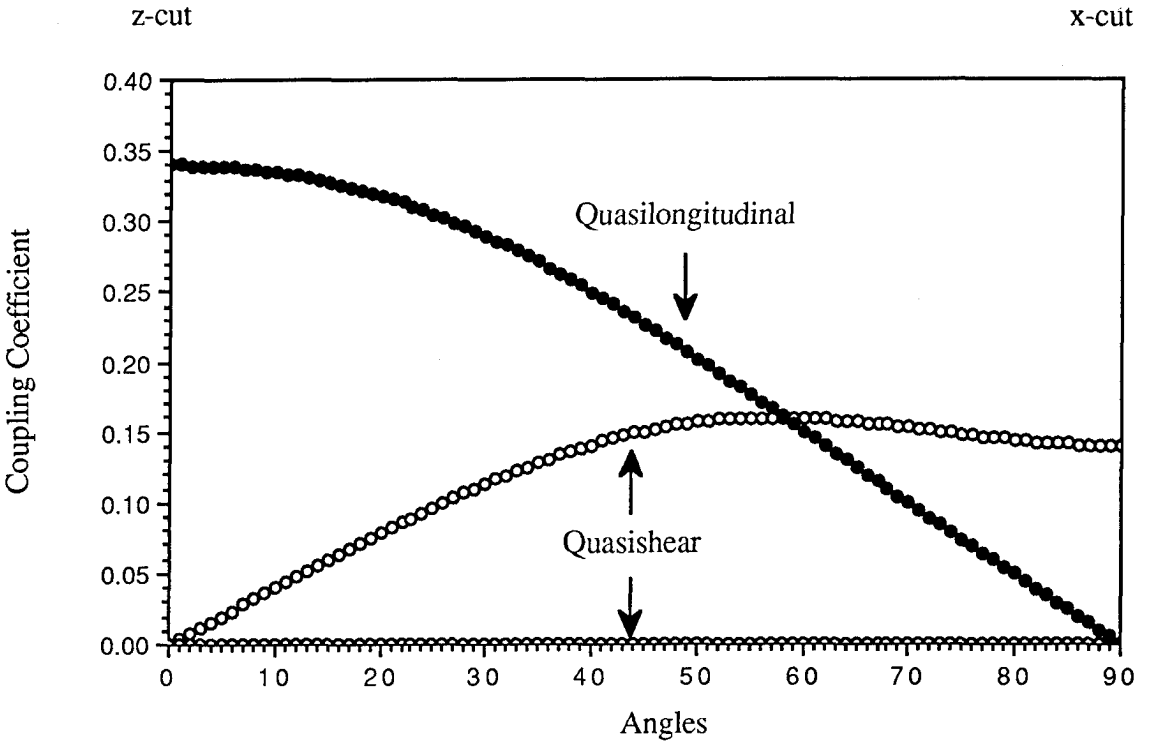


Fig. 3 Coupling coefficients vs angle of rotation for plates rotated about the y axis.

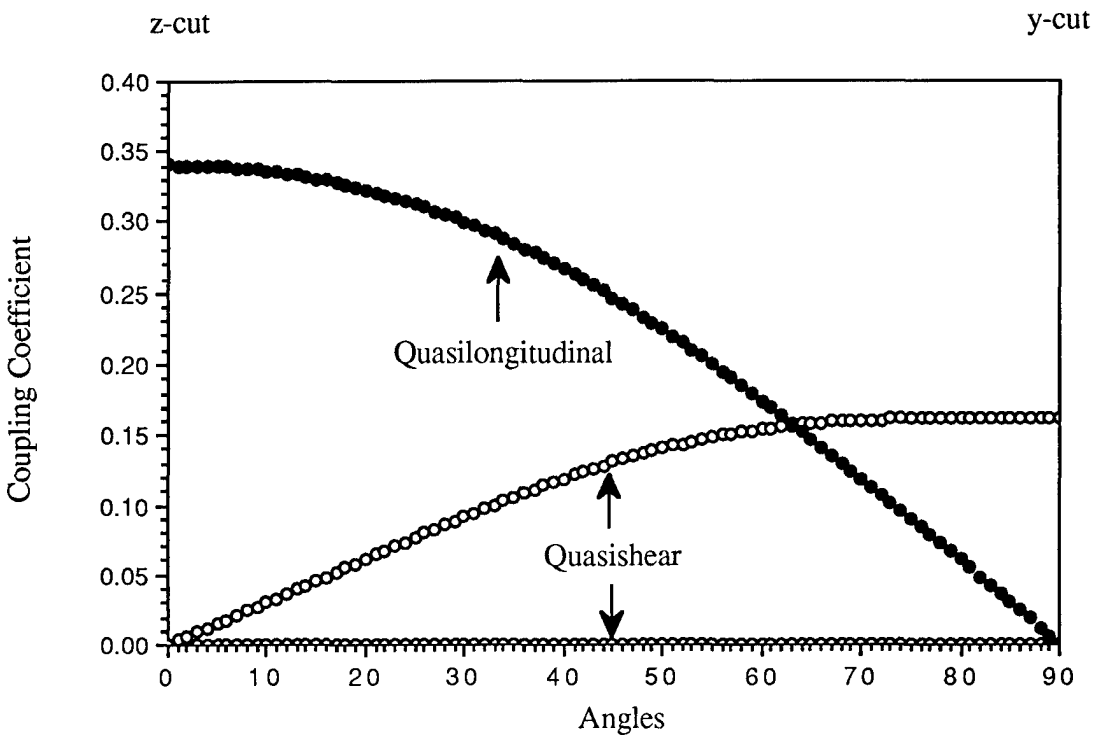


Fig. 3 Coupling coefficients vs angle of rotation for plates rotated about the x axis.

Table VI. Curie temperatures of KTP and its isomorphs (±1°C) measured at 10 KHz

Crystal	Curie Temperature (T_c)
KTP (HT)	955°C
KTA	873°C
RTP	785°C
RTA	750°C
CTA	637°C

Table VII. Barium concentration analysis of different composition baths.

Bath Composition	Barium mole percent (x)
Untreated HTHT KTP	~0
3% $BaNO_3$, 97% KNO_3	0.14
6% $BaNO_3$, 94% KNO_3	0.33
12% $BaNO_3$, 88% KNO_3	0.95
20% $BaNO_3$, 80% KNO_3	1.87

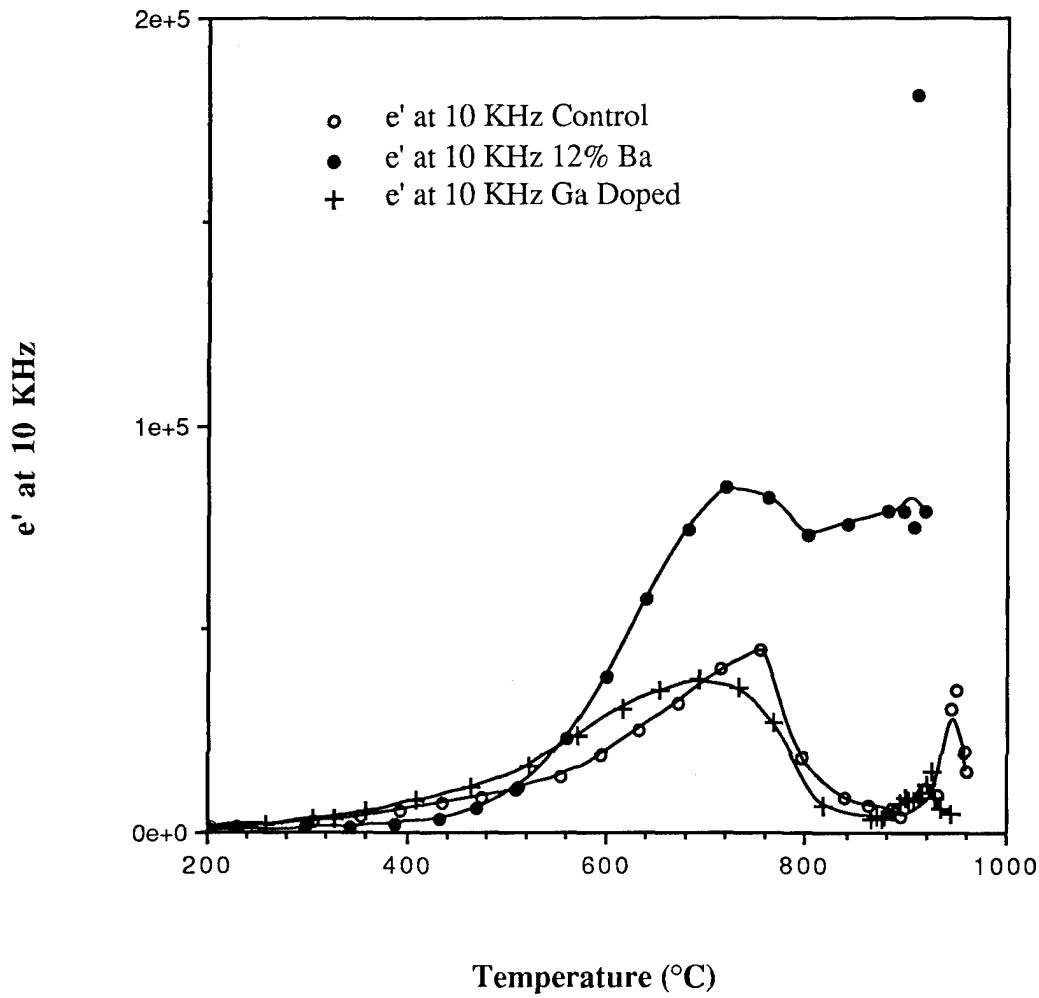
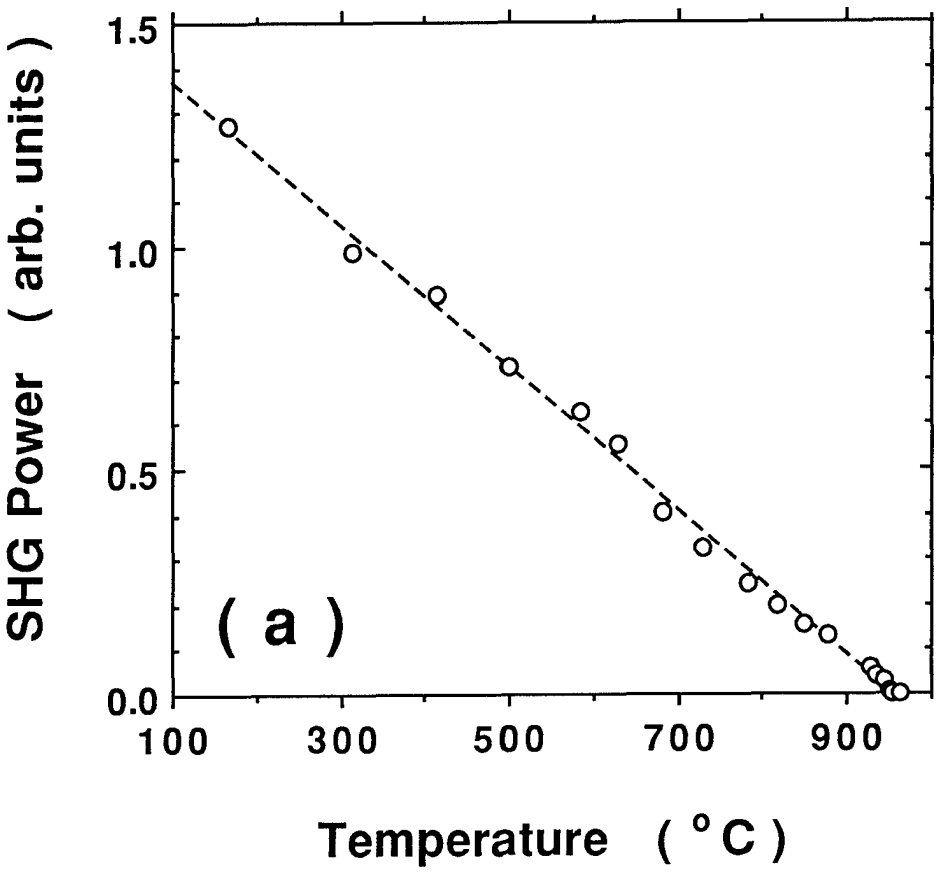
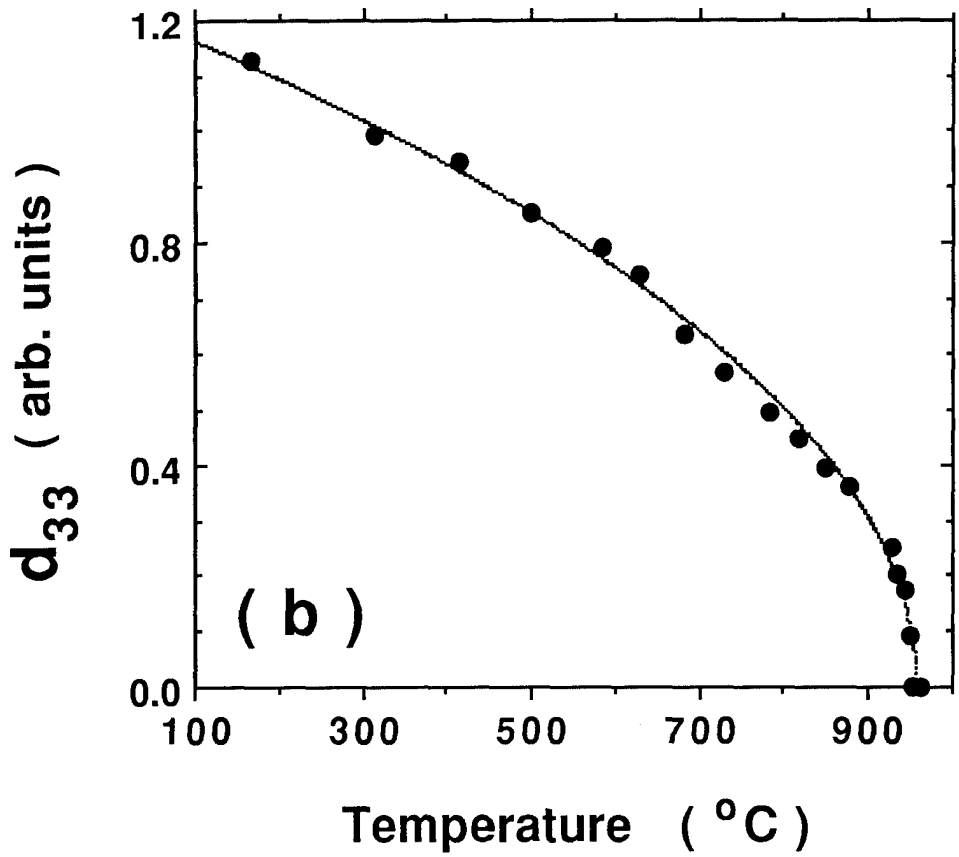


Fig.4 Influences of dopants (Ga, Ba) on real part of dielectric constant as a function of temperature at 10 KHz.


SHG Power (arb. units)
1.5
1.0
0.5
0.0
100
300
500
700
900
(a)
Temperature (°C)

(a)



(b)

Fig.5 (a) Second-harmonic power as a function of temperature. (b) Nonlinear optical coefficient d_{33} as a function of temperature.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

NEW APPROACH OF CRYSTAL GROWTH AND CHARACTERIZATION OF A QUARTZ AND BERLINITE ISOMORPH : $GaPO_4$

E. Philippot[1], A. Ibanez[1], A. Goiffon[1], A. Zarka[2], J. Schwartzel[3] and J. Détaint[3]

1 LPCMS, CNRS-URA D0407, case 003, F-34095 Montpellier cédex 5, France
2 Lab. Minér. Cristallog., Univ. Paris VI, F-75252 Paris cédex, France
3 CNET, PAB/BAG/MCT, F-92220 Bagneux, France

Abstract : This investigation concerns $GaPO_4$ crystal growth in sulphuric and phosphoric acid media through the slow heating and vertical reverse temperature gradient methods. Systematic study of growth parameters shows a Vx growth rate always much greater than the two other ones, Vz and Vy. The most interesting result is the ability to do $GaPO_4$ epitaxy on large berlinite seeds in sulphuric acid and, after that, to use them for $GaPO_4$ crystal growth in phosphoric acid. The good epitaxy process and the crystalline quality have been checked by X-ray topography. The -OH content, followed by infrared spectrometry, seems to prove a lower OH concentration than in the berlinite case of approximatively the same crystal growth conditions. First piezoelectric characterizations for resonators near the AT-cut show $GaPO_4$ to be a very promizing piezoelectric material with a large coupling coefficient near 16%, a quartz like thermal stability and a Q factor already nearly sufficient for the applications.

Introduction

Since the work of Chang and Barsch [1], many works have been devoted to growing and characterizing, first berlinite and recently $GaPO_4$ crystals, and to modelling and experiment of bulk and surface wave devices made with these materials. It was successively shown that these materials were close piezoelectric analogs of quartz, and possess similar cuts with zero frequency temperature coefficients, either for bulk or for surface wave applications [2-8].

The crystal growth investigation of this material is justified, at least, for two reasons, Table 1. Indeed, for the same AT cut, first results obtained for $GaPO_4$ crystals [9] [10] seem to indicate a value of coupling coefficient ($k>0.14$) higher than those of quartz (0.085) and berlinite (0.11).This induces the possibility to obtain larger band-pass of filters. On the other hand, there is no α---β quartz transition for this material and, then, it can be used in a large range of temperature, T>1000°C, for filter and sensor devices. We present here new crystal growth approachs for $GaPO_4$ in connection with some results of scanning electron microscopy, infrared spectrometry, X-ray topography investigations and resonator measurements.

Crystal Growth Studies

Powder and nutrient synthesis.

Gallium phosphate being not commercially available, it has been synthesized from gallium hydroxide with modest purity or gallium 6N metal. In the first case,the reaction :

$$H_3PO_4 + GaO(OH) \longrightarrow GaPO_4 + 2\,H_2O$$

Table 1: Comparison of some quartz, berlinite and $GaPO_4$ physical characteristics

	Quartz	Berlinite	$GaPO_4$
Coupling coefficient k % for AT-cut	8.5	11.0	>14.0
$\alpha \rightleftharpoons \beta$ transition °C	573	586	no

is carried out with an excess of phosphoric acid. As berlinite, gallium phosphate shows a retrograde solubility in all acid media, figure 1. The crystallization is obtained in tantalum lined autoclave. To eliminate the major part of the initial impurities, it is needful to use a re-crystallization experiment. Operating conditions are : solution 15M in H_3PO_4 saturated with gallium hydroxide, filling 80%, final temperature near 240°C. The heating rate is chosen following the wanted granularity : the slower one (# 20°C/h) for nutrient synthesis and the faster one (#100°C/h) to obtain fine powder, useful for growth solution preparations.

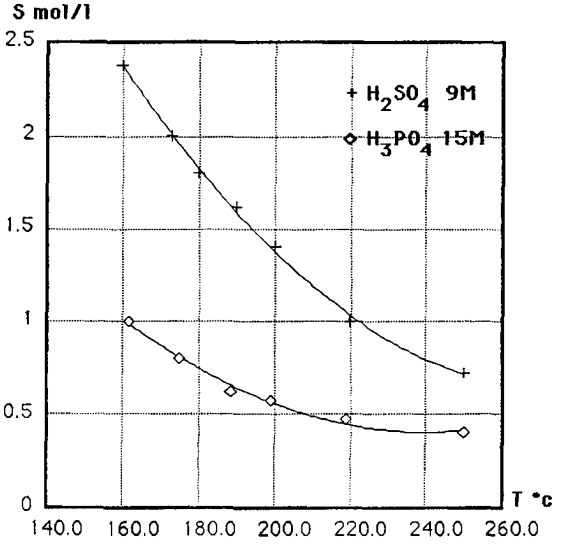


Figure 1. Berlinite solubility in sulphuric and phosphoric acids.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The second reaction needs two steps :

$$Ga + 4\,HNO_3 \longrightarrow Ga(NO_3)_3 + NO + 2\,H_2O$$

$$Ga(NO_3)_3 + H_3PO_4 \longrightarrow GaPO_4 + 3\,HNO_3.$$

The nitric solution of $GaPO_4$ so obtained is then evaporated by moderate heating and, in this case, finely powdered gallium phosphate can be obtained with an yield higher than 90% and a good purity.

GaPO4 crystal growth

The $GaPO_4$ crystal growth has been carried out through the three methods already used for berlinite [2] [8] :
- the slow heating method (SHT);
- the horizontal composite gradient method (HTG);
- the vertical reverse temperature gradient method (VTG).

Only new advances on crystal growth technique will be described here:
- VTG method in H_2SO_4 acid;
- SHT and composite gradient method in H_3PO_4 acid.

GaPO4 crystal growth by VTG method in H_2SO_4 acid: To obtain large seeds of gallium phosphate, it would be interesting to use directly large berlinite seeds five to eight centimeters long now available. Indeed, it must be kept in mind the continuous solid solution and the cell volume similarity for the whole pseudobinary system $Al_{(1-x)}Ga_xPO_4$, $V=(230\pm1)Å^3$ [11]. This being so, the growth study of $GaPO_4$ crystals from berlinite seeds was undertaken.

In a first step, some experiments have been carried out in phosphoric acid medium. Unfortunately, the dissolution kinetics of $AlPO_4$ in this solvent is faster than the growth restart one of $GaPO_4$, and so, berlinite seeds are dissolved.

Therefore, we have tried this crystal growth in sulphuric acid medium. Indeed, although all MXO_4 quartz isomorphs present their highest solubility in this solvent (#3 times higher than that one in phosphoric acid), a much lower dissolution kinetics has been demonstrated[12]. On the other hand, as for berlinite crystals, growth temperature must be chosen greater than 200°C to avoid major part of -OH impurity. In this case, for too high pressure purposes, we cannot use the SHT method in glass vessel, but the VTG one.

In order to determine the crystal growth conditions of $GaPO_4$ in sulphuric acid solvent, we have undertaken a systematic study in terms of :

- crystal growth temperature, Tc (°C);
- temperature difference between cold and hot zone, ΔT (°C);
- acid concentration, M ($mol.l^{-1}$).

We present here the first results in sulphuric acid solutions taking into account the following conditions :

- X and Z berlinite seeds;
- autoclave filling #80%;
- acid sulphuric solution, 6M, saturated in GaPO4 at the chosen growth temperature, Tc;
- growth temperature ranging from 200 to 240°C;
- vertical temperature difference (or temperature gradient): $1<\Delta T<6$°C;
- boring ratio of the baffle between cold and hot zones : 7,5%;
- nutrient mesh : 2-5 mm.

Although the experiments are still in progress, some interesting results can be emphasized. The most important features are :
- the corroboration of the crystal growth restart in $GaPO_4$ from berlinite seeds;
- the confirmation of kinetic results for berlinite controlled dissolution in both phosphoric and sulphuric acid media.

On figure 2 are represented the growth rate results as a function of ΔT at Tc= 210 and 235 °C. As for berlinite crystal growth, Vy growth rate is always close to 0.005 mm/d.face. On the other hand, another growth rate, Vz, is very weak, $0.01<Vz<0.02$ mm/d.face and independant of Tc value.

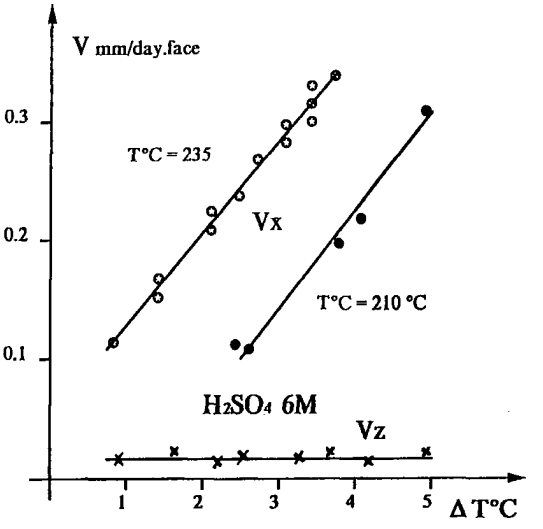


Figure 2. Vx and Vz growth rates for Tc= 210 and 235 °C in sulphuric acid medium , 6M.

This Vx and Vz growth rate disparity explains the strongly different morphology of crystals obtained from X or Z berlinite seeds. From a general point of view, from Vx berlinite seeds, all crystals are cracked along a preferential direction parallel to (1012) planes, during the air cooling of autoclaves at the end of the crystal growth. Photo 1 compares such a crystal with its crystal structure: the planes (1012) correspond to the maximum atom density of MXO_4 packing. These cracks are due to strong strains between $AlPO_4$ and $GaPO_4$ structures. Indeed the GaO_4 tetrahedra are larger than the AlO_4 ones for the same cell volume and, then, explain these strong strains.

The most interesting results are obtained from Vz seeds. In this case, Vx being the only significant growth rate, flat crystals are obtained with a very thin epitaxy of $GaPO_4$ in the Z direction. Photos 2 and 3 give an idea of such crystal at Tc=210 and 235 °C. At lower temperature, berlinite seeds are less dissolved than at the higher one. Both kinds of seeds can be used for further growth experiments in phosphoric acid medium, where $Vz \neq 0$, through the SHT and composite gradient methods.

GaPO4 crystal growth by SHT and composite gradient method in phosphoric acid: The SHT method has been also used for crystal growth from nucleations, but only the experiments using seeds obtained from the previous large Z

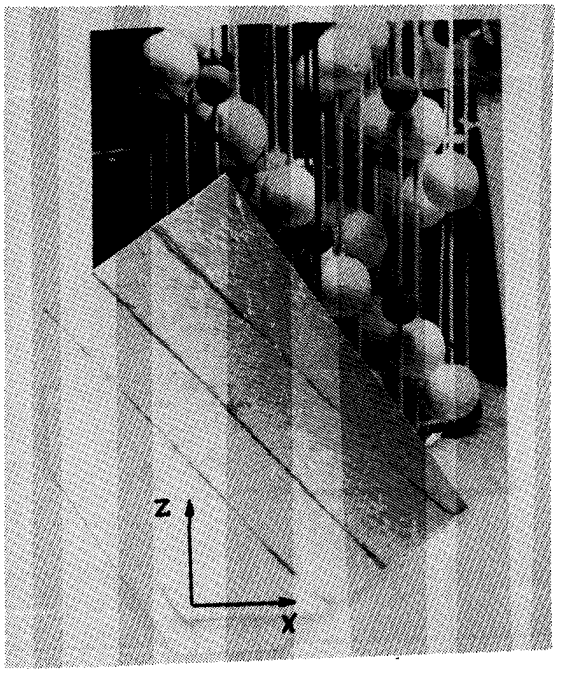


Photo 1. Cracked GaPO4 crystal from berlinite X-seed with its schematic crystal structure.

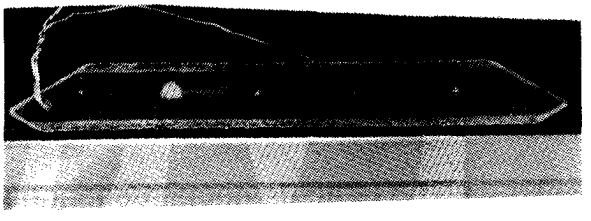


Photo 2. Flat crystal obtained from berlinite Z-seed at Tc=210°C in sulphuric acid medium, (VTG method).

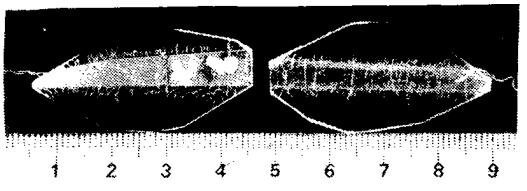


Photo 3. Flat crystal obtained from berlinite Z-seed at Tc=235°C in sulphuric acid medium, (VTG method).

crystals will be described. The horizontal glass autoclave, containing the seeds, is filled with a quasisaturated solution. A symmetrical horizontal temperature gradient is maintained during all the experiment, fig.3. In these conditions, solution circulation is favoured and avoids gradient concentration near the growing crystal interface. In this modification, 150 cm^3 horizontal glass vessels are used. Three experimental growth conditions have been chosen with some identical characteristics:

- berlinite seeds shielded against dissolution by thin epitaxy of GaPO4;
- filling # 80%;
- acid phosphoric solution saturated with GaPO4 at the initial temperature;
- horizontal temperature gradient # 5°C.

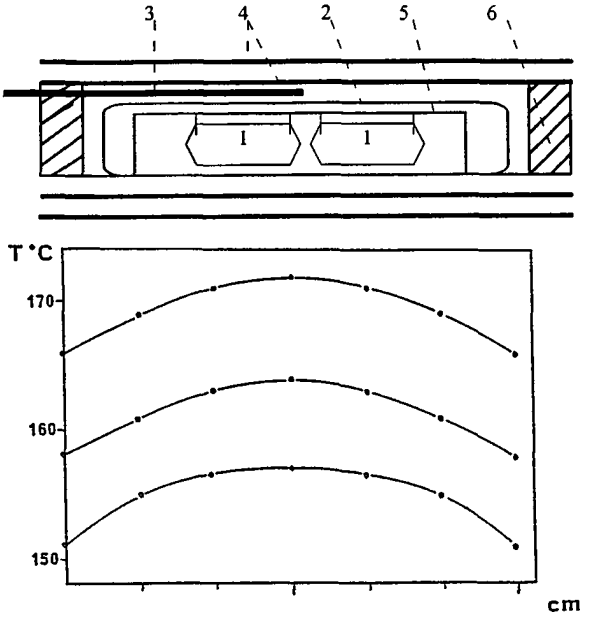


Figure 3. Schematic representation of the glass tube autoclaves, and their temperature profile, used for SHT method : 1 = seeds, 2 = glass vessel, 3 = regulation and control thermocouples, 4 = resistance-wound furnace between two transparent silica tubes, 5 = platinum rod, 6 = PTFE plugs.

The first possibility is the well known SHT method where the temperature is slowly increased, 0.5 to 2 °C/day, from #150 to 180 °C. The major drawback of this method is the impossibility of obtaining crystals of sufficient size in one operation due to the limited quantity of gallium phosphate available in the solution.

Then we have developed a "composite" gradient method with nutrient in the colder side of the glass autoclaves, fig.4.

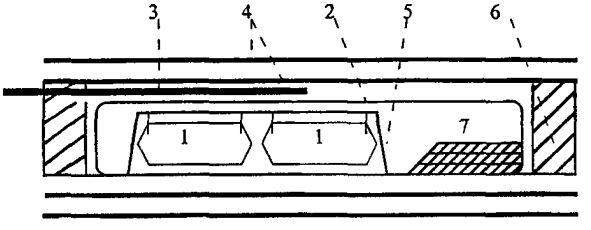


Figure 4. Same experimental device as shown in figure 3 with nutrient, 7, in the coldest side of the glass vessel.

The GaPO4 enriched solution in this part is swept along by the horizontal temperature gradient towards the seeds in the growth zone. At this step, two possibilities have been checked for the crystal growth:

- only fluid circulation from cold to hot zone due to the horizontal temperature gradient at constant T°C;
- fluid circulation coupled to a slow heating rate of 0.25 to 1.0 °C/day, fig.5.

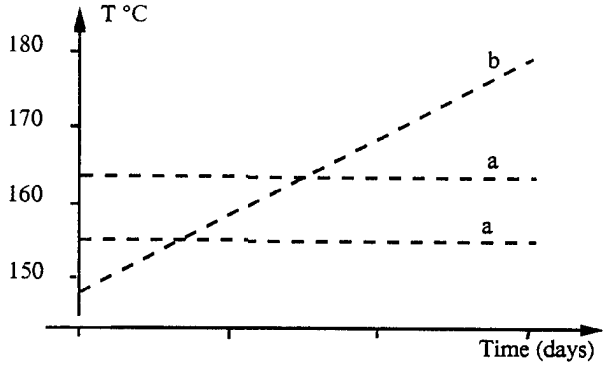
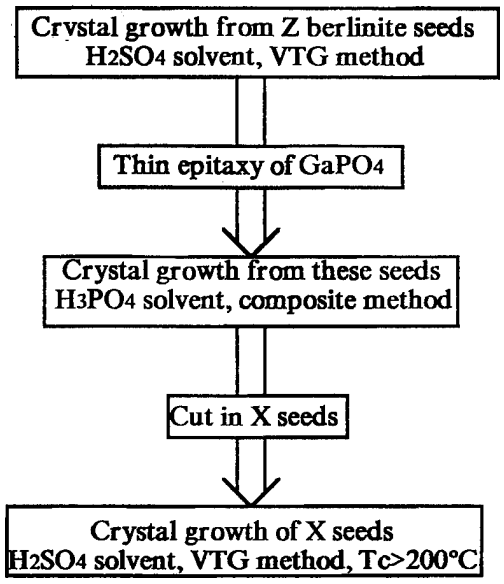


Figure 5. Schematic representation of crystal growth experiments by "composite" gradient method:
a) - constant temperature;
b) - slow temperature heating.

In both cases, quite good results are obtained for crystal growth experiments and the major drawback of the SHT method is avoided. One example is given in photo 4 where the seed is the flat crystal shown in photo 2. Whereas the Vz growth rate becomes significant, its value remains weaker than the Vx one. We observe that the Vx/Vz growth rate ratio is approximatively reverse to that of berlinite, 1/3 for $AlPO_4$ and 3/1 for $GaPO_4$.

Nevertheless, the Z-thickness is adequate to cut, in these crystals, Z-seeds which can be used in sulphuric acid, through the VTG method at higher temperature, to give crystals with low -OH content, as it is summarized below.



Several alternatives to this crystal growth process, using successively VTG method in sulphuric acid and SHT or composite method in phosphoric acid are in progress.

At this step, some preliminary characterizations have been undertaken to check :
- the M (Al,Ga) concentration evolution by scanning electron microscopy;
- the -OH content by infrared spectrometry;
- the crystalline quality by X-ray topography;
- the piezoelectric characteristics by resonator device measurements.

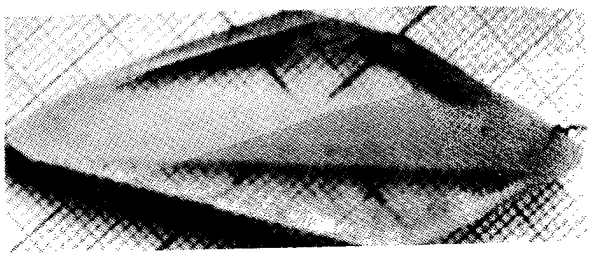


Photo 4. $GaPO_4$ crystal obtained in phosphoric acid medium through the SHT method between 150 and 180 °C.

Characterizations of $GaPO_4$ Crystals

M(Ga,Al) concentration evolution by scanning electron microscopy.

Photos 5 give the X-ray emission profiles, characteristic of gallium and aluminium Kα line, of a plate sawn perpendicular to the berlinite seed. It can be observed, in both cases, strong discontinuities on a level with growth restart and, then, no aluminium diffusion into the $GaPO_4$ growth part.

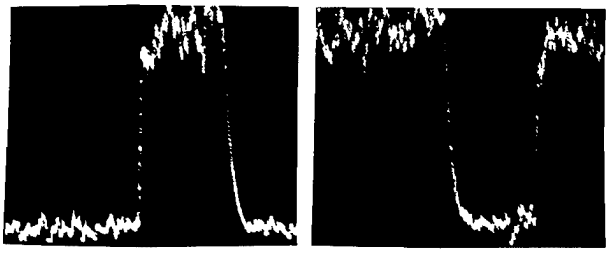


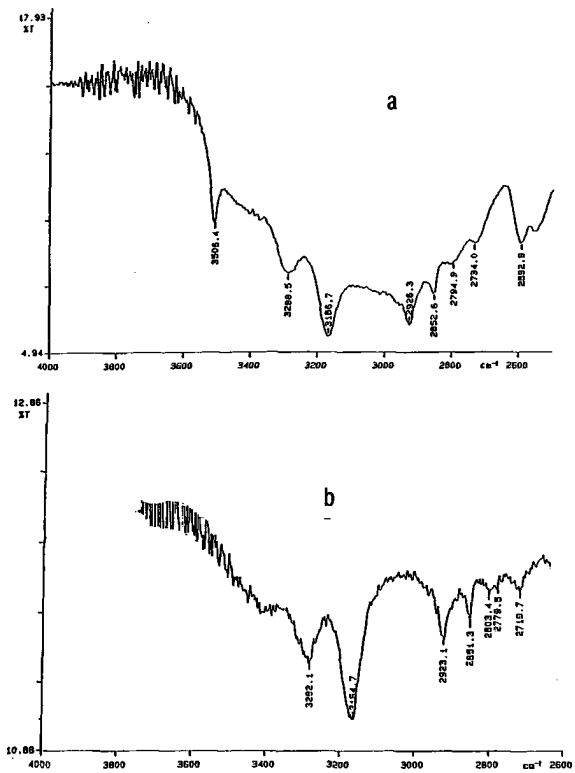
Photo 5. X-ray emission profiles of a plate sawn perpendicular to the berlinite seed: a = Al Kα line, b=Ga Kα line.

-OH content determination by infrared spectrometry.

This method, as for quartz and berlinite, is the most appropriate one for the -OH content determination and can be related to the surtension coefficient. Figures 6 give an idea of transmittance infrared spectra obtained from our crystals, in the range 4000-2600 cm^{-1} characteristic of -OH absorption. The spectrum of Fig. 6a is representative of low temperature crystal growth, SHT method from 150 to 180°C, whereas in Fig. 6b is shown that of higher temperature crystal growth, VTG method above 200°C.

In both cases, we have calculated the absorption coefficient, α, from the expression [13] :

$$\alpha = 1/d_{cm} \log(T_{3800}/T_{3167})$$



Figures 6. Infrared transmission of GaPO4 samples :
a = crystal from SHT method, 150<T<180°C;
b = crystal from VTG method, Tc>200°C.

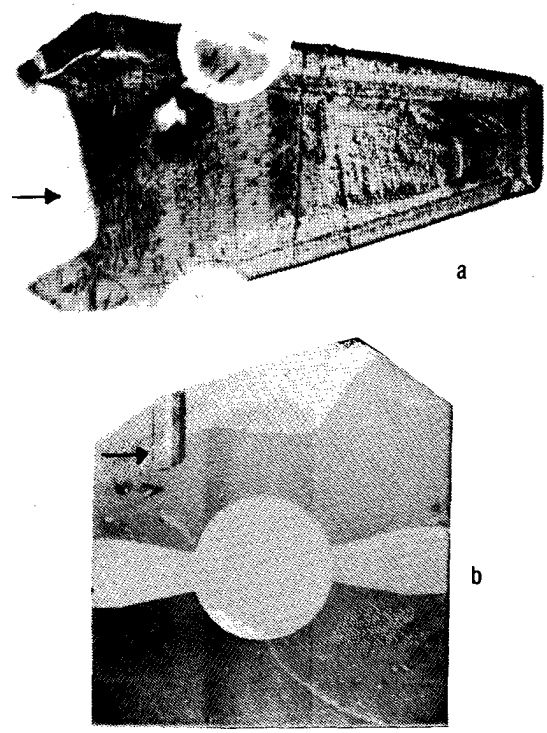
where d is the sample thickness in cm and where the strongest absorption band at 3167cm^{-1} is probably due to the Ga-OH one (for berlinite, it was used the Al-OH band at 3291 cm^{-1}). For the low temperature sample, Fig. 6a, the α value is 1.78 whereas it becomes 0.04 for the highest temperature one, Fig. 6b. Therefore, it seems to exist a strong difference between berlinite and gallium phosphate behaviour. Indeed, if we compare these α values to those ones for berlinite, for approximatively the same crystal growth conditions, it can be observed a diminution of a factor 10 for gallium phosphate [8]. This result presents, of course, a major interest, but must be confirmed.

Characterization of growth defects by X-ray topography.

Several plates of GaPO4 were studied by X-ray transmission topography using the white beam delivered by the DCI storage ring at LURE (Orsay, France) or by conventional X-ray apparatus.

In photos 6 are given typical examples of the observation made using etched plates from crystals grown through the SHT method. The cut of photo 6a contains one isolated electrical twin (Dauphiné law) probably induced by the fixation of the seed. The right part of this plate is of quite good quality with some dislocations at the growth restart. The plate represented in photo 6b contains one small optical twin (Brazil law). Twinning was much more frequently found in the first crystals obtained by successive growth of small nucleations than in recent crystals obtained by the new growth methods starting with high quality berlinite seeds, where only small isolated twins of the Brazil type can be observed in some rare instances.

Z plates obtained by VTG method in sulphuric acid medium from berlinite seeds were most often observed to be twin free. In this topograph (photo 7) a large contrast, corresponding to a highly distorted region existing between the berlinite seed and the GaPO4 lateral overgrowth, can be observed. The GaPO4 growth appears to have a very good crystalline quality.



Photos 6 :

a) - X ray topograph showing an electrical twin in first grown crystals.

b) - Electroded plate showing a small optical twin (crystal of the second generation).

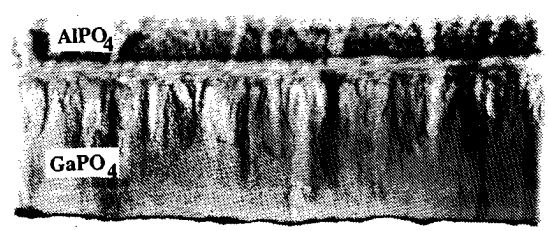


Photo 7. Topograph of a Z plate obtained by VTG method in sulphuric acid medium from berlinite seed.

Piezoelectric Characterizations.

Bulk piezoelectric wave propagation in the Y rotated cuts of GaPO4: As for quartz and berlinite, the piezoelectrically excited pure shear mode of the Y rotated cut of GaPO4 appears to be very attractive since quite simple orientations giving large coupling coefficients and a good thermal behaviour exist among these cuts. Two Y rotated compensated orientations which are of the same nature as the AT and BT cuts of quartz were reported by Krempl et al [9] [14] [15]. Several authors

have characterized $GaPO_4$ crystal using dielectric constant measurements [16]. Huart [17] and Krempl et al. [9] [14] [15] have determined several sets of elastic and piezoelectric constants of $GaPO_4$.

Using these constants, we have determined the main properties of Y rotated thickness shear resonators. For these computations we use the one dimensional model of an infinite plate with massless electrodes [18] [19]. On figures 7 and 8 are represented computed properties of the fast shear and slow shear mode as a function of the rotation angle theta. On these figures, V is the phase velocity of the mode ($V=\sqrt{C/\rho}$), k is the electromechanical coupling coefficient and Tfr, Tfr3, Tfr5, Tfa are respectively the first order temperature coefficient of the fundamental resonance frequency, of the 3rd overtone resonance frequency, of the 5th overtone resonance frequency and of the antiresonance frequency. The results given in figures 7 and 8 are mostly based on the constants given in reference [17]; the calculation of the temperature coefficients is very tentative since the available data [9] constitute an incomplete set which was completed with values measured for berlinite [20].

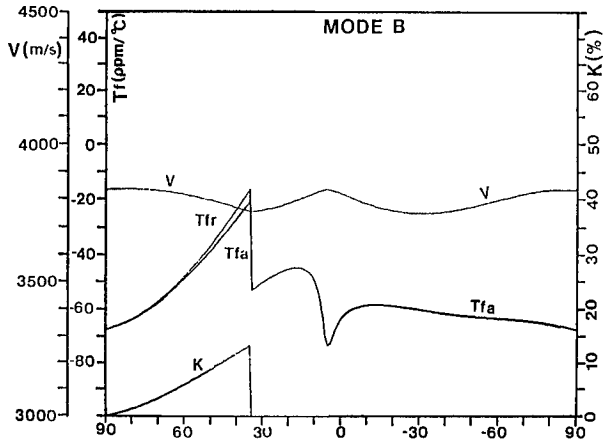


Figure 7. Computed properties of the fast shear mode.

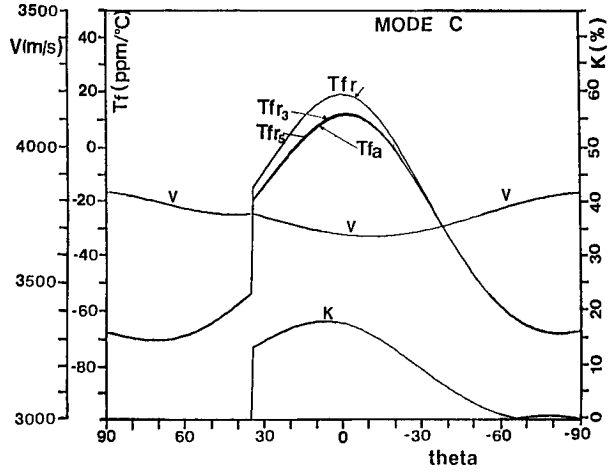


Figure 8. Computed properties of the slow shear mode.

The obtained results for V and k indicate that gallium phosphate is a material characterized by a quite small phase velocity of the shear wave and by rather large value of the coupling coefficient. Despite the some how arbitrary hypothesis made in choosing the missing temperature coefficients (TC) of the constants, the computed temperature coefficients are found to be in good agreement with all the known experimental values. Particularly the AT and BT cuts are found at angles very close to those given in [9] and [15]. Among the computed temperature coefficients, those relative to the antiresonance frequencies are the closest from the experimental values. This is probably a consequence of the formula used in [9] to extract the TC of the constants. The difference between the different temperature coefficients represented in figures 7 and 8 is probably a little too important. This result depends mostly of one hypothesis made concerning the temperature coefficients of the piezoelectric constants (supposed to have the same values that for $AlPO_4$).

Other calculations were made using the constants given in [15] and other hypothesis concerning the unknown

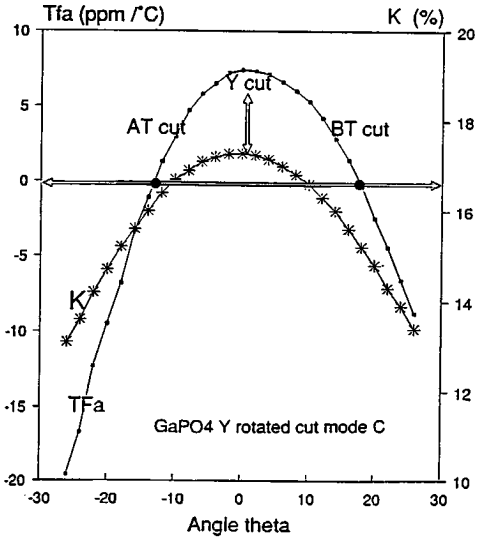


Figure 9. Computed 1st order AFTC and K for the slow shear mode in the vicinity of the AT and BT cuts.

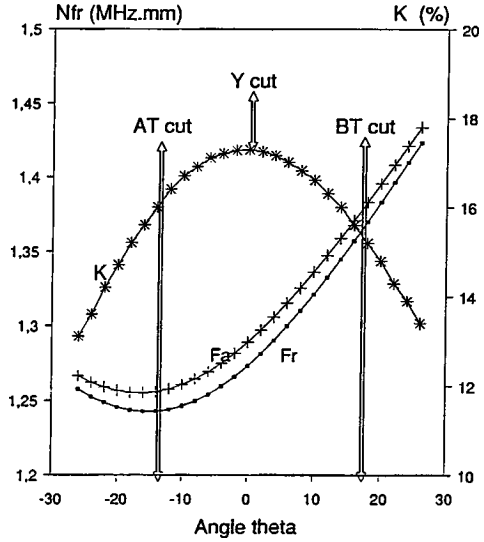


Figure 10. Computed frequency constants and K for the slow shear mode in the vicinity of the AT and BT cuts.

temperature coefficients. Two examples are given in figures 9 and 10 for the slow shear mode of cuts in the vicinity of the AT and of the BT cuts. The results are very similar to those given in figure 8. In this figure we can observe that in $GaPO_4$, the BT cut occurs at an angle for which the slow shear mode is the piezoelectrically excited mode whereas for quartz the mode of the BT cut is the fast shear one. The cut angle for the AT and the BT cuts are much smaller than those existing in quartz and in berlinite. This fact is an important advantage since the value of the coupling coefficient is then a larger fraction of the maximum possible value which, as in other quartz like materials, corresponds to the Y cut.

If we compare the properties computed for $GaPO_4$ for Y rotated cuts to those observed for the other quartz like materials, the main difference beside that concerning the coupling coefficient, is the much lower TC of the resonance frequency for the Y cut than in quartz or berlinite. As in these materials, the AT cut is situated very near a minimum of the phase velocity as a function of the angle.

Experimental results: Resonators were made from plates cut in crystals grown by the SHT method:

- by successive growth of spontaneous nucleations;
- from berlinite seeds with $GaPO_4$ epitaxy.

For these measurements, plates free of twins, or having large regions free of twins, were selected using X-ray topography and the optical examination of etch figures obtained in an acid solution (5% HF and 10% HCl), photo 6b. The plates were electroded using Cr/Au evaporated thin films with a mass loading ranging from 0.6% to 0.9%. Electrical measurements were performed using network analysers and a high precision oven monitored by a quartz thermometer.

-Resonators cut with the minor rhomboedral orientation: Plates parallel to the rhomboedral minor face (Y-21°) were obtained from one of the first crystals grown over a seed obtained from nucleation. About ten plates of sufficient size were obtained, 6 were determined as being free of twins or having sufficiently large area free of twins by etch figure examination. The electrical response of this kind of resonator is very clean and similar to the previously published results, nevertheless, the Q factor is still quite modest (5000 at 5 MHz). The thermal variations of the resonance frequencies of some of these resonators are displayed in figure 11. Important variations of the Q factors as a function of temperature were observed during this study (see figure 14 below). This fact is to be related to the large OH concentration existing in this type of crystal [8].

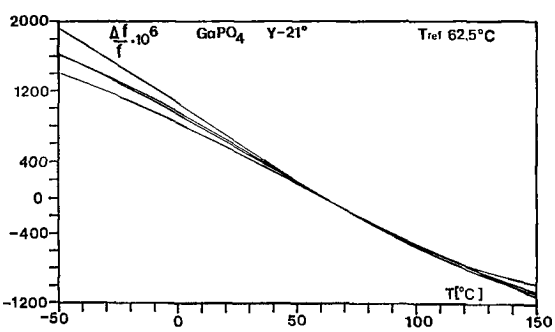


Figure 11. Thermal variations of the resonance frequencies of resonators made from the first generation GaPO4 crystals.

-Properties of resonators near the AT cut: About twenty plates having two orientations around AT (Y-15° and Y-12°30') were obtained from a crystal more recently grown from a seed obtained by lateral epitaxy over berlinite using improved growth conditions. In figure 12 is represented a typical electrical response observed for these resonators. The Q factor of this resonator is 10000 at room temperature, which corresponds to about 12000 at the usual reference frequency of 5 MHz (with the probably justified hypothesis of 1/f variations of Q). A lower dispersion than for the Y-21° cut was observed on this quantity between resonators. The large value of the relative difference between the resonance and antiresonance frequencies can be directly related to the value of the coupling coefficient which was found very near 16%. In many instances, unwanted responses corresponding to antisymmetrical anharmonics, were observed to appear between the resonance and the antiresonance. A detailed examination of this phenomena has allowed to retain only two possible explanations: a greater sensibility to defect of parallelism than for quartz or the existence of in plane variations of the acoustical properties of the material. The second explanation hold much probably for resonators having the electrodes deposited on a region close to the seed.

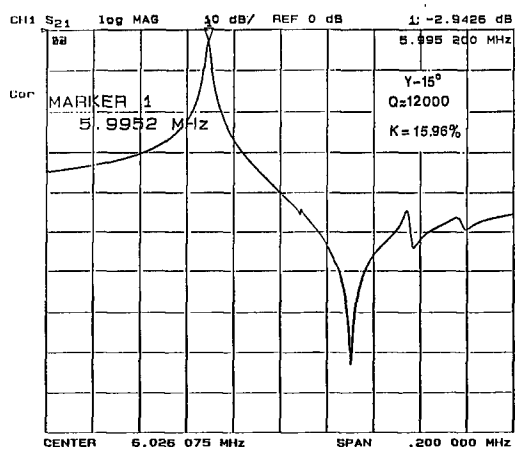


Figure 12. Typical electrical response observed for resonators from crystals grown from berlinite seeds.

On figure 13 are represented the thermal variations of the resonance frequencies observed for resonators of the two cuts near AT. The dispersion of the first order temperature coefficient is in both case of the order of 1 ppm/°C at the reference temperature.

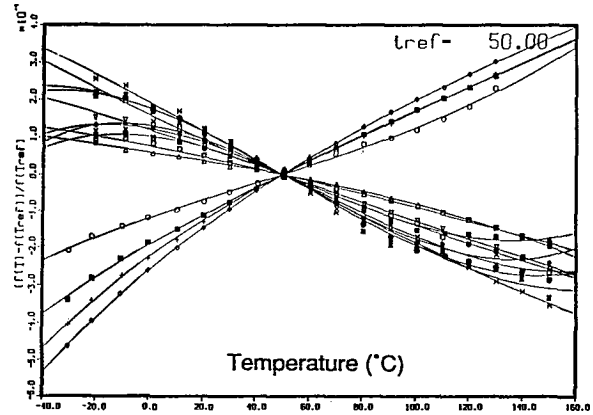


Figure 13. Thermal variations of the resonance frequencies of the second generation GaPO4 crystals.

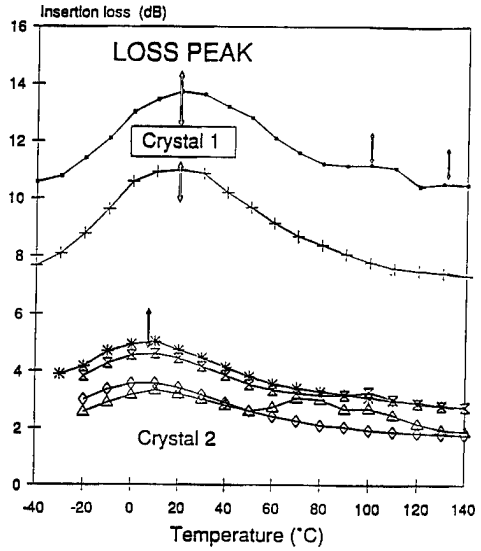


Figure 14. Compared thermal variations of the insertion losses measured for the two types of crystals.

On figure 14 we have compared the thermal variations of the insertion losses measured for both crystals used. In this figure we can observe a large difference of behaviour between these two crystals. For the crystal grown from nucleation, an important loss peak appears in the vicinity of the room temperature whereas the second crystal displays a much smaller and more constant resistance as a function of temperature ($R=100(10^{A/20}-1)$ where A is the insertion loss). This situation was already encountered for berlinite crystals for which the existence of a loss peak at room temperature was related to the OH concentration [8] [21]. A typical evolution of the resonance curve (near the resonance frequency) with temperature is representated in figure 15. We can observe that the insertion loss decreases when the temperature is raised so that the Q factor is much larger at high temperature that at room temperature. The response curve obtained at 150°C for the same resonator that in figure 12 is represented in figure 16.

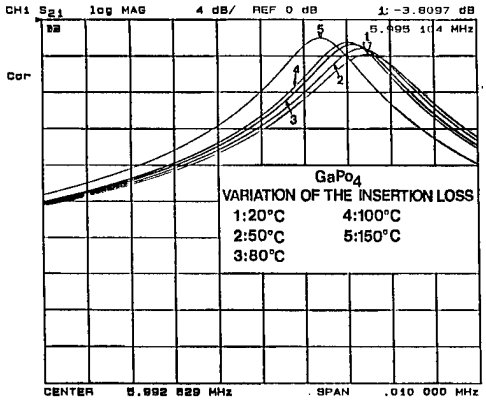


Figure 15. Typical evolution of the resonance curve with temperature.

Comparison of experimental and computed results, discussion: Measured and computed results are compared in Table 2. In this table, a generally good agreement is found between experimental and computed values. It must be

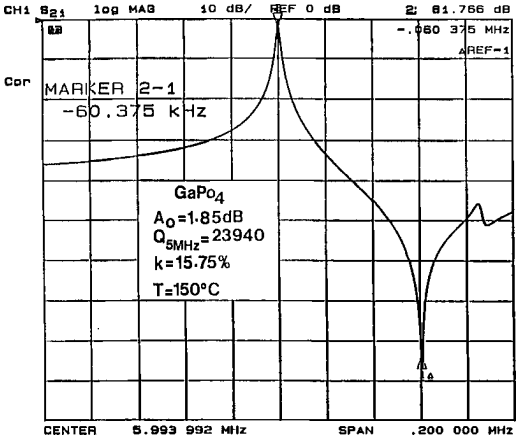


Figure 16. Electrical response at 150°C of the same resonator shown in figure 12.

Table 2: Computed (C) and experimental (E) values of some GaPO4 resonator characteristics.
* Values corrected of the mass loading
** maximum value observed for this cut

cut	NFr^{x} kHz.mm	NFa^{x} kHz.mm	K^{xx} %	TFr ppm/°C	TFa ppm/°C	Q at 5MHz
E Y-21	1258.3	1266.4	14.2	- 14.1	-15.4	5000
C Y-21	1246.9	1257.7	14.5	-	-11.7	
E Y-15	1261.0	1274.1	15.97	-2.39	-3.66	12800
C Y-15	1242.5	1255.4	15.86	+1.90	-2.13	
E Y-12.5	1255.1	1269.7	15.98	+3.37	-	10100
C Y-12.5	1243.6	1257.3	16.31	+4.48	+2.60	

emphasized that the computed values of the coupling coefficient are attained in practically all experimental cases without any corrections for the existing parasitic capacitance of the resonators. The experimental Q factors, which are already nearly sufficients for the applications for the second crystal, underestimate probably by a factor 2 to 3 the material intrinsic Q factor that can only be measured using planoconvex overtone resonators. Somehow, higher values were measured for the overtones modes of the present plane resonators. An important observation can be made relatively to the experimental frequency constants (Nfa=V/2): they are always higher than the computed values (the accuracy of this measurement is better than 0.1%); this fact is probably significative of a much lower OH concentration in the present samples than in those used for the constant measurements. It can be observed that this difference is greater for the most recent crystal for which a greater Q factor was measured. A similar augmentation of the elastic constants was already observed with berlinite when the OH concentration was reduced by modification of the growth conditions. A more detailed examination of the measurement results concerning the frequency constants and the temperature coefficients has revealed a small dispersion between resonators (about ±3kHz.mm or ±1.10-6 /°C for the second crystal, more for the other), and also, for the same plate, a slight dependence of these quantities on the position of the electrodes in the plate. These small dispersions are probably related to small variations in OH concentration from plate to plate and inside the plates.The Q factors measured for the second crystal are, to our knowledge, the best results known for this material. It is

expected that higher values will be soon obtained. The overall performances observed are very promizing for applications to filters, voltage controlled oscillators and transducers.

Conclusion

In this investigation, it has been demonstrated the feasability of $GaPO_4$ epitaxy on large berlinite seeds in sulphuric acid medium. From these new seeds, it can be directly undertaken the crystal growth of large $GaPO_4$ crystals in phosphoric acid medium. At this time, the use of both solvents is needful because only Vx growth rate has a significative value in sulphuric acid medium. Unfortunately, the crystal growth process, using alternately VTG method in sulphuric acid and SHT method in phosphoric acid, seems not easily used at an industrial scale. For this reason, other crystal growth conditions are in progress.

From first characterizations, it can be asserted the good quality of $GaPO_4$ epitaxy on berlinite seeds and the good crystalline quality of crystals obtained from these seeds. On the other hand, the OH content has been followed by infrared spectrometry. If we compare to berlinite crystals, it seems that the OH content is approximatively lowered by a factor 10 for the same growth conditions.

The piezoelectric measurements made on resonators cut in two types of crystal have shown that a large improvement of the properties of resonators were obtained with the refinement of the growth conditions and that a decisive advance has been made with the use of berlinite seeds. The reduction of the OH concentration and the improvement of the crystalline quality has permitted a significative gain for the Q factor and a great reduction of the thermal variations of the equivalent resistance. On the whole these results are a confirmation of the large interest of a material having simultaneously a large coupling coefficient, a quartz like thermal stability and the possibility to lead to large Q factors.

References

[1] Chang Y.P. & Barsch G.H., IEEE Trans. Sonics Ultrasonics vol. SU-23, pp127-135, 1976.

[2] Jumas J.C., Goiffon A., Capelle B., Zarka A., Doukhan J.C., Schwartzel J., Détaint J. & Philippot E., J. Crystal Growth, vol. 80, pp. 133-148, 1987.

[3] Kolb E.D. & Laudise R.A., J. Crystal Growth, vol.43, pp.313-319, 1978.

[4] Croxall D.F., Christie T.R.A., Isherwood B.J., Todd A.J., & Birch, 2nd Conf. on Crystal Growth, Lancaster, 1979.

[5] Chai B.H.T., Shand M.L., Buehler E. & Gilleo M.L. , IEEE Ultrasonics Symp., 1979, pp.577-583.

[6] Nagai K., Ogawa T., Okuda J., Asahara J. & Taki S., 1st Inter. Symp. on Hydrothermal Reactions, 1982, pp. 496-508.

[7] Philippot E., Jumas J.C., Goiffon A. & Astier R., French Patent 2,559,166, 1984.

[8] Philippot E., Goiffon A., Maurin M., Détaint J., Schwartzel J., Toudic Y,. Capelle B. & Zarka A., J. Crystal Growth, vol. 104, pp. 713-726, 1990.

[9] Engel G., Krempl P. & Stadler J., 3rd Eur. Time and Freq. Forum, 1989, pp. 50-56.

[10] Engel G. & Krempl P., 2nd Eur. Time and Freq. Forum, 1988, pp. 827-832.

[11] Bennazha J., Cachau D., Ibanez A., Goiffon A. & Philippot E, sub. to Europ. J. Solid State and Inorg. Chem. 1992.

[12] Cambon O., Goiffon A., Ibanez A. & Philippot E. , Europ. J. Solid State and Inorg. Chem.,in press, 1992.

[13] Steinberg R.F., Roy M.K., Estes A.K., Chai B.H.T. & Morris R.C., IEEE Ultrasonics Symp., 1984, pp. 279-284.

[14] Stadler J., Krempl P. 1 Engel G. , Proc. 4th Eur. Time and Freq. Forum, Neuchatel, 1990, pp. 567-575.

[15] Krempl P., Stadler J., Wallnöfer W., Ellmeyer W. & Selic R., Proc. 5th Eur. Time and Freq. Forum, Besançon, 1991, pp. 143-147.

[16] Hirano S., Kim P.C., Orihara H., Umeda H. & Ishibashi H., J. Mat.Sci., vol. 25, pp. 2800-2804, 1990.

[17] Huard F. 1985, Thèse Université de Montpellier , 27 Sep. 1985.

[18] Glowinski A., Ann.Telecom. , vol. 27, pp. 147-157, 1972.

[19] Detaint J., Feldmann M., Henaff J.,Poignant Y. & Toudic Y., Proc. 33rd Annual Frequency Control Symp. 1979, pp. 70-79.

[20] Bailey D.S., Andle J.C.,Lee D.L., Soluch W., Vetelino J.F.&Chai B.H.T., Proc. Ultrasonics Symposium , 1983, pp. 335-340.

[21] Philippot E., Goiffon A., Jumas J.C., Avinens C., Détaint J., Schwartzel J. & Zarka A., Proc. 42nd Annual Frequency Control Symposium , 1988, pp. 138-145.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

Robust Solder Joint Attachment of Coaxial Cable Leads to Piezoelectric Ceramic Electrodes

Paul T. Vianco

Physical and Joining Metallurgy Department
Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185

Abstract

A technique was developed for the solder attachment of coaxial cable leads to the silver-bearing thick film electrodes on piezoelectric ceramics. Soldering the cable leads directly to the thick film caused bonds with low mechanical strength due to poor solder joint geometry. A barrier coating of 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn deposited on the thick film layer improved the strength of the solder joints by eliminating the adsorption of Ag from the thick film which was responsible for the improper solder joint geometry. The procedure does not require special preparation of the electrode surface and is cost effective due to the use of non-precious metal films and the batch processing capabilities of the electron beam deposition technique.

1. Introduction

The capability of controlling the operation of equipment through real-time signal feedback and microprocessor-based system control can be used to optimize numerous industrial processes. For example, real-time signal feedback and control is used in the petrochemical and steel making industries to ensure consistent quality in manufactured products at a minimum cost. Signal feedback arrays have been used to monitor the performance of automobile engines and larger machinery such as stationary steam and gas turbines.

The generation of feedback signals depends upon the use of sensors to track pertinent operating conditions in the system. Many applications require that the sensor(s) be attached to structures such as reactor vessels or engine blocks. The operation of these systems creates harsh service environments characterized by temperature extremes, humidity, and vibration. Moreover, the electrical environment may require the use of shielded conductors for noise-free signal transmission.

A process was developed for the attachment of RG196 grade coaxial cable to the silver-bearing, thick film electrodes on lead zirconate-titanate (PZT) piezoelectric ceramic, using a near eutectic tin-lead-silver solder. A tin-lead-indium solder and several lead-based alloys with lower and higher melting temperatures, respectively, were also examined as part of the analysis. Solder joints were made to the bare electrode surface or to thin film coatings deposited onto the thick film layer by electron beam evaporation. Solder joint performance was based upon manufacturability costs, solderability properties, and the mechanical strength of the joint. The solder joints were required to be mechanically robust, not only to ensure operational reliability in service, but also to withstand handling during device assembly. The mechanical strength of the solder joints is strongly dependent upon the wetting properties of the solder on the thick film surface. Previous studies had confirmed that the solder joints did not jeopardize the resonator's piezoelectric characteristics as defined for the particular application.

2. Experimental Procedures

A. RG196 cable.

The selected conductor was RG196 coaxial cable. A schematic diagram of the cable appears in Fig. 1a. The center conductor is composed of seven silver-coated, copper-plated steel strands. A Teflon insulator separated the shielding braid from the center conductor. The configuration of the cable used in the actual device is shown in Fig. 1b in which the shielding was formed into the second voltage lead. A single lead configuration (Fig. 1c) was selected for mechanical strength measurements in order to eliminate ambiguous data arising from the simultaneous testing of two solder joints.

B. PZT ceramic.

Two manufacturers of PZT ceramic were used in this study; they are denoted as "Source A" and "Source B." The selection of these particular materials was based upon their piezoelectric properties. The configuration of the PZT ceramic for the intended device is shown in Fig. 2a. The pull tests were performed on larger disks (Fig. 2b) which permitted a greater number of tests for more efficient use of the material. Although the test configuration of the PZT disks had a larger diameter than that of the actual device, the same thickness was maintained because the ceramic thickness affects the extent of residual stresses generated in the solder, sub-surface PZT, and the joint/substrate interface.

0-7803-0476-4/92 $3.00 © 1992 IEEE

The electrode material was a silver-bearing, thick film which was printed onto the ceramic and fired at an elevated temperature to promote adhesion. A glassy phase formed the bonding element between the conductive film and the PZT ceramic. The thickness of the films (which included wetting of the glassy phase into the pores of the ceramic) was 10.5 μm for the material from Source A and 5.7 μm for Source B ceramic. The poling temperatures of both brands of materials were >350°C.

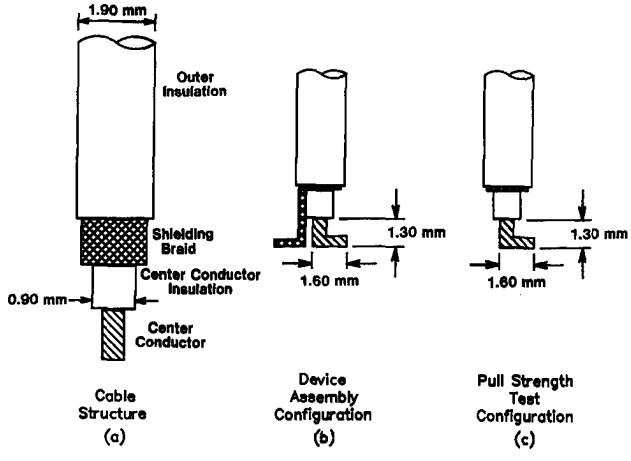


Figure 1. Configuration of the RG196 cable: (a) Structural components of the cable. (b) Lead geometry for the assembly of actual devices. (c) The lead configuration used for solderability and pull strength tests.

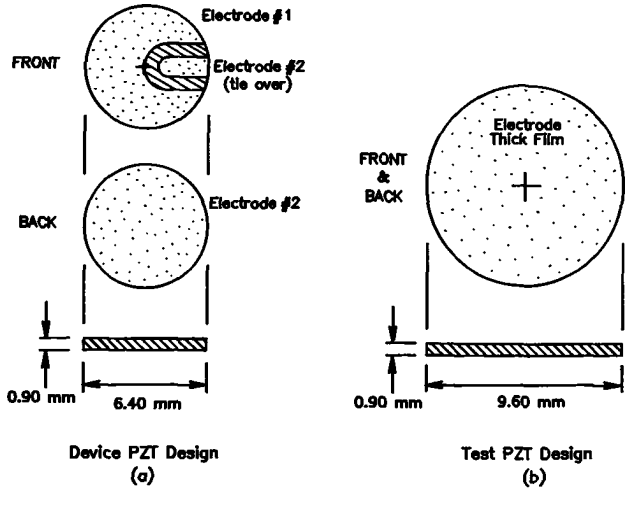


Figure 2. Thick film electrode patterns on the PZT ceramic: (a) The device units. (b) Disks used in the testing program.

C. Solder, flux, and assembly process details.

The solder alloy selected for device assembly was 62.5Sn-36.1Pb-1.4Ag (wt.%). The melting (eutectic) temperature of this solder is 179°C. The solder was supplied to the joint as a preform. The quantity of solder contained in the preform was specified by criteria for solder joint inspectability: Too little solder jeopardizes joint strength while too much may hide wetting problems between the solder and the lead or substrate. The optimum amount was determined experimentally to be 7 to 10 mg. The preform dimensions were 0.32 × 0.12 × 0.025 cm.

A rosin-based, mildly activated (RMA) flux was used in the soldering process. The flux was diluted 1:1 by volume with isopropyl alcohol.

The solder assembly process is described with the assistance of Fig. 3. No cleaning measures were used on the leads, PZT surface, or the preform prior to assembly. The PZT ceramic disks were preheated to 100°C by placing them on a hot plate. The region of the electrode to which the lead was to be soldered was coated with a layer of flux. The solder preform was then placed on the fluxed region. Next, the formed lead was coated with flux and placed on top of the preform. Soldering was performed by hand, using a temperature-controlled soldering iron with a 0.32 cm wide chisel tip at 316°C. (The assembly process could be easily adapted to mass solder techniques such as belt furnace or vapor phase reflow. Although the working temperatures would not exceed the poling temperature, fracture to the ceramic caused by thermal shock would be a concern in a process development.) After the lead had been attached, the test specimens were cleaned of flux residues using a semi-aqueous terpene-based system followed by a hot water rinse, a rinse in isopropyl alcohol, and then dried in a nitrogen gas steam.

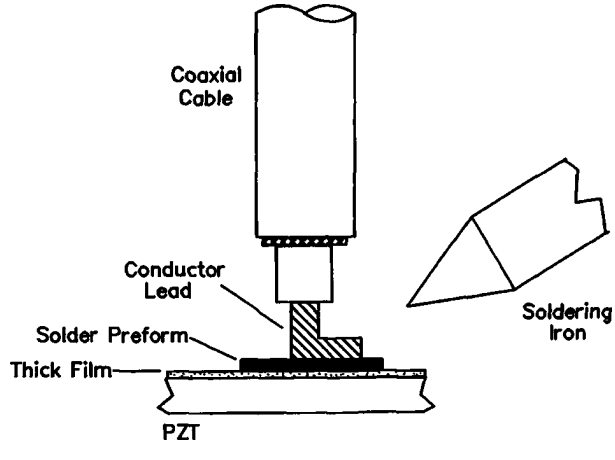


Figure 3. Component stacking for the solder joint assembly process.

D. Solderability analysis.

The solderability of the conductor/PZT electrode joint was assessed through the "soldering iron hold time" experiment. Four solder joints were assembled on a single disk in the manner described in Section C. On each of the four joints, the soldering iron was kept in contact with the molten solder for <2 s, 2 s, 4 s, or 6 s. Then, the solder joints were examined by low magnification stereo microscopy to assess the extent of wetting, non-wetting, or dewetting of the substrate and lead surfaces as well as the final configuration of the solder fillet. Dewetting describes the phenomenon in which the solder initially wets the surface; but then, a metallurgical reaction between the substrate and the solder produces a non-wettable surface. This surface condition causes the solder to re-coalesce, leaving an area with isolated mounds of solder separated by a very thin surface coating of solder. Sample solder joints were also examined by metallographic cross sections and optical microscopy in order to assess the integrity of the solder-substrate interface.

E. Mechanical tests.

The solder joint strength was tested by pulling the lead from the PZT in a testing frame at a displacement rate of 2 mm/min. The test geometry at the joint simulated a peel test configuration. The strength measurement was documented as the highest load value prior to failure. The fracture surface morphology was also recorded. Pull strength measurements were performed on units in the as-fabricated condition as well as on samples which had been exposed to thermal cycling or thermal shock test conditions. The thermal cycling conditions were (1) temperature limits of −55°C and 125°C; (2) 6°C/min ramp rates; (3) 120 min hold period at the temperature limits; and (4) a total of 100 cycles. The thermal shock conditions were (1) temperature limits of −55°C and 125°C; (2) liquid-to-liquid sample transfer; (3) 10 min hold periods at the temperature limits; and (4) a total of 100 cycles.

3. Results and Discussion

A. Solderability of the bare electrode surface.

The initial task of the process development was to determine the feasibility of soldering the coaxial cable leads directly to the thick film electrode surface. Solderability was examined via the soldering iron hold time tests. Shown in Fig. 4a is a test disk with four solder joints fabricated with 62.5Sn-36.1Pb-1.4Ag solder and the iron hold times noted in the photograph. A higher magnification of the 6 s hold time joint appears in Fig. 4b and clearly illustrates the dewetting of the solder which decreased as the hold time was lessened. Dewetting was caused by the scavenging of silver from the thick film layer by the tin component of the solder. Removal of the silver leaves the solder exposed to the glassy binder phase which it cannot readily wet; hence, it re-coalesced into isolated mounds. It is apparent that adding 1.4 wt.% silver to the solder did not suppress the scavenging phenomena.

The photograph in Fig. 4a showed that hold times of <2 s would limit the degree of dewetting. Therefore, sample solder joints were fabricated with a soldering iron hold time limited to <2 s in order to assess the mechanical strength of the joint made to the bare electrode. As-fabricated solder joints as well as those exposed to thermal cycle and thermal shock conditions were subjected to pull testing. Shown in Table 1 is the mean strength and ± one standard deviation on those results as well as the number of tests and the failure mode. The data from each of the tests on PZT material from Sources A and B have been combined; no significant difference in the pull strength was distinguishable for the two substrate types. The strength of the as-fabricated units represents a lower limit value due to a poor position of the wire within the solder mound which caused it to pull out prematurely. Thermal shock and thermal cycling tests caused a slight decrease in the mean strength of the joints; however, the significance of this difference was limited by the data scatter. In the latter two tests, failure was caused by

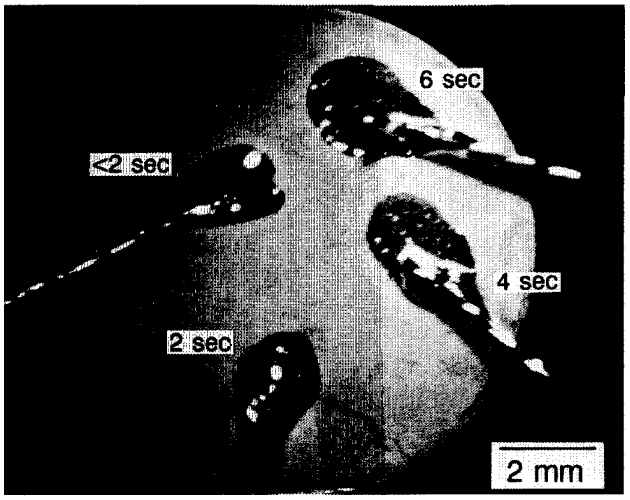


Figure 4a. Solder joints (62.5Sn-36.1Pb-1.4Ag) made as part of the soldering iron hold time experiment on bare thick film electrode.

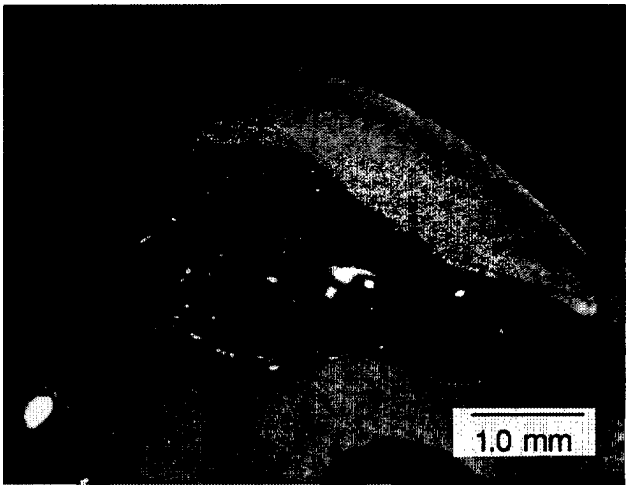


Figure 4b. The solder joint fabricated with a soldering iron hold time of 6 s, showing complete dewetting by the solder.

separation along the PZT/thick film interface. In summary, although the as-fabricated strength was a lower limit value, the mean strengths after thermal shock and cycling were unacceptably low to meet the application. The robustness of the solder joints was further limited by the data scatter which was a particular reliability problem, given the very low mean values.

Two approaches were followed in an attempt to improve the solderability of the bare electrode surface. First, a lower melting temperature solder was examined which would allow quicker fillet formation in an effort to reduce the extent of silver scavenging. Proposed service conditions and cost considerations resulted in the selection and testing of the solder composition 70Sn-18Pb-12In which has a melting point of 162°C. Soldering iron hold time experiments were conducted. Dewetting was also a problem with this solder alloy, although to a slightly lesser degree than with the 62.5Sn-36.1Pb-1.4Ag material. Mechanical tests were performed on joints fabricated with a soldering iron hold time of <2 s. The data appears in Table 2. The mean strength levels were nearly identical to those of the 62.5Sn-36.1Pb-1.4Ag solder joints. Also, the mean strength dropped for units exposed to thermal cycling and thermal shock although, as in Table 1, the data scatter precluded any significance to the trend. The failure mode for all of the tests was fracture and separation along the PZT/thick film interface. As was the case with the 62.5Sn-36.1Pb-1.4Ag solder joints, the mean strength levels and associated data scatter were excessive.

The second approach was to examine solders without the tin component responsible for silver scavenging. The lead-based, indium containing solders: 58Pb-42In (T_l = liquidus temperature = 225°C, T_s = solidus temperature = 195°C); 75Pb-25In (T_l = 264°C, T_s 250°C); and 92.5Pb-5In-2.5Ag (T_m = 300°C) were examined by the soldering iron hold time test for solderability. Solder dewetting was still prevalent with the Pb-In binary alloys. The lower tendency for indium to scavenge precious metals (as compared to that of tin) was countered by an enhanced dissolution rate at the higher working temperatures. Dewetting was significantly reduced with the Pb-In-Ag alloy. Pull strength tests were performed on 11 as-fabricated units using the Pb-In-Ag solder. The strength data was 540±230 grams, which is slightly lower than values of similar tests with 62.5Sn-36.1Pb-1.4Ag. All samples showed the PZT/thick film interface failure mode. The longer soldering iron contact times required by these alloys caused the flux to thermally degrade prior to melting of the solder. A soldering process using these

Table 1. Pull strength of RG196 center conductor soldered to the bare electrode thick film with 62.5Sn-36.1Pb-1.4Ag.

Sample Description	Pull Strength (grams)	Number of Tests	Failure Mode
As-fabricated	870±130	12	Wire pull-out from solder
Thermal cycle	550±360	6	PZT/thick film interface separation
Thermal shock	500±210	6	PZT/thick film interface separation

Table 2. Pull strength of RG196 center conductor soldered to the bare electrode thick film with 70Sn-18Pb-12In.

Sample Description	Pull Strength (grams)	Number of Tests	Failure Mode
As-fabricated	860±360	12	PZT/thick film interface separation
Thermal cycle	540±210	6	PZT/thick film interface separation
Thermal shock	680±420	6	PZT/thick film interface separation

higher melting temperature alloys would be difficult to control.

Further analysis of the data in Tables 1 and 2 was conducted to determine the cause of the poor joint strength results, despite measures used to reduce the extent of solder dewetting. A possible source of the low strength values and extensive data scatter was the formation of voids resulting from dewetting underneath what appeared to be sound fillets. An example is shown in Fig. 5. However, these voids were rare and certainly not observed to an extent which would significantly affect the pull strength of joints made with either solder alloy.

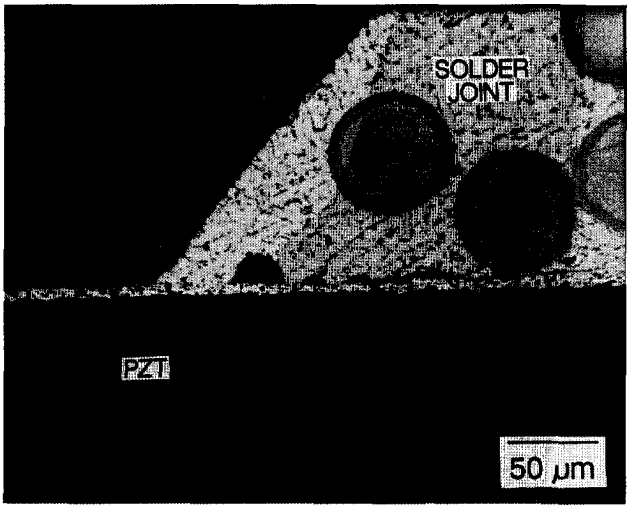


Figure 5. Optical micrograph of voids under the solder of a joint formed with 70Sn-18Pb-12In and a soldering iron hold time of <2 s.

Another source of the poor pull test performance was determined by combining the pull strength data with the failure mode observation from Tables 1 and 2. The failure mode of pull tests on the 70Sn-18Pb-12In joints suggested that the weak segment of the assembly was the PZT/thick film interface. If such was the case, then the similar pull strength levels of the 62.5Sn-18Pb-12In solder would have at least caused a mixture of failure modes between pullout of the wire and PZT/thick film interface separation; the absence of the PZT/thick film failure mode suggested that the solder joint fillet geometry prevented this mechanism. Conversely, the solder joints other than those of the as-fabricated, 62.5Sn-36.1Pb-1.4Ag solder promoted the PZT/thick film interface failure.

Additional observations were made of the shape of the as-fabricated solder joints made with the two solder alloys. It was noted that the 62.5Sn-36.1Pb-1.4Ag fillets formed a heel in the fillet due to the abnormal position of the wire. Poorly formed fillet heels were noted on the 70Sn-18Pb-12In joints. These results are diagrammed schematically in Fig. 6 which illustrates the fillet configuration with a heel at the wire bend and that with poor heel formation. A consequence of the fillet heel absence was that the application of the pulling force was modified into a "peel" load on the joint. In the peel configuration, the tensile forces of the pull test are distributed over the very small area along which the materials are separating apart. Conversely, the presence of a heel at the bend of the lead causes the resultant tensile loads to be distributed over a much larger area of the substrate, thereby increasing the load capacity of the joint. The presence of the heel geometry is a function of the joint formation process. The short soldering time used to limit dewetting did not permit the heel to form. Upon contact of the soldering iron to the joint, the solder preform melts and is drawn onto the iron tip. Then, as the lead and PZT substrate heat up, the solder flows back to the joint and wets the PZT surface and the conductor lead. Later experiments confirmed the time period required for this process to take place as 2 to 4 s.

In summary, dewetting of the bare electrode surface by the tin-containing solder required a very short soldering iron hold time (<2 s). However, the short soldering time did not permit the formation of a solder fillet configuration which possessed adequate strength.

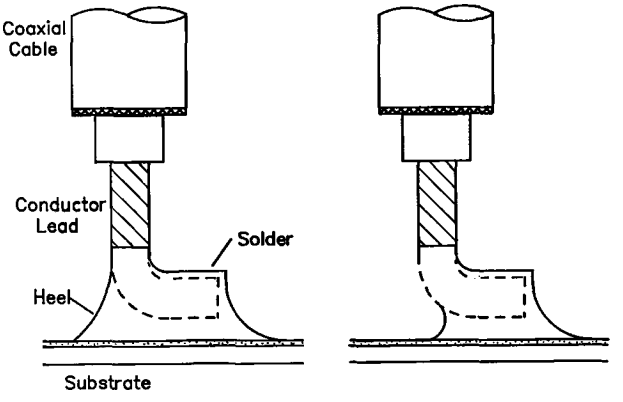


Figure 6. Schematic diagram of the solder joint cross section showing the fillet with an adequate heel configuration (left) and the case of poor heel formation (right).

B. Solderability of barrier coatings on the electrode.

It was clear that soldering the cable conductor directly to the bare electrode surface would not produce a robust solder joint. Therefore, a barrier coating between the solder and the thick film would be required which had the following properties: (1) The film was solderable and did not lose its solderability under typical storage conditions. (2) The coating had adequate adhesion to the thick film electrode surface without the need for elaborate surface preparation procedures prior to the soldering process. (3) The film could be deposited by batch processing to limit costs. (4) Finally, the coating should be composed of one or more non-precious metals.

Electron beam evaporation was selected as the film deposition technique. The PZT disks are exposed to only a nominal temperature excursion (typically <45°C) during the deposition process. The equipment is suited to batch processing with simple fixturing. The film pattern, which was defined onto the electrode surfaces by metal aperture masks, was a filled circle, 3.18 mm in diameter. Four patterns were put on each test disk (Fig. 2b).

The first barrier coating examined was 2.0 μm Ni on the electrode surface followed by 2.0 μm of Sn. The Ni layer was the actual surface to which the solder wetted. The Sn layer served to protect the surface of the Ni film from excessive oxidation or contamination; the Sn layer was dissolved into the molten solder during joint formation. Ni was selected because it is readily wet by the 62.5Sn-36.1Pb-1.4Ag solder. Also, Ni and the Sn component of the solder exhibit very slow intermetallic compound formation so that very thin layers can be deposited without consumption of the Ni film by the solder. Shown in Fig. 7 is the solder joint formed on the Ni/Sn barrier coating. The soldering iron hold time was 6 s. Excellent wetting of the Ni film was observed with no indication of dewetting.

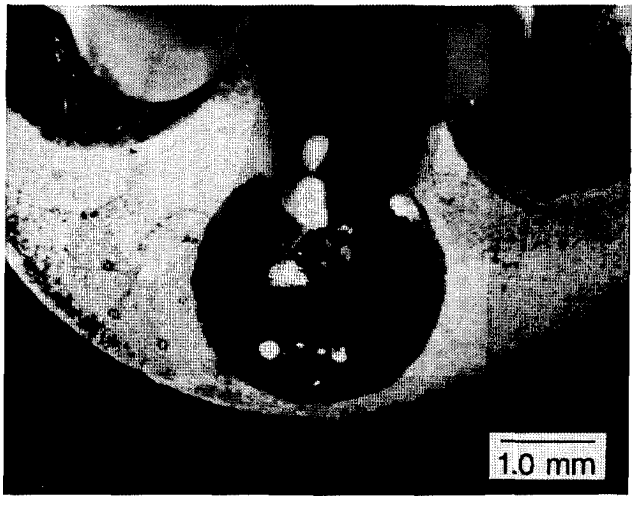


Figure 7. Stereo micrograph of the solder joint formed on the 2.0 μm Ni/2.0 μm Sn barrier coating using the 62.5Sn-36.1Pb-1.4Ag solder and an iron hold time of 6 s.

However, the Ni layer had poor adhesion to the thick film surface. Shown in Fig. 8 is an optical micrograph of the solder/Ni/thick film/PZT interface. The Ni layer has lost adhesion with the thick film electrode surface as indicated by voids between the two surfaces.

A third layer was sought which would adhere to the Ni film as well as to the thick film substrate. The successful material was Cu. Although the Cu film was also an excellent solderable surface, the growth of Cu-Sn intermetallic compounds under elevated temperature aging conditions may completely consume the Cu layer, causing dewetting as the solder spreads on the exposed, bare thick film surface.

The first three layer system evaluated was 2.0 μm Cu/2.0 μm Ni/1.0 μm Sn. The solder joints were observed (by low magnification stereo microscopy) to have excellent solderability including joints formed with the 6 s soldering iron hold time. Optical micrographs of solder joint cross sections (Fig. 9) confirmed complete wetting of the Ni film by the solder. No loss of adhesion was recorded at the Ni/Cu and Cu/thick film interfaces.

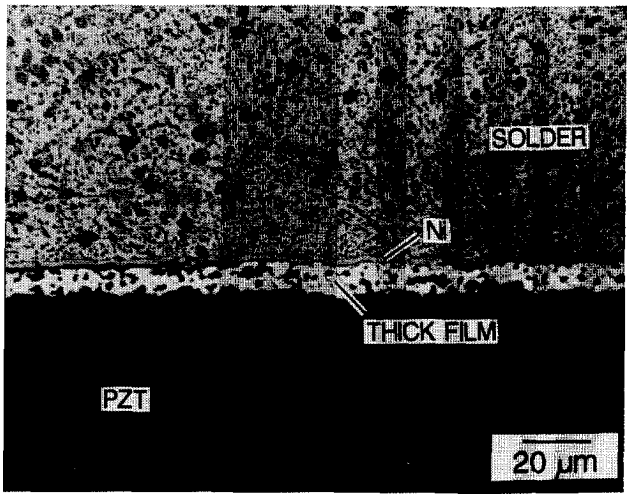


Figure 8. Optical micrograph of the 62.5Sn-36.1Pb-1.4Ag solder/Ni/thick film/PZT interface showing separation of the Ni layer from the thick film. Soldering iron hold time: <2 s.

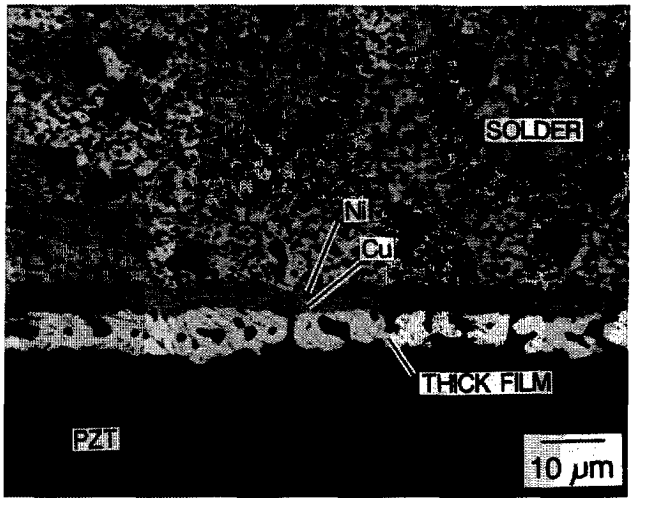


Figure 9. Optical micrograph of the PZT/solder interface of the electrode surface coated with 2.0 μm Cu/2.0 μm Ni/1.0 μm Sn.

Soldering iron hold time experiments were also conducted on reduced film thicknesses: (A) 0.25 μm Cu/0.5 μm Ni/1.0 μm Sn; (B) 0.5 μm Cu/1.0 μm Ni/1.0 μm Sn; and (C) 1.0 μm Cu/1.0 μm Ni/1.0 μm Sn. Some slight dewetting was observed for the 6 s soldering iron hold times on films (A) and (B) only. It was noted that the absolute thicknesses of the Ni and Cu films in sequences (A) and (B) were expected to support wetting by the solder without being consumed as intermetallic compound. However, the porosity of the thick film caused the coverage by the Cu/Ni coating to be further thinned and discontinuous. The area of non-wettable gaps was enlarged as the solder consumed areas of thin Cu/Ni film and the electrode material underneath. The remaining wettable surface could not support the solder film, causing it to slightly dewet.

The solderability tests outlined above established the preferred Cu and Ni layer thicknesses at 1.5 μm each. The Sn layer was maintained at 1.0 μm. Soldering iron hold time experiments indicated excellent wetting and fillet configuration for all time periods.

Oxidation protection by the 1.0 μm Sn overcoat was assessed by annealing two sample disks in air at 135°C for 250 and 1690 min. Soldering iron hold time experiments were conducted on the samples. Excellent wetting was observed on the disk exposed for 250 min. Solder spreading on the surfaces of units annealed for 1690 min was much slower; yet contact angles were <90° as noted by visual observation.

C. Mechanical tests—solder joints with Cu/Ni/Sn barrier layer.

The barrier layer permitted longer soldering times to fabricate the joint. In all cases, the hold time was 2 to 4 s which was sufficient to produce an adequate heel in the joint. Pull tests were performed on the 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn standard architecture and on film stackings with thinner individual layers described in the previous section. The results appear in Table 3. No significant difference in the pull strength was noted between the two PZT substrates (A and B). However, the data scatter was less for the material from Source A than for that from Source B.

The failure modes of solder joints made to the 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn were largely fracture of the stranded lead; the fracture in most cases took place within the insulator, away from the solder joint. Wire breaks also dominated the failure of the thinner barrier layers. The as-fabricated samples, which were not exhibiting a wire failure at the time of the test, had their tests interrupted prior to complete separation of the lead (plus solder) from the PZT disk. These specimens were cross sectioned for optical microscopy analysis to identify the exact location of the failure path. An optical micrograph appears in Fig. 10. The crack propagated partially through the PZT just under the thick film and partially along the PZT/thick film interface. More importantly, the coating layers remained adherent to one another as well as to the solder and the thick film.

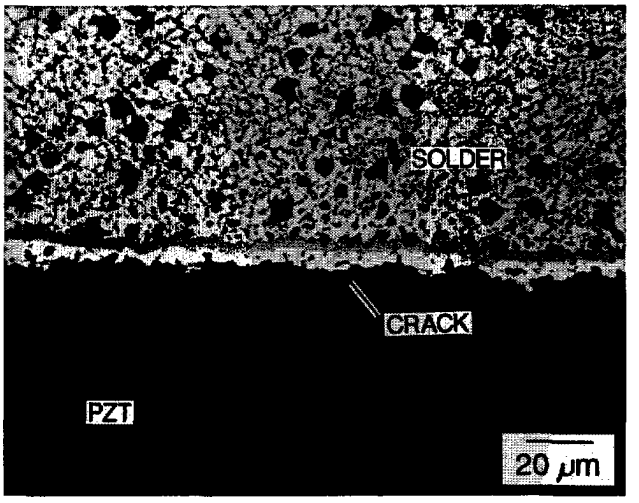


Figure 10. Optical micrograph of the PZT/solder interface with 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn exhibiting crack propagation after an interrupted pull test.

Table 3. Pull strength of RG196 center conductor soldered to the "x" μm Cu/"y" μm Ni/1.0 μm Sn barrier coating with 62.5Sn-36.1Pb-1.4Ag.

Sample Description	Pull Strength (grams)	Number of Tests	Failure Mode
Source A 1.5 μm Cu/ 1.5 μm Ni/ 1.0 μm Sn	2490±20	9	(N=2) PZT/thick film (N=7) wire failure
Source B 1.5 μm Cu/ 1.5 μm Ni/ 1.0 μm Sn	2390±240	9	(N=1) PZT/thick film (N=8) wire failure
Source A .25 μm Cu/ .50 μm Ni/ 1.0 μm Sn	2500±30	4	(N=4) wire failure
Source A .50 μm Cu/ 1.0 μm Ni/ 1.0 μm Sn	2370±90	4	(N=2) PZT/thick film (N=2) wire failure
Source A 1.0 μm Cu/ 1.0 μm Ni/ 1.0 μm Sn	2460±20	4	(N=4) wire failure

Shown in Table 4 are the pull strength data for the as-fabricated parts and the units exposed to thermal shock and thermal cycle environments. Although the two sources of PZT material had similar pull strengths in the as-fabricated condition, values of the thermal shock and thermal cycle exposure were different with the load values of Source A weaker than those of Source B. The variations in the mean strengths were significant after thermal cycling, but not as strongly so after thermal shock. The failure modes were generally different for the three categories of data. Breaking of the cable lead characterized the failure of the as-fabricated solder joints. Thermal shock tests exhibited a failure mode that began with fracture of the bulk PZT followed by separation of the PZT/thick film interface (Fig. 11). PZT/thick film failures characterized a large number of the fracture modes after thermal cycling. It is clear that both thermal shock and thermal cycling weakened the joints differently; thermal shock deteriorated the bulk PZT strength while thermal cycling weakened the PZT/thick film interface. Flaws created in the PZT by thermal shock also widened the data scatter as is customary in the failure of bulk ceramics. The resilience of Source A units to the thermal environment was less than for the Source B components.

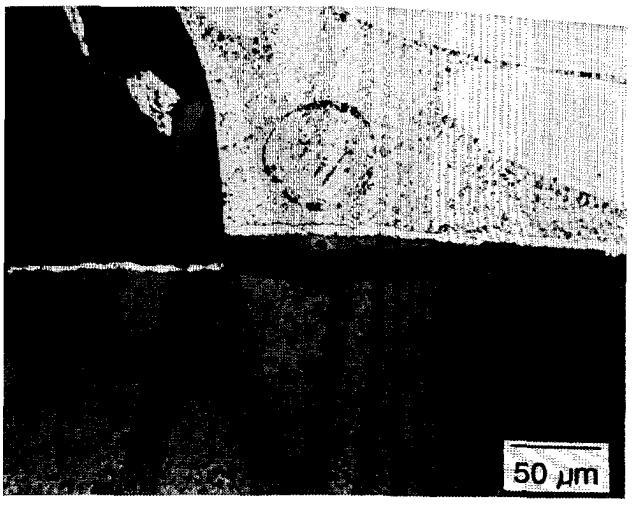


Figure 11. Optical micrograph of the PZT/solder interface showing crack initiation by the pullout of a piece of bulk PZT.

Table 4. Pull strength of RG196 center conductor soldered to the 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn barrier coating with 62.5Sn-36.1Pb-1.4Ag.

Sample Description	Pull Strength (grams)	Number of Tests	Failure Mode
Source A As-fabricated	2490±20	9	(N=2) PZT/thick film (N=7) wire failure
Source B As-fabricated	2390±240	9	(N=1) PZT/thick film (N=8) wire failure
A + B As-fabricated	2440±180	18	
Source A 100 Th. Shock	1630±830	4	(N=4) PZT fracture
Source B 100 Th. Shock	2170±240	2	(N=1) PZT fracture (N=1) wire failure
A + B 100 Th. Shock	1810±710	6	
Source A 100 Th. Cycle	1760±280	3	(N=3) PZT/thick film
Source B 100 Th. Cycle	2530±200	4	(N=2) PZT/thick film (N=2) wire failure
A + B 100 Th. Cycle	2200±460	7	

A study was also performed to determine whether the improved solder joint strength performance with the barrier coating (Tables 3 and 4) was derived from a reduced degree of dewetting, or the formation of the solder fillet (heel) configuration. Solder joints were made to the barrier coating with a soldering iron hold time of <2 s in order to prevent the adequate formation of the heel. The solder joints were then pull tested in the as-fabricated condition. Shown in Table 5 are the pull strength data. The mean strength values are approximately one-half those in Table 4. On the other hand, the strength values in Table 5 are similar to those of the as-fabricated joints made to the bare electrode surface (Table 1). In conclusion, the incomplete formation of the fillet heel in the solder joints caused by the short soldering time needed to restrict dewetting, was the primary source of the poor joint pull strengths on the bare electrode surface.

4. Conclusions

1. A robust solder attachment of coaxial conductor leads to piezoelectric ceramics can extend the introduction of sensors into service environments for real-time process control or system operation.
2. Soldering RG196 cable leads directly to the silver-bearing, thick film electrodes on PZT ceramic using 62.5Sn-36.1Pb-1.4Ag solder produced significant dewetting. Restricting the soldering time in order to reduce the dewetting problem did not allow for the complete formation of a solder fillet, thereby significantly weakening the joint's mechanical strength.
3. A barrier coating of 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn, deposited by electron beam evaporation, prevented silver scavenging by the solder and subsequent dewetting of the joint. Longer soldering times permitted the formation of the fillet heel, thereby nearly doubling the solder joint pull strength.
4. The barrier film layers were composed of non-precious metal. The deposition technique lent itself to batch processing of the PZT disks. Both attributes limited the cost of this technique.

Acknowledgments

The author thanks M. Stavig and W. Aldrich who performed the mechanical tests, A. Kilgo who prepared the metallographic samples, and M. Despain for supporting this study. This work was funded by the U.S. Dept. of Energy under contract number DE-AC04-76DP00789.

Table 5. Pull strength of RG196 center conductor soldered to the 1.5 μm Cu/1.5 μm Ni/1.0 μm Sn barrier coating with 62.5Sn-36.1Pb-1.4Ag. Poor heel formation: Soldering time <2 s.

Sample Description	Pull Strength (grams)	Number of Tests	Failure Mode
Source A As-fabricated	940±370	18	(N=18) PZT/thick film
Source B As-fabricated	1070±370	18	(N=18) PZT/thick film

INTRODUCTION TO GUIDELINES FOR THE MEASUREMENT OF ENVIRONMENTAL SENSITIVITIES OF PRECISION OSCILLATORS

Helmut Hellwig
Air Force Office of Scientific Research
Bolling Air Force Base DC 20332-6448

The following six papers are the result of work stimulated and coordinated by the IEEE Standards Coordinating Committee 27 on Time and Frequency authorized by the IEEE under Project P1193, Guidelines for Test Methods for Environmental Sensitivities of Standard Frequency Generators. Standard frequency generators include all precision oscillators such as atomic frequency standards and precision quartz crystal oscillators. Excluded from our considerations are all oscillators not conforming to this definition as well as all other active or passive electronic equipment such as receivers, amplifiers and filters.

The purpose of the proposed guidelines is the description of methods to evaluate, quantify and report the sensitivity of the frequency of precision oscillators under environmental influences such as magnetic fields, atmospheric pressure, humidity, shock, vibration, acceleration, temperature, ionizing radiation and intermittent operation. The guidelines are intended to ensure consistency and repeatability of environmental sensitivity measurements and the portability of results on particular frequency sources between the various segments of the time and frequency community.

There are three distinctly different areas of concern for environmental testing:

a. Fitness for specific user needs and actual environments (tests attempt to mimic the anticipated environments)

b. Characterization of the unit (tests attempt to provide "pure" coefficients for the various environments)

c. Reliability and survival (tests attempt to stress the unit by either going to extremes of operating ranges or by repeated application of stimuli, i.e., cycling)

The guidelines cover all three of the above areas with emphasis on the second, characterization of the unit. They provide a conceptual framework rather than an attempt to write "procedures." They emphasize proper methodology and practice and caution against pitfalls. In summary, the guidelines are not envisioned to be a specification document but rather a resource document for assisting in the development of specification statements and test methods.

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

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ACCELERATION, VIBRATION AND SHOCK EFFECTS - IEEE STANDARDS PROJECT P1193

John R. Vig, U. S. Army Electronics Technology and Devices Laboratory; Claude Audoin, Université Paris-Sud, Leonard S. Cutler, Hewlett-Packard Co.; Michael M. Driscoll, Westinghouse Defense and Electronic Systems Center; Errol P. EerNisse, Quartztronics, Inc.; Raymond L. Filler, U. S. Army Electronics Technology and Devices Laboratory; R. Michael Garvey, Frequency and Time Systems, Inc.; William J. Riley, EG&G Frequency Products; Robert C. Smythe, Piezo Technology, Inc.; Rolf D. Weglein, Consultant.

Abstract

The effects of acceleration on quartz and atomic frequency sources are discussed. Guidelines are provided for the specification and testing of oscillator acceleration sensitivities. The discussion includes: steady-state acceleration effects, gravity change effects, shock effects, vibration effects, frequency multiplication effects, and crystal filter effects. The vibration effects section includes: sinusoidal vibration, random vibration, acoustic noise, the effects on short term stability, and spectral responses at other than the vibration frequency.

Description of the Phenomenon

General

Precision frequency sources, both quartz and atomic, are affected by acceleration. The magnitude of a quartz crystal oscillator's acceleration-induced frequency shift is proportional to the magnitude of the acceleration, it also depends on the direction of the acceleration, and on the acceleration sensitivity of the oscillator. It has been shown, empirically, that the acceleration sensitivity of a quartz crystal oscillator is a vector quantity [1]. Therefore, the frequency during acceleration can be written as a function of the scalar product of two vectors

$$f(\vec{a}) = f_o(1 + \vec{\Gamma} \cdot \vec{a}) \qquad (1)$$

where $f(\vec{a})$ is the resonant frequency of the oscillator experiencing acceleration $\vec{a}$, f_o is the frequency with no acceleration (often called the "carrier frequency"), and $\vec{\Gamma}$ is the acceleration-sensitivity vector of the oscillator. The frequency of an accelerating oscillator is a maximum when the acceleration is parallel to the acceleration-sensitivity vector. The frequency shift, $f(\vec{a}) - f_o = \Delta f$, is zero for any acceleration in the plane normal to the acceleration-sensitivity vector, and it is negative when the acceleration is antiparallel to the acceleration-sensitivity vector. The frequency change due to acceleration is usually expressed as a normalized frequency change, where, it follows from Equation (1) that

$$\frac{\Delta f}{f_o} = \vec{\Gamma} \cdot \vec{a}. \qquad (2)$$

Typical values of $|\vec{\Gamma}|$ for precision crystal oscillators are in the range of 10^{-9} per g to 10^{-10} per g. $|\vec{\Gamma}|$ is independent of acceleration amplitude for the commonly encountered acceleration levels (i.e., at least up to 20g); however, high acceleration levels can result in changes in the crystal unit (e.g., in the mounting structure) that can lead to $\vec{\Gamma}$ being a function of acceleration. $\vec{\Gamma}$ can also be a function of temperature [2].

In an atomic frequency standard, the frequency of a voltage controlled crystal oscillator (VCXO) is multiplied and locked to the frequency of an atomic resonator. The effects of acceleration on atomic standards can be divided into crystal oscillator effects, atomic resonator effects, and servo loop effects. The extent to which an acceleration-induced VCXO frequency shift affects the output frequency of the atomic standard depends on the rate of change of acceleration relative to the atomic resonator-to-crystal oscillator servo loop time constant, t_o. Fast acceleration changes ($f_{vib} >> 1/2\pi t_o$) will cause the atomic standard's acceleration sensitivity to be that of the VCXO, since the servo loop will not be fast enough to correct the VCXO. Slow acceleration changes ($f_{vib} << 1/2\pi t_o$) will have little effect on the output frequency to the extent that the servo loop gain is sufficient to correct the VCXO frequency to that of the atomic resonator. However, it has been observed that a constant frequency offset may appear in the case where the atomic standard is submitted to a periodic vibration, even if its frequency is lower than $1/2\pi t_o$. As is shown in Appendix A, this is a servo loop effect and it occurs when the acceleration induced frequency change shows a distortion, with a component at twice the acceleration frequency in addition to that at the fundamental vibration frequency. In the case $f_{vib} << 1/2\pi t_o$, the spurious constant frequency offset is very efficiently attenuated when the value of t_o is decreased.

In general, vibration near[1] the servo loop modulation frequency, f_{mod}, (or near multiples or submultiples of f_{mod}) can cause significant frequency offsets in passive atomic

[1] i.e., within the servo loop bandwidth.

0-7803-0476-4/92 $3.00 © 1992 IEEE

frequency standards if the mechanical structure produces responses at harmonics of the vibration frequency. In the case of rubidium frequency standards, vibration of the physics package at f_{mod} can modulate the light beam, producing a spurious signal that can confuse the servo system and, thereby, cause a frequency offset (in a manner identical to that for cesium standards described below). Vibration of the VCXO at even multiples of f_{mod} produces sidebands on the microwave excitation to the Rb physics package that causes a frequency offset via an intermodulation effect [3-8]. A loss of microwave excitation power can occur at low vibration frequencies, when the sideband to carrier power ratio becomes large [1].

At low acceleration levels, properly designed atomic resonators possess very low acceleration sensitivities; however, high acceleration levels (e.g., >10g) can produce significant effects. For example, in a rubidium standard [3,4], the acceleration can change the location of the molten Rb inside the Rb lamp, and it can cause mechanical changes which result in deflection of the light beam. Both effects can result in a change in light and signal output, that, due to light shift and servo offset mechanisms, can cause a frequency shift. Mechanical damage can cause rf power changes that, due to the rf power shift effect, can cause a frequency shift.

In a cesium standard [8,9], high acceleration levels can affect the accuracy and stability of the output frequency through mechanisms which modify the position of the atomic trajectory with respect to the tube structure. This is most serious when the vibration frequency is near the servo modulation frequency. The vibration modulates the amplitude of the detected beam signal. The net effect of this phenomenon is normally of no consequence due to the fact that the perturbing vibration must be at or very near the servo modulation frequency and must be stable in frequency as well. When the acceleration is very near the servo modulation frequency, the vibration induced amplitude perturbation of the detected beam can be synchronously detected, leading to large output frequency errors. It should be noted that this problem is of minimal concern in actual applications due to the requirements on the precision and stability of the frequency of the perturbing acceleration.

A more subtle problem arises from the effects on the position of the beam with respect to the microwave interrogating cavity via distributed cavity phase shift effects. Another subtle effect arises from the potential to modify the detected velocity distribution. The magnitude of these effects are small when compared to the vibration induced amplitude modulation perturbations described above.

Acceleration effects can also cause frequency offsets in cesium frequency standards via degradation of the amplitude of the interrogating microwave signal. This can happen as a result of detuning of the frequency multipliers. Good mechanical design and thorough qualification of the design will minimize problems in this area.

In hydrogen masers, the most acceleration sensitive part is the microwave cavity [10]. A deformation of the cavity structure causes a shift of the cavity resonant frequency. This induces a change of the maser frequency via the cavity pulling effect. A cavity autotuning system is able to suppress this effect if the acceleration change is sufficiently smaller than $1/2\pi t_0$, where t_0 is the cavity servo loop time constant.

Types of Acceleration Effects

Steady-State Acceleration: When an oscillator is subjected to steady-state acceleration, the normalized frequency shifts by $\vec{\Gamma} \cdot \vec{a}$, per Equation (2). Steady-state acceleration occurs, for example, during the launching of a rocket, in an orbiting satellite, in a centrifuge, and in a gravitational field.

Gravity Change Effects: The frequency shift described in Equation (2) is also induced by the acceleration due to gravity. One manifestation occurs when an oscillator is turned upside down (on earth). This is commonly referred to as "2g-tipover." During 2g-tipover, the magnitude of the gravity field is 1g in the direction towards the center of the earth.[2] As is discussed in Appendix B, use of Equations (1) and (2) for gravitational field effects necessitates defining the acceleration of gravity as pointing away from the center of the earth so that the direction of $\vec{\Gamma}$ is consistent with the direction one obtains for conventional acceleration.

When an oscillator is rotated 180° about a horizontal axis, the scalar product of the gravitational field and the unit vector normal to the initial "top" of the oscillator changes from -1g to +1g, i.e., by 2g. Figure 1 shows actual data of the fractional frequency shifts of an oscillator when the oscillator was rotated about three mutually perpendicular axes in the earth's gravitational field. For each curve, the axis of rotation was horizontal. The sinusoidal shape of each curve is a consequence of the scalar product being proportional to the cosine of the angle between the acceleration-sensitivity vector and the

[2] The magnitude of acceleration is in units of g, i.e., the earth's gravitational acceleration at sea level, 980 cm/sec^2.

acceleration due to gravity, as is explained in more detail in Appendix B.

Another type of gravity change effect occurs when, for example, a spacecraft containing an oscillator is sent into space. The oscillator's frequency will, again, change in accordance with Equation (1).

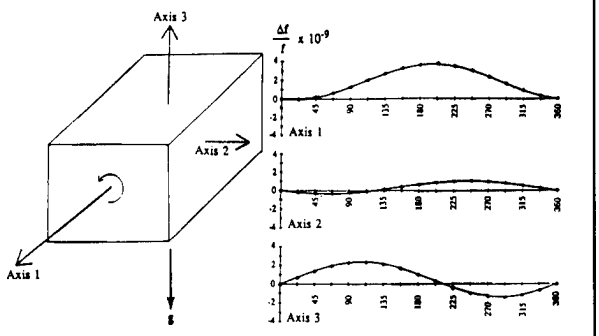


Fig. 1. "2-g Tipover," Frequency Change Versus Rotation in Earth's Gravitational Field for Three Mutually Perpendicular Axes.

Vibration Effects: The effects of vibration on frequency stability have been reviewed by Filler [1]. The effects can be summarized as follows:

1. Sinusoidal Vibration: For a small modulation index, $\beta = \Delta f/f_v = (\vec{\Gamma} \bullet \vec{A})f_o/f_v < 0.1$, sinusoidal vibration produces spectral lines at $\pm f_v$ from the carrier, where f_v is the vibration frequency:

$$\mathcal{L}'(f_v) = 20 \log \frac{(\vec{\Gamma} \bullet \vec{A} f_o)}{2 f_v}. \qquad (3)$$

(NOTE: *$\mathcal{L}'(f)$ are spectral lines (i.e., delta functions) not spectral densities.*) Most of the power is in the carrier, a small amount is in the first spectral line pair, and the higher order spectral lines are negligible.

2. Random Vibration: For a small modulation index, random vibration's contribution to phase noise is given by:

$$\mathcal{L}(f) = 20 \log \left(\frac{\vec{\Gamma} \bullet \vec{A} f_o}{2f} \right), \qquad (4)$$

where $|\vec{A}| = [(2)(PSD)]^{1/2}$ and PSD is the power spectral density of the vibration. The use of $\mathcal{L}(f)$ is in conformance with IEEE-1139-1988 [11]. A consequence of equation 4 is that vibrating platforms can cause severe phase noise degradation.

Not only does random vibration degrade the spectrum, but the time errors due to random vibration also accumulate. The time (or phase) errors do not average out because the white frequency noise is integrated to produce random walk of the phase. The noise of an oscillator produces time prediction errors of $\sim\tau\sigma_y(\tau)$ for prediction intervals of τ [12].

3. Acoustic Noise: MIL-STD-810D describes the effects of acoustic noise as follows: "Acoustic noise can produce vibration in equipment similar to that produced by mechanically transmitted vibration. In an acoustic noise field, pressure fluctuations impinge directly on the equipment. The attenuation effects of mechanical transmission are missing and the response of the equipment can be significantly greater. Further, components which are effectively isolated from mechanical transmission will be excited directly. Examples of acoustically induced problems:

a. Failure of microelectronic component lead wires.
b. Chafing of wires.
c. Cracking of printed circuit boards..."

In addition to these problems, the response of an oscillator to acoustic-noise-induced acceleration is the same as the response to any other type of vibration, i.e., the acoustic noise modulates the oscillator's frequency [13]. The modulation (or phase noise degradation) is a function of the acoustic-noise-induced vibration's amplitudes, directions, and frequencies.

Acoustic noise can have a broad spectrum. For example, in a missile environment, it may extend to frequencies above 50 kHz. An effect of such noise, e.g., in 100 MHz 5th overtone resonators, may be the excitation of flexural modes (microphonics) in the crystal plate [14]. These flexural modes in turn can produce undesirable spectral lines in the phase noise spectrum. The magnitudes of these lines are independent of the resonator's $\vec{\Gamma}$ and depend chiefly on the plate geometry and mounting structure.

Acoustic noise can be especially troublesome in certain applications. For example, when an extremely low noise oscillator was required in an aircraft radar application, after the system designers built a three-level vibration isolation system to isolate the oscillator from the vibration of the aircraft, they discovered that the isolation system failed to deliver the expected phase noise of the oscillator because the isolation system failed to deal with the acoustic noise in the aircraft; i.e., the isolation system was effective in isolating the oscillator from the vibrations of the airframe, but it was ineffective in blocking the intense sound waves that impinged on the oscillator.

4. Oscillator Sustaining Stage Vibration: When a crystal oscillator is subjected to vibration, the primary cause of the resultant output signal frequency modulation is the acceleration sensitivity of the quartz crystal resonator. However, vibration-induced mechanical motion in other circuit components, and in the circuit board itself, can also result in output signal frequency and/or phase modulation. In general, these effects are more pronounced in higher frequency (i.e., VHF) oscillators due to a combination of: (1) increased circuit signal phase sensitivity to mechanical motion (i.e., increased phase shift for a given amount of circuit reactance variation due to a larger resonator C_1), and (2) decreased crystal resonator Q.

If the vibration-induced circuit phase shift occurs inside the oscillator feedback loop, there will be a conversion of phase-to-frequency modulation for vibration frequencies within the resonator half-bandwidth. The phase-to-frequency conversion is related to the resonator group delay (i.e., loaded Q). For vibration frequencies in excess of the resonator half-bandwidth, the vibration-induced phase modulation sidebands may or may not be further attenuated, depending on whether or not the induced modulation is occurring in a portion of the circuit signal path which is subject to resonator frequency selectivity.

Methods for minimizing these effects include use of multiple circuit board chassis mounting points, circuit potting, wire and cable tie down, elimination of adjustable components or post-tuning cementing in place of adjustors, avoidance of non-potted and non-shielded inductors, and avoidance of very high circuit nodal impedances that are sensitive to nodal capacitance variation. As an example, measurements on a 100 MHz SC-cut crystal oscillator indicate that, when these precautions are taken, sustaining stage carrier signal phase shift sensitivity to vibration on the order of 10^{-6} radians per g can typically be obtained. This represents a situation where sustaining stage vibration-induced signal phase modulation becomes dominant (as compared to resonator frequency modulation effects) only at vibration frequencies in excess of approximately 50 kHz.

5. Frequency Multiplication: Upon frequency multiplication by a factor N, the vibration frequency f_v is unaffected since it is an external influence. The peak frequency change due to vibration, Δf, however, becomes

$$\Delta f = (\vec{\Gamma} \cdot \vec{A}) N f_o . \qquad (5)$$

The modulation index β is therefore increased by the factor N. Expressed in decibels, frequency multiplication by a factor N increases the phase noise by 20 log N.

When exposed to the same vibration, the relationship between the vibration-induced phase noise of two oscillators with the same vibration sensitivity and different carrier frequencies is

$$\mathcal{L}_B(f) = \mathcal{L}_A(f) + 20 \log (f_B / f_A) , \qquad (6)$$

where $\mathcal{L}_A(f)$ is the sideband level, in dBc/Hz (or dBc for sinusoidal vibration), of the oscillator at frequency f_A, and $\mathcal{L}_B(f)$ is the sideband level of the oscillator at frequency f_B. For the same acceleration sensitivity, vibration frequency and output frequency, the sidebands are identical, whether the output frequency is obtained by multiplication from a lower frequency or by direct generation at the higher frequency. For example, when a 2 x 10^{-9}/g sensitivity 5.0 MHz oscillator's frequency is multiplied by a factor of 315 to generate a frequency of 1575 MHz, its output will contain vibration-induced sidebands which are identical to those of a 1575 MHz SAW oscillator which has the same 2 x 10^{-9}/g sensitivity.

6. Large Modulation Index: A large modulation index, i.e., $\beta > 0.1$, can occur in UHF and higher frequency systems, and at low vibration frequencies. When the modulation index is large, it is possible for the sidebands to be larger than the carrier. At the values of β where $J_o(\beta) = 0$, e.g., at $\beta = 2.4$, the sidebands-to-carrier power ratio goes to infinity [1], which means that all of the power is in the sidebands, none is in the carrier. Such "carrier collapse" can produce catastrophic problems in some applications.

7. Two-Sample Deviation: The two-sample deviation [11] (or square-root of the Allan variance) $\sigma_y(\tau)$ is degraded by vibration because the vibration modulates the oscillator's output frequency. The typical degradation due to sinusoidal vibration varies with averaging time, as shown in Figure 2. Since a full sine wave averages to zero, the degradation is zero for averaging times that are integer multiples of the period of vibration. The peaks occur at averaging times that are odd multiples of half the period of vibration. The $\sigma_y(\tau)$ due to a single-frequency vibration is:

$$\sigma_y(\tau) = \frac{\vec{\Gamma} \bullet \vec{a}}{\pi} \frac{\tau_v}{\tau} \sin^2 (\pi \frac{\tau}{\tau_v}) , \qquad (7)$$

where τ_v is the period of vibration, τ is the measurement averaging time, Γ is the acceleration sensitivity vector, and $\vec{a}$ is the acceleration.

8. Integrated Phase Noise, Phase Excursions, Jitter and Wander: Specialists in crystal resonators and oscillators generally characterize phase noise by $S_\phi(f)$ or $\mathcal{L}(f)$ [11]. Many users of crystal oscillators, however, characterize phase noise in terms of "phase jitter."

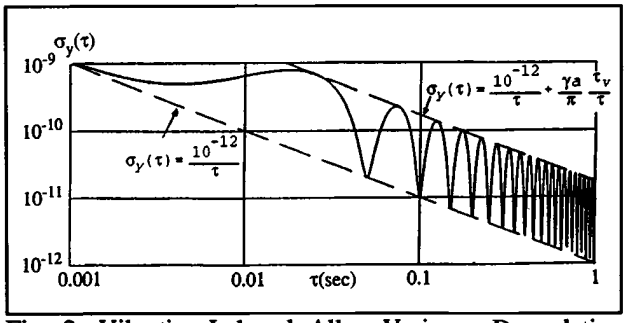


Fig. 2. Vibration-Induced Allan Variance Degradation Example (f_v = 20 Hz, |a| = 1.0g, $|\vec{\Gamma}| = 1 \times 10^{-9}$/g).

In digital communications, the terms jitter and wander are used in characterizing timing instabilities. Jitter refers to the high-frequency timing variations of a digital signal and wander refers to the low-frequency variations. The dividing line between the two is often taken to be 10 Hz. Wander and jitter, whether caused by vibration or otherwise, can be characterized by the appropriate measurement of the RMS time error of the clock. A 10 Hz low pass filter should be used to remove the effects of jitter if necessary. For very high Fourier frequencies or short integration times, it may be necessary to calculate the jitter from the spectrum rather than to measure it directly. For example, the mean-square phase jitter during a one-tenth second interval is the integral of the phase instability $S_\phi(f)$ over the Fourier frequency range from 10 Hz to infinity [12].

The integrated phase noise for the bandwidth f_1 to f_2 is

$$\Delta\phi_i^2 = \int_{f_1}^{f_2} S_\phi(f)\,df. \qquad (8)$$

For random vibration, it can be shown that

$$S_\phi(f) = \frac{1}{2f_v^2}(PSD)(|\vec{\Gamma}|\,f_o)^2. \qquad (9)$$

The phase of a vibration modulated signal is [1]:

$$\phi(t) = 2\pi f_o t + \left(\frac{\Delta f}{f_v}\right) \qquad (10)$$

The peak phase excursion is:

$$\Delta\phi_{peak} = \Delta f / f_v. \qquad (11)$$

In a phase-locked-loop, for example, the magnitude of the phase excursion determines whether or not the loop will break lock under vibration. For example, if a 10 MHz, 1 x 10^{-9}/g oscillator is subjected to a 10 Hz sinusoidal vibration of amplitude 1 g, the peak vibration-induced phase excursion is 1 x 10^{-3} radians. If this oscillator is used as the reference oscillator in a 10 GHz radar system, the peak phase excursion at 10 GHz will be 1 radian. Such a large phase excursion can be catastrophic to the performance of many systems, such as those which employ phase-locked-loops (PLL) or phase shift keying (PSK).

9. Spectral Responses at Other than the Vibration Frequency: Spectral responses at other than the vibration frequency may arise from a nominally sinusoidal vibration source of frequency f_v, if the source is not a pure sinusoid [14]. This usually occurs when the source is driven hard to generate vibration near its maximum output power, so that it operates in the nonlinear regime. Under these circumstances, the spectrum of the vibration source itself will contain not only the spectral line at frequency f_v but also lines at harmonic frequencies, $2f_v$, $3f_v$, etc. A spectrum check of the vibration source is recommended in such cases.

Even if the vibration source is a pure sinusoid at frequency f_v, it is still possible to excite oscillator vibration responses at harmonically related vibration frequencies $2f_v$, $3f_v$, etc. if the vibration level is excessive so as to drive materials in the oscillator into the nonlinear range. This situation is readily identified by observing the effect of reducing the vibration source amplitude.

Oscillator spectral responses at other than the vibration frequency have also been observed in cases where the oscillator is subjected to a random vibration spectrum. These responses are excited at frequencies much higher than the exciting spectrum and are the result of nonlinear phenomena in the crystal plate and/or the oscillator. The responses are in the form of spectral lines at carrier offset frequencies that correspond to the flexural modes of the crystal plate. The flexural mode frequencies are determined, in decreasing order of importance, by the number of crystal plate support posts, by the plate thickness and by the crystal cut. For example, the typical spectral response range in a 100 MHz, 4-post supported crystal plate extends upward from 12.2 kHz, the fundamental mode flexural frequency [14]. It has been observed that these responses are minimized in a crystal plate that is compliantly supported rather than hard mounted.

Shock Effects: When a crystal oscillator experiences a shock, frequency (and phase) excursions result which, in a properly designed oscillator, are due primarily to the quartz resonator's stress sensitivity. The magnitude of the excursion is a function of resonator design, and of the shock-induced stresses on the resonator. (Resonances in the mounting structure will amplify the stresses.) A permanent frequency offset usually results which can be

due to: shock-induced stress changes (when some elastic limits in the resonator structure are exceeded), the removal of (particulate) contamination from the resonator surfaces, and changes in the oscillator circuitry, e.g., due to changes in stray capacitances.

The shock-produced phase excursions can be calculated from Equation (10), with the proviso that at high acceleration levels, $\vec{\Gamma}$ may be a function of $\vec{a}$. For example, for a half-sine shock pulse of duration D

$$\Delta\phi_{peak} = 2 D f_o \vec{\Gamma} \cdot \vec{a} . \qquad (12)$$

Upon frequency multiplication by N, the $\Delta\phi_{peak}$ becomes N times larger, so in systems where the frequency is multiplied to microwave (or higher) frequencies, the shock-induced phase excursion can cause serious problems, such as loss of lock in PLL systems, and bit errors in PSK systems.

Survival under shock (and under vibration) is primarily a function of resonator surface imperfections. Even minute scratches on the surfaces of the quartz plate result in orders of magnitude reductions in the resonator's shock resistance [15]. Chemical-polishing-produced scratch-free resonators have survived shocks of up to 36,000g in air gun tests, and have survived the shocks due to being fired from a 155mm howitzer (16,000g, 12ms duration) [16].

In atomic frequency standards, a shock-induced phase excursion of the VCXO can result in a transitory loss of lock, however, this is not a problem when the shock duration is smaller than the servo loop time constant, which is often the case. When a loss of lock does occur, the recovery time is a function of the servo loop time constant. The phase excursions of the VCXO, and of the output crystal filter, can have a significant effect on the clock output, which can disturb the host system. Shock-induced mechanical damage can cause changes (e.g., light and rf power changes in Rb standards) that can produce a permanent frequency offset. For atomic standards employing Ramsey interrogation, large (2×10^{-8}) shock-induced permanent frequency shifts of the VCXO can cause false lock acquisition to the satellite peaks in the Ramsey resonance.

Acceleration Effects in Crystal Filters: Some frequency sources, such as synthesizers, atomic frequency standards, and precision crystal oscillators with post-filters, contain crystal filters. In these applications they are often called spectrum cleanup filters; however, under vibration such filters modulate the signals passing through them, adding as well as removing vibration-related sidebands. Hence, spectrum cleanup filters must be used with great care in systems subject to vibration.

It is well known that the principal vibration effect in crystal filters is phase modulation, although some amplitude modulation may also occur [17,18]. Appendix C contains a further discussion and examples of calculated results for the total vibration-induced phase modulation of frequency source with post-filters.

Specifications

Acceleration sensitivity can be specified several ways. The appropriate specification method depends on the application and on the types of accelerations that can be expected to adversely influence the oscillator's performance. The general military specification for crystal oscillators [19], MIL-O-55310, attempts to address all aspects of acceleration sensitivity. The paragraphs dealing with acceleration sensitivity are reproduced in Appendix D. Although MIL-O-55310 was written for crystal oscillators, because the output frequencies of atomic frequency standards originate from crystal oscillators, and because no comparable document exists that addresses atomic standards specifically, MIL-O-55310 can also serve as a suitable guide to specifying atomic standards.

Test Methods

2-g Tipover Test

In the past, the 2-g tipover test has often been used by manufacturers (and researchers) to characterize oscillators' acceleration sensitivity. This test method is deceptively simple because if not used carefully, it can yield false and misleading results.

The simple 2-g tipover test consists of measuring the frequency changes when an oscillator is turned upside down three times, about three mutually perpendicular axes. The magnitude of the acceleration sensitivity is then the vector sum (square-root of the sum of the squares) of the three frequency changes per g (where, for each axis, the frequency change per g is one-half of the measured frequency change).

There are some serious problems with using the 2-g tipover test: 1) the test is applicable only to high quality oven-controlled oscillators because in non-temperature-controlled oscillators, the frequency-shifts due to ambient temperature changes will exceed the acceleration-induced frequency changes, and, thereby, make the test results worthless; 2) many oven-controlled oscillators are not suitable for characterization by the 2-g tipover test, because rotation of the oscillator results in temperature changes (due to air convection) inside the oven that can

mask the effects due to acceleration changes; similarly, in atomic standards, changes in internal thermal distribution resulting from the tipover will mask acceleration effects; 3) the results are poor indicators of performance under vibration when the vibration frequencies of interest include resonances (see section IV); 4) since magnetic fields can change the frequencies of crystal oscillators ~ 10^{-11} to 10^{-10} per gauss [20], rotation in the earth's magnetic field can produce significant errors while measuring (unshielded) low acceleration sensitivity crystal oscillators; and 5) in atomic frequency standards, effects of the earth's magnetic field can dominate the results, and for the reasons discussed above, the results will be irrelevant to the performance under vibration if during 2-g tipover testing the acceleration changes with a time constant that is larger than the servo-loop time constant.

A 2-g tipover test that is far more reliable than the simple test described above consists of measuring the fractional frequency changes corresponding to small changes in orientation with respect to the earth's gravitational field, e.g., as shown in Figure 1. The oscillator is first rotated, e.g., in 22.5° increments, 360° about an axis (which is usually one of the major axes of the oscillator). From the frequency changes during this rotation, one can determine two out of the three components of $\vec{\Gamma}$ (keeping in mind that $\vec{a} = -\vec{g}$ in Equations (1) and (2), as is discussed in Appendix B). The oscillator is then similarly rotated 360° about a second axis that is perpendicular to the first. From the frequency changes during this second rotation, one can determine the third component of $\vec{\Gamma}$, and, simultaneously, obtain a self-consistency check for one of the other two components (i.e., the one that is normal to both the first and second axes of rotation). The calculation details are shown in Appendix B, where it is also shown that the frequency change vs. orientation, $\Delta f(\theta)$ vs. θ, is a sinusoidal function. The frequency changes during rotation about the third axis can provide additional self-consistency checks for the two components of $\vec{\Gamma}$ that are normal to the third axis. If the measurements are not self-consistent, and if there are large deviations in the $\Delta f(\theta)$ vs. θ from the best fit to a sinusoidal function, as will generally be the case if, for example, the resonator's temperature changes during the test, then the 2-g tipover test result is unreliable. It should be noted that, in this 2-g tipover test too, the earth's magnetic field can produce significant errors if the oscillator is unshielded and if the oscillator possesses low-acceleration-sensitivity.

Vibration Tests

The sidebands generated by sinusoidal vibration can be used to measure the acceleration sensitivity. From Equation (3)

$$\Gamma_i = \left(\frac{2f_v}{A_i f_o}\right) 10^B, \text{ where } B = \mathcal{L}'_i(f_v)/20 \qquad (13)$$

and where $\vec{\Gamma}_i$ and A_i are the components of the acceleration sensitivity vector and of the acceleration, respectively, in the $\hat{i}$ direction. Measurements, along three mutually perpendicular axes are required to characterize $\vec{\Gamma}$, which becomes

$$\vec{\Gamma} = \Gamma_i \hat{i} + \Gamma_j \hat{j} + \Gamma_k \hat{k} \qquad (14)$$

with a magnitude of

$$|\vec{\Gamma}| = (\Gamma_i^2 + \Gamma_j^2 + \Gamma_k^2)^{1/2}. \qquad (15)$$

One scheme for measuring $\vec{\Gamma}$ is shown in Figure 3. The local oscillator is used to mix the carrier frequency down to the range of the spectrum analyzer. If the local oscillator is not modulated, the relative sideband levels are unchanged by mixing. The frequency multiplier is used to overcome dynamic range limitations of the spectrum analyzer, using the "20 log N" enhancement discussed previously. The measured sideband levels must be adjusted for the multiplication factor prior to insertion into Equation (12). It must be stressed that Equation (12) is valid only if $\beta < 0.1$. A sample measurement output and calculation is shown in Figure 4.

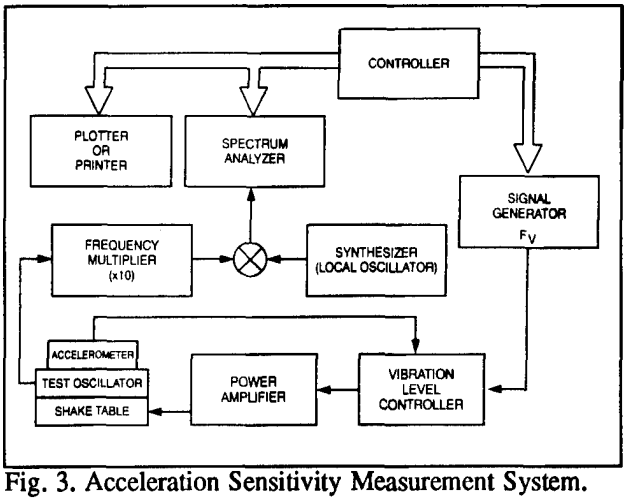


Fig. 3. Acceleration Sensitivity Measurement System.

In order to detect frequency sensitivities, e.g., due to vibration resonances, the sideband levels need to be measured at multiple vibration frequencies - e.g., see paragraph 4.9.38.2 of MIL-O-55310C [19]. An alternative to using a series of vibration frequencies is to use random vibration, e.g., as described by D. J. Healy, III, et al. [21].

Refinements of the vibration-induced-sideband method of measuring acceleration sensitivity have been described

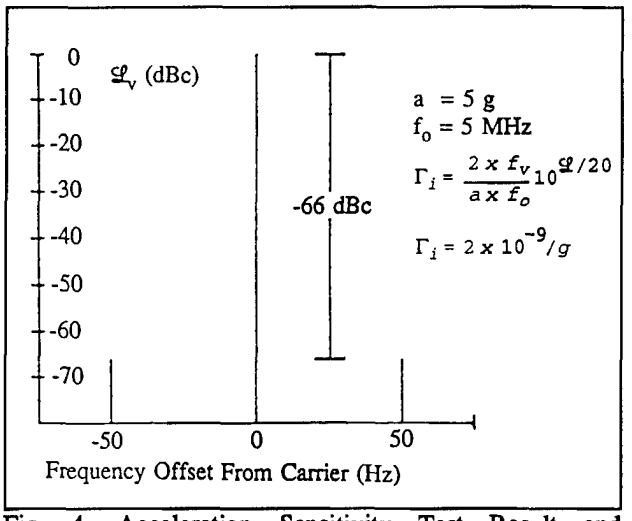


Fig. 4. Acceleration Sensitivity Test Result and Calculation Example.

by M. M. Driscoll [22], and by M. H. Watts, et al. [23]. Driscoll's method provides for minimization of measurement errors due to cable vibration. The method also allows measurement of acceleration sensitivity of the resonator alone. The resonator is mounted on the shake table and is connected to the oscillator circuitry via a quarter wavelength cable. The oscillator circuitry remains at rest while the resonator is vibrated.

In the result of the vibration-induced-sideband method of measuring acceleration sensitivity, there is a 180° ambiguity in the direction of $\vec{\Gamma}$, i.e., the results cannot distinguish between two oscillators, the $\vec{\Gamma}$'s of which are antiparallel. In the method of Watts, et al., the sensitivity of doubly rotated quartz resonators to voltages applied to the electrodes is used to resolve the ambiguity. When the proper magnitude applied voltage is in-phase with the applied acceleration, the sidebands are increased. When the applied voltage is 180° out of phase with the acceleration, the sidebands are decreased. The method allows not only the determination of the sign of $\vec{\Gamma}$, but also the elimination of cable vibration effects.

Shock Tests

The shock testing of a frequency source generally consists of measuring the frequency or phase of the source before and after exposing the device to the specified shock. The phase deviation resulting from the shock (which is the time integral of the fractional frequency change) can provide useful information about the frequency excursion during the shock (including the possible cessation of operation).

Safety Issues

During acceleration sensitivity testing, one must ensure both the operator's and the equipment's safety. Exposing the operator to high intensity noise may cause permanent hearing loss. In the U.S.A., OSHA [24] regulations require employers to provide protection against the effects of high noise exposures, and to administer "a continuing and effective hearing conservation program," whenever the noise exposures exceed specified levels. The permissible noise exposures are functions of both the sound levels and the exposure durations.

For both the operator's and the equipment's safety, all parts subjected to testing must be securely fastened. The forces generated during vibration testing can be high enough to shear the bolts that hold down the equipment.

Interactions and Pitfalls

The two major influences that can interact with the effects of acceleration during testing are thermal effects and magnetic field effects. If the oscillator's temperature changes during acceleration-sensitivity testing, then the temperature-induced frequency shifts can interfere with measurement of the acceleration-induced frequency shifts, as is discussed above in the "2-g Tipover Test" section, for example. Another example of interference by thermal effects is the cooling due to increased air flow during testing in a centrifuge.

Similarly, ac magnetic fields can produce sidebands which can interfere with the vibration-induced sidebands, and dc magnetic fields can produce frequency offsets in atomic frequency standards. Two sources of magnetic fields are the earth's field, and coils in shake-tables and centrifuges. Induced ac voltages due to motion in the earth's magnetic field and due to magnetic fields of the shake table can affect, e.g., varactors, AGC circuits, and power supplies. Since the frequency of a vibration-induced ac voltage is the vibration frequency, the sidebands due to ac voltages are superimposed on the vibration-induced sidebands. One solution to shake-table-produced magnetic fields is to use hydraulic shakers. Such devices are less commonly available than electrodynamic shakers, and have a lower frequency range.

Resonance phenomena can lead to other pitfalls in the determination and specification of acceleration sensitivity. Resonances can occur not only within the oscillator, but also in the test setup and in the platform where the oscillator is to be mounted. Figure 5 shows test results for an oscillator that had a resonance at 424 Hz. (The resonance was traced to a flexible circuit board within the

oscillator.) The resonance amplified the acceleration sensitivity at 424 Hz by a factor of 17. It is, therefore, important to test oscillators at multiple vibration frequencies (with either a series of sinusoidal vibration frequencies or with random vibration) in order to reveal resonances. It is also important to determine the resonances in the platform where the oscillator is to be mounted, and to take that information into account during the specification of acceleration sensitivity.

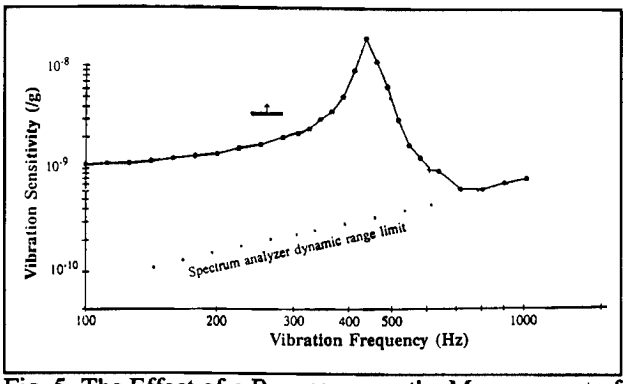


Fig. 5. The Effect of a Resonance on the Measurement of Acceleration Sensitivity vs. Vibration Frequency.

The accelerometers used in vibration-sensitivity testing have non-ideal frequency responses, usually at both low (near DC) and high frequencies. The useful frequency range at the high end is limited by resonances in the accelerometer. The limitations of the accelerometer can be measured, and can also usually be obtained from the manufacturer. The limitations should be taken into account during acceleration-sensitivity testing. Similarly, the limitations of other components in the test setup, e.g., the spectrum analyzer, the signal generator and the shake table must be taken into account (e.g., the shake table may produce vibrations transverse to the intended direction). Another factor to consider is that spectral responses at other than the vibration frequency can occur, as was discussed earlier.

Vibration isolation has been proposed as the "fix" for the acceleration sensitivity of frequency sources. The pitfalls of using such a "fix" are: 1) isolation systems have a limited frequency range of usefulness; outside this range, the isolation systems amplify the problem; 2) a single isolator isolates the vibration primarily along a single direction; 3) isolation systems add size, weight and cost; and 4) most isolation systems are ineffective against acoustic noise. Figure 6 illustrates the frequency response of a typical passive vibration isolator. It shows that, although such a device can be effective at high frequencies, it amplifies the problem at low frequencies, in the region of the isolator's resonant frequency.

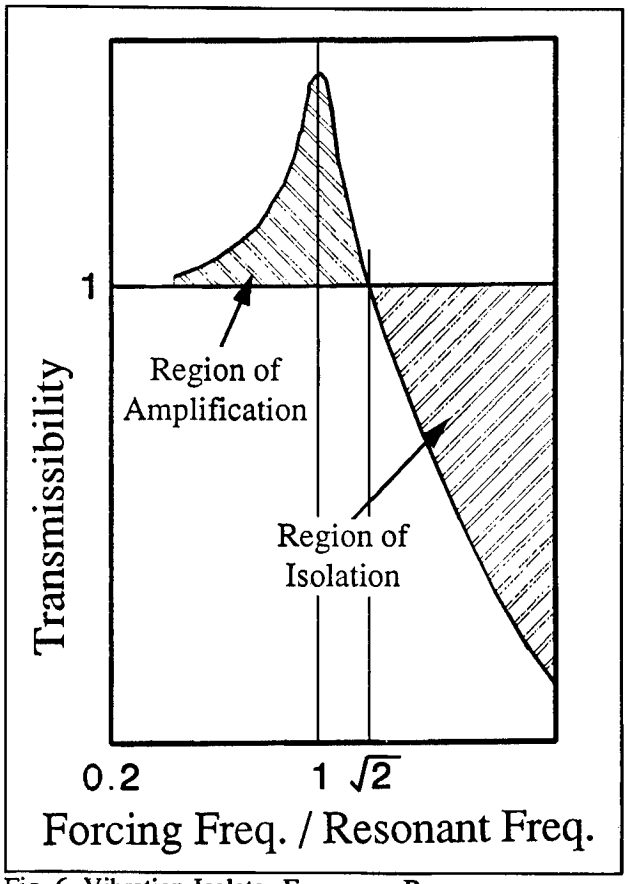


Fig. 6. Vibration Isolator Frequency Response.

Acceleration Related Specifications and General References

Documents, in addition to MIL-O-55310, which are frequently called out in the specification of the acceleration sensitivity testing of frequency sources are MIL-STD-202, "Test Methods for Electronic and Electrical Component Parts," and MIL-STD-810, "Environmental Test Methods and Engineering Guidelines." Some of the acceleration-related paragraphs in these standards are used for survival tests. General information on shock and vibration testing can be found in books [25-27].

Appendix A
Frequency Shift in Atomic Beam Standards due to "Rocking" (Low Frequency Sinusoidal Angular or Linear Motion)

The following analysis addresses an average frequency offset that is observed in cesium beam frequency standards caused by low frequency periodic motion. This motion can be either linear or angular in nature and for the purposes of this discussion is called rocking.

The only components of the cesium standard that are appreciably affected by rocking are the cesium beam tube and the quartz flywheel oscillator. Conventional beam tubes are sensitive to linear acceleration in the direction that is normal to both the beam and beam ribbon. The sensitivity in this direction is due to the fact that the deflection of the beam caused by the acceleration is parallel to the desired deflections produced by the state separation magnets. The beam tube consequently is also very sensitive to angular motion about an axis that is normal to the beam but in the plane of the beam ribbon. This is due to the effective acceleration associated with the Coriolis effect on the moving atoms in the beam when viewed in a coordinate system fixed to the tube. This effective acceleration is also parallel to the desired deflections. Periodic accelerations of both these types evidence themselves as amplitude modulation of the recovered signal, background modulation, and linewidth modulation. Angular rocking, as will be shown later, can lead to large effective accelerations. Also there is a $\pi/2$ shift in the phase of the effective acceleration for low angular rocking frequencies. All the phase angles are essential elements of the analysis.

The main effect on the quartz flywheel oscillator at these low frequencies is frequency modulation caused by change in orientation of the crystal with respect to the gravitational g vector due to angular rocking. This is an important effect at the present state of the technology of quartz resonators. The frequency modulation that is important in cesium beam standards is the residual modulation that remains with the standard's frequency control loop operating. At frequencies below the loop cutoff frequency, the frequency modulation of the free running oscillator is reduced and its phase is shifted by the attenuation characteristics of the servo loop.

An assumption used in this first analysis is that the intentional modulation used to find the line center is sinusoidal frequency or phase modulation at a low frequency. Then the normalized beam tube output signal under modulation and rocking but neglecting any constant background can be approximated as

$$I(t) = B(t) + \frac{1 + A(t)}{2}\left[1 + \cos\left(\frac{k\,F(t)}{1+L(t)}\right)\right] \qquad \text{(A.1)}$$

where

$k = \pi/f_L$,

f_L = static beam tube linewidth, Hz,

$F(t) = \Delta f + M\cos(m) + f1\cos(v + \phi_{f1}) + f2\cos(2v + \phi_{f2})$,

Δf = average frequency departure from line center, Hz,

M = peak frequency modulation of the intentional line-center-finding-modulation, Hz,

$m = \omega_m t$, where ω_m is the modulation angular frequency, radians/sec,

f1 = residual rocking peak frequency amplitude of quartz flywheel (multiplied to cesium frequency) in Hz at the rocking frequency,

f2 = residual rocking peak frequency amplitude of quartz flywheel (multiplied to cesium frequency) in Hz at second harmonic of rocking frequency,

$v = \omega_v t$, where ω_v is the rocking angular frequency,

ϕ_x = phase angles for the time responses to rocking relative to the phase of the rocking position function (this applies to either linear or angular rocking position),

$L(t) = lw\cos(v + \phi_{lw}))$, fractional dynamic beam tube linewidth modulation,

lw = peak fractional change in linewidth due to rocking,

$A(t) = a\cos(v + \phi_a)$, fractional dynamic amplitude modulation,

a = peak fractional amplitude change due to rocking,

$B(t) = b\cos(v + \phi_b)$, dynamic background signal,

b = background amplitude due to rocking,

f_o = line center frequency, Hz, and

$\omega_o = 2\pi f_o$, radians/sec.

Note that residual rocking frequency peak amplitudes, f1 and f2, are the values with the cesium servo loop operating (closed-loop values), not the free-running oscillator values. The servo loop acts as a high-pass filter that attenuates the free-running amplitudes when the rocking frequency is below the loop cutoff frequency, 20 dB/decade for a single integrator loop and 40 dB/decade for a double integrator loop. There is also phase shift associated with the attenuation, $\pi/2$ radians for a single integrator loop and π radians for a double integrator loop when the rocking frequency is far below servo loop cutoff.

The calculation is done as follows. The servo loop of

The calculation is done as follows. The servo loop of the cesium standard always forces the amplitude of the fundamental modulation frequency component of I(t) (as sensed by a synchronous detector with the modulation signal as reference) to be zero with a response time constant determined by the loop characteristics. This is accomplished by controlling the average microwave excitation frequency to the cesium beam tube. The microwave frequency is synthesized from the quartz flywheel oscillator. If the intentional modulation is distortion free, the microwave spectrum free of extraneous sidebands, and there is no rocking, then the average frequency will be at the center of the line corresponding to $\Delta f = 0$ where we have neglected the very small error due to the Bloch-Siegert effect and any error associated with the small asymmetry caused by the special relativistic effects due to the velocity spread in the beam.

To represent this behavior mathematically, I(t) is first multiplied by cos(m) to perform the synchronous detection and the result is expanded with the assumption that Δf, f1, f2, and lw are small but that M is large. It is also assumed that the frequency of the rocking is incoherent with and lower than the modulation frequency. The expansion is done to second order in f1 (for reasons to be explained later) and to first order in Δf, f2, and lw; a and b are exactly first order already. The zero frequency term, which contains all the parts that are free of periodic functions of m and/or v, is set equal to zero and the resulting equation is solved for Δf, the average frequency offset in Hz caused by the rocking. The final result is again expanded to keep only first order terms in a, b, f2, and lw and second order terms in f1.

The calculations are fairly long and tedious so we will present only the results here. The average fractional frequency offset due to the lowest order terms of the rocking effects is:

$$\frac{\Delta f}{f_o} = \frac{f1\ lw\ \cos(\phi_f 1 - \phi_{lw})}{2f_o} - \frac{f1\ a\ \cos(\phi_f 1 - \phi_a)}{2f_o} + \frac{f1^2\ f2\ k^2\ \cos(2\phi_f 1 - \phi_f 2)}{8f_o} \qquad (A.2)$$

Several things are apparent. The modulated background term, B(t), contributes nothing since b is absent. The two remaining tube effects, linewidth and amplitude modulations, are multiplied by f1. They can cancel if their phases are the same, that is, if $\phi_{lw} = \phi_a$. This turns out to be the case for many tube designs and the relative amplitudes a and lw can be controlled to a fair degree, giving the capability of moderately good cancellation. Also, since the tube and crystal effects are vectorial in nature, the frequency shift can be reduced by choosing the sensitive directions to be orthogonal in the mechanical design of the whole cesium standard.

As mentioned earlier, the atomic cesium beam experiences a Coriolis effect when there is angular rocking. If the rocking frequency is very low compared with the reciprocal of the beam transit time, the effective linear acceleration of the beam tube undergoing rocking is approximately:

$$acc(t) = 2\ \theta_p\ \omega_v\ v_{beam}\ \cos(\omega_v\ t + \pi/2) \qquad (A.3)$$

when the rocking angle is

$$\theta_{rocking}(t) = \theta_p\ \cos(\omega_v\ t) \qquad (A.4)$$

where

θ_p is the peak rocking angle, radians,

v_{beam} is the velocity of the atoms in the beam, cm/sec, and

t is the time, seconds.

(Acceleration will be cm/sec^2 with the units used above.)

A reasonable value for v_{beam} in a commercial beam tube is $v_{beam} = 1.3 \times 10^4$ cm/sec. If we take $\omega_v = 0.6$ radians/sec (a rocking period of 10 seconds) and $\theta_p = 0.3$ radians (17°), we get a peak effective acceleration of about 4680 cm/sec^2 or 4.8 g. This illustrates that the Coriolis effect can easily lead to large effective accelerations.

Note the phase shift mentioned earlier of $\pi/2$ radians between the effective acceleration and the rocking angle. For linear rocking there is a phase shift of π radians between the rocking angle and the acceleration. These phase angles affect the frequency shift through the cosine factors in Equation (A.2).

The third term in Equation (A.2) is independent of the beam tube. It involves the quartz flywheel oscillator only and consequently is second order in f1 and first order in f2. This is the reason why second order terms in f1 were kept in the analysis.

The same type of analysis was done also for slow square-wave frequency modulation. The results to lowest orders are identical to those given by Equation (A.2).

To carry the analysis further requires a functional form for the servo loop frequency response of the cesium

single integrator loop, and a second order or double integrator loop. The representation used here for the first order loop frequency response to oscillator frequency change is

$$f1_1 = f1_0 \frac{j\omega_v}{j\omega_v + \omega_{cl}} \qquad (A.5)$$

$f1_0$ is the amplitude of the free-running oscillator response to the acceleration,

$f1_1$ is the closed-loop response amplitude,

ω_v is the angular rocking frequency,

ω_{cl} is the angular cutoff frequency of the servo loop, and

$j = (-1)^{1/2}$, the square root of (-1).

This has magnitude $\omega_v / (\omega_v^2 + \omega_{cl}^2)^{1/2}$ and phase angle $\pi/2 - \tan^{-1}(\omega_v / \omega_{cl})$.

The representation for the second order loop frequency response, chosen to have two real poles in the s-plane, is

$$f1_2 = f1_0 \frac{-\omega_v^2}{-\omega_v^2 + \frac{j\,5\,\omega_v\,\omega_{cl}}{2} + \omega_{cl}^2} \qquad (A.6)$$

This has magnitude $\omega_v^2/[(\omega_v^2 - \omega_{cl}^2)^2 + 25\,\omega_v^2\,\omega_{cl}^2 / 4]^{1/2}$ and phase angle $\pi - \tan^{-1}[2.5\,\omega_v\,\omega_{cl} / (\omega_{cl}^2 - \omega_v^2)]$. Note that this phase angle goes from 0 at very high frequencies through $\pi/2$ at $\omega_v = \omega_{cl}$ to π at very low frequencies.

$f2_1$ and $f2_2$, the second harmonic amplitudes, are obtained from Equations (A.5) and (A.6) with ω_v replaced everywhere by $2\,\omega_v$ and f1 replaced by f2.

Figure A.1 shows the absolute value of the results from Equation (A.2) for angular rocking with the servo loop characteristics given above. The parameters chosen for the plot are:

$\omega_{cl} = 0.3$ rad/sec,

peak rocking angle = 0.3 rad,

tube amplitude or linewidth sensitivity = 0.001 per g,

$f1_0 / f_0 = f2_0 / f_0 = 3 \times 10^{-10}$ (oscillator open-loop fractional frequency sensitivities to the rocking assuming about 1×10^{-9} per g),

$f_l = 370$ Hz (tube static linewidth).

The plot is for the worst cases where there is no cancellation of the beam tube sensitivities and the most sensitive axis of the tube is parallel to the most sensitive axis of the quartz crystal. The frequency axis has been normalized to the loop cutoff frequency. The curve labeled "t-o 1" is for the tube-oscillator interaction with a first order loop (first and second terms in Equation (A.2)); "t-o 2" is for the second order loop. The curve labeled o-o 1 is for the oscillator-only sensitivities, fundamental and second harmonic (third term in Equation (A.2)); "o-o 2" is the corresponding curve for the second order loop. The null at $\omega_v / \omega_{cl} = 1.7$ is caused by the argument of the cosine in Equation (A.2) going through $\pi/2$. The sign of the shift changes on going through the null. The absolute value of the low frequency dependence of the shifts on ω_v / ω_{cl} is given in Table A.1.

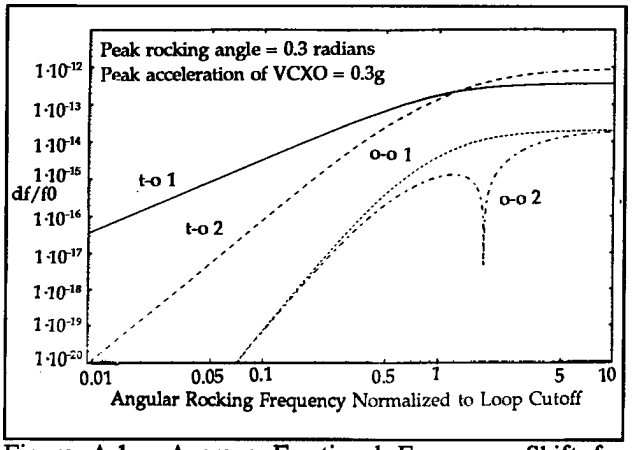


Figure A.1. Average Fractional Frequency Shift for Angular Rocking vs. Normalized Rocking Angular Frequency.

Figure A.2 shows the absolute value of the results from Equation (A.2) for linear rocking. Only the tube-oscillator results are plotted since the oscillator-only results are similar to those for angular rocking except for the fact that the crystal response to the rocking is linear so that any second harmonic must be present in the rocking source. Note that the low frequency dependence for the first and second order loops is the same. They both have $(\omega_v / \omega_{cl})^2$ behavior. This is due to the difference in phase of the effective accelerations for the angular versus linear rocking. Again, the null is caused by the argument of the cosine going to zero at $\omega_v / \omega_{cl} = 1$; 1 g peak acceleration has been assumed corresponding to a peak amplitude of about 25 cm at 1 Hz rocking frequency or 25 m at 0.1 Hz. Thus, linear rocking at these low frequencies is not as important as angular rocking.

For the tube-oscillator interactions, the shift scales as (acceleration)^2. For the oscillator-only effect, the shift scales as (acceleration)^3.

Table A.1
Low Frequency Dependence of Shifts on
ω_v / ω_{cl}.

Interaction Type in Figure A.1	Low Frequency Dependence
t-o 1	$(\omega_v / \omega_{cl})^2$
t-o 2	$(\omega_v / \omega_{cl})^4$
o-o 1	$(\omega_v / \omega_{cl})^6$
o-o 2	$(\omega_v / \omega_{cl})^6$

It is clear that, to keep the shifts from rocking small the loop cutoff frequency should be made as high as possible (loop time constant should be short) so that ω_v/ω_{cl} is small.

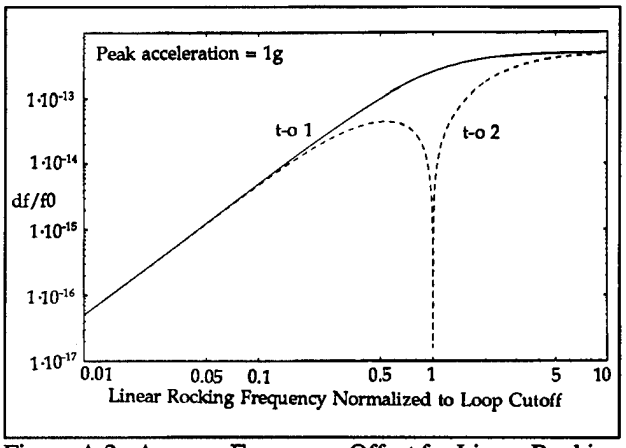


Figure A.2. Average Frequency Offset for Linear Rocking vs. Normalized Rocking Angular Frequency.

If the amplitude and/or frequency of the angular rocking gets large the tube behavior can become very nonlinear and this analysis will no longer be valid. Putting effort into good tube design to get cancellation or compensation of the acceleration effects, along with proper care with regard to the relative orientations of the tube and oscillator, can reduce the offset by a factor of ten or more. This is important if the shifts are to be kept less than 1 x 10^{-13} since the worst case results can be about 1 x 10^{-12} for 0.3 radians peak rocking angle at a frequency several times the loop cutoff frequency.

Vibration at the modulation frequency is not covered by this calculation. The situation is more complex and depends much more on the details of the modulation. If the vibration frequency exactly equals the modulation frequency the shifts can be very large and the magnitude and sign will depend on the relative phase of the modulation and vibration. If the vibration frequency is close to the modulation frequency, there will be large frequency excursions at the difference frequency and these could drive the loop into nonlinear behavior and give an average frequency offset.

Appendix B - Analysis of 2-g Tipover Test Results

The effect of a gravitational field on an oscillator is a body force acting upon the internal components. Acceleration is defined as the second derivative of displacement with time. The effect of a gravitational field is opposite in sense to an acceleration effect. Figure B.1 shows this relationship between a gravitational field and acceleration. One way to handle this effect is to use $\vec{a} = -\vec{g}$.

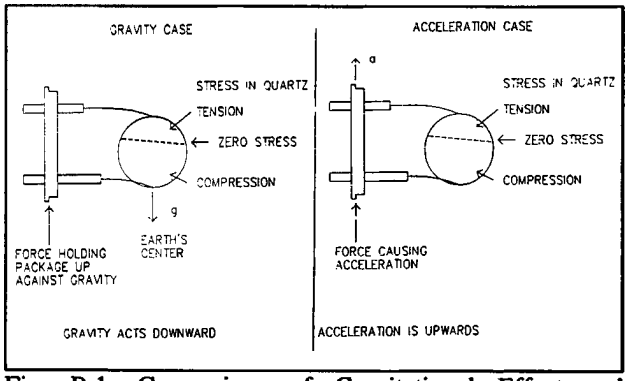


Fig. B.1. Comparison of Gravitational Effect and Acceleration Effect Showing that Opposite Direction Definition Causes the Same Stress Pattern in the Quartz and, thus, the Same Frequency Shift.

Referring to Figure B.2, let $\hat{x}$, $\hat{y}$ and $\hat{z}$ be the unit vectors fixed to the major axes X, Y, and Z of the oscillator; $\hat{i}$, $\hat{j}$ and $\hat{k}$ be the unit vectors fixed to the test bench; and $\vec{g}$ be gravity, of magnitude g and direction $-\hat{j}$. Also, let γ_x, γ_y and γ_z be the components of $\vec{\Gamma}$ along $\hat{x}$, $\hat{y}$ and $\hat{z}$ so that

$$\vec{\Gamma} = \gamma_x \hat{x} + \gamma_y \hat{y} + \gamma_z \hat{z} \qquad \text{(B.1)}$$

In the 2-g tipover test, $f(\vec{a}) = f(\vec{g})$, which can be expressed as (using $\vec{a} = -\vec{g}$)

$$\begin{aligned} f(\vec{g}) &= f_o(1 - \vec{\Gamma} \cdot \vec{g}) \\ &= f_o[1 + g\gamma_x(\hat{j}\cdot\hat{x}) + g\gamma_y(\hat{j}\cdot\hat{y}) + g\gamma_z(\hat{j}\cdot\hat{z})] \end{aligned} \qquad \text{(B.2)}$$

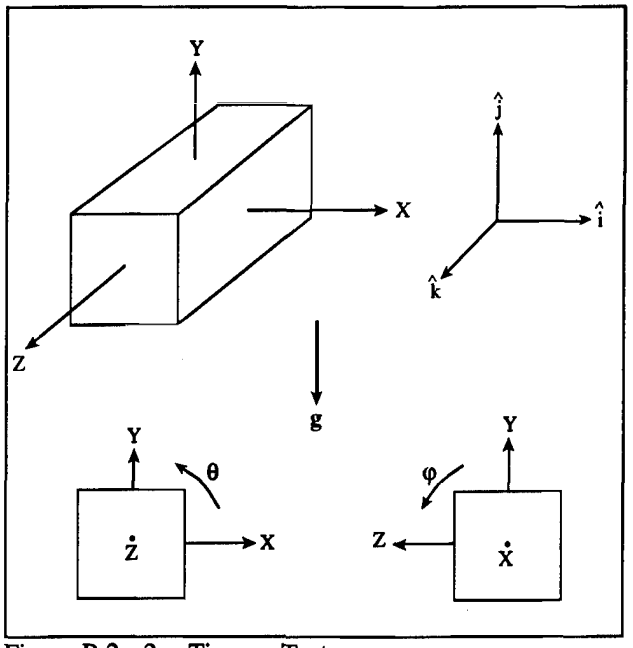


Figure B.2. 2-g Tipover Test

If, initially, $\hat{x}$, $\hat{y}$ and $\hat{z}$ are parallel to $\hat{i}$, $\hat{j}$ and $\hat{k}$, and if we start the test by rotating the oscillator counterclockwise about the $\hat{z}$ (or $\hat{k}$) axis, and if we define the angle between $\hat{i}$ and $\hat{x}$ as θ, then

$$\bar{a} = \bar{A} \cos(2\pi f_v t), \tag{B.3}$$

and $f(\vec{g})$ becomes a function of θ, where

$$\hat{j}\cdot\hat{x} = \cos(90°-\theta) = \sin\theta, \tag{B.4}$$

$$\hat{j}\cdot\hat{y} = \cos\theta, \tag{B.5}$$

$$\hat{j}\cdot\hat{z} = \cos(90°) = 0, \tag{B.6}$$

$$f(\theta) = f_o(1+g\gamma_x\sin\theta + g\gamma_y\cos\theta). \tag{B.7}$$

The normalized frequency change during rotation, per g, is

$$\begin{aligned}\frac{\Delta f(\theta)}{f_o g} &= \frac{f(\theta)-f(0)}{f_o g}\\ &= (1+\gamma_x\sin\theta+\gamma_y\cos\theta)-(1+\gamma_y)\\ &= \gamma_x\sin\theta+\gamma_y(\cos\theta-1)\end{aligned} \tag{B.8}$$

which can be written as

$$\frac{\Delta f(\theta)}{f_o g} = M\sin(\theta+D) - M\sin D \tag{B.9}$$

where

$$M^2 = \gamma_x^2+\gamma_y^2;\quad D = \tan^{-1}\left(\frac{\gamma_y}{\gamma_x}\right);\quad \gamma_x = M\cos D;\quad \gamma_y = M\sin D. \tag{B.10}$$

Equation (B.8) shows that $\Delta f(\theta)$ vs. θ is a sinusoidal function.

From a least-squares fit to the $\Delta f(\theta)$ vs. θ data (e.g., to the 16 data points while rotating 360° in 22.5° steps about one axis, as shown in Figure 1), we can determine γ_x and γ_y. To determine γ_z, let us rotate the oscillator about the $\hat{x}$ axis, starting with $\hat{x}$ parallel to $\hat{k}$, $\hat{y}$ parallel to $\hat{j}$, and $\hat{z}$ parallel to $-\hat{i}$. If ϕ is the angle between $-\hat{z}$ and $\hat{i}$, then

$$\hat{j}\cdot\hat{x} = 0 \tag{B.11}$$

$$\hat{j}\cdot\hat{y} = \cos\phi \tag{B.12}$$

$$\hat{j}\cdot\hat{z} = \cos(\phi+90°) \tag{B.13}$$

and

$$\frac{\Delta f(\phi)}{f_o g} = \gamma_y(\cos\phi-1) - \gamma_z\sin\phi. \tag{B.14}$$

From a least-squares fit to the $\Delta f(\phi)$ vs. ϕ data, we can determine γ_z, and obtain a self-consistency check on γ_y. Similarly, from a rotation about the y-axis, we can obtain a self-consistency check for γ_x and γ_z.

Appendix C - Acceleration Effects in Crystal Filters

General

Crystal filters are used in frequency sources to improve the output spectrum by attenuating far-out amplitude and phase noise as well as harmonics and other spurious outputs. In this application, they are often called spectrum cleanup filters; however, it is well-known that under vibration such filters modulate the signals passing through them, adding as well as removing vibration-related sidebands. That is, the filter adds its own vibration-induced sidebands to those of the signals it passes. Accordingly, the use of a cleanup filter, or post-filter, is not the solution to every phase-noise problem [18]. In this section, we explore these effects in greater detail.

In a properly designed and constructed crystal filter, the crystal resonators are the principal vibration-sensitive components. To ensure that this is so, care must be given to the selection of other components, particularly inductors, transformers, and variable capacitors, and to mechanical design. These considerations are not trivial, but detailed discussion is beyond the scope of this document, which will consider only crystal-related effects.

Consider a frequency source whose nominally sinusoidal output is frequency-modulated due to vibration

and then passed through a filter which is also subject to vibration. (The two vibration environments will be assumed to be the same, although this is not necessarily the case; e.g., if either the source or the filter is protected by vibration isolators.) The filter, in turn, modulates both the carrier and the vibration sidebands of the source; however, modulation of the latter, being a second-order effect, can ordinarily be neglected. Hence, the vibration sidebands of the filter output are essentially the vector sum of the sidebands of the source, attenuated and phase-shifted by the filter, and the sidebands due to the modulation of the carrier by the filter.

Analytically, a filter under vibration is a time-varying linear system. At vibration frequencies which are much less than one-half the filter passband width, the modulation due to vibration of the filter can be estimated from a quasi-static analysis. In the quasi-static case, if all of the resonators have identical $\vec{\Gamma}$'s, acceleration produces a shift of the filter center frequency equal to the frequency shift of the resonators, $\Delta f = \vec{\Gamma} \bullet \vec{A} \bullet f_c$. This modulates the phase of the carrier by $2\pi\Delta f \bullet Tg(f_c)$, where $Tg(f_c)$ is the filter group delay at the carrier frequency. When the resonators have different $\vec{\Gamma}$'s, a quasi-static analysis can still be performed by considering the sensitivity of the filter phase characteristic to each resonator frequency.

While the quasi-static viewpoint is useful for some signal-processing applications, in spectrum cleanup applications very narrow filters are often used for which the quasi-static analysis is inadequate. In the next section, an expression for the vibration-induced sidebands in a one-pole filter, valid for all vibration frequencies, is given; then the source's sidebands are added in to obtain the net effect of the filter. Normalized results for a two-pole Butterworth post-filter are also discussed.

Single-Resonator Filter

The inadequacy of the quasi-static point of view was shown for a single-resonator (one-pole) filter by Horton and Morley [28] who observed that for sinusoidal vibration at frequency f_v and a sinusoidal input at a carrier frequency, f_c, the output is a phase-modulated signal[3] for which, for small phase deviation,

$$\mathcal{L}'(f_v) = 20 \log \left(\frac{\vec{\Gamma} \cdot \vec{A} \cdot Q_L}{[1 + (2Q_L f_v / f_c)^2]^{1/2}} \right). \quad (C.1)$$

Since the 3 dB bandwidth of a one-pole filter is

$$BW = f_c / Q_L, \quad (C.2)$$

where Q_L is the loaded Q, Equation (C.1) can be written

$$\mathcal{L}'(f_v) = 20 \log \left(\frac{\vec{\Gamma} \cdot \vec{A} \cdot (f_c / BW)}{[1 + (f_v / BW/2)^2]^{1/2}} \right). \quad (C.3)$$

From Equations (C.1) and (C.3) it can be seen that the phase modulation exhibits a frequency dependence corresponding to a one-pole filter gain characteristic. For vibration frequencies within the filter passband, the denominator in Equation (C.3) is approximately 1 and the vibration-induced phase noise is nearly constant at the level predicted by quasi-static analysis. On the other hand, for vibration frequencies far removed from the passband, Equation (C.1) is asymptotic to Equation (4). Thus, at low vibration frequencies the filter's contribution to vibration-induced phase noise is less than that of a crystal oscillator using the same resonator (or, in general, any frequency source having the same acceleration sensitivity) while at high vibration frequencies the two contributions are the same. This is illustrated in Figure C.1. The curve labelled "filter" represents the phase modulation of a sinusoidal carrier by the filter; "oscillator," the phase modulation present on the unfiltered oscillator output; and "filter + oscillator" the phase modulation at the filter output, discussed below. The vibration frequency is normalized to one-half the 3 dB bandwidth, BW, of the filter; the sideband levels are shown relative to the unfiltered oscillator sideband level at BW/2.

The curve labelled "Filter + Oscillator" represents the vector sum of the sidebands of the source, attenuated and phase-shifted by the filter, and the sidebands added by the filter.[4] It can be seen that for identical gamma vectors identically oriented, the single-pole filter does not improve the output phase modulation, but rather degrades it by as much as 2.5 dB in the vicinity of the passband edges. (However, if the gamma vectors of the source and filter were oriented anti-parallel instead of parallel, some improvement would occur in the same region.) For

[3] There is, of course, some amplitude modulation as well. Because the amplitude response of the filter is an even fuction of frequency, while the phase response is odd, the amount of AM is generally negligible. Further, if the filter center frequency and the carrier frequency are the same, then the AM sidebands occur only at $2f_v$, $4f_v$, etc.

[4] There is a further $\pi/2$ phase shift between the sidebands of the frequency source, which arise from a frequency modulation process, and the sidebands due to the filter, which are produced by phase modulation.

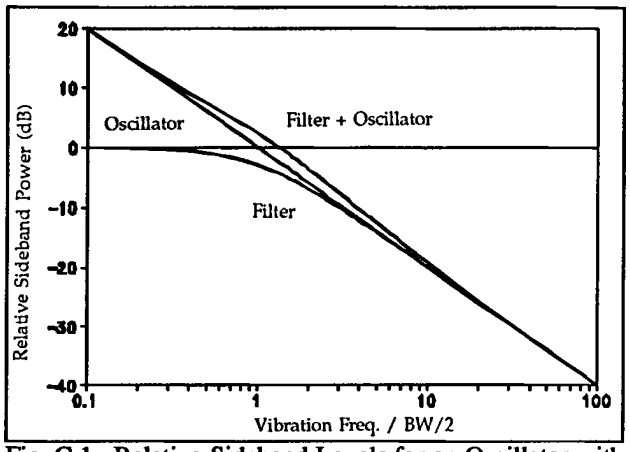


Fig. C.1. Relative Sideband Levels for an Oscillator with and without a One-pole Post-Filter when the Filter Resonator has the Same Acceleration Sensitivity as the Oscillator.

vibration frequencies much greater than or much less than BW/2, the vibration-induced sideband level is essentially unchanged by the filter. This means that the filter may still play a useful role in cleaning up spurious responses, etc., but in order to reduce the vibration-related sidebands the filter resonator's acceleration sensitivity must be less than the frequency source's.

An example is given in Figure C.2 for a 10 MHz oscillator having a g-sensitivity of 5 x 10^{-10}/g with a single-pole post-filter whose resonator has the same g-sensitivity as the oscillator. For this example, the acceleration is taken as 1 g in the direction of the gamma vector. The filter 3 dB bandwidth is 100 Hz (BW/2 = 50 Hz), corresponding to a loaded Q of 1 x 10^5.

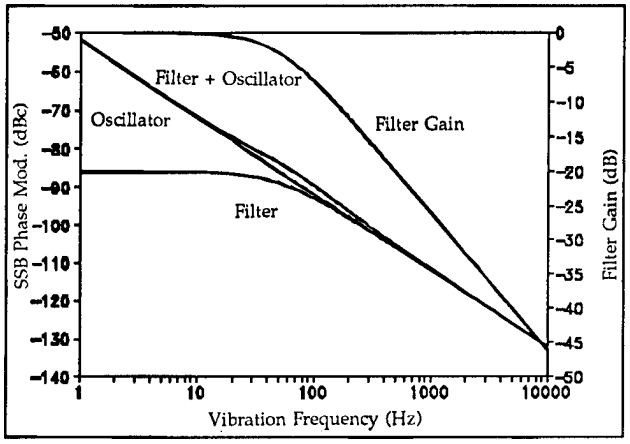


Fig. C.2. Sideband Levels for a 10 MHz Oscillator, $|\vec{\Gamma}|$ = 5 x 10^{-10}/g, with and without a One-pole Post-filter (BW/2 = 50 Hz) Whose Resonator has the same Acceleration Sensitivity as the Oscillator.

To help understand the trade-offs, consider two vibration frequencies. At f_v = 50 Hz, $\mathcal{L}'_v$ is -86 dBc for the oscillator alone, -89 dBc for the filter alone, and -83.7 dBc for the filtered output. If the filter's g-sensitivity were small enough to be neglected (or if it is provided with vibration isolation), then the filtered output would be -89 dBc, since the filter provides 3 dB of attenuation. Similarly, at f_v = 500 Hz, $\mathcal{L}'_v$ is -106 dBc for the oscillator alone, -106 dBc for the filter alone, and -105.2 dBc for the filtered output. If, again, the filter's g-sensitivity were small enough to be neglected, then the filtered output would be -126 dBc, corresponding to the 20 dB of attenuation provided by the filter.

Multiresonator Filters

For the one-pole filter, the vibration-induced phase modulation turned out to be attenuated by the filter gain. For filters of higher order, things are not so simple, even for filters having identical resonators. Since the resonators, each of which is modulating the signal, are distributed throughout the filter network, their respective contributions to the filter's acceleration sensitivity are not the same. At low vibration frequencies, quasi-static analysis is valid; otherwise, a detailed analysis must be performed [18,28].

In general, the vibration-induced resonator frequency changes may differ in magnitude and sign, so that the net effect is to produce both amplitude and phase modulation, although ordinarily, for signals in the passband of the filter, the principal effect is phase modulation. (Further, if the filter assembly is not adequately rigid, the vibration-induced motion of the several resonators may differ in magnitude and phase, in which case analysis is not likely to be practical, since it must take account of structural effects.)

Figure C.3 shows the sideband levels for a two-pole Butterworth filter when the acceleration sensitivities of the two filter resonators are the same in magnitude and orientation as that of the oscillator, and the center frequency of the filter equals the oscillator frequency. It is seen that the filter provides no net improvement in the vibration-induced sidebands and a degradation of 3 dB at the filter passband edge.

The net effect of the filter is more easily seen in Figure C.4, which shows the vibration sideband reduction for the two-pole Butterworth as a function of normalized vibration frequency, with the gamma ratio, defined as $|\vec{\Gamma}_{\text{filter crystal}}|/|\vec{\Gamma}_{\text{source}}|$, as a parameter. Three curves are shown. A gamma ratio of 1 corresponds to the equal-gamma case, Figure C.3; a ratio of 0.5, to the case in which the acceleration sensitivity of the filter crystals is

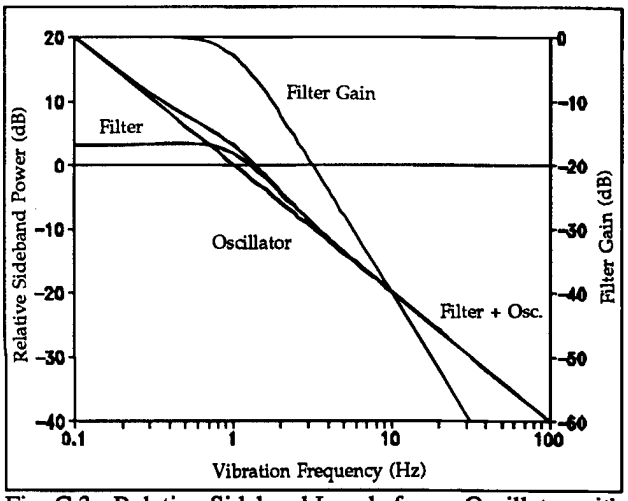


Fig. C.3. Relative Sideband Levels for an Oscillator with and without a Two-Pole Post-filter when the Filter Resonators have the Same Acceleration Sensitivity as the Oscillator.

one-half that of the source; and a ratio of 0.2 to the case in which the acceleration sensitivity of the filter crystals is one-fifth that of the source.

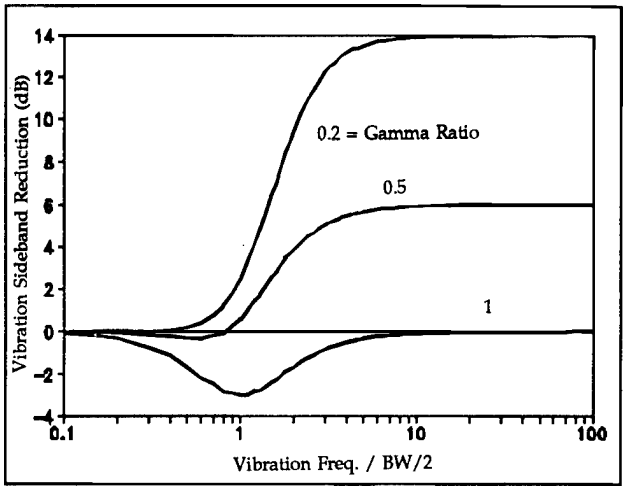


Fig. C.4. Reduction in Vibration Sideband Levels for a Source with a Two-pole Butterworth Post-filter, for Three Values of the Gamma Ratio, $|\Gamma_{filter\ crystal}|/|\Gamma_{source}|$.

The relative contributions of the filter resonators can only be touched upon here, since a given filter characteristic can be realized by many different circuit configurations. If our two-pole example is realized as a ladder network or as two one-pole half-lattice sections in tandem, then, at low vibration frequencies, both resonators contribute equally, while at frequencies in the stopband the resonator nearer the output port is the primary contributor to vibration-induced modulation. Additionally, the relative contributions of the resonators may be expected to change as a function of the frequency of the source relative to the filter center frequency, so that any cancellation technique will be limited to a narrow range of vibration frequencies.

Conclusions About the Use of Crystal Filters for Spectrum Cleanup

While spectrum cleanup filters play a valuable role in the reduction of spurious outputs in frequency sources, they do not solve the problem of acceleration sensitivity. The examples given here are representative: in order for a post-filter to reduce the vibration-induced sidebands of a frequency source, it must have either mechanical isolation or resonators whose g-sensitivity is better than that of the frequency source. Although the treatment has been in terms of sinusoidal vibration, the extension to other vibration spectra is obvious.

Acceleration effects in crystal filters are also of importance in a variety of signal-processing applications which are beyond the scope of this document. The considerations for these are somewhat different than for spectrum cleanup, although the basic relations are the same.

Appendix D

Acceleration-Relevant Paragraphs from MIL-O-55310C[5]

3.7.16 Short-term stability.

3.7.16.2 Phase noise, random vibration.

3.7.16.3 Phase noise, acoustic.

3.7.17 Acceleration sensitivity.

3.7.17.1 Acceleration sensitivity, steady-state.

3.7.17.2 Acceleration sensitivity, 2g tip-over.

3.7.17.3 Acceleration sensitivity, vibration.

3.7.31 Accumulated time error.

3.7.32 Clock accuracy.

3.7.37 Vibration.

[5] In MIL-O-55310, section 3 (paragraphs 3.1 to 3.9.2.6) contains the "Requirements" and section 4, the corresponding "Quality Assurance Provisions." Section 6 contains definitions.

3.7.37.1 Vibration, sinusoidal (nonoperating).

3.7.37.2 Vibration, sinusoidal (operating).

3.7.37.3 Vibration, random (operating).

3.7.37.4 Vibration, random (nonoperating).

3.7.38 Acoustic noise.

3.7.39 Shock.

3.7.40 Acceleration.

3.7.40.1 Acceleration (nonoperating).

3.7.40.2 Acceleration (operating).

References

[1] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, pp. 297-305, 1988.

[2] W. P. Hanson, and T. E. Wickard, "Acceleration Sensitivity as a Function of Temperature," Proc. of the 43rd Ann. Symp. on Frequency Control, pp. 427-432, 1989, IEEE Catalog No. 89CH2690-6.

[3] W. J. Riley, Jr., "The Physics of the Environmental Sensitivity of Rubidium Gas Cell Atomic Frequency Standards," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 39, pp. 232-240, 1992.

[4] T. J. Lynch, and W. J. Riley, "Tactical Rubidium Frequency Standard (TRFS)," Final Technical Report, RADC-TR-87-166, Vol. 1, 1987, AD-A192981.

[5] T. M. Kwon, and T. Hahn, "Improved Vibration Performance in Passive Atomic Frequency Standards by Servo-Loop Control," Proc. of the 37th Ann. Symp. on Frequency Control, pp. 18-20, 1983, AD-A136673.

[6] T. M. Kwon, B. C. Grover, and H. E. Williams, "Rubidium Frequency Standard Study," RADC-TR-83-230, Final Technical Report, Rome Air Development Center, Oct. 1983, AD-A134713.

[7] T. M. Kwon, R. Dagle, W. Debley, H. Dellamano, T. Hahn, J. Horste, L. K. Lam, R. Magnuson, and T. McClellan, "A Miniature Tactical Rb Frequency Standard," Proc. of the 16th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 143-155, Nov. 1984.

[8] C. Audoin, V. Candelier, and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards due to an Intermodulation Effect," IEEE Transactions on Instrumentation and Measurement, Vol. 40, pp. 121-125, April 1991.

[9] C. Audoin, N. Dimarcq, and V. Giordano, "Physical Origin of the Frequency Shifts in Cesium Beam Frequency Standards Related Environmental Sensitivity," Proc. of the 22nd Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 419-440, 1990, AD-A239372.

[10] E. M. Mattison, "Physics of Systematic Frequency Variations in Hydrogen Masers," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, pp. 250-255, 1992.

[11] "IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology," IEEE Std 1139-1988, The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017, USA.

[12] S. R. Stein and J. R. Vig, "Communications Frequency Standards," in: The Froehlich/Kent Encyclopedia of Telecommunications, Vol. 3 (F. E. Froehlich and A. Kent, eds.), Marcel Dekker, Inc., New York, 1992, pp. 445-500. A reprint of this chapter is available under the title "Frequency Standards for Communications," as U.S. Army Laboratory Command Technical Report SLCET-TR-91-2 (Rev. 1), October 1991, NTIS Accession No. AD-A243211.

[13] P. Renoult, E. Girardet, and L. Bidart, "Mechanical and Acoustic Effects in Low Phase Noise Piezoelectric Oscillators," Proc. of the 43rd Ann. Symp. on Frequency Control, pp. 439-446, 1989, IEEE Catalog No. 89CH2690-6.

[14] R. D. Weglein, "The Vibration-Induced Phase Noise of a Visco-Elastically Supported Crystal Resonator," Proc. of the 43rd Ann. Symp. on Frequency Control, pp. 433-438, 1989, IEEE Catalog No. 89CH2690-6.

[15] J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemically Polished Quartz," Proc. of the 31st Ann. Symp. on Frequency Control, pp. 131-143, 1977, AD-A088221.

[16] R. L. Filler, L. J. Keres, T. M. Snowden, and J. R. Vig, "Ceramic Flatpack Enclosed AT and SC-cut Resonators," Proc. of the 1980 IEEE Ultrasonics Symposium, pp. 819-824, 1980.

[17] R. L. Clark and M. K. Yurtseven, "Spurious Signals Induced by Vibration of Crystal Filters," Proc. IEEE Ultrasonics Symp., pp. 365-368, 1988, IEEE Catalog No. 88CH2578-3.

[18] R. C. Smythe, "Acceleration Effects in Crystal Filters - a Tutorial," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 39, pp. 335-340, May 1992.

[19] "MIL-O-55310, Military Specification, Oscillators, Crystal, General Specification for," the latest revision is available from: Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

[20] R. Brendel, C. El Hasani, M. Brunet, and E. Roberts, "Influence of Magnetic Field on Quartz Crystal Oscillators," Proc. of the 43rd Ann. Symp. on Frequency Control, pp. 268-274, 1989, IEEE Cat. No. 89CH2690-6.

[21] D. J. Healy, III, H. Hahn, and S. Powell, "A Measurement Technique for Determination of Frequency vs. Acceleration Characteristics of Quartz Crystal Units," Proc. of the 37th Ann. Symp. on Frequency Control, pp. 284-289, 1983, AD-A136673.

[22] M. M. Driscoll, "Quartz Crystal Resonator G Sensitivity Measurement Methods and Recent Results," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 37, No. 5. pp. 386-392, September 1990.

[23] M. H. Watts, E. P. EerNisse, R. W. Ward, and R. B. Wiggins, "Technique for Measuring the Acceleration Sensitivity of SC-cut Quartz Resonators," Proc. of the 42nd Ann. Symp. on Frequency Control, pp. 442-446, 1988, AD-A217275.

[24] Occupational Safety and Health Administration, 29 CFR Ch. XVII, para. 1910.95, "Occupational noise exposure," 1 July 1987.

[25] W. Tustin, and R. Mercado, Random Vibration in Perspective, Tustin Technical Institute, Inc., 1984.

[26] D. S. Steinberg, Vibration Analysis for Electronic Equipment, John Wiley & Sons, 1988.

[27] C. M. Harris, ed., Shock and Vibration Handbook, 3rd ed., McGraw-Hill, NY, 1988.

[28] T. S. Payne, "Improved Tactical Miniature Crystal Oscillator," First Interim Report, SLCET-TR-86-0011-1, U.S. Army Laboratory Command, Fort Monmouth, New Jersey, November 1987, AD-B124235. The section of this report that deals with the vibration sensitivity of crystal filters was prepared by W. Horton and P. Morley.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

PRECISION OSCILLATORS: DEPENDENCE OF FREQUENCY ON TEMPERATURE, HUMIDITY AND PRESSURE

WORKING GROUP 3 OF THE IEEE SCC27 COMMITTEE

Membership as follows:

David W. Allan, Chairman
James Barnes
Franco Cordara
Michael Garvey
William Hanson
Robert Kinsman
Jack Kusters
Robert Smythe
Fred L. Walls

Abstract

Excluding vibration effects, variations in temperature, humidity and pressure (THP) are the most common environmental perturbations on precision oscillators. Thus the modeling, measurement and understanding of how these variations affect the frequency outputs of such devices are very important -- particularly for long-term behavior. In general, the effects of THP on frequency are non-linear and interdependent. For example, the temperature coefficient of a frequency standard is often directly dependent on the level of humidity. Hence, to obtain a complete model for even a single device, let alone a whole model line, would be incredibly complex.

Our purpose here is to arrive at tractable (non-burdensome) guidelines, standards, and precautions for test methods used in determining the dependence of the output frequency of precision oscillators on temperature, humidity, and pressure. Over specification, under specification, or the lack of proper specification will miscommunicate. We offer a perspective for the manufacturer and the designer, as well as the user so that clear understanding and communication can occur. The guidelines, standards and precautions encourage consistency and repeatability for measurement and specification of these environmental sensitivities. We believe very large cost savings will be appreciated if these guidelines, standards and precautions are followed.

Contribution of the U. S. Government, not subject to copyright.

INTRODUCTION AND DEFINITIONS

Excluding vibration effects, variations in temperature, humidity and pressure (THP) are the most common environmental perturbations on precision oscillators. Thus the modeling, measurement and understanding of how these variations affect the frequency outputs of such devices are very important -- particularly for long-term behavior. [1-6] In general the effects of THP on frequency are non-linear and interdependent. For example, the temperature coefficient of a frequency standard is often directly dependent on the level of humidity.

Given this complexity and nonlinear interdependence of these three environmental parameters, it is generally not possible to obtain a complete model for even a single device. Obtaining a general model for just the THP parametric dependence for a given type of device would be incredibly complex. There is also the danger that specifications might indicate that all is known in this regard. This would be very misleading. Our purpose here is to arrive at tractable (non-burdensome) guidelines, standards, and precautions for test methods used in determining the dependence of the output frequency of precision oscillators on temperature, humidity, and pressure.

0-7803-0476-4/92 $3.00 © 1992 IEEE

This is not a specification document for these parameters, but a resource document which gives guidelines for deriving specification statements and for developing methodology that will be efficient and cost effective for the manufacturer, designer and user. In addition, the information herein is intended to help the researcher, those developing precision oscillators and those wanting to get the best possible performance out of a precision oscillator.

We define, in the usual way, y(t) as the relative frequency (ref IEEE Standard 1139-88). This is the actual time dependent frequency minus the nominal frequency all divided by the nominal frequency. Hence, y(t) is a dimensionless number describing the instantaneous frequency offset from the nominal at time t.

We define the THP frequency dependence as the causal effect on y(t):

$$y(t) = f(T,H,P). \qquad (1)$$

We recognize that there may be other environmental parameters which are dependent on one or more of the T, H or P parameters. This problem is addressed later. Nonetheless, equation (1) is a useful conceptual model. In addition, the actual y(t) of an oscillator, of course, will be driven by internal effects. These other effects need to be adequately understood, be held constant or be sorted out in some appropriate way as the effects of T, H and/or P are studied.

In this document "quartz" will typically mean a precision quartz-crystal oscillator, "rubidium" will mean a rubidium gas-cell frequency standard, "cesium" will mean a cesium-beam frequency standard, and "hydrogen" will mean an active hydrogen-maser frequency standard. We now list a set of specific issues before we discuss general issues, guidelines, precautions and design issues.

* In quartz crystal oscillators, "activity dips" may occur. An activity dip occurs when a y(t) versus temperature (T) curve of an unwanted mode intersects the y(t) versus T curve of the wanted mode. Such activity dips are usually highly dependent on the way the quartz crystal resonator is being driven and on the load reactance in the oscillator circuit.

* Thermal hysteresis is also found in quartz crystal oscillators. This means that the frequency versus temperature plot generated as the temperature is increased will not be the same as that generated while decreasing the temperature.

* The pressure effect on a clock should not be confused with an altitude effect. For a high-accuracy clock, frequency will change with altitude due to the gravitational, relativistic "redshift" as measured against a clock at a fixed gravitational potential. This effect is small--about 1×10^{-16} per meter.

* It is preferable to measure the absolute humidity. Since often the relative humidity is given, the temperature must also be known.

* Because of the interdependence of the frequency of precision oscillators on various environmental parameters, it is always good practice to record all relevant data during a measurement. It is also wise to record what may seem to be trivial experimental conditions. Little things, like fans in a room moving the air around an experiment can make a big difference in time constants and apparent temperature and humidity responses of the output frequency.

* Most buyers of a product specify worst-case performance. It is also true that most design specifications are not necessarily written on the same basis. Our general practice is to use a statistical approach where the expected performance plus a guard band is listed as the specification. Companies differ widely in the size of the guard band. The specification is usually derived from a root-mean-square (rms) approach for all of the expected environmental parameters. In other words, in determining the effect of THP on an oscillator, it is not uncommon to assume that the full environmental range will never be met and that it is suitable to take the environmental influence on overall accuracy as the rms value of the individual environmental effects. In some cases this may result in a conflict between the designer and user. For a view of this, observe the data sheets of most typical quartz oscillators.

* We recommend the arithmetic-sum, worst-case approach, as we believe it is more realistic and honest. Its major fault is that it assumes linear dependence and no interaction between external effects. For frequency standards where interactions between outside influences dominate performance, this approach might be too aggressive. This approach also places additional financial burden on the manufacturer to do additional testing, but it should result in happier designers and users. Of course, the best approach lies in characterizing the standard as completely as possible.

GENERAL ISSUES

Because the environment is so enormously important to the long-term performance of precision oscillators and in some cases even to the short-term performance, it is very desirable to quantify the dependence of the oscillator frequency on the relevant environmental parameters. In this regard, it will be useful to come up with models describing these dependencies given the environmental perturbations. To handle all of them in a single model would be very difficult. It is better to break the problem into pieces and consider only those items that are important to the manufacturer, designer and user.

In the following equation we write a general description of the fractional frequency dependence of a precision oscillator as a function of temperature, humidity and pressure:

$$y(T,H,P) = \alpha_0 + \alpha_1 T + \alpha_2 H + \alpha_3 P + \alpha_4 TH + \alpha_5 TP + \alpha_6 HP + \alpha_7 THP. \quad (2)$$

From this we obtain:

$$\begin{aligned} dy = {} & \alpha_1 dT + \alpha_2 dH + \alpha_3 dP \\ & + \alpha_4 TdH + \alpha_4 HdT \\ & + \alpha_5 TdP + \alpha_5 PdT \\ & + \alpha_6 HdP + \alpha_6 PdH \\ & + \alpha_7 THdP + \alpha_7 TPdH + \alpha_7 HPdT. \end{aligned} \quad (3)$$

If the parameter cannot be varied independently, then covariance term must be included [7]. The manufacturer has the responsibility to state which ones of these coefficients are most important and for which type of oscillator. Although the assumption of linearity is almost always useful over small ranges, one of the problems in current commercial specifications of any of the THP parameters is the assumption that the parameters are linear over a large range. We will make some suggestions regarding this problem.

The above equations deal with coefficients which may depend upon the values of parameters other than T, H, or P, and hence will not be constant coefficients. In cases where there is a significant dependency of T, H or P on some other environmental condition, that needs to be stated.

We could write another set of equations that would describe the dynamics. For example, we could use a set of coefficients that are functions of the rates of change of each of the above parameters. Furthermore, in most precision oscillators there will be more than one time constant, hence the frequency-temperature dependence will be a very complex function as it involves the dynamics of the environment. If a non-standard model is used, there should be good motivation (e.g., unusual dependencies) to do so.

Often at "turn on" or during certain transient situations, the output frequency behaves in an exponential way. Where this is the case, the time constant of a particular model gives another method of describing the frequency behavior. In general, we need to distinguish between turn-on or transient behavior and steady-state behavior.

Although mathematical models can be powerful tools, we encourage practical (measurable) specifications and simple models and measurement methods.

Most users tend to prefer specification of worst-case performance; i.e., outer limits which must be met over a range of environmental parameters. However, this often leads to over-specification and undue expenditure. We strongly encourage a dialogue between user and designer so that the final oscillator meets the actual needs at affordable costs of manufacture and test.

There are always those who wish to carry the analysis and the modeling as far as they can. For those so inclined, we will give cautions, guidelines and suggestions that, hopefully, will be helpful in their analysis. However, we will primarily address the typical user, designer, developer and manufacturer.

GUIDELINES

Given the complexity of the problem, as much as practical, it is important to develop methods that will keep all parameters constant, except the one under test. In the case of humidity in precision quartz oscillators, it may be impossible to specify temperature and other dependencies in an open environment where the humidity is not held constant.

In general, the temperature coefficient of quartz oscillators is a strong function of humidity. If a quartz crystal oscillator is sealed against changes in humidity, then the temperature coefficient can be reasonably obtained. Condensed H_2O in a unit can cause drastic changes in performance and should be avoided. This can occur as significant temperature cycling occurs. Another problem with units open to the atmosphere is that the time constant associated with humidity change can be very long. Also, pressure changes can alter the mechanical stress on internal components. To cover this potential sensitivity, measurements of the pressure effect at one temperature are probably required, and this can probably only be done for a sealed unit.

In addition to holding other parameters constant, it is important to achieve steady-state conditions after any change in the parameter under test. Time constants for achieving steady state can vary enormously, but unless steady state is achieved, the transient effects can seriously cloud the estimates of the dependence on a particular environmental parameter.

Obtaining and separating these time constants is one of the first challenges in properly characterizing the effects of the environment. Time constants range from minutes in quartz oscillators (thermal transients lasting up to hours may be a dominant effect in non-SC-cut oscillators) to days in some atomic standards.

In order to separate and determine the dominant effect,we suggest using methods related to statistically controlled design of experiments. One of the simplest views of this is outlined in Bhote [8]. In this book, the dominant effect is known as the "Red X." Assuming that the constants are properly understood, in a relatively crude experiment the major effects can be determined. In this methodology, the variables, T, H, and P, are allowed to assume high and low values in a defined pattern. No real precision is sought, rather the goal is to determine the magnitude of the largest effect whether a single environmental parameter or a combination of them is important.

Once we know the dominant effect, we can then systematically explore it. In some cases there may be no single dominant effect, at least within some measurement precision. At this point, the next step may be to explore each of the variables separately with more precision.

The following are a general set of guidelines:

* Do a crude experiment to determine the dominant effect.

* Examine the dominant effect variable to determine its time constant (careful plotting and analysis are necessary to determine whether there is more than one time-dependent process present).

* Measure both dynamic and static responses to changes in the dominant effect with all other variables held constant. Once valid data are obtained, follow statistical procedures to eliminate the effect of the existing dominant effect, and find the next most significant factor, measure it, eliminate it, and continue to iterate as required.

* Write a specification that outlines the major environmental effects, test conditions and responses. Graphs here are essential. Three-dimensional graphing software make this relatively easy.

* Define an overall accuracy specification and the terms under which it was derived. Identify it explicitly as an rms specification, additive worst-case specification, or whatever other type of specification is being defined.

In general, in equations (2) and (3) above we should identify those terms that are most important and least important for different categories of oscillators. We can make some general statements about the different types of oscillators, but, as with all general statements, there will be exceptions. For example, temperature effects are the dominant factor in most, but not all, quartz crystal oscillators. In rubidium gas-cell frequency standards, T, H, and P are all important, with another caveat regarding the constituents of the atmospheric gasses. With the improper choice of glass, trace helium content becomes a dominant effect in apparent aging. Figure 1 is a classic example of environmental effects on a rubidium gas-cell frequency standard. Cesium-beam frequency standards, prior to recent optimization concepts, were dependent on both temperature and humidity.

RUBIDIUM FREQUENCY STABILITY WITH ENVIRONMENTAL CHANGES

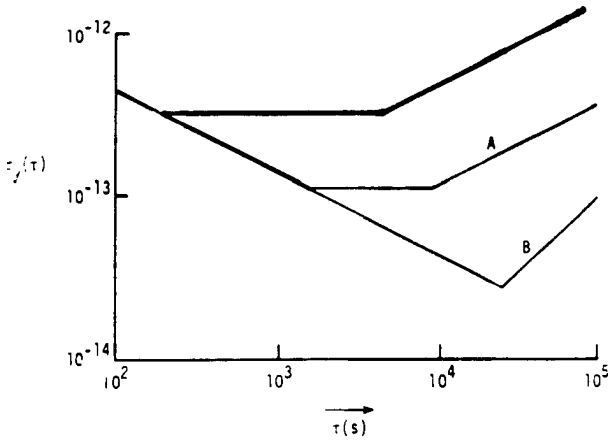


Figure 1. A plot of the random instabilities of a commercial rubidium gas-cell frequency standard with changing environmental conditions. The better the environmental parameters were controlled the better was the long-term frequency stability performance. The top curve is the frequency stability for nominal laboratory environment. Curve (a) additional magnetic shielding and temperature compensation was provided. Curve (b) was with the additional magnetic shielding, temperature compensation and with barometric sealing (pressure control). A frequency drift of approximately 1 part in 10^{13} per day was removed from the data before analyzing the stability.

From most experience, it seems that it is not relative humidity, but absolute humidity that is more important. It has only been in recent years that the humidity dependence on both atomic standards and quartz oscillators has been recognized as a very significant environmental perturber. (see Appendix)

Given the nonlinearities occurring, especially with temperature, it has been suggested that a coefficient be expressed such that the slope, dy/dT would not exceed some value over some specification range. If such significant nonlinearities exist, then the appropriate time constants need to be respected as the effects of these nonlinearities are determined. In general, nonlinearities will be very different between the different types of oscillators.

Since, in most cases, a frequency standard will have more than one thermal time constant, a measurement of steady-state thermal effects should not be made until waiting twice as long as the longest time constant, which takes it 86% of the way to its final value. If more precision is wanted, then the metrologist should wait as long as needed to accomplish the goal.

In some applications the dynamic effects may be more important. In space applications of clocks, because of expense and non-repeatability of the situation, it is best to simulate the dynamics of the space environment over some appropriate range of temperature.

For those pushing the performance of a standard as far as they reasonably can, it is very important to first know the part of the oscillator which is most sensitive to T, H or P, and then control the system with a T, H or P sensor at that point. If this is not practical, then temperature gradients in the environment should be minimized.

THP effects frequently are a cause of long-term frequency drift. Polynomial modeling can be misleading in estimating frequency drift [9]. Having an accurate measurement of frequency drift in precision oscillators is very important for both the manufacturer and the user. If an efficient estimator of the drift is used, this can save large amounts of time and money in the manufacturing process. It has been shown that misleading estimates are all too often obtained using a quadratic least-squares fit to the phase or a linear least-squares fit to the frequency, given the kinds of long-term random variations that are superimposed on top of the drift. Long-term random spectral density models for the frequency modulation are usually 1/f or random-walk in character. For these kinds of random residuals, a

second-difference estimator for the drift is typically more efficient than the two methods mentioned above [10].

The mean second-difference estimate may be somewhat contaminated by higher Fourier frequencies than from the pure random-walk or 1/f model. For example, if white PM or white FM are also present (as they often are) these noise processes can significantly degrade the confidence of the mean second-difference drift estimate. However, a simple second-difference estimate using the first, middle and end data point from the time or phase residuals is very close to an optimum estimator for the above cases. This approach gives a much better confidence on the estimate of the drift as well [10].

White FM tends to be the predominant noise model for integration times of the order of a second out to several thousand seconds in both cesium and rubidium frequency standards. In the case of cesium this model may be appropriate for integration times of the order of a day and even longer. Measuring THP coefficients in the presence of this kind of noise presents a practical signal-to-noise problem. Since the optimum estimate of the mean of a white process is the simple mean, when measuring the effect of changes in these environmental parameters it is best to hold them constant and average the frequency for an interval such that the $\sigma_y(\tau)$ curve starts changing from $\tau^{-1/2}$ toward a flattening (flicker floor) where τ is the integration time over which the frequency is averaged. Then change the environmental parameter being evaluated and repeat the integration time to measure the frequency change. We may improve the precision with which we can determine the change in frequency with a change in an environmental parameter by reiterating the above process several times -- following the above rule for integration time. In principle, if "N" is the number of changes back and forth, then the confidence on the frequency change is the value of $\sigma_y(\tau)$ times $1/\sqrt{N}$. One must respect the settling times after changing a parameter's value as well as other systematics affecting the measurement.

PRECAUTIONS

Setting up proper measurement configurations is critically important. It will always be important to measure the "real" temperature. The exteriors of most frequency sources are not isothermal. Gradients depend on conductive and convective heat transfer; convection especially depends on the presence or absence of forced circulation in the surrounding atmosphere. Most testing in environmental chambers ignores this. In addition, the size of a unit is very important. Typically, the smaller the unit the less important will be the effect of temperature gradients.

The temperature at (and within) a device depends on the interplay between the external heat (or cooling) source and its conductive paths and the internal heat-generating mechanisms (internal ovens, electrical losses, etc.), and their respective conductive paths. Furthermore, the degree of coupling of both external and internal sources determines the various time constants, and thus is critical when one attempts to define quasi-static conditions.

In some instances temperature gradients may be more important than the actual temperature coefficient. For example, if a commercial cesium beam standard is turned upside down to measure the effects of a 2-g tip-over, what seems to be the stronger effect on the frequency is the change in temperature gradients since the convection currents flow in the opposite direction for many of the components. These effects can often be separated because of the difference in time constants.

The effect of thermal gradients in rubidium can easily be the dominant effect in its performance. The frequency drift rate is a strong function of the temperature gradients inside the physics package. As gradients change, the drift rate can change in magnitude and even in sign.

Oscillators that are hermetically sealed should show a totally different character than those that are open or sealed with gaskets that are permeable to moisture or different gasses (especially helium). Manufacturers should specify whether an oscillator is or is not sealed against changes in pressure, humidity and helium. Most plastic and rubber gaskets are permeable to moisture and helium. Open units and those with permeable gaskets will show many nonlinear and transient effects that are not present in sealed units. At high humidity, moisture can condense inside the unit and alter many of the electrical parameters. This effect will persist long after the high humidity has been removed because of the high heat capacity and relatively low vapor pressure of water.

Any change in the orientation of the oscillator during testing can invalidate the data since

frequency changes due to acceleration and magnetic field can range up to 10^{-8}. Orientation of the oscillator under test also enters into the characterization. Physically inverting a quartz crystal oscillator has the potential of significantly changing the test results via temperature gradient, magnetic field and gravitational field. Therefore, repeating tests in different orientations may be necessary.

Polynomial modeling has its drawbacks and is not universally recommended. Actual devices may exhibit polynomial behavior in one property, exponential behavior in another, and something else in a third. In other words, the mathematical model chosen is probably as important as the coefficients used in understanding environmental coefficients. Polynomial modeling with too many coefficients may make the model too device dependant. The number of model parameters should be kept as low as practicable and still provide useful quantitative information.

Intuitively, we might think that we can learn the frequency dependence of a clock by exposing it to a white spectrum of temperature variations. By taking the Fourier transform of the frequency variations cross-correlated with the known temperature spectrum, we might hope to determine the impulse response function. Unfortunately, as clever as this experiment seems, it will fail because the coefficients are usually nonlinear. In addition, there are typically several different thermal time constants for almost any precision oscillator which complicates reaching the end goal.

For many standards, it appears that there is a maximum frequency shift with pressure change followed by a relaxation period. This may depend upon the time rate of change of pressure. Perhaps a maximum allowed shift for a specific pressure change could be easily measured. This might avoid some of the nonlinear characterization problems.

Particular concerns associated with each type of frequency standard are as follows:

Quartz

Orientation, temperature, pressure, humidity, magnetic field and gravitational field all produce significant frequency shifts. Proper sealing of the case can reduce pressure and humidity shifts, but might actually exacerbate orientationally dependent thermal-gradient effects. Over the years, some of these thermal gradient effects have been reduced by relocating heaters and thermistors, and rerouting high-current leads. "Activity dips" in quartz crystal oscillators can cause adverse temperature dependence. If not properly included in a manufacturer's specifications, a system could fail, when in fact the specifications might indicate that the oscillator should work in a normal temperature-dependent fashion over some range of temperature. For example, when an activity dip occurs, there are cases where the magnitude of the temperature coefficient increases by as much as an order of magnitude and even changes sign. The oscillation amplitude may also change, and in the worst case the device might cease oscillation.

Measurement of "Activity dips" is adequately addressed by software designed to implement ANSI/EIA-512 [11]. Following this software design, one scans over a range of frequencies near the desired resonance thus determining the unwanted modes. The additional requirement is that the crystal should be monitored on a temperature-controlled stage. Several years ago activity dips were reported which occurred over millidegree temperature ranges. Therefore, it becomes critical that the temperature-controlled stage be of high precision.

Manufacturers obviously have a responsibility to indicate the presence of activity dips and to specify their impact on the oscillators performance as well as the range of temperatures over which they might occur. If it is believed that none are present over some range of performance, then that also should be stated.

In quartz oscillators, pressure effects should be very small for all sealed units. Oscillators that are sealed in vacuum should have the best pressure performance since outside pressure changes will not have an effect on the internal pressure. Units with small dimensions and/or a strong enclosure should show little response to changing pressure. There is the possibility that pressure changes could alter the mechanical

stress on internal components. To cover this potential sensitivity, measurements of the pressure effect at one temperature are probably required.

Rubidium

These oscillators exhibit all of the effects of quartz, most of which can be reduced by proper sealing. The impact of all effects are reduced by one or two orders of magnitude over that in quartz because of the fundamental use of an atomic resonance in addition to a much higher line-Q. One unique problem for rubidium is that the effect of atmospheric gasses (diffusing into the gas cell) may show up as an unwanted change in frequency drift. The presence of helium in the ambient atmosphere can be devastating to rubidium standards.

Cesium

Cesium standards typically have long time constants. The Ramsey cavity is usually copper or a copper alloy. At one end is an oven at roughly 100° C. At the other end is a hot-wire ionizer at 950-1100° C. Full thermal equilibrium of the cavity may take many hours to reach. Temperature probably affects the physics package more than any other variable. Humidity affects high-impedance, beam-current, output amplifiers. All three parameters (THP) affect power delivered by the harmonic generator, although newer electronic circuit designs reduce these effects.

For cesium beam frequency standards, the temperature coefficient for the harmonic generator can be much larger than that for the power supply controlling the cesium oven, and each will have very different time constants.

The limited experience that we have with the frequency dependence on humidity in cesium standards indicates that the coefficients may be quite linear, and the coefficients seem to not be production line dependant; e.g. two units with adjacent serial numbers can have very different coefficients and these may be even of opposite sign.

Hydrogen

Since cavity pulling is a significant concern in hydrogen-masers, much effort has gone into stabilizing these cavities. The cavities must be stable to about 10^{-8} cm in order to have a frequency stability of the order of 1×10^{-14}. Much progress has been made in this area and frequency stabilities of the order of 1×10^{-15} are common for averaging times of the order of 100 s and longer -- in some cases as long as several days. Both temperature and pressure can detune the cavity. Long thermal time constants, similar to those in cesium, are often present in masers.

DESIGN ISSUES

Some specific design issues for the different types of precision oscillators are given here.

Quartz

Seal the device. Use insulation liberally to reduce convection inside the oscillator. Use SC-cut quartz. Evacuate the can. Keep it small.

Rubidium

If possible, seal the unit. Pre-age thermistors. Reduce convection currents if possible. Evacuate the unit if possible. Use a good quartz oscillator and second order integration loops to minimize the effect of changes in the quartz oscillator on the overall rubidium accuracy. If the unit is not sealed, use a glass for the storage bulb that has low permeability to helium and to the gasses in the bulb.

Cesium

Proper design can optimize performance to minimize most environmental effects. Use a good quartz oscillator and second order integration loops. Special attention should be given to THP offsets on the output level of the microwave power generation.

APPENDIX

Shown in Figure 2 is a year's worth of recent data comparing (through GPS) the various national standards and two commercial cesium standards in a temperature-controlled chamber. In

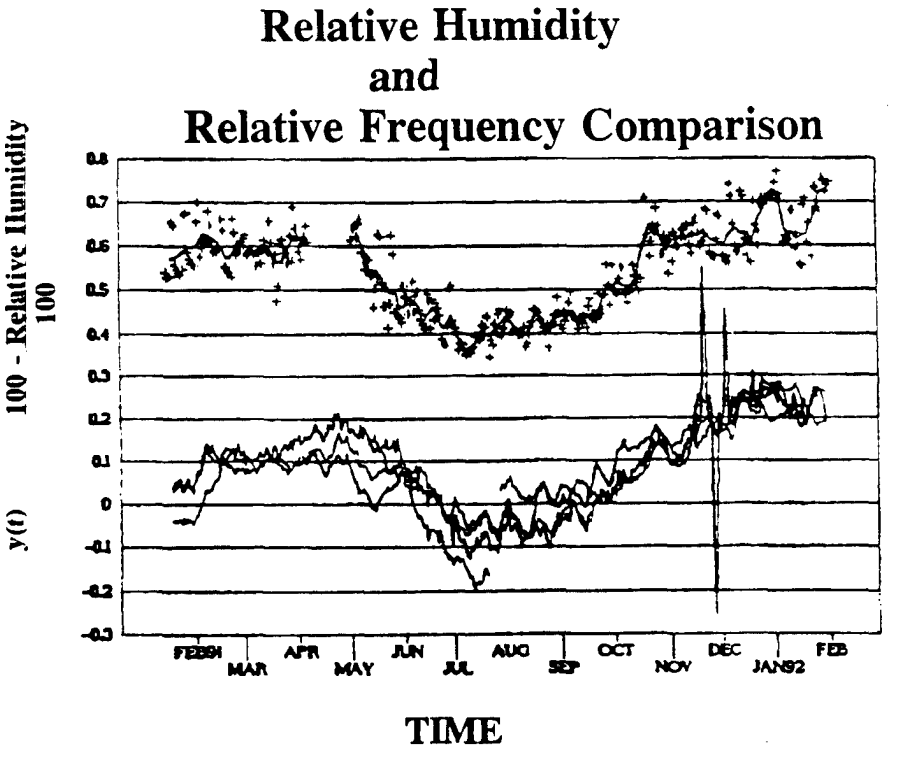


Figure 2. The plot is of both frequency, y(t), and the negative of the relative humidity versus time as measured in the laboratory. The top set of curves "+" are the negative of the relative humidity with a smoothed average plotted as well. The smoothing is a seven day average. Given that the temperature was relatively constant (to about ±0.5°C.) the change in relative humidity is directly proportional to the absolute humidity. The bottom set of four curves are of two commercial cesium-beam frequency standards, located in Geneva Switzerland, being compared with three independent timing centers (PTB, OP and USNO) via GPS in the common-view mode. The correlation is self-evident. The coefficient is about -1 x 10^{-14} per percent of relative humidity change. The environmental temperature was about 21°C.

this case the two commercial standards (both the same model number) had similar humidity coefficients. That is not always the case. Two other units with the same model numbers and differing by only one in their serial numbers had quite different humidity coefficients.

Figures 3 and 4 show the nonlinearity of frequency vs. temperature for AT-cut and SC-cut quartz crystal resonators. The advantages of the SC cut are evident. A variety of ways have been found to take advantage of these profiles. It is beyond the scope of this document to cover these, except to mention, since our main goal is to reduce environmental sensitivity, that controlling the temperature to a place where dy/dT is zero is obviously ideal. Not only does the SC-cut crystal have a larger plateau where this condition is approximately met, but the linear coefficient is usually about an order of magnitude smaller and the dynamic temperature coefficient is approximately one hundred times smaller. The turnover temperature for SC-cut resonance is also more desirable for many applications.

The development of the SC-cut crystal illustrates an important point. The best way to improve performance is to reduce the sensitivity of the basic frequency determining element rather than trying to provide better and better environmental control. The latter can be more and more expensive and also impractical for many field applications.

A related situation exists in cesium standards. One of the reasons for a temperature coefficient in cesium is that temperature affects the electronics causing a change in the micro-wave power output feeding the Ramsey cavity. The transition

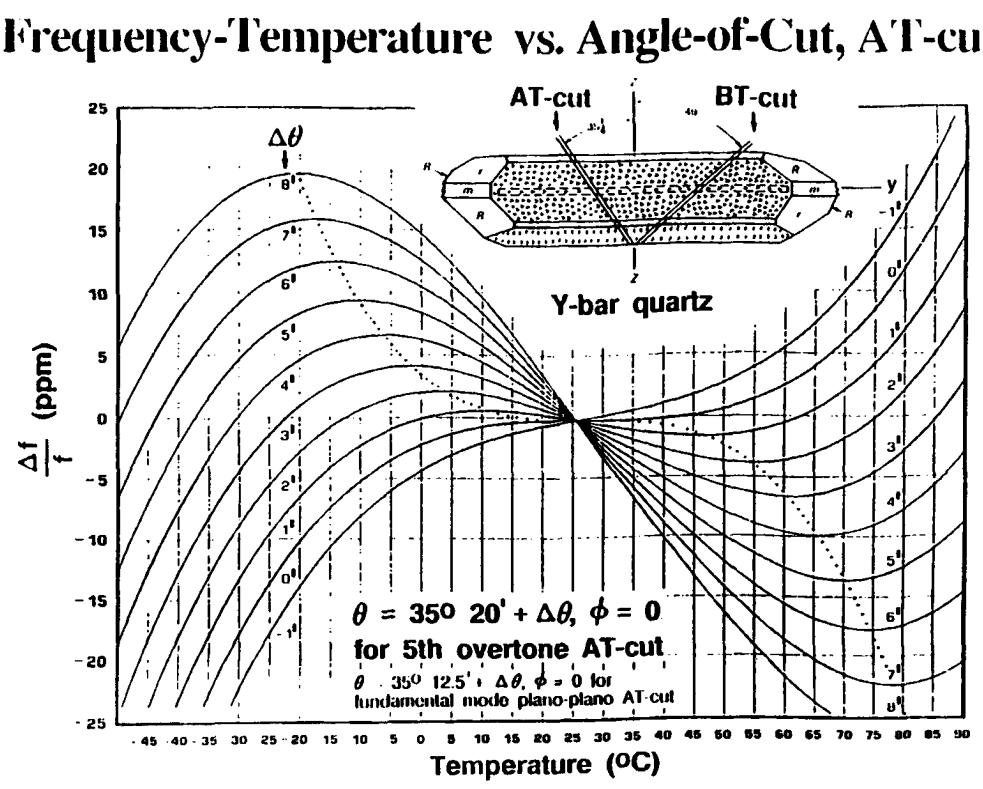


Figure 3. A family of plots of frequency versus temperature for AT-cut quartz-crystal resonators with differing angles of cut.

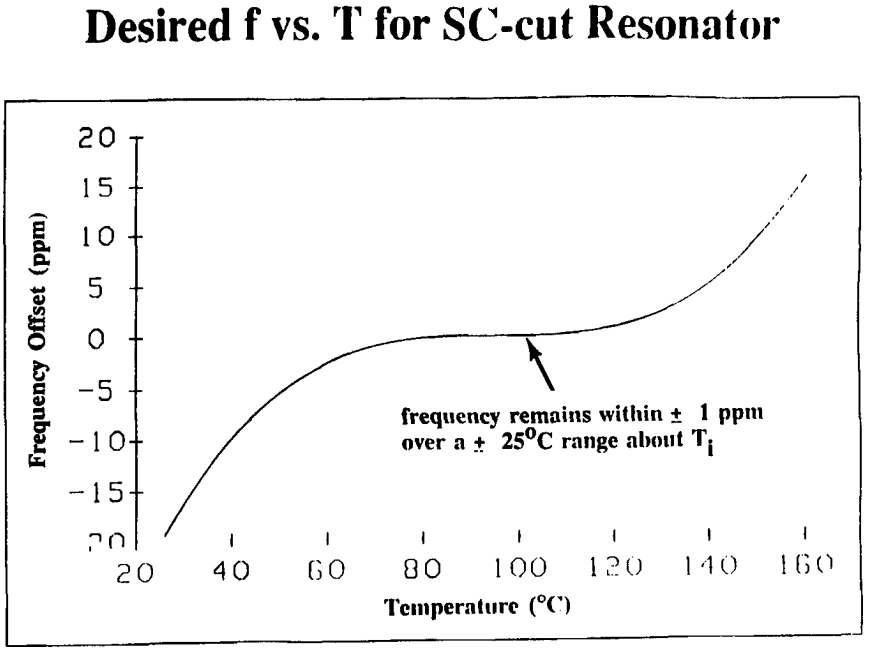


Figure 4. A plot showing the reduced temperature dependence of SC-cut quartz crystal resonators. Not only can a wider and more useful range of temperature operation be achieved, but the coefficient is about an order of magnitude smaller. The dynamic temperature coefficient is approximately 100 times smaller.

probability is a function of power, hence different velocities of atoms will be chosen with such change. If there is a non-zero cavity phase-shift across the Ramsey cavity (which there always is), this causes a frequency shift. As the velocity changes, the frequency shift changes. Figure 5 illustrates this sensitivity. The first sidelobe of the Ramsey spectrum is displaced by v/ℓ, where v is the nominal atom velocity and ℓ is the length of the Ramsey cavity. Hence, the microwave power can be servoed to a velocity corresponding to optimum power by sampling this sidelobe. By adapting such control, the overall temperature sensitivity is reduced [12,13].

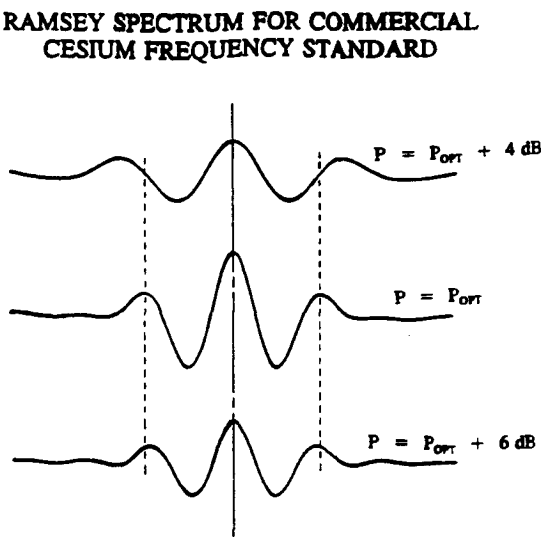


Figure 5. A plot of the Ramsey spectrum at three different microwave power levels in a commercial cesium standard. The velocity selection changes dramatically with changes in microwave power, causing a change in the spectrum. The value of P_{opt} is chosen as that power which gives best signal-to-noise.

ACKNOWLEDGEMENT

The authors wish to express appreciation to the several people who helped bring the material together for this paper and who helped with very useful comments and suggestions. These include the other members of the IEEE SCC27 committee, chaired by Dr. Helmut Hellwig, as well as Dr. Donald Sullivan, William Riley, and the several people at each of the author's organizations who provided significant support.

REFERENCES

[1] E. Bava, F. Cordara, V. Pettiti, and P. Tavella, "Analysis of the Seasonal Effects on a Cesium Clock to Improve the Long-Term Stability of a Times Scale," Proc. of the 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Redondo Beach, CA, Dec. 1-3, 1987, pp. 185-202.

[2] S. Iijima, K. Fujiwara, H. Kobayashi, and T. Kato: "Effect of environmental conditions on the rate of a Cs Clock-Annals of Tokyo Astronomical Observatory," vol. XVIII n1.

[3] G. Becker: Zeitskalenprobleme: jahreszeitliche Gangswankugen von atomuhren, PTB Mitteilungen 92 2/82.

[4] K. Dorenwendt: Das verhalten kommerzielle caesium atomuhren und ide zeitskalen der PTB-PTB Mitt. 94.

[5] J.-J. Gagnepain, Characterization Methods for the Sensitivity of Quartz Oscillators to the Environment, Proc. of 43rd Annual Symposium on Frequency Control, 1989, pp. 242-247.

[6] Fred L. Walls, and J.-J. Gagnepain, "Environmental Sensitivities of Quartz Crystal Oscillators," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 39, pp. 241-249, March 1992.

[7] L. A. Breakiron, L. S. Cutler, H. W. Hellwig, J. R. Vig, and G. M. R. Winkler, "Environmental Metrology for Precision Oscillators," these proceedings.

[8] Keki, R. Bhote, World Class Quality, AMA Membership Publication #2334, American Management Association, New York, 1988.

[9] J. A. Barnes, "The Measurement of Linear Frequency Drift in Oscillators," in NIST Technical Note 1337, pp. 264-295, 1990.

[10] M. A. Weiss, D. W. Allan, and D. A. Howe, "Confidence on the Second Difference Estimation of Drift," these proceedings.

[11] ANSI/EIA-512, Standards Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985. Order from: Naval Publication & Form, Center or General Services Administration Business Service Centers in major U. S. Cities. See also Military Standard MIL-0-55310B,

Oscillators, Crystal, Gen Spec for. Order from: Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

[12] D. W. Allan and M. Garvey, "Microwave Power Level Stabilizing Circuit for Cesium Beam Frequency Standards," U. S. Patent No. 4,331,933, issued May 25, 1982.

[13] L. S. Cutler and R. P. Giffard, "Architecture and Algorithms for New Cesium Beam Frequency Standard Electronics," these proceedings.

IEEE Standards Project P 1193 Electric and Magnetic Fields and Effects[1]

Richard L Sydnor
Jet Propulsion Laboratory, California Institute of Technology

R. Brendel
Centre National de la Recherche Scientifique

W. Wiedemann
Efratom/Ball Corporation

Abstract

This paper discusses techniques for measuring the magnetic field sensitivity of frequency standards with particular emphasis on problems and pitfalls. Several alternative procedures are discussed with their relative merits and drawbacks. The paper also discusses the necessary modifications to standard EMI tests to make them suitable for the measurement of the susceptibility of frequency standards to ambient electromagnetic fields. An example is given.

Objective: To formulate guidelines for test methods leading to practical and factual specifications for electromagnetic effects on standard frequency generators. Electromagnetic effects include, for the purpose of this guideline electric fields, magnetic fields, electromagnetic interference (EMI).

In all cases the effect is to be considered a "black box" effect, i.e. the internal response of various components to the stimulus is not to be considered, only the overall response of the device is the question to be considered.

Magnetic Field (DC or low-frequency) Sensitivity

I. Description

This effect results in the change in the frequency of a frequency standard due to changes in the ambient magnetic field. For static fields, such as may be caused by the proximity of magnetic material (racks, vehicles, etc.) the effect is an offset in the operating frequency from the unperturbed frequency. For dynamic fields (caused by leakage fields from power supplies fluorescent lighting ballasts and the like) the effect results in frequency modulation of the standard at the frequency of the interfering field.

II. Magnetic Field Effects

A) Definition

The static (DC) magnetic field sensitivity of a device is the change in frequency due to a change in the applied magnetic field. The magnetic field is not measured at the device, but is instead the free space field in the same location as the device but with the device removed. Since the sensitivities are different in different directions, the measurements should be made and results given in the three directions, *x*, *y*, and *z*. The utility of the static sensitivity measurement is the determination of the effect of differing magnetic environments on the performance of the device.

The dynamic magnetic field sensitivity of a device has the same units as the static, but the measurement technique and analysis of the data are, of necessity, different. The applied field is characterized in the same manner as in the static case, but the field is now an alternating field and the measurements are usually made using a spectrum analyzer. The value of the dynamic sensitivity measurement is the ability to ascertain the effect of power line and related fields on the device.

1) This work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

0-7803-0476-4/92 $3.00 © 1992 IEEE

B) Discussion

1) Non-linearities and Hysteresis

Since the changes in frequency of a device are the result of interaction of the applied field either with atomic states in an atomic device or with magnetic materials in the electronics of all devices, it is to be expected that the effect will be non-linear and will exhibit some hysteresis. Enough measurement points should be made that an adequate characterization of the effect may be made. It is important that the range of fields be traversed several times so that hysteresis effects may be characterized.

2) Effects of other environmental parameters

The permeability of typical magnetic materials (mu metal, moly-permalloy, ferrite) varies with temperature. If it is necessary to completely characterize the magnetic susceptibility of the device, the magnetic sensitivities may be measured over the range of temperatures which are considered normal for the application. For a normal room temperature of 25°C, measurements at 20°, 25° and 30° may be adequate.

3) Time effects

There is some indication that the shielding efficiency of some magnetic materials improves slightly with time. Probably no measurements need be made of this effect, but awareness of it may be useful.

III. Test Methods

A. Steady State (DC) Field Tests

The traditional test uses a Helmholtz coil as shown in Figure 1. As a rule of thumb, the Helmholtz coils should have a diameter at least twice as large as the maximum dimension of the device under test (DUT), three or four times larger is better. A description of the Helmholtz coil design and construction is given in the ASTM Document A 698 - 74.[1] The standard spacing between the coils is one radius. The earth's field is on the order of 40 microteslas (400 milligauss) and varies in direction depending on the location on the globe. Since the purposes of measuring the effect of magnetic field are to 1) determine the changes in frequency that will occur in a given environment and 2) determine the frequency that the standard will produce in a given environment, a single Helmholtz coil is inadequate to completely characterize the standard. A three-axis Helmholtz coil as shown in Figure 4 is recommended. The field in the center may be set to zero, nullifying the earth's field, and then varied around this point by large enough changes in the applied field to cover the expected amount that the DUT may experience in actual operation. A variation of $\pm$ 200 microteslas (2 gauss) is usually adequate. The DUT must be kept in the same orientation for all tests (preferably the normal operating position) to eliminate the effects of varying gravitational forces and changes in thermal gradients which may also effect the frequency standard and give erroneous results for the magnetic field sensitivity. The length of the measurement period for a given magnetic field is governed by the Allan Deviation curve for the standard. A measurement period should be chosen that corresponds to an averaging time in the area of the Allan Deviation curve that is near the minimum. Typical times might be 100 seconds for a quartz crystal standard, 2000 seconds for a hydrogen maser and 10^5 second for a cesium beam standard.

If there are non-linearities and hysteresis, the effect of a magnetic field in one direction is affected by the magnitude of the field in other directions. For most frequency standards this is not the case for fields in the normal range. Some standards have magnetic shielding and this shielding may exhibit such effects. In this case the characterization of the standard's sensitivity is very complicated and the measurements should be made with the impressed field corresponding to the ambient field at the location wherein the standard is to be located.

B. Alternating Field Tests

If the Helmholtz coil used for the Static Field Tests is made on a non-metallic form, it may also be used for the dynamic field tests. One calculates the applied field in a manner similar to the calculation of the static field, except that the current will be an alternating current. For the purposes of standardization, peak reading instruments will be assumed in the calibration of the impressed field. If other instruments are used, proper conversion to peak values should be applied. The applied field is given by:

$$B = (4\pi\ 10^{-7}\ 2\ NI \sin \Omega t)/(5\sqrt{5}\ R),$$

where:

I is the peak value of the current in the coils (expressed in Amperes),
N is the number of turns in each coil,
R is the radius of the coils in meters, and
B is the field in teslas.

By examining the resulting spectrum of the output signal, sidebands may be observed at the excitation

frequency (and possibly at its harmonics). Since the field is known and the expression for the amplitude of the sidebands is:

$$P=20\log(\alpha/(2\sqrt{2}))$$

dBc, where $\alpha = Bs_m$ and B is the magnetic field in teslas and s_m is the magnetic sensitivity of the oscillator in relative frequency change per tesla. The desired equation for the sensitivity is then:

$$s_m = 2\sqrt{2}\ 10^{(P/20)}\ B.$$

Since the magnetic field sensitivities are non-linear functions of the field intensity, the test should be made over a range of fields that might be expected in practice. In a standard relay rack which is full of various electronic equipments, the field can be of the order of 5 microteslas (50 milligauss) at the power line frequency or its second harmonic. Ordinary laboratory environments may have power line related field intensities on the order 0.5 microteslas, while areas in vicinity of large motors or transformers can have field intensities as high as 100 microteslas.

Alternative Methods

An alternative method for pseudo-static tests utilizes a very slowly varying field with a sinusoidal variation of period commensurate with the lowest noise portion of the Allan Deviation curve for the standard in question. The data could then be analyzed with a spectral analysis computer program to achieve improved sensitivity. This method has been employed by R. Brendel[2] with excellent results.

Electromagnetic Interference

I. Description

This effect results the change in frequency of a frequency standard due to impressed electromagnetic field. Such fields may leak through the joints in an instrument case or be conducted into the interior of the frequency standard via power lines or cables. The resulting effect may be either a change in the frequency of the standard, or production of sidebands on the standard frequency output.

II. Specification

A) Definition

By impressing a specified amplitude EM (electromagnetic) field on the frequency standard and determining the effect of the field on the standard, a measure is made of the sensitivity of the standard to the applied field. In general, measurement of the susceptibility is the same as the standard EMI test applied to other devices, i.e., a field is produced in the location of the standard and changes in the output are detected. The normal means of performing this test uses a swept frequency field generator which spans a large segment of the RF spectrum. The entire test is performed in an RF anechoic chamber to prevent erroneous results due to standing waves. The spectrum, frequency and phase of the output is monitored during the test to determine what effect the applied field has on the device.

The purpose is to determine what electrical leakage exists, what internal resonances exist that can be excited by the leakage, and what effect this combination of leakage and resonances has on the operation of the device.

An additional test is necessary in the case of frequency standards, since they may have susceptibilities which occur in extremely narrow bands and will be missed during the fast sweep of frequencies that is normal for EMI tests. A very slow sweep through these bands is necessary in order to determine the susceptibility of the frequency standard in these frequency ranges. The sweep should cover a band approximately 10 times the resonator bandwidth and should be at a rate approximately $f/100Q$ of the resonator . For example, a 5 MHz oscillator with a resonator Q of 2×10^6 would have a resonator bandwidth of $\approx$ 2 Hz. The sweep should cover 5 MHz $\pm$20 Hz and should be at a rate of $\leq$ 0.02 Hz/second. An example of the results of such a test are given in Figure 2. The variation in the slope of the phase-time curve indicates that the oscillator was being affected by the impressed electromagnetic field.

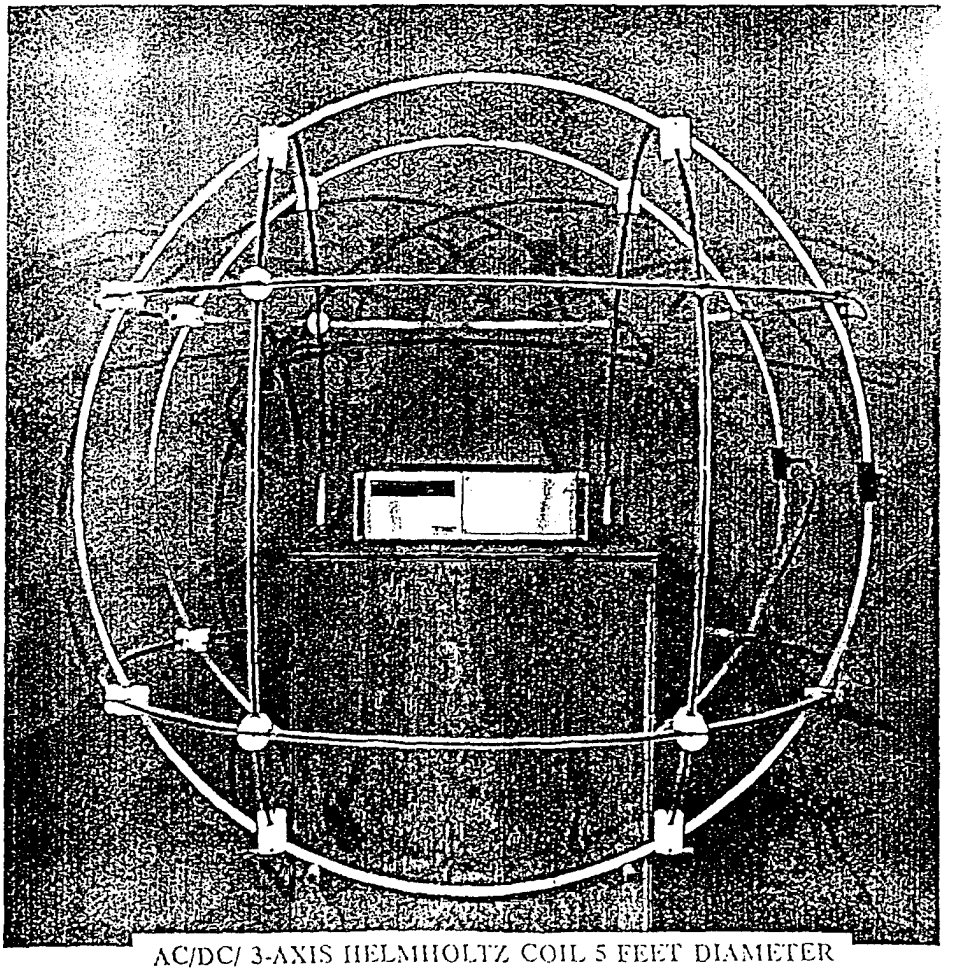
The actual bands that must be investigated depend on the type and design of the frequency standard: a quartz crystal oscillator should be tested in a band around the crystal resonances and submultiples of the resonant frequencies, a passive frequency standard (e.g. cesium beam) should be tested at the crystal oscillator frequency and the various frequencies used in the frequency multiplier as well as the frequency modulator frequency.

There does not exist a standard level of excitation for radiating the device under test. Usually a level is used that is specified based on the expected operational environment. The required level or equipment which is to be used in a field radar installation is different from the required level for laboratory equipment. Levels on the order of one volt per meter are in the general range of testing. Figure 3 is an example of an EMI test requirement for a spacecraft oscillator.

References

[1] American Society for Testing and Materials, A 698 - 74, "Standard Method of Test for Magnetic Shield Efficiency in Attenuating Alternating Magnetic Materials"

[2] R. Brendel, C. El Hassani, P. Krux, and E. Robert, Magnetic Sensitivity of Oscillator Components, Fifth European Frequency and Time Forum, 1991



AC/DC/ 3-AXIS HELMHOLTZ COIL 5 FEET DIAMETER

Figure 4

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

THE EFFECTS OF IONIZING AND PARTICLE RADIATION ON PRECISION FREQUENCY SOURCES
(PROPOSAL FOR IEEE STANDARDS PROJECT P1193)

Joseph J. Suter, Richard H. Maurer, and James D. Kinnison

Johns Hopkins University, Applied Physics Laboratory
Laurel, Maryland 20723-6099

Raymond Besson

Ecole Nationale Supérieure de Mécanique et des Microtechniques
25030 Besancon, France

John R. Vig

U.S. Army Laboratory Command
Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey 07703-5000

Andy Koehler

Harvard Cyclotron Laboratory
Harvard University
44 Oxford Street
Cambridge, Massachusetts 02138

ABSTRACT

The susceptibility of quartz oscillators and atomic frequency standards to natural and enhanced ionizing and particle radiation is an important parameter in predicting the short- and long-term performance of these standards in spacecraft. Characterization of the response of frequency standards to ionizing and particle radiation should be based on a thorough understanding of the radiation environment (proton, electron, neutron, flash X-ray radiation, and single-event upset) and radiation scenarios (dose/anneal cycle, combined environments). A discussion of radiation environments (proton, electron, neutron, flash X ray, gamma, and single-event upset) is presented. Selection criteria for radiation sources, including dosimetry and procedures for the radiation testing of frequency standards, are discussed. By examining results of numerous previous studies, including radiation environments, this paper , which is a proposal for IEEE Standard Project P1193, provides direction for the methods by which the radiation sensitivity of frequency standards can be assessed.

INTRODUCTION

The susceptibility of quartz oscillators and atomic frequency standards to natural and enhanced ionizing and particle radiation is an important parameter in predicting the short- and long-term performance of these standards in spacecraft. Many studies have been conducted to establish a relationship between the radiation-sensitive components, like quartz, and the radiation response of resonators made from quartz.[1-7] These studies involved ionizing and particle radiation at specified doses and dose rates employing gamma radiation from cobalt 60 sources (1.25 MeV photons) and electron, proton, flash X-ray, and neutron radiation. The primary goal of these investigations was to establish a base of experimental data mainly for use in determining the radiation hardening requirements for specific spacecraft missions. It has often been difficult, therefore, to extend the results of these radiation tests to a more general understanding of the radiation susceptibility of frequency standards. Studies have revealed that susceptibility of frequency standards, in particular quartz crystal oscillators, to low levels of ionizing radiation (<100 rad [Si]) is not determined by the electronic components in the oscillator circuit. Most electronic components are capable of surviving, without any significant degradation, accumulated doses up to several krad (Si). The response of these standards to low-dose radiation as encountered in Earth orbits, therefore, is controlled by the frequency-determining element (quartz resonator and physics package). On the other hand, the radiation

0-7803-0476-4/92 $3.00 © 1992 IEEE

sensitivity of atomic frequency standards and quartz oscillators to enhanced environments, such as accumulated doses greater than 1 krad (Si), peak prompt dose rates of 10^{10} rad (Si)/s, and neutron fluences (1 MeV) on the order of 10^{11} to 10^{13} neutrons/cm^2, is not only determined by the quartz resonator or physics package but also by the hardness of the oscillator electronics (components in the oscillator circuit, voltage-controlled oscillators, tuning circuitry, and thermal control). Design limits must be established to minimize the effects of radiation on these devices. Order-of-magnitude estimates of radiation-induced frequency shifts can be established by assuming that these limits increase linearly with electrical parameter changes and that they act independently.[8]

Frequency standards on spacecraft are, in general, intermittently exposed to both natural and enhanced radiation. In the case of low-Earth orbits, a significant amount of radiation is delivered to the frequency standard during that part of the orbit in which the satellite passes through the South Atlantic Anomaly and to a lesser degree when it passes through the Northern and Southern Horns.[9,10] Annealing effects during radiation-free orbit segments should be studied carefully in any radiation hardening study because they affect the long-term stability and frequency offset in different ways. The effects of enhanced radiation also need to be studied with scrutiny, since multiple radiation events may occur spaced over different time intervals, resulting in cumulative effects and increasing the likelihood of timing errors in clocks.

Furthermore, depending on the ephemeras of the satellite, single-event upsets (SEUs) and latchup may occur because of high-energy particles traveling through integrated circuits.[11] The induced current pulses can create SEUs and latchups, resulting in data errors caused by malfunctioning of voltage-controlled oscillator circuitry and microprocessor systems or a change in information in storage registers.

Consequently, a comprehensive study, both theoretical and experimental, of the radiation sensitivity of frequency standards by type or even a particular individual unit is recommended to evaluate the responses to both natural and enhanced ionizing and particle radiation environments. This characterization should entail an evaluation of the short- and long-term stabilities and aging rates of the standard. The guidelines presented herein are proposed as a foundation on which subsequent studies and tests can build. By examining results from previous studies, including radiation environments, this paper provides direction for the methods by which the radiation sensitivity of frequency standards can be assessed.

TEST METHODOLOGY

Effects due to total radiation dose are simulated using sources of ionizing radiation. The most common such sources in the natural environment are X rays, gamma rays, electrons, and protons. Since the electrons are not very massive and the protons prefer to lose their energy (refer to Table 1) in a dense media via ionization, researchers have found that cobalt 60 gamma rays can be used to simulate the natural space environment. Low dose rate (10^{-4} to 10^{-2} rad [Si]/s) sources are recommended for simulating the space environment, as time or frequency shifts may have dose rate as well as accumulated total dose dependence.

In contrast, for weapons or burst environments, high dose rate sources such as flash X-ray machines should be employed to deposit the total radiation dose. In practice, the ionizing dose rate is not high enough to cause any thermal effects such as an increase in temperature of a quartz crystal resonator.

Radiation Interaction

Ionization is the creation of electron-hole pairs in a target material by some incident particle. These particles are either photons or charged particles. Although the results are the same, photons and charged particles create the electron-hole pairs through different processes. Photonic ionization results from three different mechanisms: the photoelectric effect, Compton scattering, and electron–positron pair production. Each of these mechanisms dominates in a different photon energy regime.

Charged particles create electron-hole pairs mainly through ionization, but nuclear interactions can also occur, which create charged secondaries. Rutherford (or Coulomb) scattering is the primary mechanism by which charged particles interact with materials. The nucleus of an atom is much smaller than the total size of the atom, so charged particles that encounter atoms are mostly influenced by the electrons surrounding the atomic nucleus. Energy is transferred to the electrons by electromagnetic interaction. Ionization occurs when this energy is sufficient to free an electron from the atom. The energy lost by the incident

Table 1. Radiation Interaction Mechanisms.

Radiation Type	Interaction Mechanisms
Gamma radiation and X rays	Photoelectric effect (E_k^* < 10 MeV) Compton scattering (1 MeV < E_k < 10 MeV) Pair production (E_k > 10 MeV)
Protons, electrons, ions	Rutherford scattering Straggling Nuclear reactions Electronic collisions Atomic collisions Bremsstrahlung (electrons and heavy ions) Ionization (secondary electrons) Capture (electrons)
Neutrons	(In)elastic nuclear collisions

*E_k = Kinetic energy.

particle is usually a small fraction of its total energy, and thus the particle continues on its path after the interaction. As the particle travels in the material, it leaves a track of electrons and ionized atoms, or holes, along its path. The incident particle stops when it loses all its energy to the target atoms.

The effect of each of the interactions discussed above is to produce electron-hole pairs in the target material. In space, many of these pairs are created in a solid-state device. Once created, pairs may recombine; the number of pairs at a given time is determined by the dose rate, which gives the pair generation rate, and the recombination rate, which is a function of the electron and hole mobilities and the bias of the material.

The rate of radiation dose accumulation in space is quite low. In addition, mission lifetimes are on the order of years. When testing devices for ionization effects, therefore, the dose rate is usually increased such that the same total dose that a device would receive in orbit is delivered to the test sample in a few minutes or hours. The rate at which positive charge is built up depends on the dose rate. The higher the dose rate, the faster the charge buildup. Annealing occurs during exposure as well, however, with a rate that is not dependent on dose rate. For a given total dose, a device will show more damage over short time intervals when irradiated at a high dose rate than when irradiated at a low dose rate. Therefore, total dose testing should be performed with a dose rate as close to the mission dose rate as possible.

Radiation Tests, Sources, and Dosimetry

Cobalt 60 facilities often have the advantage of being relatively inexpensive and offer the possibility of adjusting the radiation dose rate either by using shielding or varying the distance between the device under test and the source (refer to Table 3). Electron beams can be used in either a continuous mode for steady-state total dose or a pulsed mode primarily to simulate burst phenomena. Proton beams produce a continuous flux through the device under test, but an energy spectrum simulating a low-Earth orbit environment can be produced by a modulating wheel so that the experimenter is not confined to a monoenergetic source.

Neutrons should be used to simulate displacement damage effects, since they carry no electrical charge and create no primary ionization effects. As mentioned previously, protons prefer to lose their energy by ionization, and thus their lattice displacement effects are dominated by ionization. To study displacement damage effects requires the use of pulsed research reactors. Such test sites are rare and expensive. When testing for the weapons environment, however, bombarding the precision frequency source with neutrons is the only reliable simulation. Since the research reactor and weapon neutron energy spectra are not usually the same, neutron energy fluences are expressed in terms of 1 MeV equivalent neutrons.

Table 2. Test Facilities (Selected).

Radiation Type	Facility	Characteristics*
Gamma	Co^{60}—usually local, NASA/GSFC	1.17 MeV (50%) and 1.33 MeV (50%) photons
Proton	Harvard University	E_{max} = 154 MeV Φ_{max} = 10^{10} protons/cm²/min d_{max} = 12 cm Low-Earth orbit simulation
	University of California at Davis	E_{max} = 65 MeV d_{max} = 10 cm
Electron	NASA/GSFC NRL	E_{max} = 3 MeV E_{max} = 65 MeV Φ_{max} = 10^{11} rad(Si)/s d_{max} = 0.4 cm
Neutron	Aberdeen Proving Ground	Φ_{max} = 8×10^{14} neutrons/cm² E > 10 keV Gamma ray dose 3.9×10^5 rad (Si) per pulse
X rays	HDL Febatron	Dose = 100 rad/2 cm² Pulse width = 4 ns V_{max} = 600 keV

*E_{max}, Φ_{max}, d_{max}, and V_{max} are the maximum kinetic energy, flux, beam diameter, and maximum voltage, respectively.

Low-Earth Orbit Tests. For the natural space environment, it is also important to study the dose/anneal cycle that will occur—particularly in low-Earth orbits. For such missions, accurate frequency standards and clocks are necessary to carry out reliable surveillance. Because of the contamination inherent in some precision frequency sources, additional dose periods following annealing periods can reveal greater sensitivity to radiation, since some mobile contaminant ions free themselves during the anneal/recovery period and can interact again with subsequent ionizing radiation.

Figure 1 shows the total dose rate per orbit for the TOPEX mission. The largest amount of radiation is accumulated in the South Atlantic anomaly (refer to Figure 2).

High Dose Rate Tests. High dose rate (flash X-ray or gamma ray sources) testing is done to simulate the nuclear burst environment. Primarily, one is looking for transient upsets or interruptions of the frequency source. Such upsets may cause a loss of time for a spaceborne clock. Also of intense interest are the quickness and completeness of the recovery from such an upset.

Precision frequency sources and their associated electronic circuitry can be expected to exhibit thresholds with respect to the dose rate causing upset and to be dependent on the dose rate beyond the threshold with respect to the amplitude of the upset. In addition to upset thresholds, one also worries about burnout thresholds caused by photocurrents generated in the associated electronics at very high dose rates.[12]

Finally, in a weapons environment combining both high dose rate ionization and neutron displacement damage pulses, the dose rate effects on frequency shifts are expected to be of greater amplitude but shorter duration than the induced effects due to neutrons (refer to Figure 3).

Single-Event Tests. In the case of precision frequency sources, single-event phenomena are expected to occur only in the associated digital electronic circuitry or in the power MOSFETs. Individual cosmic ray ions or solar or trapped protons can produce enough ionization near critical nodes in solid-state circuitry to cause a change of the logic state. In bulk CMOS integrated circuits, a destructive burnout caused by latchup can occur. Single-event induced burnout has also been observed in power MOSFETs.

In most cases, it should be possible to screen the proposed parts list for integrated circuits susceptible to single-event phenomena and to replace them with integrated circuits immune to such effects. Hardened MOSFETs and digital logic families are now readily available but at substantially increased cost.

Testing could be performed at the system level to verify the immunity of the frequency reference system to single-event effects. Unlike data handling systems, in which some upsets can be tolerated, it is recommended that critical frequency reference and clock systems be immune to single-event effects.

Sources and Dosimetry. Many facilities offer sources for X rays, gamma rays, protons, electrons, neutrons, and combined environments.[13] Table 2 presents a summary of some of the dosimetry techniques for each type of radiation.

Electron and proton beams from accelerators are generally monitored by Faraday cup techniques. Since the beam flux and energy are known, the total dose in silicon or quartz can be computed readily. For cobalt 60 irradiators or gamma cells, a thermoluminescent dosimeter (TLD) is quite useful, since it can be calibrated to read dose rate directly (refer to Table 3). When dealing with cobalt 60, one finds that if the source strength in Curies is known at a given time, measurements of dose rate just verify the law of radioactive decay.

Measurement of the radiation dose deposited at critical locations with the target equipment is carried out employing TLDs. These devices are made from lithium or calcium fluoride and give up light, after irradiation, in a TLD reader. Doses monitored by TLDs can generally be measured only to a 10 to 15% accuracy.[13,14]

In order to understand the radiation doses at various locations within the frequency standard fully, it is recommended that a mechanical mock-up be developed. Dosimeters, such as TLDs, can then be located at various strategic locations, allowing for a direct experimental determination of accumulated doses at the site of critical components. A geometric model of the frequency standard should also be developed to enable the shielding characteristics of the standard to be fully understood. Several theoretical models are available to implement this analysis.[13]

Table 3. Radiation Sources and Dosimetry.

Radiation	Source	Dosimetry
1.25 MeV gamma	^{60}Co	Thermoluminescent dosimeters Law of radioactive decay PIN diodes Fibers Calorimeters
Protons	Cyclotrons, linacs Synchrotrons Van de Graaff	Radiochromic dyes PIN diodes Ionization chambers
Electrons	Linac Van de Graaff	Faraday cup Ionization chambers PIN diodes Radiochromic dyes
X rays	Van de Graaff	Film Ionization chambers
Neutrons	Nuclear reactor (fission)	Activation foils Sulphur $[S^{32}(n,p)P^{32}]$ Scintillators
Heavy ions	Van de Graaff	Faraday cup

TEST PROCEDURES

The most important parameter to be monitored during radiation exposure is the output frequency of the source both over long and short time periods. The frequency behavior over long time periods is represented by the drift rate. Over shorter periods such as 1 or 1000 s, frequency stability is measured using the Allan variance. Before and after total dose exposures, it is also important to measure and compare the single sideband phase noise. During burst environment testing, the amplitude and phase of the frequency source output must be monitored so that any interruptions of function or losses of time can be determined.

An adequate warmup period (greater than 24 h) is necessary to ensure that the precision frequency source has stabilized before radiation exposure is begun. Similarly, it is often critical that the recovery of the source be monitored after the radiation exposure has been completed. Usually one is interested in the time it takes for the frequency source to regain its original drift rate and short-term stabilities as well as whether or not a residual frequency offset remains between the time lines before and after irradiation.

In radiation testing of frequency standards, the methods used to measure changes in output frequency involve customary equipment such as reference standards, frequency multipliers, mixers, and frequency counters. As shown in Figure 4, the RF output signal from a frequency standard is

compared with that of a reference by first multiplying their output signals to a higher frequency. Multiplication before frequency mixing may yield a greater resolution of the beat frequency signal from the mixer.

Typical systems mix the two frequencies so that their beat frequency is in the audio frequency range (20 to 4000 Hz). A frequency counter, together with a data acquisition system, is then used to calculate the frequency shift and Allan Variance as functions of time and accumulated dose.

In order to observe radiation-induced frequency shifts in real time, an analog strip chart recorder may be needed to display the frequency standard's fractional frequency shift as compared with the laboratory reference. Previous experiments have indicated that radiation-induced frequency shifts (df/f) can be as small as 1.0 part in 10^{-12}/rad. Hence, the reference source should have a stability at least one order of magnitude better than the device under test.

To observe the recovery characteristics of the frequency standard after radiation exposure, the reference standard should have a drift rate at least a factor of ten better than the standard under test.

Prompt radiation can often introduce momentary outages in a frequency standard, which result in a possible loss of time. Figure 5 shows a method for measuring the magnitude of this time loss. The time difference between two clocks, one referenced to the laboratory standard and one to the device under test in terms of 1 pulse per second (pps), are input to a counter. The time difference, either a loss or gain of the device under test relative to the standard during a radiation event, is then monitored by the counter. The resolution of this system should be better than 1 ns.

The difference in phase between the test and reference standards should also be continuously monitored before, during, and after the radiation tests. For example, using a vector voltmeter, the difference in phase between the standards can be recorded by using a high-speed recorder. The phase measurement system supplies not only information on the phase difference but also the possible time loss between the test standard and reference (refer to Figure 6).

In prompt radiation tests, a fourth experimental setup is recommended. This monitoring system records not only the frequency standard's output but also flash X ray and neutron pulses as a function of time. Figure 7 shows the experimental arrangement of this test system.

SPACE ENVIRONMENT

The space radiation environment basically consists of electrons, galactic cosmic rays, and solar protons, whereas the Earth orbital environment consists of trapped electrons and protons (Van Allen Belts). For the region of the solar system within 11 astronomical units, the dominant source of relativistic electron flux with energies below 25 MeV is Jupiter.[15] Energetic electrons follow the solar magnetic field lines towards the interior of the solar system. The electron flux for electron energies between 1 and 100 MeV is shown in Figure 8, which shows experimental electron data as well as the results (solid line) predicted by a model.[15]

Galactic cosmic rays are composed of electrons and ions from all the natural elements. These particles originate outside the solar system and are isotropic outside planetary magnetospheres. Solar activity attenuates the cosmic ray flux, and so the highest concentration of cosmic rays is found during solar minimum conditions in the solar cycle. Figure 9 presents the combined Linear Energy Transfer (LET) spectra of all elements from hydrogen to iron found in cosmic rays, which is the form most commonly used in radiation effects calculations. The data in Figure 9 include the LET spectra as a function of various shielding thicknesses.

An epochal analysis[16] of the annual proton flux produced in solar events over the solar cycle shows that the cycle can be divided into two periods: four inactive years of low annual fluence and seven years with annual fluence greater than 5×10^7 protons/cm^2 (>10 MeV). The active period extends from two years before solar maximum to four years after maximum. The times of solar maximum for the 19th through the 21st cycles are, respectively, 1957.9, 1968.9, and 1979.9. Figure 10 shows the solar cycle dependence of annual fluence for the past three solar cycles. The Earth's magnetic field in the region associated with the Van Allen radiation belts can generally be described by a dipole located near the center of the Earth and directed so that the Earth's magnetic south pole is located on the surface in northern Greenland. Empirical models of the radiation belts have been compiled from satellite measurements made over the last three decades. Figure 10 clearly shows that the particle fluence is appreciable two years before the sunspot maximum and four years after this event.

Figure 11 is a plot of the integrated proton flux (energy >10 MeV) observed in circular orbits with a range of altitudes and inclinations. The peak flux corresponds to the proton belt, which shifts in altitude only a small amount as the inclination is increased.[11]

Figure 12 demonstrates the features of the flux of electrons with energy greater than 0.25 MeV averaged over several orbits at each inclination and altitude. As with the protons, belt-like structure is observed, but two regions of high concentration are present in this case. The higher flux region nearer to the Earth is called the inner zone; the belt farther from the Earth is the outer zone.[11]

SUMMARY

This paper has summarized issues pertaining to the radiation susceptibility of frequency standards as proposed for IEEE Standards Project P1193.[17] The proposal for this standard presents a detailed analysis of types of radiation,

interaction mechanisms, low-Earth orbit and interplanetary environments, nuclear burst and pulse characteristics, dosimetry, and test configurations. The proposed standard also presents information on electromagnetic pulse (EMP) and contains a detailed analysis of models for low-Earth orbit and interplanetary environments.

The susceptibility of frequency standards to radiation should be addressed in the design stage. The exact radiation environments must be characterized with respect to dose, dose rate, flux, fluence, and energy spectrum. The design of the frequency standard and the hardness of its physics package and electronics should be based on geometrical modeling so that the dose levels at various locations inside the standard will be well understood. Radiation testing of a mechanical mock-up will assist in confirming the geometrical calculations.

Actual radiation testing of atomic or quartz crystal frequency standards can only be performed after obtaining a thorough understanding of the environment, including dose and anneal cycles. Tests should be conducted using reference frequency standards with a stability of at least one order of magnitude better than the standard under test or the radiation-induced frequency shifts. Radiation experiments must measure the frequency standard's frequency shift versus dose and dose rates, recovery during annealing intervals, clock signal outages during burst, and flash X-ray and pulsed neutron events. Data should be obtained on the recovery of the frequency standard after radiation exposure. The radiation hardness of frequency standards is environment related, and generalization of test results from specific radiation tests can therefore lead to erroneous interpretations, which may cause costly overdesigns or the construction of frequency standards that are not adequately hardened. As mentioned previously, the radiation hardening of frequency standards starts with a full knowledge of the environment outside the standard as well as that which is created internally.

REFERENCES

[1]J. C. King and D. R. Koehler in *Precision Frequency Control*, vol. 1, E. A. Gerber and A. Ballato, Eds., Orlando: Academic Press, 1985, p. 147.

[2]P. Pellegrini, et al., "Steady-State and Transient Radiation Effects in Precision Quartz Oscillators," *IEEE Trans. Nucl. Sci.*, vol. NS-25, p. 1267, 1978.

[3]H. G. Lipson, et al., "Radiation Effects in Swept Premium-Q Quartz Material, Resonators, and Oscillators," in *Proc. 33rd Annual Frequency Control Symposium*, 1979, p. 122.

[4]T. M. Flanagan and T. F. Wrobel, "Radiation Effects in Swept Synthetic Quartz," *IEEE Trans. Nucl. Sci.*, vol. NS-16, p. 130, 1969.

[5]L. E. Halliburton et al., "A Study of the Defects Produced by the Irradiation of Quartz," RADC Technical Report TR-80-120, Apr. 1980.

[6]T. M. Flanagan, "Hardness Assurance in Quartz Crystal Resonators," *IEEE Trans. Nucl. Sci.*, vol. NS-21, p. 390, 1974.

[7]J. J. Suter, J. R. Norton, and J. M. Cloeren, "Gamma Ray and Proton Beam Radiation Testing of Quartz Resonators," *IEEE Trans. Nucl. Sci.*, vol. NS-31, p. 1230, 1984.

[8]D. Breuner, "Radiation Effects Analysis of EG&G Rubidium–Crystal Hybrid," JayCor Report J200-85-936A/2404, San Diego, 1986.

[9]J. J. Suter, "Low-Earth-Orbit Proton and Gamma Ionization Effects in Piezoelectric Alpha Quartz Crystal Resonators," *IEEE Trans. Nucl. Sci.*, vol. 37, no. 2, p. 524, 1990.

[10]E. G. Stassinopoulos and J. M. Barth, "Transport and Shielding Analysis of the Non-Equatorial Terrestrial Low-Altitude Charged Particle Radiation Environment," I, NASA Report X-601-84-6, 1984.

[11]J. D. Kinnison, R. H. Maurer, and T. M. Jordan, "Estimation of the Charged Particle Environment for Earth Orbits," *Johns Hopkins APL Technical Digest*, vol. 11, nos. 3 and 4, p. 300, 1990.

[12]D. Breuner, "Hydrogen Maser Radiation Hardening," JayCor Report J200-90-1809/2503, San Diego, 1990.

[13]G. C. Messenger and M. S. Ash, *The Effects of Radiation on Electronic Systems.* New York: Van Nostrand Reinhold, 1991.

[14]L. J. Verhey et al., "The Determination of Absorbed Dose in a Proton Beam for Purposes of Charged Particle Radiation Therapy," *Radiat. Res.*, vol. 79, p. 34, 1979.

[15]J. R. Letaw, "Interplanetary Ionizing Radiation on the Mars Observer Mission," SCG Report 88-03, 30 June 1988.

[16]J. Feynman et al., "New Interplanetary Proton Fluence Model," *Journal of Spacecraft and Rockets*, vol. 27, pp. 403–410, July–August 1990.

[17]J. J. Suter, R. H. Maurer, J. D. Kinnison, R. Besson, J. R. Vig, and A. Koehler, "The Effects of Ionizing and Particle Radiation on Precision Frequency Sources," Johns Hopkins University Applied Physics Laboratory Report SDO-9065, Feb. 1992.

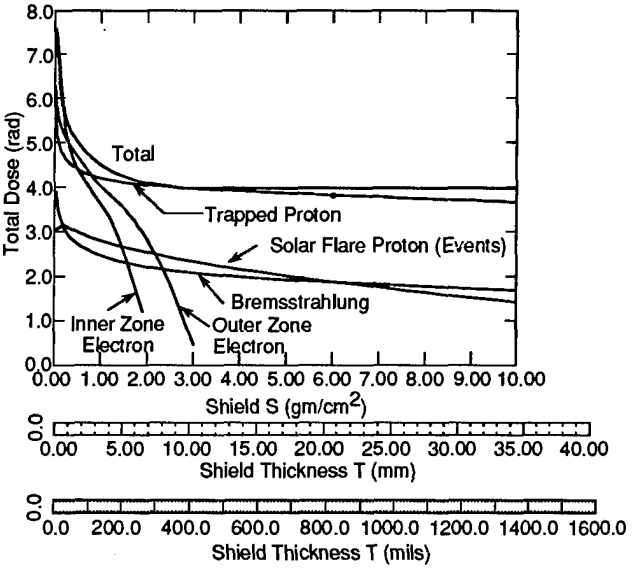


Figure 1. TOPEX Orbit Total Dose. (After Stassinopoulos and Barth, Ref. 10.)

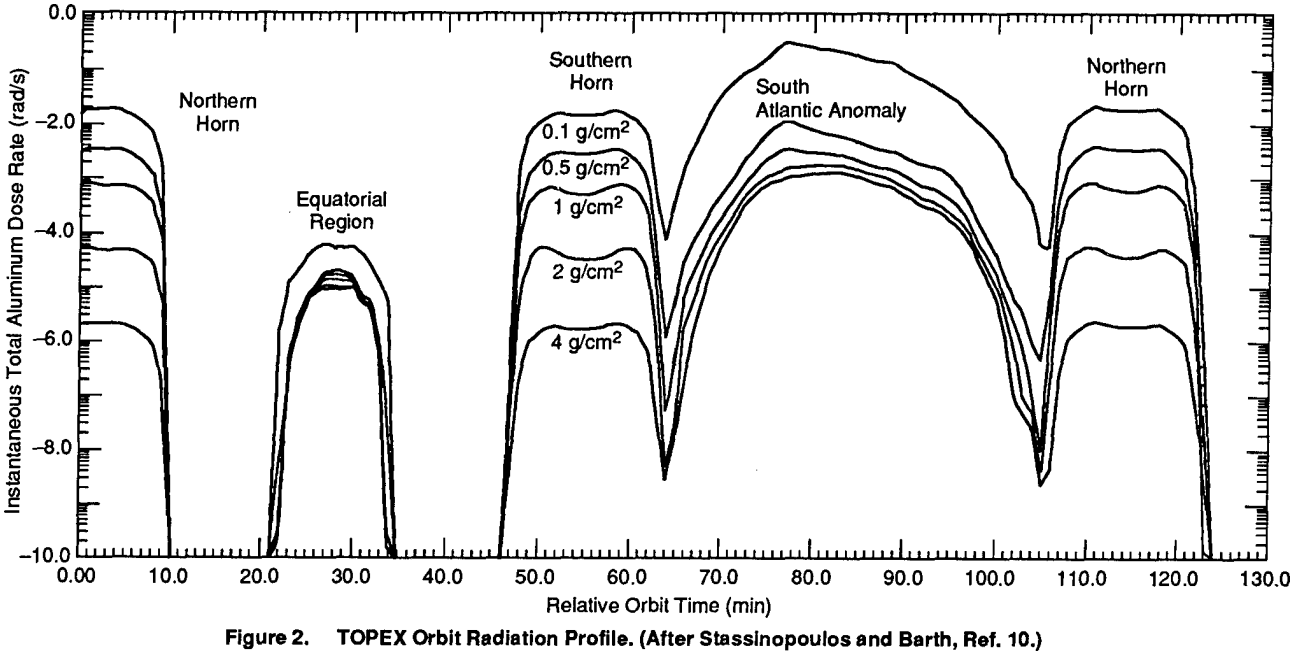


Figure 2. TOPEX Orbit Radiation Profile. (After Stassinopoulos and Barth, Ref. 10.)

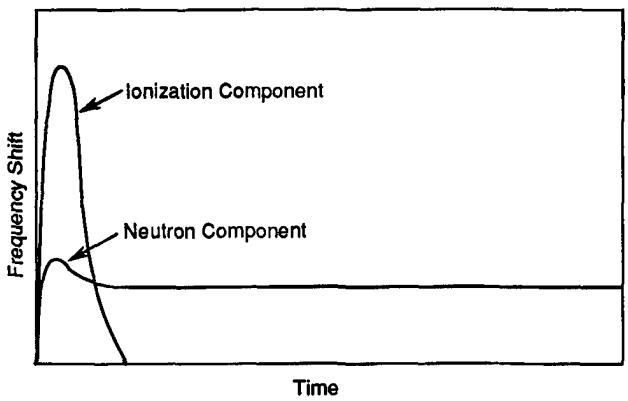


Figure 3. Weapons Environment.

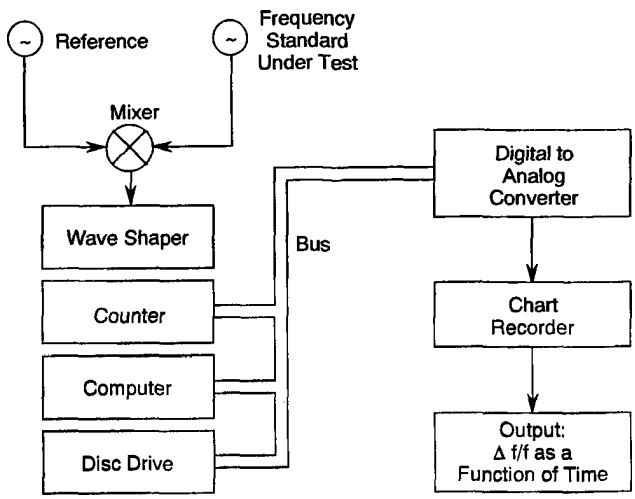


Figure 4. Frequency Shift Data Acquisition System.

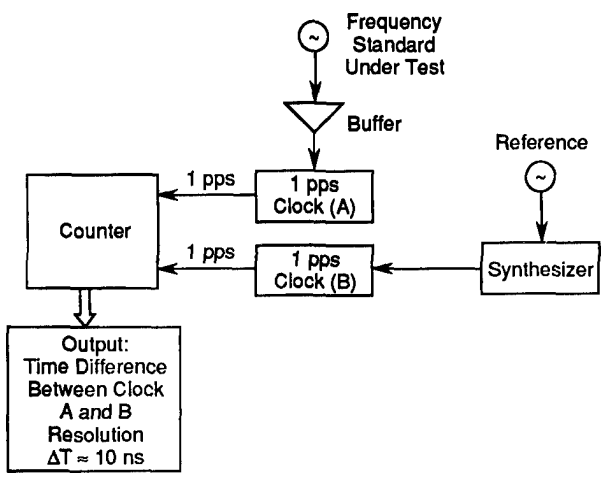


Figure 5. A Method for Measuring Time Loss Due to Momentary Outages Caused by Prompt Radiation.

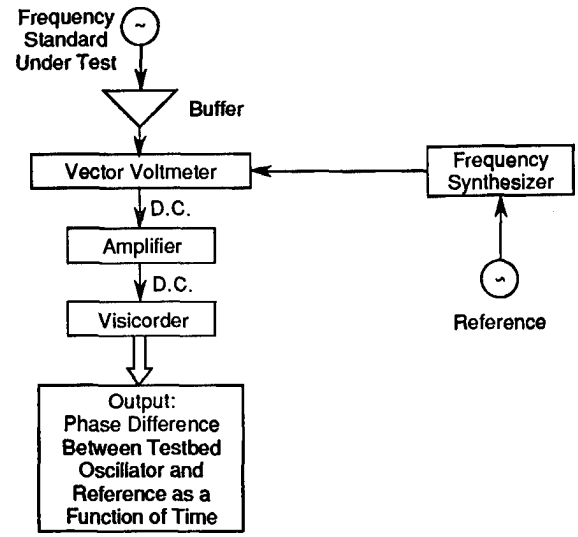


Figure 6. Phase Measurement System.

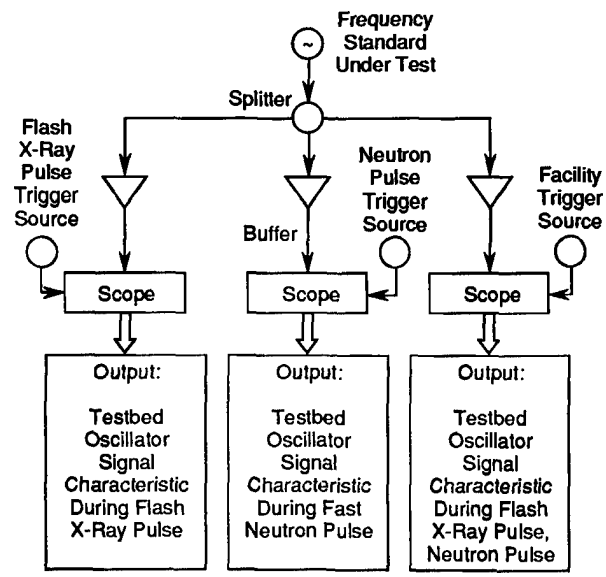


Figure 7. Test Setup for Recording Frequency Standard Output During Flash X-Ray and Neutron Pulses.

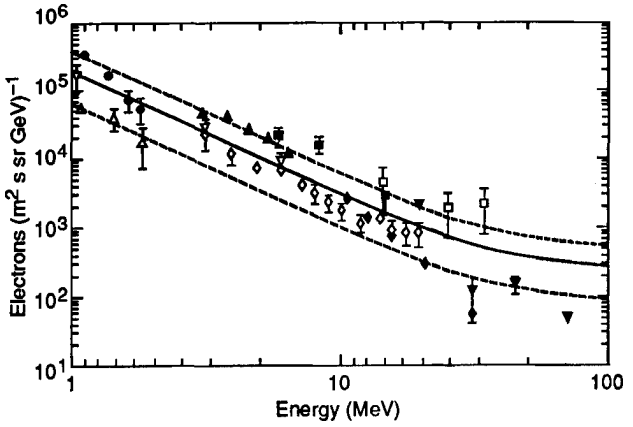


Figure 8. Low-Energy Electron Data. (After Letaw, Ref. 15.)

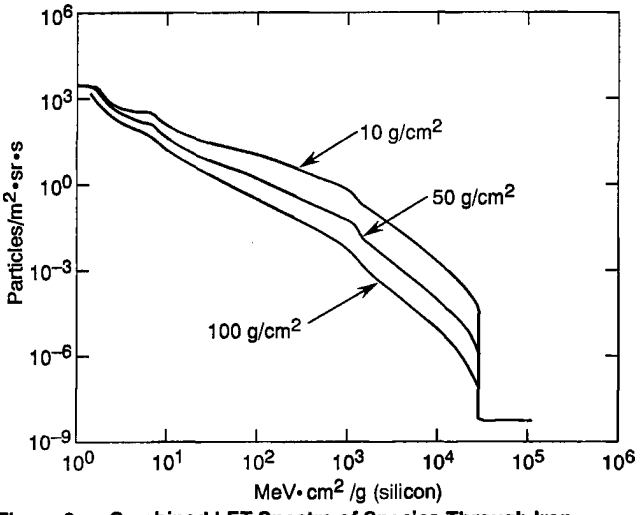


Figure 9. Combined LET Spectra of Species Through Iron Found in Cosmic Rays. (After J. R. Adams, "The Variability of Single Event Upset Rates in the Natural Environment," *IEEE Trans. Nucl. Sci.*, vol. 30, pp. 4475–4480, 1983. © 1983 by IEEE.)

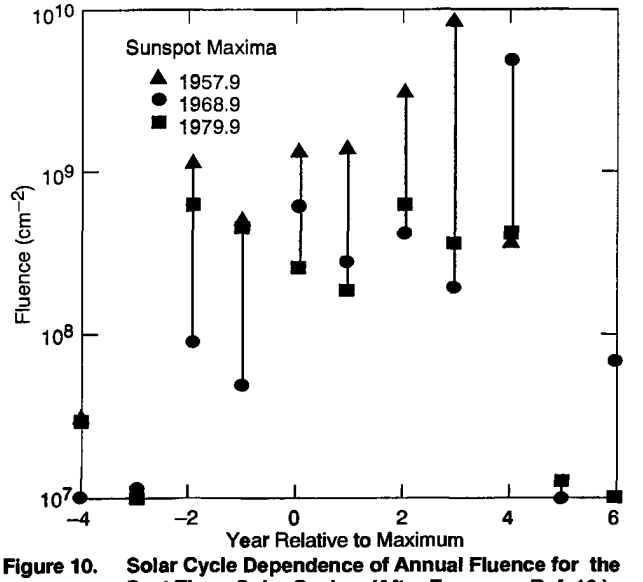


Figure 10. Solar Cycle Dependence of Annual Fluence for the Past Three Solar Cycles. (After Feynman, Ref. 16.)

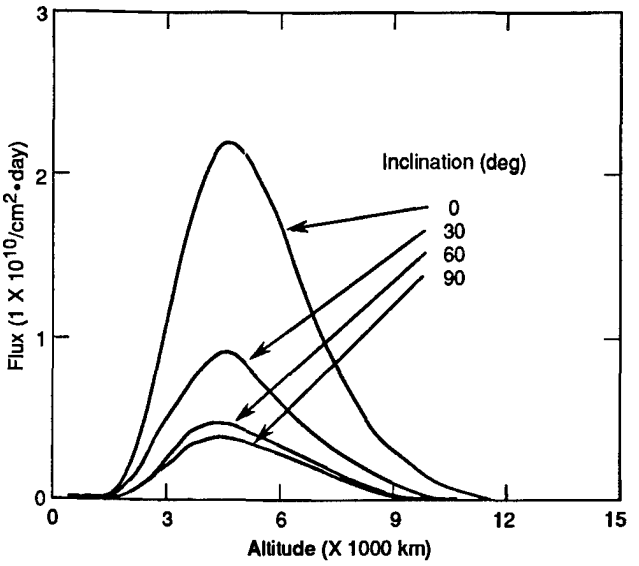


Figure 11. Plot of the Integrated Proton Flux (Energy > 10 MeV) Observed in Circular Orbit. (After Kinnison, Ref. 11.)

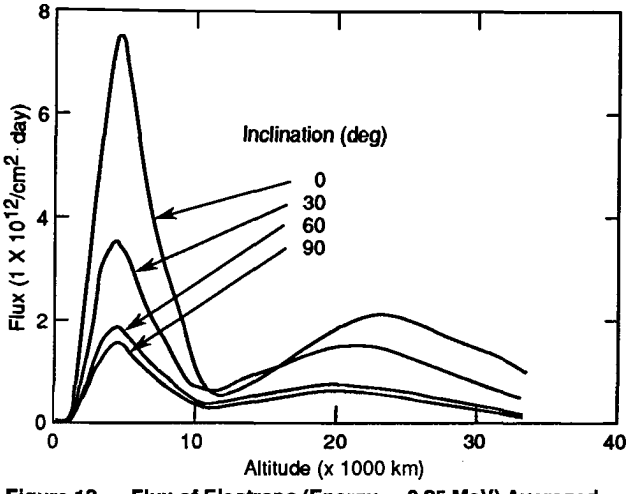


Figure 12. Flux of Electrons (Energy > 0.25 MeV) Averaged Over Several Orbits at Each Inclination and Altitude. (After Kinnison, Ref. 11.)

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

AGING, WARM-UP TIME AND RETRACE; IMPORTANT CHARACTERISTICS OF STANDARD FREQUENCY GENERATORS

(proposal for IEEE Standards, Project P1193)

J. Vanier❶, J.J. Gagnepain❷, W.J. Riley❸, F.L. Walls❹, M. Granveaud❺

❶National Research Council, Ottawa, Canada, ❷Centre National de la Recherche Scientifique, Paris, France, ❸EG&G, Salem, Mass., ❹National Institute of Standards and Technology, Boulder, Co. ❺Laboratoire Primaire du Temps et des Fréquences, Paris, France

Abstract

This report is an effort to produce an IEEE standard providing guidelines for standardized methods of defining, measuring and reporting environmental sensitivities of precision frequency generators. The present report covers the subject of aging, warm-up time and retrace.

Definitions

Aging

The frequency of a typical precision frequency source may vary with time as shown in Figure 1. In the graph, a long term change of the frequency in one direction is observed after the turn on transient.

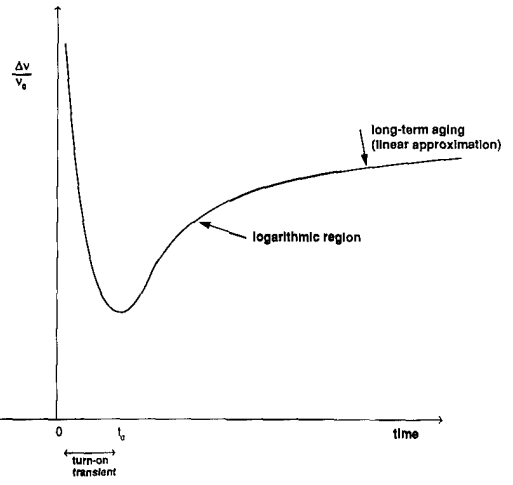


Fig. 1 Illustrations of the againg characteristic of a precision frequency source.

Shorter term fluctuations generally superimposed on the long term changes are not shown. These shorter term fluctuations may sometimes be correlated with temperature or barometric fluctuations. In such cases, their origin is known and, with adequate techniques, they can be remedied if a need is expressed. The longer term fluctuations in many cases, however, are more troublesome and it may prove to be impossible to remedy them. They may be caused for example by internal changes of the controlling element of the precision frequency source e.g. relaxation or outgassing of the quartz crystal in a quartz oscillator or the evolution of the cell in a rubidium frequency standard. In general these effects are called aging. These phenomena should be differentiated from the much broader concept of frequency drift (ref: IEEE Std 1139-1988). The frequency of an oscillator may drift with time for several reasons. For example a long term continuous change in the environment temperature may cause a long-term change in the oscillator frequency. The size of the frequency change will depend on the oscillator control element temperature sensitivity and the gain of the oscillator temperature controls. This may go on for a long period (until the temperature variations change) and, with limited measurement capability, could be mistakenly identified as "aging." We reserve the concept of aging for changes in the fundamental intrinsic properties of the control element of the oscillator, a property that characterize for example a class of oscillators fabricated by means of similar manufacturing processes. There are some areas where the distinction between aging and drift may become blurred, sometimes because of superimposed effects difficult to assess. For example a hydrogen maser's frequency may drift because of changes in its cavity frequency. If the cavity material creeps slowly with time, the frequency change observed appears as aging of the maser. However, the effect can be readily corrected by a simple re-tuning of the cavity which can be done automatically, and the "aging" disappears. On

0-7803-0476-4/92 $3.00 © 1992 IEEE

the other hand, some bulb coatings have been known to change characteristics with time with the consequence that the maser frequency varies (ages) in a continuous fashion for very long periods of time without any ability to remedy this condition operationally. Similar discussions can be held in connection with Rb frequency standards, quartz oscillators, Cs beam frequency standards.

Consequently, we define aging as follows :

"Aging is the systematic change in frequency with time due to internal changes in the oscillator."

This excludes frequency fluctuations caused by environmental changes as well as frequency fluctuations caused by malfunctioning oscillator components which could be improved or repaired. Aging in this context is thus a property of a type or class of oscillators and not of a defective unit in a changing environment.

Warm-up time (T_{wu})

An oscillator is characterized by various properties which are normally grouped under the form of a table. Similar oscillators are generally given the same fundamental and environmental specifications which include for example accuracy, frequency stability, temperature sensitivity, magnetic field sensitivity etc. These specifications are usually quoted for steady state operation reached only after a sufficient period of continuous operation as illustrated in Figure 2. This period is called warm up time and we simply define it as follows:

"Warm-up time (T_{wu}) is the time taken by an oscillator, after turn on, to reach a steady state in which the quoted specifications are met."

The concept "steady state" is important because of possible overshoot in the oscillator characteristics. In such a case damped oscillations may even be present in the characteristics after turn-on. Steady state may be claimed to exist only after these oscillations have decayed to a level such that specifications are continually met at subsequent times.

It is worth mentioning that warm-up depends on the environmental history of the oscillator before turn-on, in particular on its environmental temperature.

Retrace (R)

Oscillators are subject to many environmental changes of which the most drastic one is an interruption of operation called on/off operation. An important characteristic is the reproducibility of its specifications upon such a drastic environmental change. In general, for most oscillators, most specifications such as frequency stability, magnetic field sensitivity and several others will reproduce closely upon on/off cycling (after the appropriate warm-up period). However the frequency itself may be different after such an operation and to an extent that makes it clearly measurable, i.e. greater than the variation caused by its inherent frequency instability as illustrated in Figure 3.

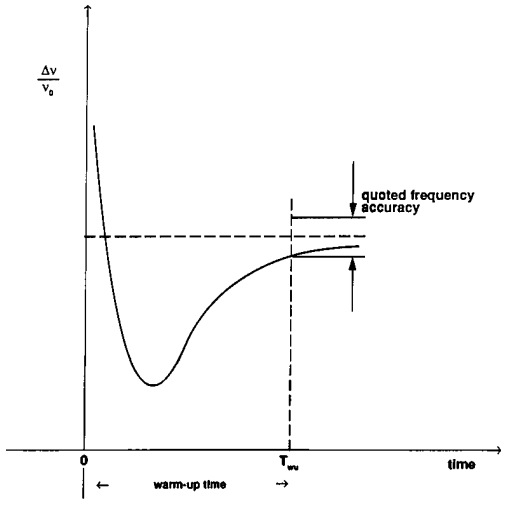


Fig. 2 Illustration of the warm-up characteristic of a precision frequency source.

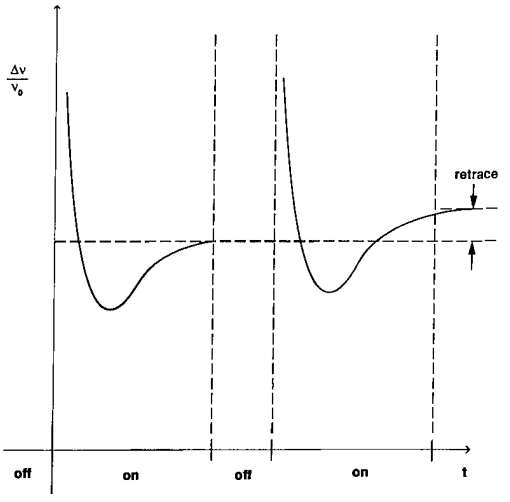


Fig. 3 Illustration of the retrace characteristic of a precision frequency source.

Thus the frequency of the oscillator is reproducible only to a certain extent during on/off operations. The following definition is given:

"Retrace (R) is the change of frequency of an oscillator after exposure to on/off operations measured after the specified warm-up time."

This is a somewhat restrictive definition. It does not include for example other cycling operations such as on/off magnetic fields and abrupt cycling changes in temperatures. The term retrace has also been used in certain instances for characterizing other phenomena such as hysteresis in the frequency/temperature characteristics of a crystal oscillator when subjected to temperature cycling. The use of the term "retrace" is discouraged for such situations.

Formulation of a standard method of reporting Aging, Warm-Up Time and Retrace

General

1. If values are specified that apply to all individual samples of a particular type of frequency source (product specification), maximum values must be given.
2. For quantities defined here it is recommended that relative frequency be used for reporting data.

Aging

In general the long term behaviour of frequency with time of a precision frequency source is non-linear. In quartz crystal oscillators, for example, the change in frequency after turn-on may be represented by a logarithmic equation. After a period of time, characteristics of the particular source, the change may be approximated by a linear equation. In most atomic frequency standards, except may be in Rubidium standards, there is not sufficient data to conclude on a given law and thus long term behaviour is normally characterized by means of a linear equation.

a) In the logarithmic model, the frequency change is represented by:

$$\frac{\Delta\nu}{\nu_0} = A \ln (B \cdot (t - t_0) + 1) \qquad (1)$$

where $\Delta\nu = \nu - \nu_0$, ν is the frequency at time t, ν_0 is the frequency at $t = t_0$, normally taken as the time at which turn-on transients have decayed to a negligible level. A and B are determined by a least-squares fit. A has no unit and B has the unit of inverse time.

b) In the linear model, the frequency change is represented by:

$$\frac{\Delta\nu}{\nu_0} = C \cdot (t - t_0) \qquad (2)$$

where C is the aging rate reported on the basis of per day, per month or per year, and is determined by a linear least-squares fit, and the other symbols are defined as above.

Effects of environmental and short term fluctuations must be removed from the specifications, i.e. they either must be controlled to be sufficiently small or they must be measured and the data correspondingly adjusted.

Warm-up Time

Warm-up time may be given in minutes, days or other units of time which appear most appropriate. The environmental temperature at which the specification applies should be given and it is assumed that the oscillator while off has reached thermal equilibrium with its environment at the time of turn on. An example for stating a warm-up time specification is:

T_{wu} = less than 12 minutes for temperatures between -25°C and 0°C
T_{wu} = less than 5 minutes for temperatures between 0°C and 25°C.

Retrace

If the off-time duration is not assumed to be much larger than the warm-up time it must be given as part of the specification. If the warm-up time is not given, the specification should include detailed data on the assumption made including the time of turn off and turn on. The specification should also give the range of environmental conditions which are relevant and over which retrace specifications apply if more limited than the environmental ranges otherwise specified, e.g. temperature, pressure, humidity, magnetic field intensity, etc. With the exception of aging, the unit may not move outside its retrace specification during continuous operation or discontinuous operation within the given environmental constraints. The specification should also mention if the unit is

sensitive to repetitive on/off cycling. For example, the specification could be:

Retrace or R = ± 5 x 10^{-11} for any turn off time greater than 6 hours. (Specification applies after operating more than 4 hours). Specification applies for temperatures between -20 and +30°C, barometric pressure between 50 kPa and 105 kPa and relative humidity between 20% and 65%. This statement is required when warm-up specification is not given.

Test methods

Aging

An aging measurement shall, to the greatest extent practical, avoid the effects of environmental sensitivities (such as temperature and barometric pressure), noise, and reference error. The aging data can be continuous or sampled. In the latter case, measurements shall be made often enough to resolve the true shape of the aging characteristic. Supplementary continuous data (or data taken at shorter intervals) may be necessary to establish the absence of environmental and other disturbances faster than, or synchronous with, the normal sampling interval. All measurements shall extend over sufficient time that the noise inherent in the measurements is smaller than the frequency changes to be measured (or specified) in cases where the aging is small. Measurements should start after all transient frequency changes caused by the turn-on of the oscillator have become negligible compared to the aging itself. The identity of the reference shall be stated, as well as the measurement averaging time, the measurement interval and any data averaging that is done.

The measurement of aging thus requires a) ascertaining the integrity of the frequency source and b) the stabilization of environmental parameters to levels below those which cause frequency changes of the order of magnitude of the observed aging (i.e. drift and aging must be identified).

The aging measurement shall extend over a period of time sufficient to show the intended information. The measurement of the aging of a frequency source over a certain period requires data over at least that length of time. Extrapolation shall not be used unless explicitly stated. For example, a statement that the aging rate is 3 x 10^{-10}/month implies that data was taken for at least 1 month, and that this is the average linear aging over a month. It should be neither the result of only one week's data nor the slope at the end of one month. The latter should be expressed as (say) 1 x 10^{-11}/day after 30 days. Continuous operation of the source is assumed.

Comments on specific sources

Crystal Oscillators (XO): The relatively low noise and high aging of a typical crystal oscillator generally results in a well-defined fit to an aging model. The aging is usually determined by such processes as stress relaxation and redistribution of contamination that decrease with time and are often described by a logarithmic model. Some units exhibit a change in sign of aging. This is because there is more than one source of aging of opposite signs.

Rubidium Gas Cell Standards (RbFS): The relatively low aging of a typical rubidium gas cell frequency standard can easily be masked by noise or environmental sensitivities. The early aging is usually determined by such stabilization processes as Rb redistribution in the cells, and a logarithmic model may be used. The later aging is usually quite low and steady, and a linear model may be used.

Rubidium-Crystal Oscillators (RbXO): The aging of a RbXO is that of its XO during low-power operation between Rb synchronizations and that of its Rb reference during full-power operation and over the long term.

Hydrogen Masers (HM): The aging of a HM is still a question of debate. For masers with automatic or continuous tuning of their cavities, sometimes a change of wall shift, yet to be identified, causes a continuous aging. A linear model is often used .

Cesium Beam Frequency Standards (CsFS): An exceptionally stable reference and control of environmental conditions is necessary to measure the aging of these devices. The relatively high noise requires a long averaging time. In general the very low aging of a cesium beam tube frequency standard is considered to be consistent with zero.

Warm-up time

The warm-up behavior may be strongly dependent on the operating conditions, particularly environmental temperature and/or electrical supply voltage. All such factors should be considered. When the warm-up depends on temperature, the device should be allowed to soak off long enough to reach complete equilibrium.

This is particularly important for a device with a well-insulated internal oven.

The warm-up characteristics of interest can span a range of second to days (or longer). The data often takes the form of strip-chart records of such variables as input current and frequency. As a minimum, these variables should be recorded at intervals not exceeding one-tenth of the specified warm-up time.

Comments on specific sources

Crystal Oscillators: the most significant XO warm-up characteristic may be attaining/regaining a certain aging rate, and the specifications should be specific on this.

Rubidium Gas Cell Standards: The warm-up of a (RbFS) is often defined as the time to atomic lock and/or a certain absolute frequency tolerance (after prior calibration). This is generally a few minutes, and depends strongly on the environmental temperature and the supply voltage. Lockup time should be stated as the time to reach and remain in atomic lock, as indicated by a status signal. The energy required to attain a certain accuracy may be important for battery-backed applications. (This is referred to "syntonization energy" for a RbXO.) The reference point for reaching a certain frequency accuracy during warm-up is most often based on the "final" frequency-the settled value after no significant additional change takes place. The warm-up criterion is usually such that a stabilized value is quickly attained. It is unusual, however, to have a warm-up requirement related to attaining/regaining a certain aging rate.

Rubidium-Crystal Oscillators: The syntonization time and energy is critical for many RbXO applications, and the specification should be clear on this question.

Hydrogen Masers: The warm-up of a HM is ordinarily very long. Generally these devices are kept in continuous operation and warm-up conditions are not necessarily given. These devices are not designed in such a way as to optimize this parameter.

Cesium Beam Frequency Standards: The warm-up of a CsFS is usually associated with its attaining a specified frequency accuracy. All other characteristics are assumed to be reached upon warm-up unless specified otherwise.

Retrace

While the retrace requirement can involve a number of environmental and other factors, retrace, by definition, implies a return to exactly the same operating conditions. For example, a typical retrace test might involve turning a device off and on a number of times while measuring the variation in stabilized frequency. The operating conditions (temperature, supply voltage, orientation, etc) of the device should not change between runs. The test should emphasize the consistency of the retrace, and whether or not the unit shows any accumulative frequency change (trend). A sufficient number of runs should be made to show this (as opposed to fewer runs under different operating conditions). Sufficient off time should be allowed so that all internal parts of the unit reach ambient temperature. Sufficient on time should be allowed for the unit to fully restabilize.

A related factor is the on-off and/or thermal-power cycling endurance of the frequency source. This is especially important for a device such as a RbXO that is intended for many on-off cycles. A design verification test consisting of thousands of on-off and thermal power cycles may be necessary to qualify such a device.

Comments on specific sources

Crystal Oscillators: Phenomena taking place within a quartz crystal when the environment is changed (temperature, rf, power) are causes which limit the retrace. Of particular importance is the hysteresis that may be present to some extent. In general the retrace will be clearly visible over the normal aging of the oscillator.

Rubidium Gas Cell Standards: The controlling element being the atomic resonance of the rubidium gas, retrace is generally a characteristic which makes rubidium frequency standards the units of choice in some specific applications. In some types of design retrace may be of the order of the aging rate, that is barely measurable. However, in some other types of design, rubidium redistribution upon turn off, cooling and warming, may cause important changes in the characteristics of the standards with resulting measurable frequency shifts.

Rubidium-Crystal Oscillators: The retrace characteristic of the Rb reference of a RbXO is critical because, for intermittent operation, it is the retrace rather than the aging that determines the long-term stability.

Hydrogen Masers: Unless it has been turned off for a long period, the question of retrace of a hydrogen maser is normally connected to the ability to reproduce or retrace the microwave cavity. For very long periods, wall shift drift may have some importance.

Cesium Beam Frequency Standards: In general cesium beam frequency standards are used for their long-term stability and as étalons in the frequency domain. Their accuracy is given and their retrace is closely connected to their quoted accuracy.

Summary

An important characteristic of a precision frequency source or a standard frequency generator is the variation of its output frequency, caused either by internal long-term changes or alterations of operating conditions, sometimes called environmental effects. The spectrum of these changes and their causes is very large. In this report, we limit ourselves to the concept of aging, warm-up time and retrace characteristics of precision oscillators. They are examined in connection with their definition, methods of reporting their size in a given device, and techniques of measurements.

Bibliography

Books

Design of Crystal and Other Harmonic Oscillators, B. Parzen with A. Ballato, 1983, John Wiley and Sons, New York.

Precision Frequency Control, Ed. E. A. Gerber, A. Ballato, 1985, Academic Press Inc. New York .

"The Quantum Physics of Atomic Frequency Standards", J. Vanier and C. Audoin, Adam Hilger, Editor, 1989, Bristol, England.

Standards

Comité Consultatif International de Radio Communication - Glossary (CCIR), International Telecommunications Union, Geneva.

IEEE Standard Definitions of Physical Quantities for Fundamental Frequency and Time Metrology, IEEE Std. 1139-1988, The Institute of Electrical and Electronics Engineers, Inc., 345 East 47th Street, New York, NY 10017, USA.

MIL Specification 0-55310 (USA), General Specification for Oscillators, Crystal, Military Specifications and Standards, 700 Robbins Ave., Bldg. 4D, Philadelphia, PA 19111-5094.

Articles

Dynamic Temperature Behavior of Quartz Crystal Units, W.H. Hicklin, Proc. 24th Annual Frequency Control Symposium, **1970**, 148-156. (Crystal, warm-up, aging)

Rubidium Frequency and Time Standard for Military Environment, M.E. Frerking and D.E. Johnson, Proc. 26th Annual Frequency Control Symposium, **1972**, 216-222. (Rb, warm-up, aging)

A Fast Warmup Quartz Crystal Oscillator, H.M. Greenhouse, R.L. McGill and D.P. Clark, Proc. 27th Annual Frequency Control Symposium, **1973**, 199-217. (Crystal, warm-up, retrace)

Evaluation of a Rubidium Standard for Satellite Application, S.A. Nichols, J.D. White, and R.B. Moore, Proc. 27th Annual Frequency Control Symposium, **1973**, 390-399. (Rb, aging)

Short and Long Term Stability Measurements Using Automatic Data Recording System, J.A. Bowman, Proc. 27th Annual Frequency Control Symposium, **1973**, 440-445. (Meas. system)

Aging Effects in Plasma Etched SAW Resonators, D.T. Bell Jr., and S.P. Miller, Proc. 30th Annual Frequency Control Symposium, **1976**, 358-362. (Saw, aging)

Aging Analysis of Quartz Crystal Units with Ti Pd Au Electrodes, G.L. Dybwad, Proc. 31st Annual Frequency Control Symposium, **1977**, 144-146c. (Crystal, aging)

Time-Dependent Frequency Shifts in the Hydrogen Maser, D. Morris, IEEE Trans. Instr. & Meas. IM-27, 339-343, **1978**. (H maser, aging)

Performance Results of an Oscillator Using the SC Cut Crystal, R. Burgoon and R.L. Wilson, 33rd Annual Frequency Control Symposium, **1979**, 406-410. (Crystal, aging, warm-up)

Tactical Miniature Crystal Oscillator, H.W. Jackson, 34th Annual Frequency Control Symposium, **1980**, 449-456. (Crystal, aging, retrace, warm-up)

A Miniature, High performance Rubidium Frequency Standard, T. Hashi, K. Chiba and C. Takeuchi, Proc. 35th Ann. Symp. Frequency Control, **1981**, 646-650. (Rb, warm-up)

Frequency Retrace of Quartz Oscillators, F. Euler and N. Yannoni, Proc. 35th Ann. Symp. Frequency Control, **1981**, 492-500. (Crystal, retrace)

Long Term Performance of VLG11 Masers, J. White and K. McDonald, Proc. 35th Ann. Symp. Frequency Control, **1981**, 657-661. (H maser, aging)

No Warm-up Crystal Oscillator, D.H. Phillips, Proc. 13th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1981**, 831-850 (Crystal, warm-up)

Performance of Compact H Masers, A. Kirk, Proc. 37th Ann. Symp. Frequency Control, **1983**, 42-48. (H maser, aging)

Rubidium Frequency Standard Study, T.M. Kwon, B.C. Grover, and H.E. Williams, RADC-TR-83-230, Final Technical Report, Rome Air Development Center, Oct. **1983**, AD-A134713. (Rb, general)

Aging Studies on Quartz Crystal Resonators and Oscillators, R.L. Filler, J.A. Kosinski, V.J. Rosati and J. R. Vig, Proc. 38th Ann. Symp. Frequency Control, **1984**, 225-232. (Crystal, aging)

High Performance from a new design of Crystal Oscillator, G. Beauvy, G. Marotel, P. Renoult,Proc. 16th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1984**, 375-384 (Crystal, aging, warm-up, retrace)

New Approach of Fast Warm-up for Crystal Resonators and Oscillators, J. P. Valentin, M.D. Decailliot and R.J. Besson,Proc. 38th Ann. Symp. Frequency Control, **1984**, 366-373. (Crystal, warm-up)

On-orbit Frequency Stability Analysis of the GPS NAVSTAR-1 Quartz Clock and NAVSTAR-6 and 8 Rubidium Clock, T.B. McCaskill, J.A. Buisson, M.M. Largay, W.G. Reid, Proc. 16th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1984**, 103-126 (Crystal,Rb, aging)

Performance Data of U.S. Naval Observatory VLG 11 Hydrogen Masers since September 1983, R.F.C. Vessot, E. M. Mattison, E.A. Imbier, Z.C. Zhai, W.J. Klepczynski, P.G. Wheeler, A.J. Kubik and G.M.R. Winkler, Proc. 16th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1984**, 375-384. (H Maser, aging)

Rubidium and Cesium Frequency Standards Status and Performance of the GPS Program, H. Bethke, D. Ringer and M. van Melle, Proc. 16th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1984**, 127-142. (Cs, Rb, aging)

The Rubidium-Crystal Oscillator Hybrid Development Program, J.R. Vig and V.J. Rosati, Proc. 16th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1984**, 157-166. (Rb, aging, retrace)

Aging Specification, Measurement and Analysis, R.L. Filler, Proc. 7th Quartz Devices Conf. & Exh. **1985**, 93-104. (Crystal, aging)

Hybrid Miniature Oven Quartz Crystal Oscillator, J. Ho,Proc. 39th Ann. Symp. Frequency Control, **1985**, 193-196. (Crystal, aging, warm-up)

Properties of Low Expansion Materials for Hydrogen Maser Cavities, E.M. Mattison, R.F.C. Vessot and S.F. Jacobs, Proc. 39th Ann. Symp. Frequency Control, **1985**, 75-79. (H maser, aging)

Results of Two Years of Hydrogen Maser Clock Operation at the U.S. Naval Observatory and Ongoing Research at the Harvard Smythsonian Centre For Astrophysics, R.F.C. Vessot, E.M., Mattison, W.J. Klepczynski, I.F. Silvera, H.P. Godfried and R.L. Walsworth Jr, Proc. 17th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1985**, 413-431 (H Maser, aging)

The Quartz Resonator Automatic Aging Measurement Facility, Proc. 39th Ann. Symp. Frequency Control, **1985**, 75-79. (Crystal, aging, meas. system)

A Disciplined Rubidium Oscillator, A. McIntyre and S.R. Stein, Proc. 40th Ann. Symp. Frequency Control, **1986**, 465-469. (Rb, general)

A Partial Analysis of Drift in the Rubidium Gas Cell Atomic Frequency Standard, J.C. Camparo, Proc. 18th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1986**, 565-588. (Rb, aging)

A Rubidium- Crystal Oscillator (RbXO), W.J. Riley Jr. and J.R. Vaccaro, Proc. 40th Ann. Symp. Frequency Control, **1986**, 452-464. (Rb, general)

Aging Measurement on Quartz Crystals in the Batch Mode, E. Hafner and H.W. Jackson, Proc. 40th Ann. Symp. Frequency Control, **1986**, 306-312. (Crystal, Meas. system)

Subminiature Rubidium Oscillator, Model FRS, W. Weidemann, Proc. 40th Ann. Symp. Frequency Control, **1986**, 470-473. (Rb, warm-up, aging)

A Rubidium-Crystal Oscillator (RbXO), W.J. Riley and J.R. Vaccaro, IEEE Trans. on Ultrason. Ferroelec. Freq. Cont. UFFC34, **1987**, 612-618. (Rb, general). See also: Rubidium-Crystal Oscillator (RbXO) Development Program, W.J. Riley and J.R. Vaccaro, Research and Development Technical Report SLCET-TR-84-0410-F US Army Laboratory Command, Fort Monmouth, New-Jersey 07703-5302.

Characteristics and Performance of Miniature NBS Passive Hydrogen Maser, F.L. Walls, IEEE Trans. Instrum. Meas. IM36, **1987**, 596-603. (H maser, aging)

Experiences with EFOS 1 and EFOS 3 at the Fundamental Station Wettzell, W, Schluter Proc. 1st Eur. Time and Frequency Forum, **1987**, 342-347. (H maser, aging

Frequency-Temperature and Aging Performance of Microcomputer Compensated Crystal Oscillators, R. Miller, J.A. Messina and V.J. Rosati, Proc. 43rd Ann. Symp. Frequency Control, **1987**, 27-33. (Crystal, aging)

Long Term Performance of the Johns Hopkins-Built Hydrogen Maser, L.G. Rueger and M.C. Chiu, IEEE Trans. Instrum. Meas. IM36, **1987**, 594-595. (H maser, aging)

New Insights into Causes and Cures of Frequency Instabilities (Drift and Long Term Noise) in Cesium Beam Frequency Standards, A. De Marchi, Proc. 41st Ann. Symp. Frequency Control, **1987**, 53-58. (Cs, aging)

Review of the Revised Military Specification of Quartz Crystal Oscillators, S. Schodowski and V.Rosati, Proc. 41st Ann. Symp. Frequency Control, **1987**, 466-470. (Crystal, aging)

Specification of Precision Oscillators, S.S. Schodowski and J.R. Vig, Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1987**, 163-174. (General)

Surface Interaction of Atomic Hydrogen with Teflon, E.M. Mattison and R.F.C. Vessot, Proc. 41st Ann. Symp. Frequency Control, **1987**, 95-98. (H maser, aging)

Tactical Rubidium Frequency Standard (TRFS), T.J. Lynch, and W.J. Riley, Final Technical Report, RADC-TR-87-166, Vol. 1, **1987**, AD-A192981. (Rb, general)

The Aging of Resonators and Oscillators under Various Conditions, R. Filler, R. Lindenmuth, J. Messina, V. Rosati and J. Vig. Proc. 41st Ann. Symp. Frequency Control, **1987**, 444-451. (Crystal, aging, retrace)

Understanding Environmental Study of Aging of Cesium Beam Frequency Standards, A. De Marchi, Proc. 1st Eur. Time and Frequency Forum, **1987**, 288-293. (Cs, aging)

Aging Prediction of Quartz Crystal Units, M. Miljkovic, G. Trifunovic, V. Brajovic,Proc. 42nd Ann. Symp. Frequency Control, **1988**, 404-411. (Crystal, aging)

Experimental Results on Aging of AT Cut Strip Resonators, J. Gebreke and R. Klawitter,Proc. 42nd Ann. Symp. Frequency Control, **1988**, 412- 418. (Crystal, aging)

Characterisation Methods for the Sensitivity of Quartz Oscillators to the Environment, J.J. Gagnepain, Proc. 43rd Ann. Symp. Frequency Control, **1989**, 242-247. (Meas. system)

Testing and Specification of Environmental Sensitivities in Cesium and Precision Quartz Signal Sources, R.M. Garvey, Proc. 43rd Ann. Symp. Frequency Control, **1989**, 236-267 (Cs, Crystal Meas. system)

The Testing of Rubidium Frequency Standards, T.J. Lynch, W.J. Riley Jr. J. R. Vaccaro. Proc. 43rd Ann. Symp. Frequency Control, **1989**, 257-262. (Rb, Meas. system)

A BVA Quartz Crystal Oscillator for Severe Environments, Proc. 44th Ann. Symp. Frequency Control, **1990**, 593-596. (Crystal, aging)

Design and Industrial Production of Frequency Standards in the USSR, N.A. Demidov, and A. A. Uljanov, Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 187-208. (Rb, H Maser, cesium, general)

Long Term Microwave Power Drift of a Cesium Frequency Standard for GPS Block IIR, W. J. Riley, Proc. 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 209-219. (Rb, aging, retrace)

On the Line Q Degradation in Hydrogen Masers, L.G. Bernier, G. Busca, H Schweda, Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 599-606, (H Maser, aging)

Perfomance of SAO Model VLG-12 Advanced Hydrogen Masers, E.M. Mattison and R.F.C. Vessot, Proc. 44th Ann. Symp. Frequency Control, **1990**, 66-69. (H maser, aging)

Performance of Soviet and U.S. Hydrogen Masers, A.A. Uljanov, N. A. Demidov, E.M. Mattison, R.F.C. Vessot, D.W. Allan and G.M.R. Winkler, Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 509- **1990**, 509-524. (, H Maser, aging)

Physical Origin of the Frequency Shifts in Cesium Beam Frequency Standards Related Environmental Sensitivity, C. Audoin, N. Dimarcq, and V.Giordano, Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 419-440. (Cs, general)

Physics of Systematic Frequency Variations in Hydrogen Masers, E. M. Mattison, Proc. 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 453-464. (H Maser, aging)

Rubidium Atomic Frequency Standards for GPS Block IIR, W.J. Riley, Proc. 22nd Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 221-227. (Rb, aging, retrace)

Sensitivity of Quartz Oscillators to the Environment: Characterization and Pitfalls, J.J. Gagnepain, IEEE Trans. Ultrason. Ferroelect. Freq. Contr. UFFC37, **1990**, 347-354. (Quartz, general)

Thermal Hysteresis in Quartz Crystal Resonators and Oscillators, R.L. Filler, Proc. 44th Ann. Symp. Frequency Control, **1990**, 176-184. (Crystal, retrace)

Thermal hysteresis in Quartz Resonators - a Review: J.A. Kusters and J.R. Vig, Proc. of the 44th Annual Symposium on Frequency Control **1990**, 165-175. (Crystal, retrace)

Special Hydrogen Maser Workshop, D. Morris, 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, **1990**, 349-354. (H maser, aging)

Acceptance Test Results for a New Space Rubidium Frequency Standard, T. McClelland, D. Silvermetz, I. Pascaru, M. Meirs, Proc. 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Advance Program, **1991**, p. 40. (Rb, aging)

Introduction to Quartz Frequency Standards, J. R. Vig, Proc. 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Advance Program, **1991**, p. 13. (Crystal, general)

The Aging of Bulk Acoustic Wave Resonators, Filters and Oscillators, J.R. Vig and T.R. Meeker, Proc. 45th Ann. Symp. Frequency Control, **1991**, 77-101. (Crystal, aging)

Communications Frequency Standards, S.R. Stein and J.R. Vig, The Froehlich/Kent Encyclopedia of Telecommunication, Editors: F.E.Froelich and A. Kent, M. Dekker Inc. New York, **1992**. (General)

Environmental Sensitivities of Quartz- Crystal-Oscillators, Fred L. Walls, J.-J. Gagnepain, Proc. UFFC, March **1992**. (Crystal, general)

Physics and Systematic Frequency Variation in Hydrogen Masers, E. Mattison, Proc. UFFC, March **1992**. (H maser, aging)

The Physics of the Environmental Sensitivity of Rubidium Gas Cell Atomic Frequency Standards, W.J. Riley, Proc. UFFC, March **1992**. (Rb, general)

Quartz Crystal Resonators and Oscillators for Frequency Control and Timing Applications, J.R. Vig, U.S. Army Electronics Technology and Devices Laboratory (LABCOM), Fort Monmouth, New Jersey 07703. (Crystal, general)

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

GENERAL CONSIDERATIONS IN THE METROLOGY OF THE ENVIRONMENTAL SENSITIVITIES OF STANDARD FREQUENCY GENERATORS

IEEE STANDARDS COMMITTEE P1193 WORKING GROUP 6 REPORT

Lee A. Breakiron*, Len S. Cutler**, Helmut W. Hellwig***,
John R. Vig****, and Gernot M. R. Winkler*

*U.S. Naval Observatory, Time Service Department, Washington, DC 20392-5100
**Hewlett-Packard Laboratories, P.O. Box 10350, MS 26M-9, Palo Alto, CA 94303-0867
***U.S. Air Force Office of Scientific Research, Bolling AFB, Washington, DC 20332-5448
****U.S. Army Electronics Technology and Devices Laboratory, SLCET-EQ, Fort Monmouth NJ 07703-5302

Abstract

In order to provide guidelines for standardized tests and methods for measuring and characterizing the environmental sensitivities of precision oscillators, the IEEE is preparing a standards document. As part of that effort (Project P1193), this paper reviews general experimental considerations and analytical techniques.

I. Introduction

The basic principle governing the physical measurement of time is demarcation of equal time intervals by observation of a repeating process. More than one such process, or clock, are required for statistical intercomparison (to achieve precision). Isolation of each clock from the rest of the universe is essential for identical process repetition (to achieve stability). Analysis of residual errors attributable to environmental influences allows corrections to a standard system (to achieve accuracy).

Resonances of macroscopic resonators like quartz crystals are determined by the type and dimensions of the material, the particular method of excitation, and other factors. However, changes in the environment (like temperature) that affect those dimensions and factors, as well as other external stimuli (like radiation), inevitably introduce frequency error at some level. Likewise, though isolated atoms at rest oscillate at fixed frequencies, any force acting on them inside an atomic oscillator (e.g. due to exciting microwaves, cavity walls, or magnetic fields) cause frequency shifts. In practice, the apparent frequency of ensembles of atoms can also vary because of such effects as dimensional changes or mistuning of the resonant cavity, imperfections in the detector, and distortions due to the electronics. If any of these effects vary with time, the result is frequency instability. Thus, the instability of a well-adjusted precision oscillator depends on a number of (perhaps interacting) factors mostly relating to the stability of the environment.

It is well known that the frequencies and accuracy of high-precision frequency standards are susceptible to changes in ambient conditions, necessitating their operation in a controlled environment. The most important effects have been found to be acceleration (including vibration) [1], temperature, humidity, barometric pressure [2], load impedance, power-supply voltage, electric and magnetic fields [3], and radiation [4]. The references just cited are the elements of an IEEE standard that provides guidelines for standardized ways of characterizing the environmental sensitivities of clocks and other precision oscillators. As part of that effort, this paper summarizes relevant general and systems considerations.

For the user, it is of the utmost importance to know the magnitudes of an oscillator's sensitivity to each external influence and what accuracy can be anticipated under all expected operating conditions. Accordingly, the manufacturer must employ test methods and evaluation criteria that are reasonably standardized, that accurately predict the performance of his product, and that clearly define the limits on its accuracy imposed by the environment. It is also necessary to understand the source of these sensitivities if ways are to be found to reduce them. For the systems designer, it is necessary to appreciate the relative importance of these sensitivities in order to make the optimal tradeoffs. Standardi-

0-7803-0476-4/92 $3.00 © 1992 IEEE

zation of environmental test methods should ease device specification, simplify test plans, clarify test results, and lead to improved oscillator performance and application. The need for greater knowledge about environmental sensitivities has increased with the increasing accuracy of timing and positioning systems, especially in the extreme environments encountered in land- and sea-mobile, airborne, and spaceborne applications.

II. Analytical Methods

Instabilities in an oscillating system should be analyzed deterministically and stochastically. In the deterministic part of the analysis, systematic effects are modelled by an analytic method like curve-fitting or Kalman filtering the time or frequency data relative to some reference, yielding the derivative (frequency or frequency drift) and higher-order or otherwise nonlinear trends. Steps may have to be removed and gaps bridged, say by cubic splines. After the systematic variation is documented and removed, random errors are evaluated by measures such as the two-sample (Allan) variance in the time domain [5,6,7] or the spectral density function in the frequency domain [8,9].

In the stochastic approach, trends such as a rate (frequency) or drift can often be removed by differentiating the function or differencing the series one or more times. For example, the instantaneous frequency y (t) of a clock is the first derivative of its phase x (t) and is approximated by the first difference in the time. Random errors may then be characterized by the autocorrelation function, the power spectrum, or Allan variance.

Given a continuous signal x (t), say a clock's phase relative to some reference, the "autocorrelation (function)" is:

$$\Phi_{xx}(k) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} x\,(t + k)\; x\,(t)\; dt$$

where t is the time and k is the lag time. Since the data are usually discrete and equally spaced, we can replace the signal with a time series and the autocorrelation function with the kth "autocorrelation coefficient":

$$p_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})\,(x_{t+k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2}$$

where $\bar{x}$ is the mean of x (t) and n is the number of data samples. We have necessarily introduced some sampling error due to our time steps and finite data length. A plot of p_k vs. k is called the "autocorrelogram."

A sinusoidal appearance to the autocorrelogram indicates the presence of periodic phenomena, either internal or environmental in origin. Internal variations (which also include spontaneous phenomena such as phase jumps and relaxation effects) plus statistical noise constitute the inherent noise of the system. Before environmental effects can be investigated, these noise processes must be calibrated (on the basis of system performance in steady state) and corrected for. Such correction is better done during postprocessing than in real time because of the superior determination of constant and periodic characteristics (through averaging or modelling, e.g. with ARIMA [10] or Fourier analysis) and clearer recognition of spontaneous changes (say, by forward and backward filtering).

Time-series analysis generally involves the assumption of ergodicity and stationarity. An "ergodic process" has the property that sample (or time) averages of observations may be used as approximations to the corresponding ensemble (or population) averages. A "stationary series" is one whose mean, variance, and higher statistical moments do not change significantly with time. Generally, approximate ergodicity, and stationarity of the mean, variance, and autocorrelation coefficients for low lag times are sufficient in practice. Approximate stationarity can often be achieved by differentiation or a change of variable, such as use of first differences or logarithms. For example, replacing the phase x (t) above with its first difference y (t) may make the series sufficently stationary. Stationarity is indicated by rapid damping of the autocorrelation coefficients. Drifts with time in the characteristics of a disturbing signal may invalidate the assumptions of ergodicity and stationarity.

Generally, the input x (t) and output y (t) of a linear time-invariant system can be related as follows:

$$y(t) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} w(t-t')\, x(t')\, dt'$$

where w (t) is called the system's "impulse response (function)." The Fourier transforms of y (t), w(t), and x (t), namely Y (t), W (t), and X (t), are related similarly:

$$Y(f) = W(f)\, X(f)$$

where f is the Fourier frequency and W (f) is referred to as the system's "frequency response (function)."

The response of a system, or a quantity under measurement, may be affected by other variables called "input (or influence) quantities", which may be internal or external. We are concerned with the latter, in particular those that depend on time and, hence, may affect our time- or frequency-measuring process. If one can deduce the environmental factors likely to be of significance and has concurrent data on their magnitude, then environmental influences may be investigated with least-squares (see Section V) or cross-correlation techniques (see Appendix). The latter can be compromised if there are different (e.g. thermal) time constants involved [2].

Providing the dependence can be modelled, a correlation coefficient (or, equivalently, a covariance) can be computed whose significance can be tested statistically. When the correlation is significant, the coefficient of dependence can usually be determined with sufficient accuracy to correct for most of the effect, in which case the residuals should be Gaussian or, in the case of frequencies, white noise, i.e. neither auto-correlated nor cross-correlated.

Nonlinear responses can greatly complicate determination of environmental sensitivities and may invalidate statistical measures of frequency stability. Examples of nonlinear behavior are: (1) gross changes in the magnitude and sense of the temperature coefficient of quartz oscillators near "activity dips"; (2) the electric field sensitivity of crystal oscillators; and (3) the magnetic field sensitivities of many types of oscillators.

After all relevant environmental influences have been identified, the "sensitivity coefficients (or factors)" can usually be determined from an analysis of the oscillator's frequency variations $\blacktriangle f$ as a linear function of the relevant driving variables z_j:

$$\frac{\blacktriangle f}{f} \approx \sum_j \frac{1}{f} \frac{\delta f}{\delta z_j} \blacktriangle z_j \qquad (1)$$

for a steady-state or very slowly varying system. This assumes a Taylor series expansion about the mean of each variable, at which the derivatives are evaluated, and keeps only the first-order terms, so significant nonlinearity would require higher-order terms.

Some variables act through transducing factors, e.g. a hydrogen maser cavity's frequency variation may be related to the cavity's dimension D and temperature T thusly:

$$\blacktriangle f = \frac{\delta f}{\delta D} \frac{\delta D}{\delta T} \blacktriangle T \qquad (2)$$

If these influences can be measured or predicted and their effects modelled, f can be corrected for them. For example, the frequency of a temperature-compensated crystal oscillator (TCXO) is automatically corrected for thermal effects on the basis of a predetermined temperature dependence [11].

III. Measurement Methods

A published measurement should always be accompanied by its associated uncertainty, and to ensure that both of these are meaningful and reproducible, the physical and statistical methods involved should be clearly documented, including all limitations, correction factors, and error sources.

One cannot measure, and must be careful not to specify, the environmental sensitivity of an oscillator more precisely than is permitted by the fundamental instability (noise) of that oscillator. Since stability generally depends on sampling time, test procedures must allow sufficient measurement time before and after application of environmental stimuli for the oscillator to reach a given stability level (e.g. the "FM flicker floor" [6,7,12]). This limitation can be overcome somewhat by repeated measurements and use of correlation techniques.

Complete, unambiguous tests should

include measurements of time and frequency offset, phase noise, and frequency stability (e.g. Allan variance) before, during, and after environmental changes over a range of frequencies for a length of time adequate to average down the noise and contain the lowest frequencies of interest [13]. The sampling (Nyquist) frequency must be greater than twice the highest frequency present in the data; otherwise, higher frequencies will be misinterpreted as lower frequencies ("aliasing")[12]. The frequency bandwidth (spectral window) should be narrow enough and properly shaped by weighting factors for sufficient resolution and minimal bias and leakage, yet wide enough for adequate smoothing [14].

The frequency stability of a precision oscillator is best measured with a computerized system consisting of a frequency reference; frequency multiplier, divider, or synthesizer; time interval and/or frequency counter or heterodyne arrangement; and spectrum analyzer. Phase noise may be measured with a double-balanced mixer and a phase-locked crystal oscillator [12,15,16]. Control of experimental conditions is critical to obtaining reproducible results. All critical equipment, environmental control, and test units should be powered by an uninterruptible power source [17].

All frequency measurements depend on a reference frequency, which should be as accurate and as fundamental as required in view of the basic performance being measured. A primary standard like most cesium clocks or hydrogen masers would be least compromising in the interpretation of data; such primary standards can be accessed through disseminated time and frequency signals and services. Any trend in the data should be removed first with curve-fitting or other techniques. Every oscillator has a characteristic behavior of Allan variance vs. sampling time according to the types of noise present [5,6,7,12,13]. Over increasingly longer sampling times, the Allan variance eventually increases due to environmental influences and aging. Over long sampling times, one must refer to a "paper time-scale" computed as a filtered average from an ensemble of clocks [e.g. 18,19,20] or utilize a time signal, say via common-view GPS [21], from an official timing center whose master clock might be steered in frequency toward a paper timescale (by no more than a few ns/day). In the latter case, the effects of steering may have to be removed by reference to TAI, though TAI is only available every ten days two months after the fact. Any reference may be expected to contain seasonal variations and other long-term systematic errors due to similarities in clock types and climatic conditions at different laboratories. TAI contains seasonal variations of 4 ns/day peak to peak [22].

Any analysis using time or frequency differences between oscillator pairs yields only the vector sum of their stabilities (neglecting intercorrelations). An "n-cornered-hat analysis" of Allan variances may be used to separate the individual oscillator errors, assuming the measurements are simultaneous and statistically independent and neglecting noise added by the measurement system. For the case of n oscillators, the variance σ_i^2 of an oscillator i may be computed from the following:

$$\sigma_i^2 = \frac{1}{n-2}\left(\sum_{j=1}^{n} \sigma_{ij}^2 - A\right)$$

where σ_{ij}^2 is the variance of the differences between oscillators i and j, $\sigma_{ii}^2 = 0$, and

$$A = \frac{1}{n-1}\sum_{i<j} \sigma_{ij}^2$$

[13,23,24]. The method works best when the oscillators are comparable in stability. Even so, intercorrelations among the oscillators and the errors associated with the variance determinations may result in the computation of negative variances [15,25], which may be avoided by taking explicit account of the intercorrelations [26]. Measurement system noise may be rejected by an equivalent method that utilizes cross-correlations [27].

Interactions between adjacent oscillators (e.g. a test unit and its local reference or a field unit and its backup) may occur as a result of RF coupling, electric field coupling, magnetic field coupling, common ground currents, feedback through buffer amplifiers, feedback through common power supplies, and feedback through control loops where circuit isolation is deficient [28,29]. These interactions are elusive to measurement except by use of three independent references in dual RF bridge circuits capable of picosecond to femtosecond phase measurements using heterodyne techniques for interactions $\leq 10^{-14}$ in frequency. Physical separation of the independent oscillators is

especially effective if the electrical or magnetic shielding of the oscillators is inadequate [29]. The ambient conditions can also cause correlated behavior whose common errors will be impossible to detect without recourse to an external reference. Any correlation between two oscillators reduces the variances obtained, resulting in their underestimation.

When a precision oscillator is under test, its resonance frequency and phase noise at a particular drive level and bandwidth are determined; then, as ambient conditions are varied, measurements are made of the oscillator's characteristics (e.g. its frequency vs. temperature curve, its electrical parameters after compensation for the impedance of the measuring system [30], etc.). Abrupt changes in ambient conditions, such as oven turn-on, have protracted effects until the oscillator stabilizes at a particular frequency.

In passive atomic frequency standards like a cesium clock or a rubidium/crystal oscillator (RbXO), a voltage-controlled quartz crystal oscillator is frequency- or phase-locked to an atomic resonator, which generates a highly stable frequency reference based on an atomic transition. Of the many transitions available, the ones selected are those least sensitive to environmental (in particular, magnetic) disturbances and to which a crystal oscillator may be conveniently locked [31]. The environmental sensitivity of such a frequency standard is theoretically given by the crystal's sensitivity modified by the transfer function (see Appendix) of the frequency- or phase-locked servo-loop or filter. In practice, only total system testing can ascertain system behavior, since the atomic resonator is environmentally sensitive as well. The short-term stability is generally determined by the crystal oscillator and the long-term stability by the atomic resonator, with the crossover behavior and the point from uncontrolled to resonator-controlled conditions being determined by the design and frequency response of the feedback control loop [31,32].

The difficulties involved in clearly relating the amount of frequency change to variations in environmental conditions may require each unit to be tested individually. The individual components of a frequency standard might be tested separately, though there will still be influences from connections and component interactions. Environmental testing of satellite clock subassemblies is often preferred by clock manufacturers over total system testing because of accidental compensations and because design weaknesses can be masked by good performance on the short term.

One difficulty is to measure effectively the property of the element under test, and not the property of the environmental sensor or transducer (e.g. oven). Consequently, the sensor and transducer should be as closely coupled to the oscillator as possible or, given limitations to that, it should be located and/or buffered so as to duplicate as closely as possible the experience of the oscillator as it will be deployed in actual field use.

IV. Interactions Among Environmental Stimuli

If more than one environmental factor proves to be of significance, all other such factors should be held constant while the dependence of each is being determined. Since the relevant factors are not always obvious and since unrecognized influences can obscure the effect of a given parameter, it is best to control as many of the ambient conditions as possible, e.g. by isolating the frequency standard in an environmental chamber with good thermal insulation, humidity control, air circulation, and magnetic shielding. Regarding the latter, the residual magnetic field inside the shields must be evaluated and the frequency possibly reset whenever changes in the external magnetic field occur, e.g. after moving the device. The magnetic shields may also have to be demagnetized.

Even so, coupling between the parameters may necessitate the use of special statistical techniques like multivariate analysis. For example:

(1) Temperature, humidity, and barometric pressure effects are interrelated. Temperature changes can cause significant changes in humidity both inside an environmental chamber and inside a frequency standard, especially if the chamber is not humidity controlled. Temperature and humidity affect the characteristics of electrical components in frequency standards [33,34,35,36]. Condensation and even ice can form inside a unit and permeate materials, which can affect electrical properties and cause shorts and other damage to sensitive electronics and other components. Condensation can be eliminated by testing under a vacuum, but the concomitant removal of convection alters the temperature distribution, and electrical

discharges may occur at very low pressures that will affect performance and may even damage components.

The time constants for humidity changes within the unit are not necessarily equal to the thermal time constants, possibly causing erratic results with regard to both humidity and temperature. The unit may have critical areas that are not well vented and, hence, have local time constants that differ substantially from the global time constants of the unit [2,37]. These problems can be circumvented by the stochastic approach mentioned in Section II, even in the presence of nonlinearities, since very small random disturbances would avoid responses in the nonlinear region.

(2) Oscillators are often hermetically sealed against humidity, e.g. the absorption cell of a rubidium standard, the resonant cavity of a hydrogen maser, or the mounting structure of a crystal oscillator. However, changes in barometric pressure or ambient magnetic field can deform this packaging, altering the heat transfer conditions and affecting the frequency [2,34,35,36].

(3) Acceleration effects are usually measured by changing the orientation of a frequency standard relative to gravity. However, since a significant portion of the heat transfer inside a unit is by air convection, such a reorientation ("2-g tipover") will alter the internal heat distribution, resulting in thermal effects. If the data desired are really for tipover, the test is still valid, even though it is not exclusively a measurement of acceleration sensitivity [1,37].

(4) Movement, acceleration, or vibration can lead to magnetic or electric effects, if such fields are present, by changing the orientation of the resonator therein.

(5) In passive atomic standards, vibrations at the servo-loop modulation frequency and its subharmonics can affect the power supply, harmonic generator, microwave power, and other electronics.

(6) The temperature of a crystal oscillator affects its sensitivity to vibration [1,38] and electric fields [39].

(7) If an oscillator receives a dose of radiation in a period of time short in comparison to its thermal time constant, the internal heating caused by the radiation will have effects that are particularly noticeable when measurements are made within a period of time less than the thermal time constant of the unit [37].

(8) Power-supply fluctuations, filtered by voltage regulators, can lead to lagged secondary effects such as changes in the internal power dissipation and, hence, heat distribution, producing thermal effects.

(9) In a rubidium standard, the microwave excitation field, RF power, temperature distribution, and pumping light intensity are interrelated, so that inhomogeneities in any of these shift the region of optimal signal, resulting in a frequency change [34].

(10) AC magnetic fields can induce voltages that interfere with vibration-sensitivity tests and can also cause thermal effects, since any transient magnetic field dissipates energy proportional to the area of the hysteresis cycle in the material. Eddy current losses also add heat. Air cooling during centrifuge tests cause thermal effects as well.

(11) The permeability of typical magnetic materials and, hence, the magnetic field sensitivities of oscillators may in general depend on temperature and the oscillator's history.

(12) Residual magnetic fields may arise from oven heaters. Magnetic effects may be induced if the oscillator is not sufficiently protected from strong AC fields associated with vibration test equipment and centrifuges.

Other factors that can complicate evaluation of a given type of sensitivity are changes due to hysteresis and, on the long term, aging and maintenance adjustments.

Interactions among influences that affect crystal oscillators, for example, are listed in Table I.

The testing of mathematical models for the environmental dependences should start with the simplest form (in number and order of terms) and proceed by the addition of single terms (in decreasing order of importance). The model chosen should be that which yields the best chi-square fit to the observed frequency behavior, provided that the reduction in the variance of the fit accompanying the incorporation of each new term exceeds the increase in the Allan variance [given by Eq. (3) below] caused by having to solve for the additional term [40]. This often, but not always, yields the same result as retaining those parameters whose coef-

TABLE I. Interactions Among Influences Affecting Crystal Oscillators

In attempting to measure the effect of a single influence, one often encounters interfering influences, the presence of which may or may not be obvious.

Measurement	Interfering Influence
Resonator aging	ΔT due to oven T (i.e., thermistor) aging Δ drive level due to osc. circuit aging
Short term stability	Vibration
Vibration sensitivity	Induced voltages due to magnetic fields
2-g tipover sensitivity	ΔT due to convection inside oven
Resonator f vs. T (static)	Thermal transient effect, humidity T-coefficient of load reactances
Radiation sensitivity	ΔT, thermal transient effect, aging

ficients exceed their associated standard errors. The most consistently successful models reflect parsimony of parameters (see Appendix).

If error residuals are normally distributed, the values of the sensitivity coefficients and their standard errors may be determined by least squares. In the case of errors that have non-Gaussian but well-determined distributions, the coefficients may be estimated by the maximum likelihood method or the method of moments.

If individual measurements differ significantly in precision, each should be weighted by the inverse of its estimated variance. If different types of measurements are combined in a solution, each type should be weighted by its respective inverse variance. In order to have least-squares normal equations that are well-conditioned against the accumulation of round-off errors, the variable involved must be linearly transformed such that the coordinate origins are in the data range. The most stable solutions are provided by the method of singular value decomposition [41]. However, the older method of orthogonal polynomials does have the advantage that the residuals from one solution may be analyzed for additional roots without affecting the values of roots already obtained.

If any of the variables are significantly correlated, the covariance matrix must be included in the solution. While multivariate analytical methods exist for the circumvention of covariance determination, the correlations are too interesting in themselves to recommend these. If a correlation between two quantities arises because they are functions of some other quantity, it might be preferable (depending on the acceptable number of parameters) to incorporate the latter quantity in the model, thereby eliminating the correlation.

In the case of nonlinearities, hysteresis, and dynamic effects, plots may be more informative than models. The extra cost of performing complicated tests, rather than simple ones, or of repeated tests to average down errors, might exceed the value of the results to the user.

V. ERROR BUDGETS

In principle, variation of all the parameters on which the results of a measurement depend should yield the total variation in the measured value. The total variance would be just the vector sum of the variances and covariances of the parameters and of any other sources of error. However, time and resource

constraints usually require that the measurement uncertainty be evaluated using a mathematical model and the law of propagation of errors. Assuming the errors are random, if a variable f is a function of other variables $z_1, z_2, \ldots, z_j, \ldots, z_p$, then the law of propagation of errors states:

$$\sigma^2(f) \approx \sum_{j=1}^{p} \left[\frac{\delta f}{\delta z_j}\right]^2 \sigma^2(z_j) + 2 \sum_{j=1}^{p} \sum_{k<j} \frac{\delta f}{\delta z_j} \frac{\delta f}{\delta z_k} \phi(z_j, z_k) \qquad (3)$$

$$\approx \sum_{j=1}^{p} \left[\frac{\delta f}{\delta z_j}\right]^2 \sigma^2(z_j) + 2 \sum_{j=1}^{p} \sum_{k<j} \frac{\delta f}{\delta z_j} \frac{\delta f}{\delta z_k} r(z_j, z_k)\, \sigma(z_j)\, \sigma(z_k)$$

where $\sigma^2(z_j)$ is the variance of variable z_j, $\phi(z_j, z_k)$ is the covariance between variables z_j and z_k, and $r(z_j, z_k)$ is the linear correlation coeffcient between z_j and z_k. This assumes a steady-state or very slowly varying system. Also, as in the case of Eq. (1), it involves a Taylor series expansion about the mean of each variable, at which the partial derivatives are evaluated, and keeps only the first-order terms, so again significant non-linearity (or nonlinear intercorrelations) would require higher-order terms.

In the special case where the derivatives are unity and the inter-correlations are zero, Eq. (3) becomes:

$$\sigma^2(f) = \frac{1}{p} \sum_{j=1}^{p} \sigma^2(z_j)$$

Further, if f is strictly a product and quotient of the input quantities, then all the variances may be expressed as relative variances:

$$\frac{\sigma^2(f)}{f} = \frac{1}{p} \sum_{j=1}^{p} \left[\frac{\sigma^2(z_j)}{z_j}\right]^2$$

In Eq. (3), let f be the frequency and $z_j, z_k, \ldots$ the environmental parameters at time t, $\sigma^2(z_j)$ the two-sample variance:

$$\sigma^2(z_j) = \frac{1}{2(n-1)} \sum_{t=1}^{n-1} [z_j(t+1) - z_j(t)]^2$$

for n observations, and $\phi(z_j, z_k)$ the two-sample covariance:

$$\phi(z_j, z_k) = \frac{1}{2(n-1)} \sum_{t=1}^{n-1} [z_j(t+1) - z_j(t)]\,[z_k(t+1) - z_k(t)]$$

One could then use Eq. (3) to predict how variations in the ambient conditions transduce into variations in the frequency, as well as how intercorrelations between the parameters affect the frequency. The two-sample $\sigma^2(f)$ is just the Allan variance. More often, the Allan variance is expressed as a relative uncertainty, *i.e.* $\sigma^2(f)/f$, usually logarithmically.

In order to properly evaluate the overall instability of a system in the presence of all sources of error that might reasonably be expected to be present, one should perform an "error (or uncertainty) budget" analysis. After identifying all such significant sources, say by correlation studies, the individual contributions to the combined uncertainty $\sigma^2(f)$ by each of the parameter variances $\sigma^2(z_j)$, $\sigma^2(z_k)$, ... and their covariances $\phi(z_j, z_k)$, ... may be evaluated by actual measurement of the variances and covariances. Sometimes the functional relationships are clear theoretically (e.g. Eq. 2) and sometimes they must be modelled empirically (e.g. Eq. 1) [42].

Some error sources are not available or amenable to repeated measurements and

statistical analysis. Still, variances for them should be estimated as well as possible on the basis of published information or experience (as should be stated) [42].

Substituting one-time parameter variations Δz_j, Δz_k, ... in Eq. (1) permits one to predict the effects of changes in the individual parameters, but is a further approximation; in such cases, the derivatives must be evaluated at the parametric values. Ideally, the effects should be modelled by varying the parameters randomly over their operational ranges, as in the stochastic method. Use of Eq. (3) would then allow the apportioning of an error budget among the relevant contributing sources. Any detailed report about uncertainties should consist of a complete list of the error components, the numerical value of each standard error, the method used to determine it, and the number of degrees of freedom. If the measured quantities are significantly correlated, the correlation coefficient matrix, the covariance matrix, or preferably both should be given [42].

While it is generally necessary to vary each parameter throughout its operational range to determine properly the coefficient of dependence, as a rule it is preferable to keep each parameter as constant as possible during routine operation rather than to attempt to correct for the effect of the variation during postprocessing, because of the uncertainties associated with the coefficients, the intercorrelations among the parameters, and dynamic effects.

The total uncertainty may be expressed either as a tolerance (a specified number of standard errors) or as a confidence limit (a specified probability percentage). For the former to be meaningful and for the latter to be determinate at all, the probability distribution of the total error residuals must be known, at least approximately. Hence, it is preferable to correct the data first for systematic (nonstochastic) effects so the residuals will be normally (Gaussian) distributed, assuming that the number of degrees of freedom is greater than about 30; for cases fewer than that, random residuals should follow Student's t distribution. In either case, the hypothesis that the residuals are randomly distributed may be tested, as with a chi-square test or probit plot. The presence of uncorrected systematic effects may be recognized in residuals that depart significantly from a normal distribution. The corrections for systematic effects have their own uncertainties that should be included in any error budget [42].

Reliance on internal errors, i.e. failure or inability to use an external reference (or at least independent determinations), risks miscalibrations and oversights of error sources that generally cause one to underestimate one's errors. If, on the other hand, such systematic effects (uncorrected departures from random Gaussian errors) are present in the measurements, the errors will generally be overestimated, especially if the errors are computed relative to the mean rather than from consecutive residuals (two-sample variances and covariances). Increasing the amount of data under a variety of conditions reduces the random errors (true white noise having a stationary mean), but risks systematic errors if parameter dependences are not properly modelled (e.g. through incorrect linear or higher-order differentials or neglect of significant parameter intercorrelations), usually resulting in underestimation of predicted frequency variations.

VI. Transient Effects and Aging

Specifications of environmental sensitivities generally assume quasi-static conditions, i.e.: (1) the frequency source has stabilized and is in equilibrium with its surroundings before the ambient change occurs; and (2) after application of the stimulus, the frequency standard has again reached a new equilibrium with its environment. Since the idealized conditions would require both infinite time periods before and after application of the stimulus, as well as infinitely slow application of the stimulus itself, acceptable procedures must be found to approximate the quasi-static condition. Furthermore, measures of frequency stability are theoretically based on infinite time averages, so enough data for a statistically meaningful result must be acquired over a period of time during which sufficently quasi-static conditions can be said to apply [37].

Measurement under nonequilibrium conditions generally occurs when the time constants involved in the transfer of the environmental stimulus into the sensitive components of the frequency source are longer than the period of time between the application of the stimulus and the measurement. In such a case, the result is highly dependent on the exact temporal and spatial profile of the stimulus (e.g. whether the profile is abrupt or ramped),

the degree of coupling between the oscillator and its environment (e.g. whether the unit is in direct contact with a heating element or separated from it by circulating air), and the transfer function of the oscillator (e.g. dependence on the internal temperature gradient). Also, if fast ramping and close coupling are involved, the final results will depend on the initial condition of the oscillator [37]. The time required by the correction cycle of a microprocessor-compensated crystal oscillator (MCXO) limits the system's ability to follow rapid temperature variations [43].

If the user is specifically interested in the effect of a dynamic environment, test methods would probably have to conform to the exact profile of the stimulus in question. In such cases, experience has shown that one cannot extrapolate reliably the effects from one profile to another. In general, use of statistical measures in nonequilibrium situations should be avoided [37]. Nonetheless, dynamic effects can be at least approximately quantified by differentiating Eq. (1) with respect to time and determining the resulting coefficients. When more than one time constant is involved, the effect on the frequency can be very complex. A slow, complete sweep and retrace through the range of an environmental parameter is necessary to characterize dynamic response and to check for local nonlinearities like crystal "activity dips".

Inability to maintain quasi-static conditions can frequently be traced to external changes like power-supply fluctuations. When filtered by voltage regulators, they lead to lagged secondary effects such as changes in the internal power dissipation and, hence, temperature distribution, setting up gradients that cause frequency variations as well as further changes in the power consumption of the major modules. The latter require further action by the voltage regulators, etc., so the frequency standard never achieves equilibrium, but always has a floor of flicker FM noise, if not a noise type of even higher variance. Flicker also results in part from environmental shocks, though the latter mainly produce random-walk FM noise [6,7,12].

In passive atomic oscillators, susceptibility to power-supply ripples is generally worst at the servo-loop modulation frequency; large frequency shifts are possible due to interference with the servo. System turn-on and turn-off transients may also cause problems, often exhibiting exponential behavior and requiring modelling with time constants [44].

Frequency measurements for environmental tests should be made when the oscillator is beyond any significant initial relaxation, but has not yet shown any significant aging effects or spontaneous changes (e.g. FM random walk) unless these are unavoidable or can be adequately corrected for. Also, unless such corrections are well known, the environmental variations should be restricted to intervals small in comparison to that over which significant aging occurs or over which a spontaneous change is likely to occur. Well-behaved frequency drift, such as that due to aging, need not limit system usage if it is measured and corrected for. Frequency drift is usually most accurately determined from the mean of nonoverlapping phase second differences rather than from a linear fit to frequency first differences or a quadratic fit to phase first differences [45]. All oscillators except crystals have a useful lifetime (varying significantly from unit to unit), at the end of which aging effects are extreme and unpredictable [46].

In addition, each type of oscillator is subject to frequency variations as the result of changes and distortions in the electronics, such as those in voltage references, power supplies and amplifiers, resistors, diodes, capacitors, etc., which in turn are sensitive to variations in ambient conditions [33,34,35,36]. Power-supply and load impedance changes affect crystal oscillator circuitry and, indirectly, the crystal's drive level and load reactance. A change in load impedance changes the amplitude or phase of the signal reflected into the oscillator loop, which changes the phase and frequency of the oscillation. The effects can be minimized through voltage regulation and the use of buffer amplifiers [11]. In cesium clocks, changes and distortions in the electronics and power supply propagate into the microwave field, the frequency servo-loop, and the slaved crystal oscillator [32,33].

VII. Summary

Environmental effects on precision oscillators may be evaluated by: (1) identification of relevant parameters and transducing factors through correlation and spectral analyses; (2) control or removal of systematic effects (through curve-fitting, differentiation, etc.); and (3) evaluation of residual random errors by means of two-sample variances and

covariances and an error budget analysis. Given an adequate measurement system, frequency reference, and control over experimental conditions, optimal data reduction involves choices as to parameter range, sampling time, averaging process, and mathematical model. Matters may be complicated by nonlinear responses, intercorrelations, different time constants, transient effects, and aging. If quasi-static conditions are not applicable, explicit account must be taken of the temporal and spatial profile of the stimulus.

References

[1] 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", 1991, Research and Development Technical Report SLCET-TR-91-3, U.S. Army Laboratory Command, Fort Monmouth, NJ; DTIC #AD-A235470 (also see elsewhere in these Proceedings).

[2] D. W. Allan, J. Barnes, F. Cordara, M. Garvey, W. Hanson, R. Kinsman, J. Kusters, R. Smythe, and F. L. Walls, "Dependence of frequency on temperature, humidity, and pressure in precision oscillators", IEEE P1193 Working Group 3 Report (in draft; also see elsewhere in these Proceedings).

[3] R. L. Sydnor, R. Brendel, M. Garvey, and W. Wiedemann, "Electromagnetic effects on precision frequency sources", IEEE P1193 Working Group 1 Report (in preparation; also see elsewhere in these Proceedings).

[4] J. J. Suter, R. H. Maurer, J. D. Kinnison, R. Besson, J. R. Vig, and A. Koehler, "The effects of ionizing and particle radiation on precision frequency sources", 1992, Proposal for IEEE Standards Project P1193, Applied Physics Laboratory, Johns Hopkins University, Baltimore, MD (also see elsewhere in these Proceedings).

[5] D. W. Allan, "Should the classical variance be used as a basic measure in standards metrology?", 1987, IEEE Transactions on Instrumentation and Measurement, vol. IM-36, pp. 646-654.

[6] D. Allan, P. Kartaschoff, J. Vanier, J. Vig, G. M. R. Winkler, and N. F. Yannoni, "Standard terminology for fundamental frequency and time metrology", Proceedings of the 42nd Annual Symposium on Frequency Control, 1-3 June, 1988, Baltimore, MD, pp. 419-425.

[7] "IEEE standard definitions of physical quantities for fundamental frequency and time metrology", 1988, Institute of Electrical and Electronics Engineers, New York, NY, IEEE Std. 1139-1988.

[8] 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, 29-31, May, 1991, Los Angeles, CA, pp. 667-678.

[9] D. B. Percival, "Characterization of frequency stability: frequency domain estimation of stability measures", 1991, Proceedings of the IEEE on Time and Frequency, vol. 79, pp. 961-972.

[10] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control, 1970, San Francisco: Holden-Day.

[11] S. R. Stein and J. R. Vig, "Frequency standards for communications", 1991, Research and Technical Report SLCET-TR-91-2 (Rev. 1), U.S. Army Laboratory Command, Fort Monmouth, NJ; DTIC #AD-A243211.

[12] D. A. Howe, D. W. Allan, and J. A. Barnes, "Properties of signal sources and measurement methods", Proceedings of the 35th Annual Symposium on Frequency Control, 27-29 May, 1981, Philadelphia, PA, pp. 1-47 = NIST Technical Note 1337 (NTIS #PB 83-103705), pp. TN14-TN60.

[13] D. W. Allan, "Time and frequency (time-domain) characterization, estimation, and prediction of precision clocks and oscillators", 1987, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. UFFC-34, pp. 647-654 = NIST Technical Note 1337 (NTIS #PB 83-103705), pp. TN121-TN128.

[14] A. V. Oppenheim and A. S. Willsky, Signals and Systems, 1983, Englewood Cliffs: Prentice-Hall, Inc.

[15] S. R. Stein, "Frequency and time - their measurement and characterization", Precision Frequency Control, 1985 (edited by E. A. Gerber and A. Ballato, New York: Academic Press), vol. 2, pp. 191-232 and 399-416 = NIST Technical Note 1337 (NTIS #PB 83-103705), pp. TN61-TN120.

[16] 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", Proceedings of the 42nd Annual Symposium on Frequency Control, 1-3 June, 1988, Baltimore, MD, pp. 432-441 = NIST Technical Note 1337 (NTIS #PB 83-103705), pp. TN129-TN138.

[17] R. L. Sydnor, T. K. Tucker, C. A. Greenhall, W. A. Diener, and L. Maleki, "Environmental tests of cesium frequency standards at the Frequency Control Laboratory of the Jet Propulsion Laboratory", Proceedings of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 28-30 November, 1989, Redondo Beach, CA, pp. 409-420.

[18] D. W. Allan and M. A. Weiss, "The NBS atomic time scale algorithm AT1", NBS Technical Note 1316 (in preparation).

[19] D. B. Percival, "The U.S. Naval Observatory clock time scales", 1978, IEEE Transactions on Instrumentation and Measurement, vol. IM-27, pp. 376-385.

[20] S. R. Stein and J. Evans, "The application of Kalman filters and ARIMA models to the study of time prediction errors of clocks for use in the Defense Communication System (DCS)", Proceedings of the 44th Annual Symposium on Frequency Control, 23-25 May 1990, Baltimore, MD, pp. 630-635.

[21] W. Lewandowski and C. Thomas, "GPS time transfer", 1991, Proceedings of the IEEE on Time and Frequency, vol. 79, pp. 991-1000.

[22] P. Tavella and C. Thomas, "Report on correlations in frequency changes among the clocks contributing to TAI", 1991, Bureau International des Poids et Mesures, Report BIPM-91/4.

[23] J. A. Barnes, "Atomic timekeeping and the statistics of precision signal generators", 1966, Proceedings of the IEEE on Time and Frequency, vol. 54, pp. 207-219.

[24] J. E. Gray and D. W. Allan, "A method for estimating the frequency stability of an individual oscillator", Proceedings of the 28th Annual Symposium on Frequency Control, 29-31 May, 1974, Atlantic City, NJ, pp. 243-246.

[25] K. Yoshimura, "Degrees of freedom of the estimate of the two-sample variance in the continuous sampling method", 1989, IEEE Transactions on Instrumentation and Measurement, vol. IM-38, pp. 1044-1049.

[26] P. Tavella and A. Premoli, "Characterization of frequency standard instability by estimating their covariance matrix", Proceedings of the 23rd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Pasadena, CA, 3-5 December, 1991 (in press).

[27] J. Groslambert, D. Fest, M. Oliver, and J. J. Gagnepain, "Characterization of frequency fluctuations by crosscorrelations and by using three or more oscillators", Proceedings of the 35th Annual Symposium on Frequency Control, Philadelphia, PA, 27-29 May, 1981, pp. 458-463.

[28] B. Jaduszliwer, R. A. Cook, and R. P. Frueholz, "Frequency shifts in a rubidium frequency standard due to coupling to another standard", Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 4-6 December, 1990, Vienna, VA, pp. 293-300.

[29] L. J. Rueger, 1992, private communication.

[30] R. Kinsman and D. Rydback, "Precision temperature test station for quartz crystals", Proceedings of the 43rd Annual Symposium on Frequency Control, 31 May-2 June, 1989, Denver, CO, pp. 300-308.

[31] C. Audoin and J. Vanier, The Quantum Physics of Atomic Frequency Standards, 1989, Bristol: Adam Hilger, vol. 2.

[32] A. De Marchi, G. D. Rovera, and A.

Premoli, "Effects of servo loop modulation in atomic beam standards employing a Ramsey cavity", 1987, IEEE Transactions on Instrumentation and Measurement, vol. IM-34, pp. 582-591.

[33] C. Audoin, N. Dimarcq, V. Giordano, and J. Viennet, "Physical origin of the frequency shifts in cesium beam frequency standards: related environmental sensitivity", Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 4-6 December, 1990, Vienna, VA, pp. 419-440.

[34] W. J. Riley, "The physics of the environmental sensitivity of rubidium gas cell atomic frequency standards", Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 4-6 December, 1990, Vienna, VA, pp. 441-452.

[35] E. M. Mattison, "Physics of systematic frequency variations in hydrogen masers", Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 4-6 December, 1990, Vienna, VA, pp. 453-464.

[36] F. L. Walls, "Environmental sensitivities of quartz crystal oscillators", Proceedings of the 22nd Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 4-6 December, 1990, Vienna, VA, pp. 465-486.

[37] H. Hellwig, "Environmental sensitivities of precision frequency sources", 1990, IEEE Transactions on Instrumentation and Measurement, vol. 39, pp. 301-306.

[38] H. P. Hanson and T. E. Wickard, "Acceleration sensitivity as a function of temperature", Proceedings of the 43rd Annual Symposium on Frequency Control, 31 May-2 June, 1989, Denver, CO, pp. 427-432.

[39] R. Brendel, "Etude des defauts ponctuels dans les resonateurs a quartz par effect electroelastique", Proceedings of the 1st European Forum on Time and Frequency, 18-20 March, 1987, Besancon, France.

[40] H. Eichhorn and C. A. Williams, "On the systematic accuracy of photographic astrometric data", The Astronomical Journal, 1963, vol. 68, pp. 221-231.

[41] G. E. Forsythe, M. M. Malcolm, and C. B. Moler, Computer Methods for Mathematical Computations, 1977, Englewood Cliffs: Prentice-Hall.

[42] "Expression of uncertainty in measurement", 1991, guide prepared by the International Organization for Standardization (ISO/TAG4/WG3), Geneva, Switzerland, for the International Committee for Weights and Measures (CIPM) (in draft).

[43] A. Benjaminson, "Factors influencing stability in the microprocessor-compensated crystal oscillator", Proceedings of the 44th Annual Symposium on Frequency Control, 23-25 May, 1990, Baltimore, MD, pp. 597-614.

[44] J.-Q. Lu and Y. Tsuzuki, "Analysis of start-up characteristics of crystal oscillators", Proceedings of the 45th Annual Symposium on Frequency Control, 29-31 May, 1991, Los Angeles, CA, pp. 360-363.

[45] J. A. Barnes, "The measurement of linear frequency drift in oscillators", Proceedings of the 15th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 6-8 December, 1983, Washington, DC, pp. 551-582 = NIST Technical Note 1337 (NTIS #PB 83-103705), pp. TN264-TN295.

[46] J. Vanier, W. J. Riley, J. J. Gagnepain, F. L. Walls, and M. Granveaud, "Time evolution effects in precision frequency sources", IEEE P1193 Working Group 2 Report (in preparation; also see elsewhere in these Proceedings).

Appendix

Cross-Correlation Techniques

The "cross-correlation (function)" of a signal x (t) [say, that of an oscillator] with another signal z (t) [say, that of an environmental sensor] is:

$$\Gamma_{xz}(k) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{+T/2} x(t)\, z(t+k)\, dt$$

For discrete data, the kth "cross-correlation coefficient" is:

$$c_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(z_{t+k} - \bar{z})}{\left[\sum_{t=1}^{n} (x_t - \bar{x})^2\right]^{1/2} \left[\sum_{t=1}^{n} (z_t - \bar{z})^2\right]^{1/2}}$$

A plot of c_k is called a "correlogram."

Significant peaks in the correlogram indicate a relationship between x and z that is probably causal if all lag times k with such peaks are nonnegative. Assuming these lags are limited in range and there are no other complicating parameters of significance, a plot of the data vs. the parameter in question may be sufficient to reveal the dependence.

If the effects appear to be caused over a range of lag times, the system is said to be "dynamic" and one must characterize its "transfer function". Frequently the inertia of the system can be adequately approximated by a linear filter:

$$\dot{y} = w_0 \dot{z}_k + w_1 \dot{z}_{k-1} + w_2 \dot{z}_{k-2} + \ldots$$

$$= (w_0 + w_1 B + w_2 B^2 + \ldots) \dot{z}_k$$

$$= w(B)\, \dot{z}$$

where $\dot{z}$ is the input (or "forcing function") and $\dot{y}$ is the output (e.g. changes in an environmental parameter and in a clock's frequency respectively), B is the "backshift operator" defined by $B \dot{z}_k = \dot{z}_{k-1}$, w (B) is the transfer function, and $w_0, w_1, w_2, \ldots$ comprise the impulse response (or "weighting") function of the system [10].

If such a system is afflicted with white noise, the transfer function can be modelled by:

$$(1 - u_1 B - u_2 B^2 - \ldots)\, \dot{y} = (v_0 - v_1 B - v_2 B^2 - \ldots)\, \dot{z}_{k-d} + a_k$$

where $u_1, u_2, \ldots$ and $v_0, v_1, v_2, \ldots$ are cross-correlation parameters, d is a delay ("dead time") parameter, and a_k is white noise. Usually only a few parameters need be solved for, and indeed the fewest necessary should be solved for if one is to maximize the accuracy of one's prediction. Such a "parsimonious" model most appropriate for one's data can be chosen using recursive estimation of the input's autocorrelation and the cross-correlation between the input and the output. If the input is a unit step function ($z(t) = 0$ for $t < 0$ and $z(t) = 1$ for $t \geq 0$), the resulting "step response" is:

$$y(t) = \int_0^t w(k)\, dk$$

If the environmental input is under one's control, it can be preferable to prewhiten, rather than postwhiten, the noise, i.e. calibrate it stochastically rather than deterministically. Considerable simplification of the model-identifying process can be achieved if the input to the system is white noise, for in this case the impulse response function is given by an orthogonal set of equations:

$$w_k = \frac{c_k(\dot{z},\dot{y})\, \sigma(\dot{y})}{\sigma(\dot{z})}$$

for lags k= 0, 1, 2, ...; $c_k(\dot{z},\dot{y})$ is the kth cross-correlation coefficient between $\dot{z}$ and $\dot{y}$; and $\sigma(\dot{z})$ and $\sigma(\dot{y})$ are the standard errors of $\dot{z}$ and $\dot{y}$ respectively, i.e. the cross-correlation between the input and output of a linear, time-invariant system is proportional to the impulse response of the system. This expression for w_k can be used to make initial estimates of the cross-correlation and delay parameters, which can then be improved recursively by nonlinear least squares [10].

In a "first-order dynamic system," the input and output are related by a differential equation of the form:

$$\frac{dy}{dt} = \frac{1}{T} [g\, z(t) - y(t)]$$

where the constants T and g are the "time constant" and "gain" of the system

respectively. The solution is:

$$y(t) = \int_0^{\infty} w(k)\, z(t - k)\, dk$$

which, assuming the usual Laplacian behavior ($z(t) = 0$ for $t < 0$), becomes:

$$y(t) = \int_0^{t} w(k)\, z(t - k)\, dk$$

where:

$$w(k) = g\, T^{-1}\, e^{-k/T}$$

Thus, the impulse response in this case undergoes simple exponential decay. If the input is a unit step function, the resulting step response is:

$$y(t) = g\,(1 - e^{-t/T})$$

Similar examples can be given for more complex inputs and for nonlinear response functions.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ENVIRONMENTAL EFFECTS IN MIXERS AND FREQUENCY DISTRIBUTION SYSTEMS

L. M. Nelson and F. L. Walls
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

This paper examines the environmental sensitivity of key elements of measurement systems, frequency synthesizers, and frequency distribution systems designed to achieve frequency stabilities of order $\sigma_y(\tau) = 5 \times 10^{-15}\ \tau^{-1/2}$ in the short-term and 3×10^{-18} in the long-term. Specifically, we investigate the timing errors in signal transmission for distances up to 100 m and the effect of changes in temperature and rf amplitude on timing (or frequency) errors in several available phase detectors operated in different configurations.

Introduction

Several new types of frequency standards have been proposed that show a potential for a short-term frequency stability of approximately $2 \times 10^{-15}\tau^{-1/2}$ and a long-term frequency stability less than 1×10^{-17} [1-4]. To measure frequency stability on this level and to construct the electronics of the local oscillator for interrogating the atomic system will require substantial improvements in key building blocks such as phase detectors within the measurement system, phase detectors used in frequency manipulation, and frequency distribution systems. To measure these new clocks, it is desirable to have a measurement system that reaches the clock performance at a measurement time no longer than 10^4 s. This requires a resolution of approximately $\sigma_y(\tau) = 2 \times 10^{-17}$ at a measurement time of 10^4 s, or a timing error of 0.2 ps. Similar timing requirements apply to the frequency synthesis within the reference source and in the frequency distribution systems. Figure 1 shows the common elements of these systems. In this paper we investigate the timing errors in signal transmission over distances up to 100 m and the effect of changes in temperature and rf amplitude on timing (or frequency) errors in several types of phase detectors operated in different configurations.

Contribution of the U. S. Government, not subject to copyright.

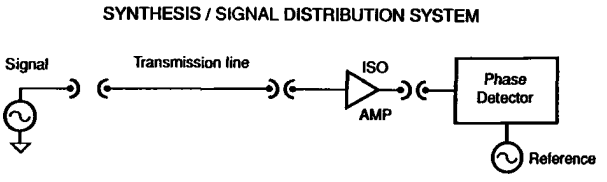


Fig. 1. Overall block diagram showing the key elements for measuring the frequency or time of a remote clock.

Timing Errors In Signal Transmission

Potentially serious timing errors can occur due to the variation of transit time delay and the phase of the standing waves on the transmission line connecting the signal (clock) to the measurement system. The transit delay of a signal is given by the length divided by the velocity of propagation. Signals transmitted to the measurement system and reflected from the load and then from the signal source lead to phase change of the signal at the load. Changes in the voltage standing wave ratio (VSWR) therefore lead to changes in timing errors. The complete timing error from transmission is

$$\delta t = L/(\beta c) + (1/(4\nu_o))\ ((\rho_l\rho_s)/\eta) \sin\phi, \quad (1)$$

where L is the equivalent free space length of the transmission line, c is the speed of light, ß is the propagation factor, ν_o is the carrier frequency, η is the round trip attenuation, ρ_l is the voltage reflection ratio from the load, ρ_s is the voltage reflection ratio from the source (ρ = (VSWR - 1)/(VSWR + 1)), and ϕ is the angle of the twice reflected wave at the

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

load relative to the primary signal. This assumes that the next order of reflected waves can be ignored ($(\rho_s\rho_l)/\eta \ll 1$). The first term in (1) is independent of frequency while the second term scales as $1/\nu_o$.

The variation in the timing error due to changes in L/(ßc) is potentially quite large for long transmission cables. Typical variations for several types of semi-rigid coaxial cables versus temperature are shown in Fig. 2 [5]. Typical insertion losses are of order 1 dB/30 m at 5 MHz and 3 dB/30 m at 100 MHz for 0.358 cm diameter cable. Cables with solid teflon dielectric exhibit large hysteresis effects (due to the mismatch between the thermal expansion coefficient between the dielectric and the conductors) for temperature variations from 10 to 20 °C. Various attempts have been made to produce cable with a smaller temperature sensitivity by reducing the amount of teflon dielectric or incorporating compensating materials (curves B-D of Fig. 2.). Curve E of Fig. 2 shows the temperature sensitivity of a cable that uses SiO_2 powder as the dielectric. The phase variation from 18.3 to 23.8 °C is less than ±0.5 ps per 30 m, while the insertion loss is 3.2 dB/30 m at 100 MHz. This generally satisfies the timing requirements for distances up to approximately 100 m. The primary limitation with this cable is the cost. Except in cases where there are large temperature variations, one of the other cables from Fig. 2 usually performs acceptably at less cost for distances up to approximately 1 m. This data is summarized in Table 1.

At 5 MHz, $1/(4\ \nu_o)$ is 50 ns and typical numbers for ρ_l and ρ_s are 1/3 (VSWR = 2). This can lead to large timing error variations since VSWR is often dependent on temperature, rf drive level, and even supply voltage. Significant improvements in the timing errors due to VSWR can be made by making ρ_l, ρ_s, and sin ϕ small. At 5 MHz, for example, the timing error due to VSWR can be reduced below 11 ps by reducing VSWR below 1.1 at the source and load, and reducing sin ϕ below 0.1. At 100 MHz this term can be made less than 1 ps by making VSWR < 1.1 and sin ϕ < 0.1. The characteristic impedance of coaxial cables changes slightly from one batch to another which might require a small adjustment in source and load impedances for the most critical applications.

An alternative approach to the problem is to use two-way transmission measurements to correct for variations in the transmission medium. This technique has been pursued at the Jet Propulsion Laboratory (JPL) where they have concentrated on AM modulation of a laser transmitted over fiber optic cables to solve both the attenuation of the signal and timing errors for cable lengths up to 25 km. The JPL results [6], shown in Fig. 3, are very impressive. Although this system is relatively costly, it is currently the only practical approach to long transmission distances due to the problem of attenuation.

Table 1. Timing/frequency errors due to variations in cable length with temperature.

BASIC TYPES OF CABLES	Loss/30 m @ 100 MHz DIAMETER	δt/K 23-30°C per 30 m	APPROXIMATE COST/30 m
SOLID DIELECTRIC PTFE common semi-rigid (A)	3.4 dB 0.358 cm	1 ps	$ 210
MICROPOROUS (B)	2.8 dB 0.483 cm	1 ps	$4000
LOW DENSITY PTFE (C)	2.4 dB 0.358 cm	0.9 ps	$ 550
AIR ARTICULATED PTFE (D)	2.7 dB 0.358 cm	0.4 ps	$ 950
COMPENSATED DIELECTRIC ISOCORE (E)	3.5 dB 0.358 cm	0.18 ps	$ 657
SiO_2 DIELECTRIC (F)	3.2 dB 0.358 cm	0.06 ps	$2200

Environmental Sensitivity of Phase Detectors

One measure of phase detector performance is the phase error $\delta\phi$ given by

$$\delta\phi = \delta V/k_d, \qquad (2)$$

where δV is the dc output voltage error and k_d is the phase detection slope in volts/rad [5]. For comparing

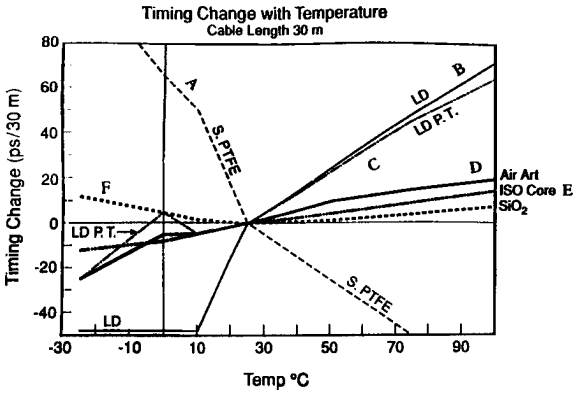


Fig. 2. Fractional phase shift versus temperature for several different semi-rigid coaxial cables. Curve A, typical .358 cm (0.141") solid teflon dielectric cable. Curve B, data for .483 cm cable with lowered teflon content. Curve C, data for .358 cm cable with lowered teflon content. Curve D, data for .358 cm air-articulated cable. Curve E, phase shift data for .358 cm cable with the temperature coefficient of the dielectric matched to the conductors. Curve F data for .358 cm cable with SiO_2 dielectric.

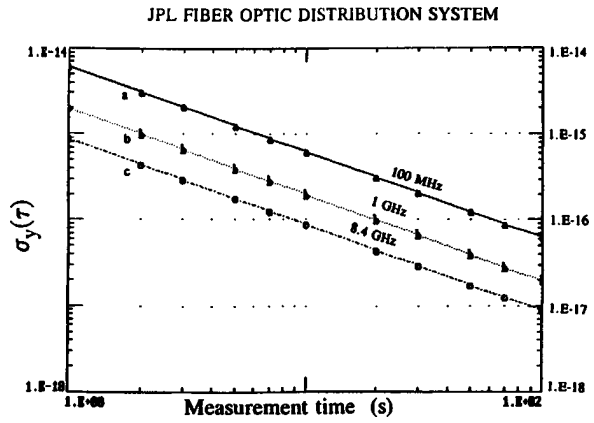


Fig. 3. Fractional timing error versus measurement time of the JPL fiber optic system for AM frequencies of 100 MHz, 1 GHz, and 8.4 GHz. Timing error is $\tau\sigma_y(\tau)$.

different phase detector circuits as a function of carrier frequency, ν_o, it is more useful to use the timing error given in seconds by

$$\delta t = (\delta V/k_d)\,(1/(2\pi\nu_o)). \qquad (3)$$

The dc voltage error originates from imbalance between the I-V curves of the diodes, imbalance in the various transformers, and asymmetric stray impedances. These imbalances are a function of the temperature and drive levels and therefore lead to changes in apparent phase or time with change in these parameters. The most important parameter that characterizes these effects is the isolation between the LO and rf ports and the IF port. Table 2 shows approximate dc offsets as a function of LO drive and isolation [7].

Table 2. DC offset voltage in mV for various combinations of LO drive level and isolation assuming a 50 Ω resistance on all ports.

ISOLATION - dB \ LO DRIVE LEVEL – dBm	0	+3	+7	+10	+13	+17	+20
15	25.3	35.8	63.7	90.0	127.1	226.0	319.2
20	14.2	20.1	35.8	50.6	71.5	127.1	179.5
25	8.0	11.3	20.1	28.5	40.2	71.5	100.9
30	4.5	6.4	11.3	16.0	22.6	40.2	56.8
35	2.5	3.6	6.4	9.0	12.7	22.6	31.9
40	1.4	2.0	3.6	5.1	7.1	12.7	17.9
45	0.80	1.1	2.0	2.9	4.0	7.1	10.1
50	0.45	0.64	1.1	1.6	2.3	4.0	5.7

Table 2. DC offset voltage in mV for various combinations of LO drive level and isolation assuming a 50 Ω resistance on all ports. From Watkins-Johnson Application Note [7].

Figure 4 shows a typical circuit diagram for evaluating the phase detectors. The dc output (IF) port is terminated with the characteristic impedance, which is typically 50 Ω for the double balanced mixers and 500 Ω for the single balanced mixers.

Figures 5 through 7 show timing errors in selected phase detectors due to variation in input power at two temperatures for a carrier frequency of 5 MHz [5]. One set of curves was taken without attenuators and the others with either 3 or 6 dB attenuators on the rf and LO inputs. At 5 MHz the mixer with the individually matched diodes has a temperature coefficient of approximately 6 ps/K when used with 6 dB pads, while the mixers with the quad diodes

exhibit a temperature coefficient of about 25 ps/K. The sensitivity to power changes is generally improved by adding 3 or 6 dB attenuators to the input

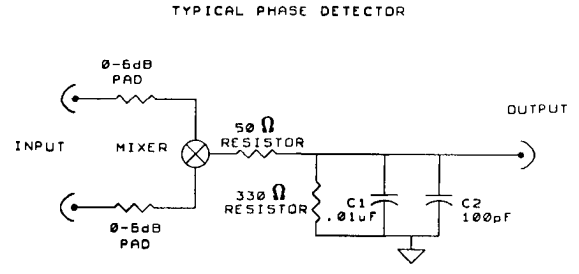


Fig. 4. Typical circuit diagram for evaluating the sensitivity of various phase detectors to variations in temperature and rf level.

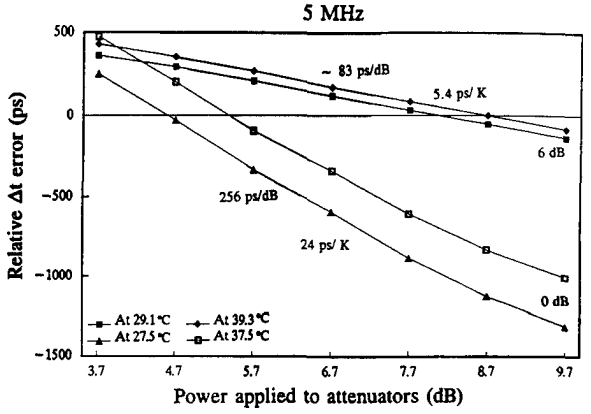


Figure 5. Timing errors in a double balanced mixer with individual discrete diodes at a carrier frequency of 5 MHz as a function of drive power, input attenuation, and temperature.

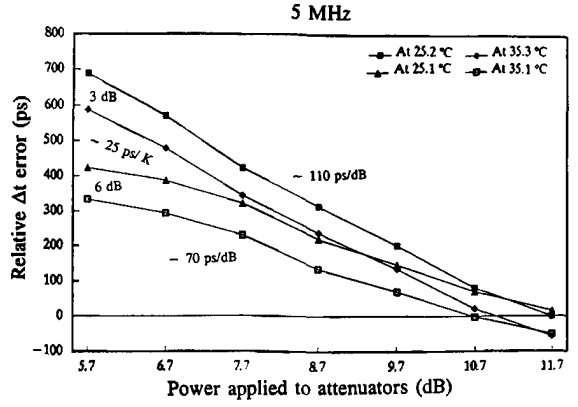


Figure 6. Timing errors in a double balanced mixer with a quad diode module, at a carrier frequency of 5 MHz as a function of drive power, input attenuation, and temperature.

ports. Without attenuators the power sensitivity is of order 1 ms/(dBν_o) or 250 ps/dB. The improvement, due to the attenuator use on the rf and LO inputs, is thought to originate from the reduction in the cable VSWR generated by the phase detector load.

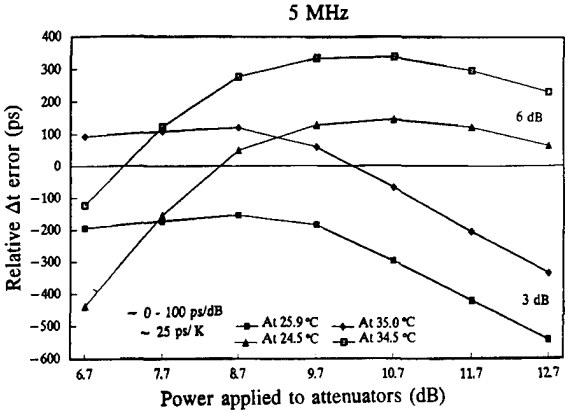


Figure 7. Timing errors in a single balanced mixer with a dual diode module, at a carrier frequency of 5 MHz as a function of drive power, input attenuation, and temperature.

Figures 8 through 10 show the timing errors in selected phase detectors due to variations in input power at two temperatures for a carrier frequency of 100 MHz [5]. One set of curves was taken without attenuators and the others with either 3 or 6 dB attenuators on the rf and LO inputs. The performance at 100 MHz of all three phase detectors versus temperature and power is improved over the results at 5 MHz by approximately the ratio of the carrier frequencies. Temperature sensitivity is of order 0.2 to 1 ps/K. The best temperature performance comes from the mixer with individually matched diodes. The power sensitivity is of order 0.1 to 1 ps/(dBν_o) or 1 to 10 ps/dB at 100 MHz depending on mixer type, input power, and attenuation.

Figures 11 and 12 show the timing errors in a double balanced phase detector with quad diodes. This detector has been approximately matched to 50 Ω by adding an extra resistor in series with a 6 dB attenuator [5]. Timing changes due to variations in input power level and two different temperatures are shown. Figure 11 shows a set of curves taken with the added resistor at a carrier frequency of 5 MHz. The mixer has a temperature coefficient of approximately 3.5 ps/K. The power sensitivity is approximately 71.1 ps/dB and is better at higher power levels. The sensitivity to temperature is

improved greatly as compared to the data obtained without the matching as shown in Figure 6.

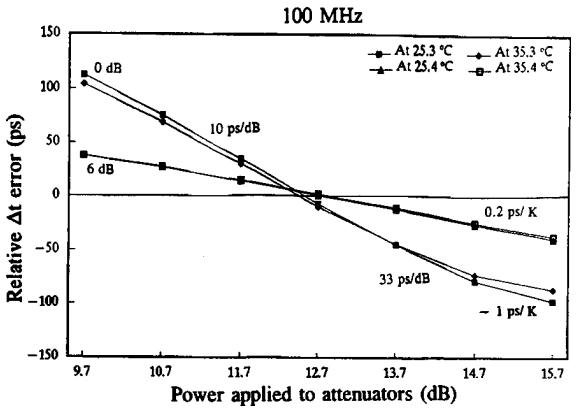


Figure 8. Timing errors in a double balanced mixer with individual discrete diodes, at a carrier frequency of 100 MHz as a function of drive power, input attenuation, and temperature.

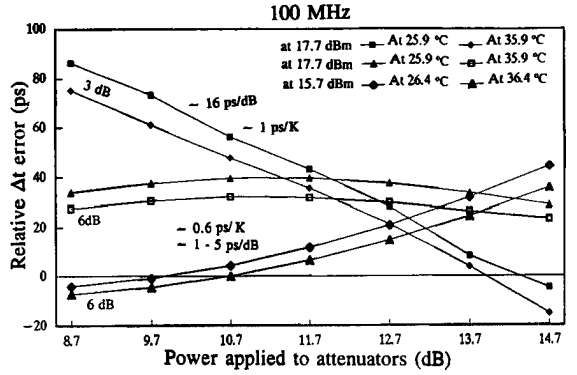


Figure 9. Timing errors in a double balanced mixer with a quad diode module, at a carrier frequency of 5 MHz as a function of drive power, input attenuation, and temperature.

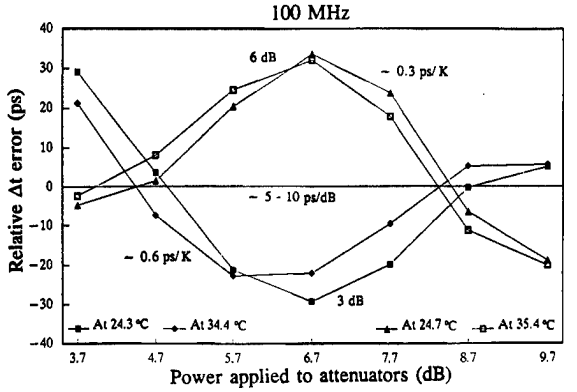


Figure 10. Timing errors in a single balanced mixer with a dual diode module, at a carrier frequency of 100 MHz as a function of drive power, input attenuation, and temperature.

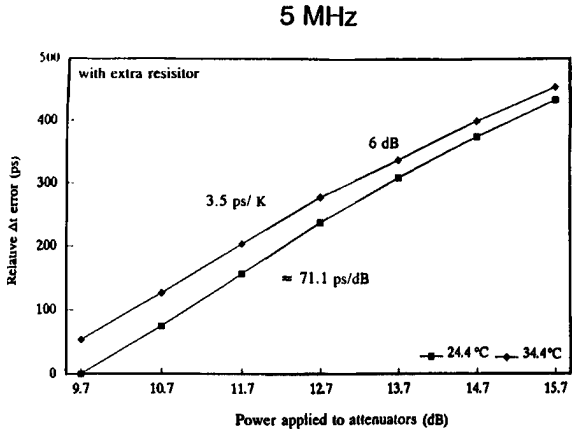


Figure 11. Timing errors in a double balanced mixer with quad diodes, at a carrier frequency of 5 MHz as a function of drive power, input attenuation (6 dB with extra resistance), and temperature.

Figure 12 shows a set of curves taken with the added resistor at a carrier frequency of 100 MHz. The mixer has a temperature coefficient of approximately 0.25 ps/K. The power sensitivity proved to be less at lower power and ranged from 1.6 ps/dB to 4 ps/dB. The extra resistor for the 50 Ω match helped to make the system less sensitive to the changes in the mixer temperature and power level. Figure 9 shows this improvement over the data taken before matching.

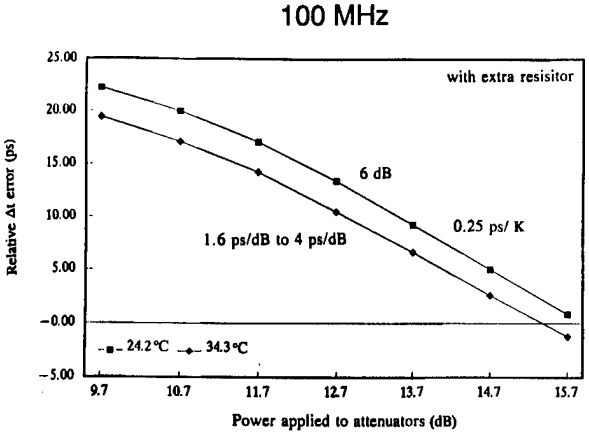


Figure 12. Timing errors in a double balanced mixer with quad diodes, at a carrier frequency of 100 MHz as a function of drive power, input attenuation (6 dB with extra resistance), and temperature.

Figures 13 and 14 show timing errors in a single balanced phase detector with an adjustable external output, a potentiometer, due to the variation in input power for two different temperatures. For a carrier frequency of 5 MHz (Figure 13), the mixer has a temperature coefficient of approximately 33 ps/K.

This is quite different from data taken for a single balanced phase detector without an adjustable external output (Figure 7). The fine tuning capabilities provide a more stable system that is less sensitive to temperature and power variations, providing a more stable system.

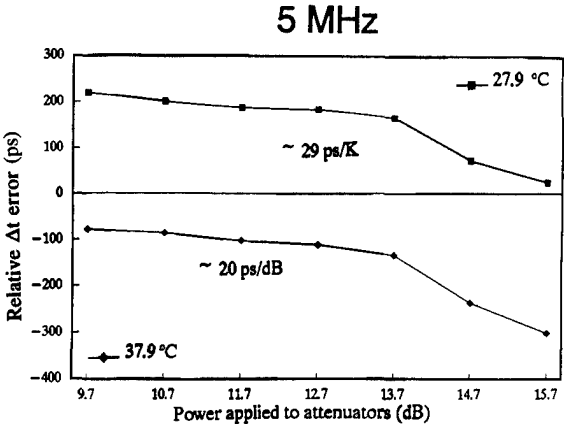


Figure 13. Timing errors in a single balanced mixer with an external adjustment, at a carrier frequency of 5 MHz as a function of drive power and temperature.

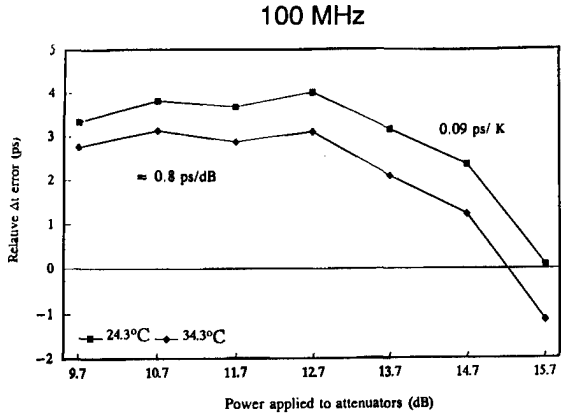


Figure 14. Timing errors in a single balanced mixer with an external adjustment, at a carrier frequency of 100 MHz as a function of drive power and temperature.

Measurements taken at 100 MHz (Figure 14) show that the phase detector has a temperature coefficient of approximately 0.09 ps/K and a power sensitivity of approximately 0.8 ps/dB. Using the external adjustment to match the diodes we were able to achieve a minimum dc offset variation between power levels which allowed for a very stable measurement of the time variation with the temperature change. Figure 10 shows data for a similar single balanced phase detector without the external adjustment for matching capabilities.

Figures 15 and 16 show the VSWR changes in the quad diode double balanced mixer with changing rf power. For a carrier frequency of 5 MHz, shown in Figure 15, the mixer VSWR exhibits a small dependence on power when it is matched to the 50 Ω characteristic system impedance with an extra resistor added in series with the 6 dB pads on each input port. The low sensitivity of VSWR to power reduces timing error within the measurement system. At 100 MHz the mixer also has low sensitivity to changes in power when it is matched with the extra resistor (Figure 16). These results also indicate that the VSWR is better at low powers as was evident at 5 MHz (Figure 15). However, we find that there is less change in the VSWR at 100 MHz than at 5 MHz and we obtain an even smaller timing error in our measurement system.

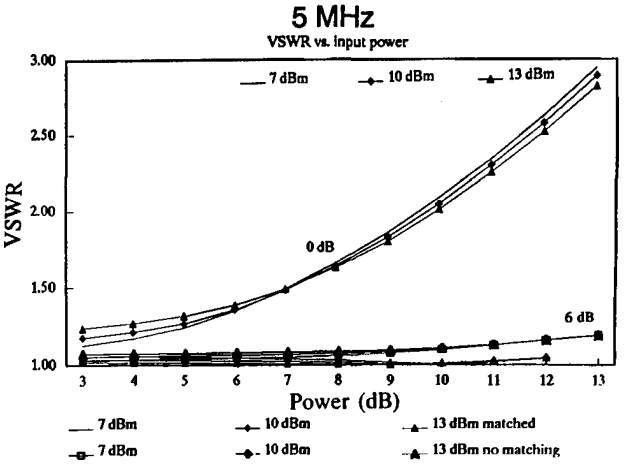


Figure 15. VSWR in a double balanced mixer with quad diodes, at a carrier frequency of 5 MHz as a function of the drive power and input attenuation (6 dB and 6 dB with extra resistance for matching).

Discussion

It is much easier to achieve the desired time domain performance of $\delta\nu/\nu_o = \delta t/t = 2 \times 10^{-17}$ at a measurement time of 10^4 s for a carrier frequency of 100 MHz than at 5 MHz due to the sensitivity of currently available phase detectors to both temperature and rf amplitude. Even at 100 MHz it will probably be necessary to stabilize the temperature of the phase detectors to about 1 K and carefully match the characteristic impedance of the

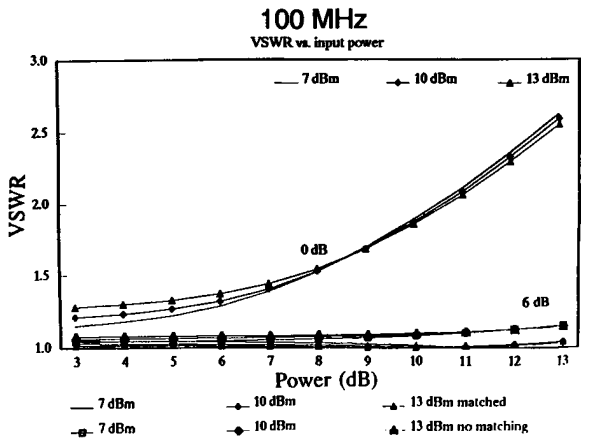


Figure 16. VSWR in a double balanced mixer with quad diodes, at a carrier frequency of 100 MHz as a function of the drive power and input attenuation (6 dB and 6 dB with extra resistance for matching).

coaxial cable. For distances of up to 100 m, existing coaxial cables are probably sufficient. For longer distances one must consider either AM modulation of a laser transmitted over fiber optic cables or using lower frequencies where the attenuation losses are smaller. The fiber optic has already demonstrated excellent performance for distances up to 25 km and is probably the best choice. Going to 5 MHz would require two-way timing measurements to correct for the time dispersion in the transmission medium and much more effort on stabilizing the operating parameters of the phase detector and lowering the effects of VSWR at the source and the load. Being able to match the diodes by using an external adjustment for tuning and matching the VSWR at the inputs seems most reasonable to accomplish the accuracy and stability that we want to achieve.

Acknowledgements

The authors would like to thank C. Osborn for help in construction details and G. Valdez for help in construction. They would also like to thank Z. Popovic for her guidance and support.

References

[1] D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano, D. J. Heinzen, S. L. Gilbert, C. H. Manney, and M. G. Raizen, "Progress at NIST toward absolute frequency standards using stored ions," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, **37**, 515-523 (1990).

[2] J. D. Prestage, G. J. Dick, and L. Maleki, "Linear ion trap based atomic frequency standard," Proc. of 44th Ann. Frequency Control Symp., 1990, pp. 82-88.

[3] S. Chu, "Laser manipulation of atoms and particles," Science, **253**, 861-866 (1991).

[4] J. L. Hall, M. Zhu, and P. Buch, "Prospects for using laser-prepared atomic fountains for optical frequency standards applications," J. Opt. Soc. Am. B, **6**, 2194-2205 (1989).

[5] Certain commercial equipment, instruments, or materials are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Institute of Standards & Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

[6] G. F. Lutes and R. T. Logan, "Status of frequency and timing reference signal transmission by fiber optics," Proc 45th Ann Frequency Control Symp., 1991, pp. 679-686.

[7] "Mixers as phase detectors," from 1985 Watkins-Johnson Catalog, pp. 666-673, 1985.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

ADVANCES IN ACCELERATION SENSITIVITY MEASUREMENT AND MODELING

John A. Kosinski[+,*] and Arthur Ballato[+]

[+]U.S. Army Research Laboratory, Ft. Monmouth, NJ 07703-5601

[*]Rutgers, the State University of New Jersey
Department of Electrical and Computer Engineering
P.O. Box 909, Piscataway NJ 08855-0909

Abstract: The effects of acceleration on "precision" frequency sources have recently been reviewed by the IEEE Standards Project P1193 Committee. Acceleration effects in "less precise" frequency sources (UHF and microwave sources such as SAW resonator and delay line oscillators, as well as VCXO modulation oscillators) are not addressed. Experience has shown that acceleration effects can be just as important and in some cases are more challenging to quantify in these other classes of devices. This paper discusses various aspects of measurement and modeling which have been discovered while working with these "less precise" devices, but do in fact have general application. The first part of this paper discusses the modeling of acceleration effects in BAW resonators and oscillators by means of equivalent electrical circuits. The second part of this paper discusses practical considerations in the measurement of UHF and microwave frequency sources. The final section of this paper presents a derivation of the general conditions for an intrinsically acceleration-insensitive crystal resonator and discusses the application of the conditions to the evaluation of the relative sensitivity of different crystal cuts, modes of motion, mounting geometries, and piezoelectric materials.

Introduction

The effects of acceleration on "precision" frequency sources have recently been reviewed by the IEEE Standards Project P1193 Committee [1]. Acceleration effects in "less precise" frequency sources (UHF and microwave sources such as SAW resonator and delay line oscillators, as well as VCXO modulation oscillators) are not addressed. Experience has shown that acceleration effects can be just as important and in some cases are more challenging to quantify in these other classes of devices. This paper discusses various aspects of measurement and modeling which have been discovered while working with these "less precise" devices, but do in fact have general application.

The first part of this paper discusses the modeling of acceleration effects in BAW resonators and oscillators by means of equivalent electrical circuits. An equivalent circuit representation of purely elastic nonlinearities in BAW resonators has been developed, and the accuracy of the model in describing the acceleration-induced perturbation in resonator transimpedance has been verified by tests in a unique centrifuge system. The equivalent circuit model has been linked to an oscillator model which includes an empirically derived model of acceleration-sensitive loop components. The calculated acceleration sensitivity of a VCXO modulation oscillator modeled in such a fashion compares favorably with observations.

The second part of this paper discusses practical considerations in the measurement of UHF and microwave frequency sources. The small modulation index approximation of $\beta < 0.1$ corresponds to a sideband-to-carrier ratio $\mathcal{L}(f) \leq -26$ dBc. This limit is regularly exceeded in testing UHF and microwave devices; thus a "moderate" modulation index approximation useful for $\mathcal{L}(f) \leq +10$ dBc has been derived. Another practical consideration concerns the effects of transverse motion in a shake-table system on measurements of the various components of the acceleration sensitivity vector. It has been found that simultaneous measurements of both intended and transverse accelerations of the device under test are required for correct interpretation of the observed frequency shifts. This requirement is particularly evident when the magnitudes of the vector components are substantially unequal.

The final section of this paper presents a derivation of the general conditions for an intrinsically acceleration-insensitive crystal resonator and discusses the application of the

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

conditions to the evaluation of the relative sensitivity of different crystal cuts, modes of motion, mounting geometries, and piezoelectric materials.

I. Equivalent Electrical Circuit Models

The vector description of the acceleration sensitivity of quartz resonators and oscillators has been widely used during the past decade. The familiar vector equation

$$f(\mathbf{a}) = f(\mathbf{a}=0)\cdot(1+\Gamma\cdot\mathbf{a}) \qquad (1)$$

has been used interchangeably to describe the changes in eigenfrequency of the quartz resonator and operating frequency of the crystal oscillator. That these two quantities are in fact different was first suggested by the data of Valdois, et al. [2], and recently shown conclusively by the measurements of voltage controlled crystal oscillators (VCXO) by Ballato, et al. [3]. This has led to the development of a modeling scheme wherein the effects of acceleration on the quartz resonator and oscillator may be examined separately [4]. That is to say, the acceleration-induced frequency perturbation in a quartz crystal oscillator is well represented by first properly modeling the effects of acceleration on the resonator transimpedance, and then examining the interaction of the perturbed resonator with the oscillator loop. The modeling procedure has been verified experimentally through a series of separate resonator and oscillator tests.

Two separate measurement systems have been employed in this work. Figure 1 shows a block diagram of the shake table system used for VCO tests including recent improvements. The capabilities of this type of system have been described elsewhere [5]. Figure 2 shows a block diagram of the unique centrifugal accelerator system used for resonator tests. This system can produce acceleration levels up to 600g. The linkage of the network analyzer with the centrifuge allows for improved data acquisition and analysis as compared to previous implementations [6]. Although improvements in the commutator and interconnections have been made, commutator related systematic errors present a challenge in measuring the acceleration effects. The accurate determination of the acceleration effects along any resonator axis requires two tests taking advantage of the change in sign associated with the acceleration-induced perturbation when the angle between the applied acceleration and the resonator sensitivity vector is changed by π radians. Figures 3 through 6 show typical test results and data analysis for a centrifugal test system measurement of

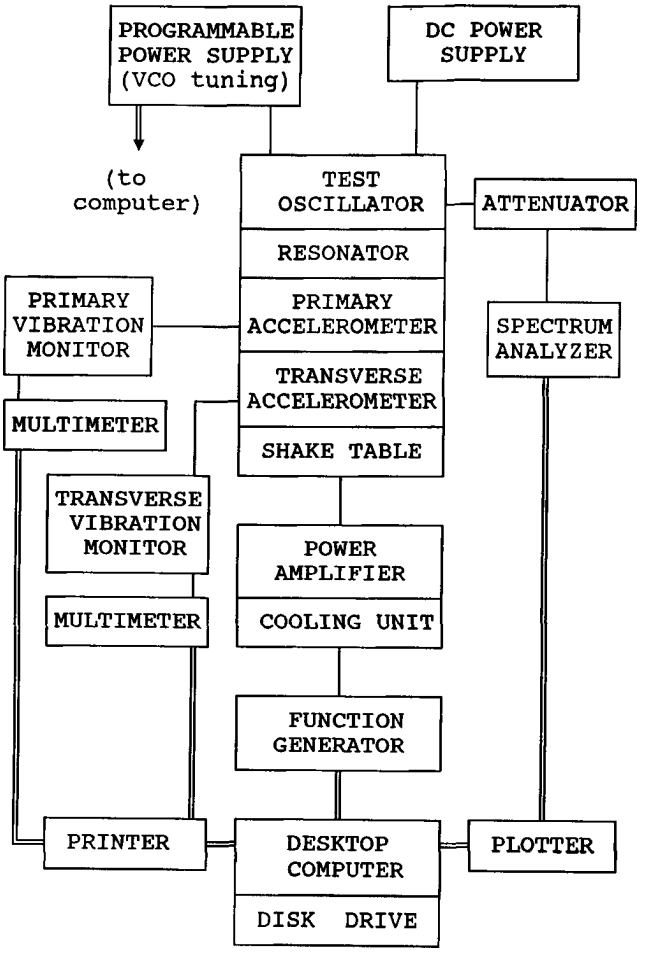


Figure 1. Shake table test system.

the perturbation in the scattering matrix element S_{11} of a 30MHz AT-cut crystal in an HC-6 type enclosure.

The Butterworth-Van Dyke single-mode lumped element approximate equivalent electrical circuit is well known. The equivalent circuit parameters may be determined from the crystal material properties using well known relations obtained using the linear piezoelectric theory for the case of the infinite flat plate:

$$C_o = \frac{\epsilon A}{2h} \qquad (2)$$

$$R_1 = \frac{n^2 \pi^2 h}{4 A e^2}\eta_s = \frac{n^2 \pi^2 h \bar{c} \tau_1}{4 A e^2} \qquad (3)$$

$$L_1 = \frac{h^3 \rho}{A e^2} \qquad (4)$$

$$C_1 = \frac{4\,A\,e^2}{n^2\,\pi^2\,h\,\bar{c}} \qquad (5)$$

The point of note is that R_1 and C_1 depend on $\bar{c}$, while C_0 and L_1 do not. Since the acceleration-induced effects are considered to arise out of purely elastic nonlinearities, one is led to consider the effect of introducing an acceleration sensitive elastic stiffness in these relations. Recognizing that the piezoelectric contribution to the stiffened elastic constant is small in the case of quartz resonators, we model the acceleration sensitive elastic stiffness as

$$\bar{c}(\mathbf{a}) = \bar{c}(\mathbf{a}=0)*(1+\gamma\cdot\mathbf{a}) \qquad (6)$$

wherein we use the lower case Greek gamma to denote the acceleration sensitivity of the elastic stiffness. The relationship between the acceleration sensitivity γ of the elastic stiffness and the acceleration sensitivity Γ of the resonance frequency is simply $\gamma \approx 2\,\Gamma$. The perturbations in the various circuit elements follow directly from the substitution of equation (6) into equations (2) through (5). We find

$$C_0(\mathbf{a}) = C_0 \qquad (7)$$

$$R_1(\mathbf{a}) = R_1(\mathbf{a}=0)*(1+\gamma\cdot\mathbf{a}) \qquad (8)$$

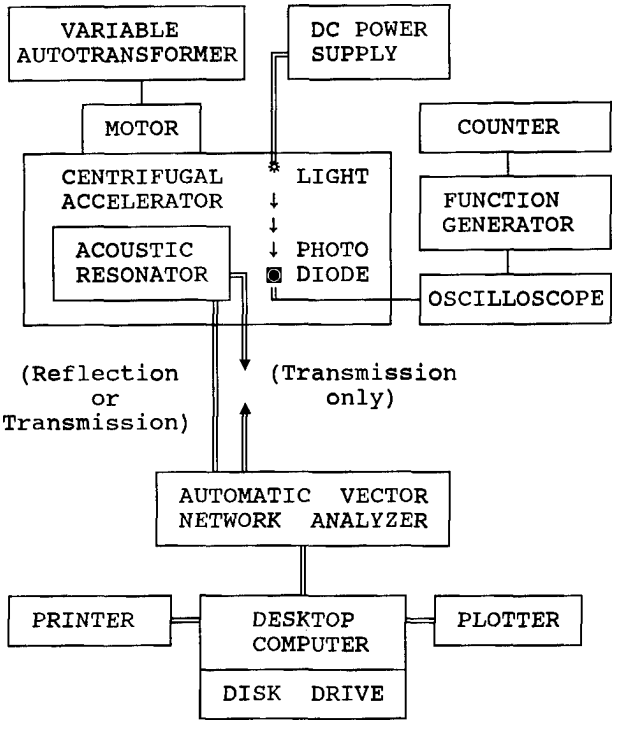


Figure 2. Centrifugal test system.

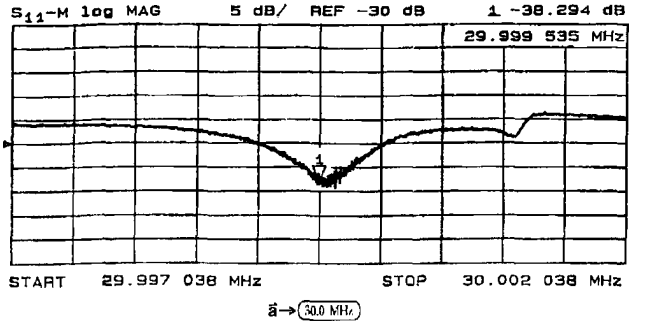


Figure 3. Perturbation in $|S_{11}|$ for an AT-cut resonator for $|\mathbf{a}| \approx 600g$ applied as shown. The marker is at the $|\mathbf{a}| = 0$ resonance frequency.

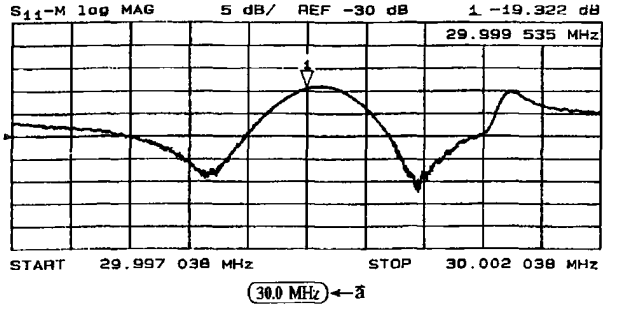


Figure 4. Perturbation in $|S_{11}|$ for the resonator of Figure 3 for $|\mathbf{a}| \approx 600g$ after reversing the resonator orientation as shown.

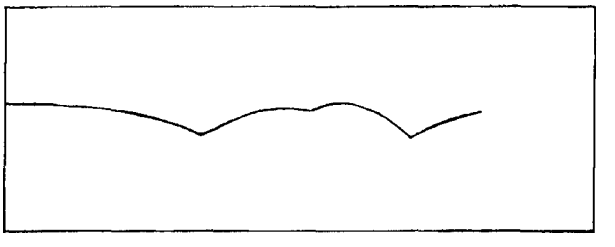


Figure 5. Systematic offset recovered from the $|S_{11}|$ data of Figures 3 and 4. Scales are the same as Figures 3 and 4.

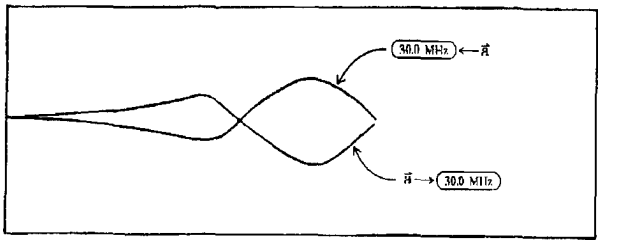


Figure 6. Perturbation in $|S_{11}|$ recovered from the data of Figures 3 and 4. Scales are the same as Figures 3 and 4.

$$L_1(\mathbf{a}) = L_1 \qquad (9)$$

$$C_1(\mathbf{a}) = C_1(\mathbf{a}=0)*(1-\gamma\cdot\mathbf{a}) \qquad (10)$$

These relations may then be used to calculate the crystal resonator transimpedance under acceleration.

The perturbation in S_{11} as calculated using this model for the one-port reflection measurement of Figures 3 through 6 is shown in Figure 7. We note that the calculations involve small differences in large numbers and are near the limits of the numerical precision for our computer. Nevertheless, the modeled perturbation shows the same shape and nearly the same bandwidth in frequency space as the data.

Once the perturbation in resonator transimpedance is known, it is a simple task to obtain the corresponding frequency shift and acceleration sensitivity of an ideal oscillator employing such a resonator as

$$|\Gamma| = -\frac{[\phi(\mathbf{a}) - \phi(\mathbf{a}=0)]}{2\,|\mathbf{a}|\,Q_L} \qquad (11)$$

where $\phi(\mathbf{a})$ and $\phi(\mathbf{a}=0)$ are the perturbed and unperturbed phase of the resonator transimpedance, and Q_L is the loaded or phase slope Q of the resonator at the operating point. Implicit in equation (11) is that no other components in the oscillator loop are perturbed by the acceleration. For actual hardware implementations, this is not always true.

The calculated acceleration sensitivity for the ideal case with no loop component contributions other than the crystal resonator is shown in Figure 8. The corresponding resonator S_{21} response is shown in Figure 9. The calculations imply that if the resonator is operated in the vicinity of resonance or antiresonance the overall oscillator acceleration sensitivity will depend on the actual operating point. We have not at this point determined whether this result accurately represents the physical process being modeled, or if it is instead an artifact of the calculational routine as implemented on our computer. We are continuing to examine this issue.

In the case where other loop components are also sensitive to the external acceleration, the effects on the oscillator loop are well represented by an ideal amplifier in series with an acceleration sensitive load reactance. Based on experimental observations, we treat the load reactance as an ideal component with an acceleration sensitive lead inductance as shown in Figure 10. The effect on the lead inductance may be modeled as a linear perturbation in similar

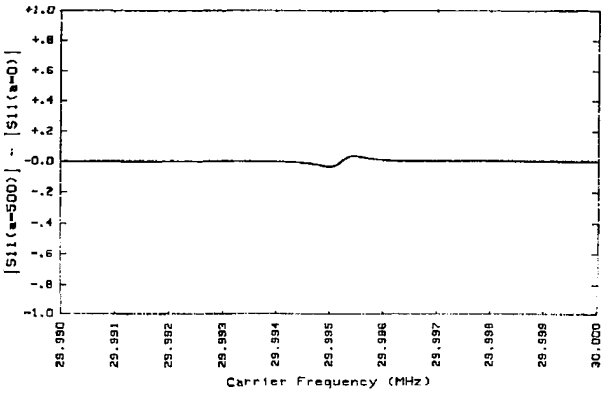


Figure 7. Calculated perturbation in the magnitude of S_{11} for the 30MHz resonator of Figures 3 through 6 with $|\Gamma|=6\times10^{-9}$/g as $|\mathbf{a}|$ varies between 0 and 500g.

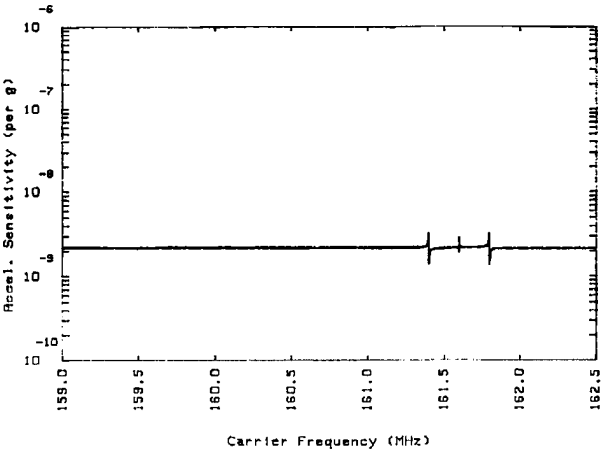


Figure 8. Acceleration sensitivity of an ideal tunable oscillator in which only the resonator transimpedance is acceleration sensitive. For this resonator, $|\Gamma|$ is nominally 2.2×10^{-9}/g.

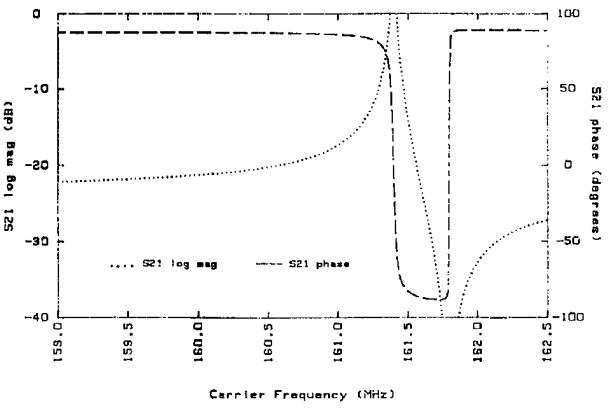


Figure 9. S_{21} response for the resonator of Figure 8.

fashion to equation (6), and the perturbation in transimpedance phase of the resonator/series load reactance combination may be employed in equation (11) to obtain the oscillator acceleration sensitivity under various operating conditions.

Figure 11 compares the application of this modeling technique to a crystal VCO incorporating the resonator of Figures 8 and 9 with the measured data obtained for the unit. The results obtained using the equivalent circuit approach are in good agreement with the measured data. The importance of maximizing the resonator loaded Q in order to minimize the effects arising from other loop components is evident in this figure. The importance of applying good construction techniques is also clearly illustrated.

In summary, an equivalent electrical circuit model of the acceleration sensitivity of piezoelectric plate resonators has been derived based on the linear piezoelectric theory, and the accuracy of the model in describing both resonator and oscillator behavior has been verified experimentally. The equivalent circuit modeling approach should prove particularly helpful in analyzing dual-resonator and other acceleration sensitivity compensation schemes. For SAW, SBAW, and similar resonators, the relations corresponding to equations (2) through (5) are more complicated, however the procedure for incorporating acceleration sensitivity in these equivalent circuit models follows directly from the derivation given here.

II. Practical Measurement Considerations

Sinusoidal vibration testing with a system similar to that of Figure 1 provides a convenient means to measure the acceleration sensitivity of a crystal oscillator. It follows directly from equation (1) and FM theory that the crystal oscillator output under sinusoidal vibration will be a frequency modulated signal with modulation index ß given by

$$ß = \frac{(\Gamma \cdot \mathbf{A}) \cdot f_o}{f_v} \qquad (12)$$

where **A** is the peak acceleration vector, f_o is the unperturbed oscillator output frequency, and f_v is the frequency of the applied sinusoidal vibration. The nth FM sideband ($-\infty \leq$ integer $n \leq +\infty$) is located at an offset frequency of $n \cdot f_v$ from the carrier and has a magnitude relative to the carrier given by

$$\mathcal{L}_v^n\{\text{decimal}\} = [J_n(ß)/J_0(ß)]^2 \qquad (13)$$

or

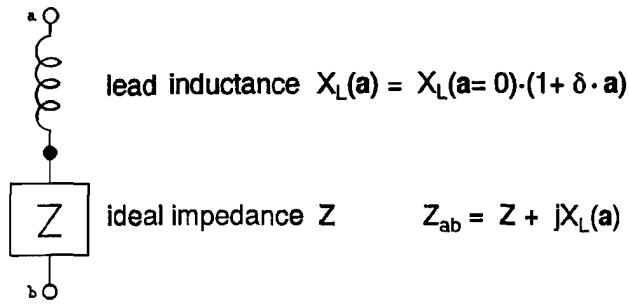


Figure 10. Empirically determined model of an acceleration sensitive loop component. The vector δ represents the acceleration sensitivity of the component leads.

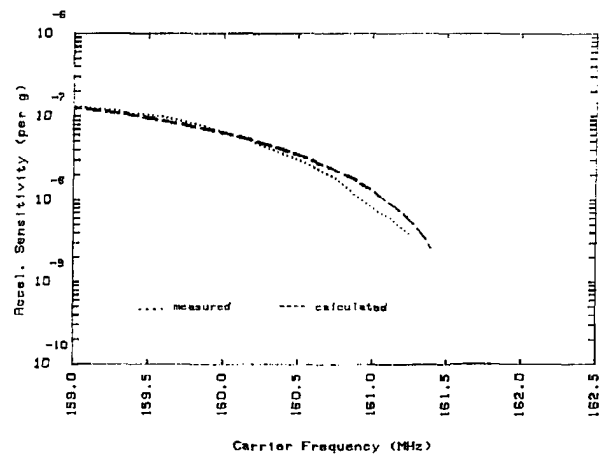


Figure 11. Measured and calculated acceleration sensitivity of an actual VCO employing the resonator of Figure 8 and 9.

$$\mathcal{L}_v^n\{\text{dBc}\} = 20\cdot\log[J_n(ß)/J_0(ß)] \qquad (14)$$

where the $J_n(ß)$ are Bessel functions of the first kind of order n and argument ß. For "precision" frequency sources in the low r.f. range, the modulation index ß is typically less than 0.1 which enables the use of the small index approximation

$$J_0(ß) = 1 \quad ; \quad ß < 0.1$$

$$J_1(ß) = ß/2 \; ; \quad ß < 0.1$$

$$J_n(ß) = 0 \quad ; \quad ß < 0.1,\ n > 2 \qquad (15)$$

as popularized by Filler [7]. The small modulation index approximation is convenient in eliminating the need to employ Bessel functions but is limited in usefulness by the fact that ß=0.1 corresponds to a first sideband-to-carrier ratio of $\mathcal{L}_v^1(f) \leq -26$ dBc. This limit is regularly exceeded in testing UHF and microwave frequency sources.

In order to develop a "moderate" index approximation useful for larger sideband levels, we turn to the work of Onoe on quotients of Bessel functions [8]. Onoe defines the quotient function of the first kind $T_\nu(z)$ as

$$T_\nu(z) \equiv \frac{z\, J_{\nu-1}(z)}{J_\nu(z)} \qquad (16)$$

and gives the continued fraction expansion as

$$T_\nu(z) = 2\nu - \cfrac{z^2}{2\cdot(\nu+1) - \cfrac{z^2}{2\cdot(\nu+2) - \cdots}} \qquad (17)$$

Substitution of $z=\beta$ and $\nu=1$ into equations (16) and (17) leads to the result that

$$\frac{J_1(\beta)}{J_0(\beta)} = \cfrac{\beta}{2 - \cfrac{\beta^2}{4 - \cfrac{\beta^2}{6 - \cfrac{\beta^2}{8 - \cdots}}}} . \qquad (18)$$

The moderate modulation index approximation is obtained by truncation of the infinite series in the denominator as

$$\frac{J_1(\beta)}{J_0(\beta)} \approx \frac{4\beta}{8 - \beta^2} . \qquad (19)$$

This approximation is typically better than an order of magnitude more accurate than the small modulation index approximation over the range -26 dBc $\leq \mathcal{L}_v^1(f) \leq +10$ dBc. We have chosen $\mathcal{L}_v^1(f) = +10$ dBc as a practical upper limit for the moderate modulation index approximation. Above this level the error in the approximation exceeds 10 per cent. It is a simple matter to extend the approximations by retaining additional terms of the series in the denominator when higher accuracy or greater range is required.

In order to use the moderate modulation index approximation in practical measurements we employ the following procedure:

(1) Measure the first sideband-to-carrier ratio $\mathcal{L}_v^1(f)$ in units of dBc using the spectrum analyzer.

(2) Convert the data to decimal format using

$$[J_1(\beta)/J_0(\beta)]_{\text{decimal}} = 10^{[\mathcal{L}_v^1(f)/20]} . \qquad (20)$$

(3) Determine the modulation index β from

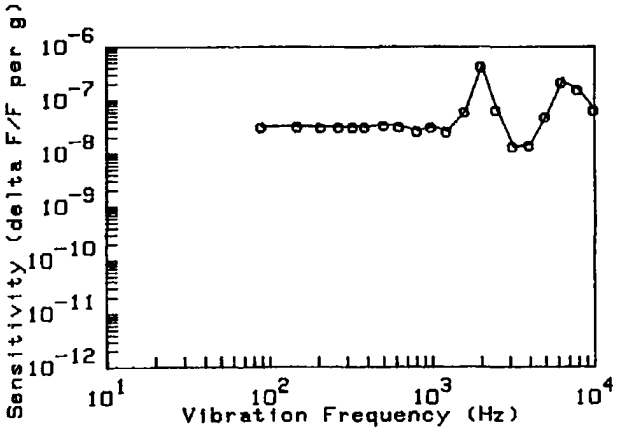


Figure 12. Measured acceleration sensitivity of a 250 MHz SAW oscillator for the vector component normal to the SAW propagation plane.

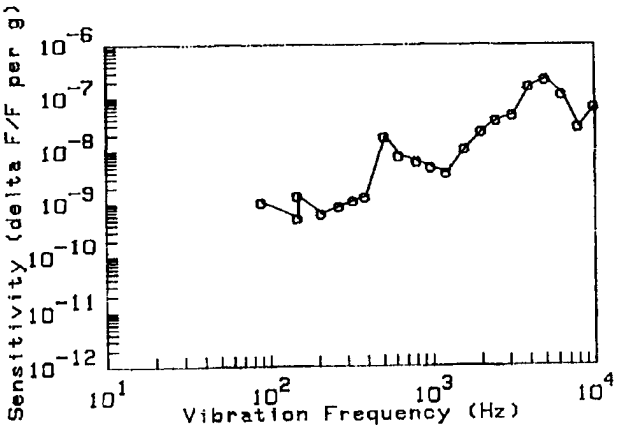


Figure 13. Measured acceleration sensitivity of a 250 MHz SAW oscillator for the in-plane vector component transverse to the SAW propagation direction.

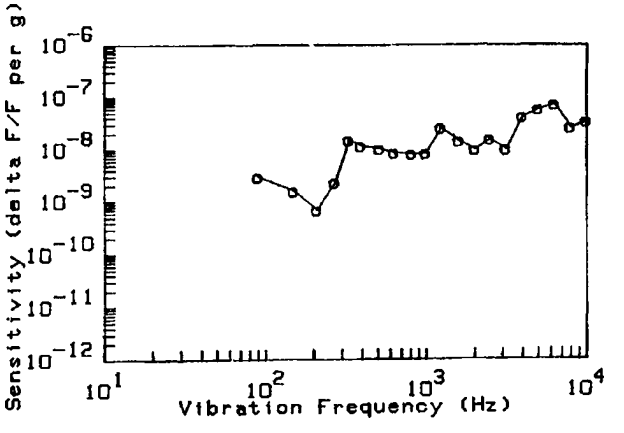


Figure 14. Measured acceleration sensitivity of a 250 MHz SAW oscillator for the in-plane vector component parallel to the SAW propagation direction.

$$\beta = \frac{-2 + (4 + 8\cdot[J_1(\beta)/J_0(\beta)]^2)^{1/2}}{[J_1(\beta)/J_0(\beta)]} \quad . \qquad (21)$$

(4) Determine the acceleration sensitivity vector component magnitude Γ_i from

$$\Gamma_i = \frac{\beta\cdot f_v}{|\mathbf{A}|\cdot f_o} \quad . \qquad (22)$$

In the majority of SAW oscillators tested by the authors, the component of the acceleration sensitivity vector normal to the SAW propagation plane is approximately an order of magnitude larger than either of the in-plane components [3,5,9,10]. In such cases, the effects of transverse motion of the device under test in the shake-table/test fixture system must be considered. This is clearly appreciated by expressing the fractional frequency shift experienced by the accelerated oscillator in component form,

$$\Delta f/f_o = \boldsymbol{\Gamma}\cdot\mathbf{A} = \Gamma_x A_x + \Gamma_y A_y + \Gamma_z A_z \qquad (23)$$

We observe that the largest product $\Gamma_i A_i$ will dominate the measured response. If one vector component is substantially larger than the others, even a small acceleration along that axis will dominate the measured response. This is illustrated in Figures 12 through 16.

Figure 12 presents the data obtained for a 250 MHz SAW oscillator, converted from sideband levels to acceleration sensitivity using equations (20) through (22), for the vector component normal to the SAW propagation plane. We have confidence in the accuracy of this data because (1) the magnitude of the vector component being measured is large, and (2) our test fixture does not experience transverse motion in this test orientation. Figures 13 and 14 present the "data" obtained for the two in-plane vector components. For these two tests, however, our test fixture experiences measurable transverse motion along an axis parallel to the normal component of the acceleration sensitivity vector of the device under test. We therefore measure the acceleration levels along both the intended test axis and the axis of transverse motion during each test, and use the normal vector component data and measured transverse acceleration in equations (12), (19), and (14) to calculate the sideband levels associated with the transverse motion. Figures 15 and 16 compare the calculated transverse acceleration sideband levels to the measured sideband levels for the tests of Figures 13 and 14. In both cases, the measured sideband levels are nearly the same as those associated with the transverse motion, and the "data" can only be taken as representing an upper bound for the vector component in question.

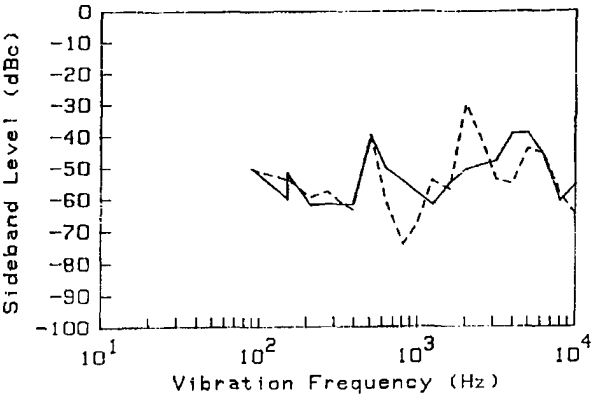


Figure 15. Comparison of measured sideband levels (solid line) and sideband levels generated by transverse motion in the shake-table/test fixture system (dashed line) for the test of Figure 13.

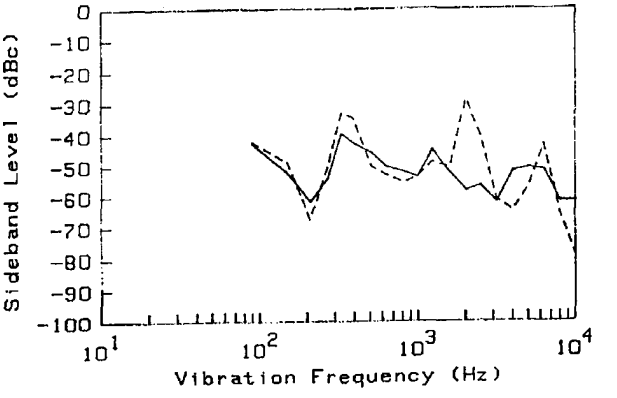


Figure 16. Comparison of measured sideband levels (solid line) and sideband levels generated by transverse motion in the shake-table/test fixture system (dashed line) for the test of Figure 14.

III. General Conditions for Intrinsically Acceleration Insensitive Crystal Resonators

The shift in eigenfrequency of a piezoelectric resonator subjected to a biasing condition may be found using the perturbation integral formulation of Tiersten [11], which we reproduce here for completeness as

$$\omega = \omega_\mu - \Delta_\mu \qquad (24)$$

$$\Delta_\mu = H_\mu/2\omega_\mu \qquad (25)$$

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha}\, g^\mu_{\alpha,M}\, g^\mu_{\gamma,L}\, dV \quad . \qquad (26)$$

In equations (24) through (26), ω and ω_μ are respectively the perturbed and unperturbed eigenfrequencies, the $\hat{c}_{L\gamma M\alpha}$ are coefficients dependent on the biasing field, g^μ_α denotes the normalized mode shape, and the integral is taken over the undeformed volume of the piezoelectric plate. If the biasing deformation and mode shape are known, the frequency shift can readily be calculated.

Recent analyses by various authors have focused on the role of symmetry in resonator mounting and mode shape in order to obtain acceleration-insensitive crystal resonator designs [12-29]. The practical limitation of such designs is how well the required symmetry conditions can be reproduced in a production environment. It would be preferable to develop a resonator design wherein acceleration insensitivity is obtained independently of both mounting and mode shape. We are thus led to examine the perturbation integral for a set of general conditions, independent of the biasing state and mode shape, wherein the resulting integral is identically zero.

We begin our analysis by noting that for any mode of motion, there must exist some non-zero mode shape g^μ_α. Thus any zeros of the perturbation integral must be associated with the biasing coefficients $\hat{c}_{L\gamma M\alpha}$. The defining relation for the biasing coefficients is

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM}\,\delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN}\,E^1_{KN} + c_{2L\gamma KM}\,w_{\alpha,K} + c_{2LKM\alpha}\,w_{\gamma,K}\,. \quad (27)$$

In equation (27), the $c_{L\gamma KM}$ and $c_{LKM\alpha}$ are the "ordinary" second order elastic stiffnesses while the $c_{L\gamma M\alpha KN}$ are the "non-linear" third order stiffnesses. The terms $w_{\alpha,K}$ and $w_{\gamma,K}$ represent the biasing deformation gradients, and describe the physical deformation of the piezoelectric plate in response to the bias. T^1_{LM} and E^1_{KN} represent respectively the static biasing stress and strain and are defined as

$$T^1_{LM} = c_{2LMKN}\,E^1_{KN} \quad (28)$$

and

$$E^1_{KN} = \frac{1}{2}\,(w_{K,N} + w_{N,K}) \quad (29)$$

Equation (29) is substituted into equation (28) to obtain

$$T^1_{LM} = \frac{1}{2}\,c_{2LMKN}\,(w_{K,N} + w_{N,K}) \quad (30)$$

and equations (29) and (30) are substituted into equation (27) to obtain

$$\hat{c}_{L\gamma M\alpha} = \frac{1}{2}\,c_{2LMKN}\,(w_{K,N} + w_{N,K})\,\delta_{\gamma\alpha} + \frac{1}{2}\,c_{3L\gamma M\alpha KN}\,(w_{K,N} + w_{N,K}) + c_{2L\gamma KM}\,w_{\alpha,K} + c_{2LKM\alpha}\,w_{\gamma,K}\,. \quad (31)$$

We can combine terms and employ the symmetry of the stiffness tensors to rewrite equation (31) as

$$\hat{c}_{L\gamma M\alpha} = w_{N,K}\,[c_{3L\gamma M\alpha KN} + c_{2LMKN}\,\delta_{\gamma\alpha} + c_{2L\gamma KM}\,\delta_{\alpha N} + c_{2LKM\alpha}\,\delta_{\gamma N}]\,. \quad (32)$$

The bracketed terms represent a set of stiffness cofactors dependent only upon the material properties of the piezoelectric plate. Defining the stiffness cofactors $k_{L\gamma M\alpha KN}$ as

$$k_{L\gamma M\alpha KN} = [c_{3L\gamma M\alpha KN} + c_{2LMKN}\,\delta_{\gamma\alpha} + c_{2L\gamma KM}\,\delta_{\alpha N} + c_{2LKM\alpha}\,\delta_{\gamma N}]\,, \quad (33)$$

we may rewrite equation (32) as

$$\hat{c}_{L\gamma M\alpha} = w_{N,K}\,k_{L\gamma M\alpha KN}\,. \quad (34)$$

The general conditions for an intrinsically acceleration-insensitive crystal resonator are now clearly understood:

If for each coefficient $\hat{c}_{L\gamma M\alpha}$, the stiffness cofactors $k_{L\gamma M\alpha KN}$ of the deformation gradients $w_{N,K}$ are zero, then the eigenfrequency of the resonator will be independent of the biasing state and mode shape.

In the most general case, there are 81 $\hat{c}_{L\gamma M\alpha}$ each of which contains 9 stiffness cofactors for a total of 729 independent conditions which must be met for the resonator to be intrinsically acceleration-insensitive. The number of independent conditions which must be met for any particular resonator will be reduced by the symmetry class of the piezoelectric material, and may be further reduced by both mounting structure and mode shape.

As an example of the mounting structure reducing the number of required conditions, we consider the case of a SAW substrate rigidly and uniformly supported along its back surface. The substrate deformation in response to an acceleration is then limited to thickness extension or thickness shear only. In this case, the

number of allowed deformation gradients is reduced from nine to three.

As an example of the mode shape reducing the number of required conditions, we consider the case of rotated Y-cut quartz BAW resonators operating in the slow thickness-shear mode. The mode shape varies slowly enough in the lateral dimensions that the only mode displacement gradient required for the perturbation integral is $g^{u}_{1,2}$ and thus the only biasing coefficient required is $\hat{c}_{1212}$ [17]. There are then only nine conditions required to achieve intrinsic acceleration insensitivity for an arbitrarily mounted resonator of this type. In Figures 17 through 21, the five non-zero k_{1212KN} are plotted as a function of orientation angle θ for the family of rotated Y-cut quartz resonators.

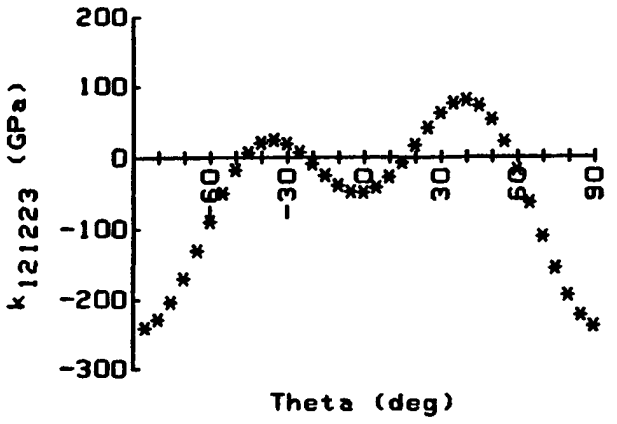


Figure 19. The cofactor k_{121223} versus orientation angle for rotated Y-cut quartz.

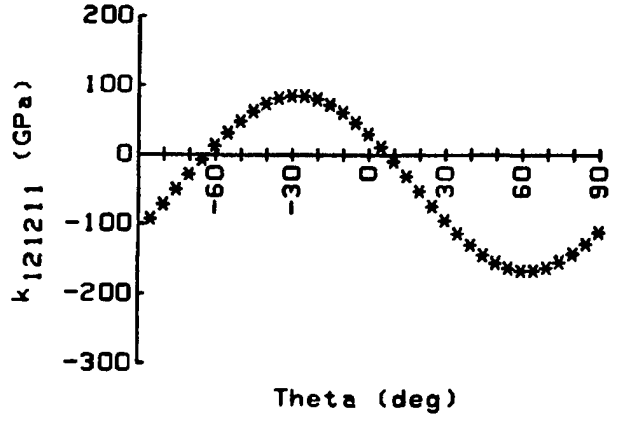


Figure 17. The cofactor k_{121211} versus orientation angle for rotated Y-cut quartz.

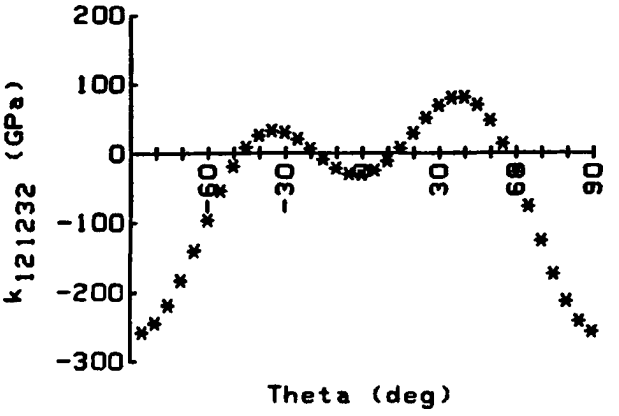


Figure 20. The cofactor k_{121232} versus orientation angle for rotated Y-cut quartz.

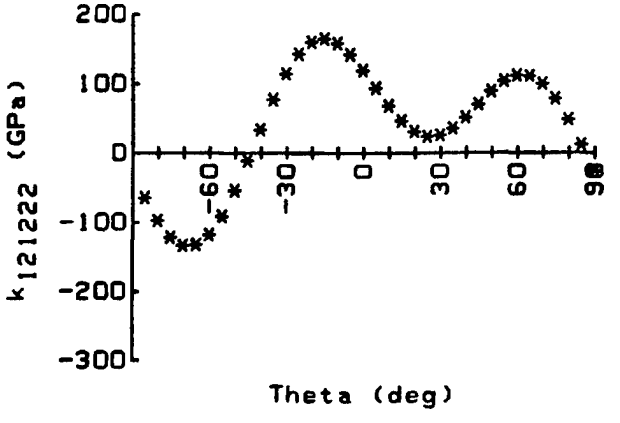


Figure 18. The cofactor k_{121222} versus orientation angle for rotated Y-cut quartz.

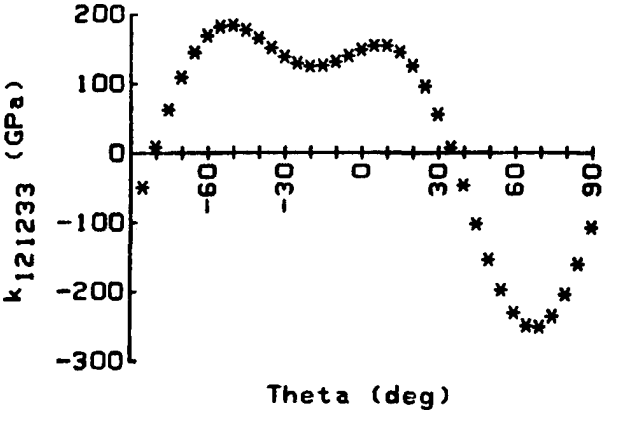


Figure 21. The cofactor k_{121233} versus orientation angle for rotated Y-cut quartz.

There are a number of observations to be made regarding the intrinsic insensitivity conditions:

(1) The statement of the general conditions represents a set of criteria for which a material may be searched. It does not constitute proof that the conditions can be met.

(2) This analysis is valid only to the order of the approximations underlying the perturbation theory. The conditions as stated should be sufficient to determine the existence of zeros in a region of crystal orientations, however a higher order analysis may be required to precisely determine exact loci.

(3) The general conditions have been derived for an arbitrary static biasing deformation and an arbitrary mode shape. They are applicable not only to the problem of acceleration sensitivity, but to other stress- and deformation-induced phenomena such as the contributions of electrode and mounting stress relief to aging.

(4) The general conditions shed light on the misconception that acceleration sensitivity arises from the fact that the piezoelectric material is nonlinear. The cofactors $k_{L\gamma M\alpha KN}$ contain sums of both second- and third-order stiffnesses. In a perfectly linear material, the third-order stiffness terms vanish but not the second-order terms. Thus while material nonlinearity is a significant part of acceleration sensitivity, there is also a "linear" part arising solely from the physical deformation of the resonator shape.

Acknowledgment

The authors wish to thank Professor S. Mallikarjun of Monmouth College, and our coworkers T. Lukaszek, J. Himmel, and R. McGowan for their assistance in obtaining the experimental data presented here. The authors would also like to thank Dr. D. Stevens of AT&T for helpful discussions regarding Section III of this work. One author (J.K.) would like to thank Professor Yicheng Lu of Rutgers University whose lectures on perturbation theory as applied to semiconductor energy band calculations provided the basic understanding from which Section III was generated.

References

[1] 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 (Proposed for IEEE Standards Project P1193)," U. S. Army LABCOM Research and Development Technical Report SLCET-TR-91-3, March 1991, 45 pp.

[2] M. Valdois, J. Besson, and J. Gagnepain, "Influence of Environment Conditions on a Quartz Resonator," Proc. 28th Annual Symposium on Frequency Control (ASFC), May 1974, pp. 19-32.

[3] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, and R. McGowan,"Electronic Desen-sitization of Resonators to Accelerations," Proc. 44th ASFC, May 1990, pp. 444-451.

[4] J. Kosinski, A. Ballato, T. Lukaszek, and S. Mallakarjun, "A New Approach to Modeling Quartz Resonator and Oscillator Acceleration Sensitivity," Proc. 5th European Frequency and Time Forum, March 1991, pp. 25-29.

[5] J. Kosinski, A. Ballato, and T. Lukaszek, "Measurements of Acceleration-Induced Phase Noise in Surface Acoustic Wave Devices," Proc. 44th ASFC, May 1990, pp. 488-492.

[6] M. Nakazawa, T. Lukaszek, and A. Ballato, "Force- and Acceleration-Frequency Effects in Grooved and Ring-Supported Resonators," Proc. 35th ASFC, May 1981, pp. 71-91.

[7] R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-35, No. 3, pp. 297-305, May 1988.

[8] Morio Onoe, Tables of Modified Quotients of Bessel Functions of the First Kind for Real and Imaginary Arguments. New York: Columbia University Press, 1958, pp. 1-13.

[9] A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, R. McGowan, and K. Klohn, "Acceleration Sensitivity Reduction in SAW and BAW Resonators by Electronic Means," Proc. 1990 IEEE Ultrasonics Symposium, December 1990, pp. 573-576.

[10] J. Himmel, R. McGowan, J. Kosinski, and T. Lukaszek, "Market Survey of Acceleration-Insensitive SAW Oscillators," elsewhere in these Proceedings.

[11] H. F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," Journal of the Acoustical Society of America, Vol. 64, No. 3, pp. 832-837, September 1978.

[12] M. H. Watts, E. P. EerNisse, R. W. Ward, and R. B. Wiggins, "Technique for Measuring the Acceleration Sensitivity of SC-cut Quartz Resonators," Proc. 42nd ASFC, June 1988, pp.442-446.

[13] E. P. EerNisse, L. D. Clayton, and M. H. Watts, "Variational Method for Modeling Static and Dynamic Stresses in a Resonator Disc With Mounts," Proc. 43rd ASFC, May-June 1989, pp. 377-387.

[14] E. P. EerNisse, R. W. Ward, and O. L. Wood, "Acceleration-Induced Frequency Shifts in Quartz Resonators," Proc. 43rd ASFC, May-June 1989, pp. 388-395.

[15] E. P. EerNisse and R. W. Ward, "Crystal Resonator With Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Patent 4,837,475, June 6, 1989.

[16] E. P. EerNisse, R. W. Ward, M. H. Watts, R. B. Wiggins, and O. L. Wood, "Experimental Evidence for Mode Shape Influence on Acceleration-Induced Frequency Shifts in Quartz Resonators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-37, No. 6, pp. 566-570, November 1990.

[17] E. P. EerNisse, L. D. Clayton, and M. H. Watts, "Distortions of Thickness Shear Mode Shapes in Plano-Convex Quartz Resonators with Mass Perturbations," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-37, No. 6, pp. 571-576, November 1990.

[18] L. D. Clayton and E. P. EerNisse, "Frequency Shift Calculations For Quartz Resonators," Proc. 45th ASFC, May 1991, pp. 309-320.

[19] P. C. Y. Lee and X. Guo, "Effect of Electrodes on the Acceleration Sensitivity of Crystal Resonators," Proc. 43rd ASFC, May-June 1989, pp. 416-418.

[20] P. C. Y. Lee and X. Guo, "Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes," Proc. 44th ASFC, May 1990, pp. 468-473.

[21] P. C. Y. Lee and X. Guo, "Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. UFFC-38, No. 4, pp. 358-365, July 1991.

[22] Y. S. Zhou, D. S. Shick, and H. F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Simple Rectangular Supports," 1989 Ultrasonics Symposium Proceedings, October 1989, pp. 393-400.

[23] Y. S. Zhou and H. F. Tiersten, "On the Influence of a Fabrication Imperfection on the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," Proc. 44th ASFC, May 1990, pp. 452-460.

[24] H. F. Tiersten and Y. S. Zhou, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports," Proc. 44th ASFC, May 1990, pp. 461-467.

[25] H. F. Tiersten and D. V. Shick, "On the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along Rectangular Edges," Journal of Applied Physics, Vol. 67, pp. 60-67, 1990.

[26] H. F. Tiersten and D. V. Shick, "On the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," Journal of Applied Physics, Vol. 67, pp. 2554-2566, 1990.

[27] 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," Journal of Applied Physics, Vol. 69, pp. 2862-2870, 1991.

[28] H. F. Tiersten and Y. S. Zhou, "The Increase in the In-Plane Acceleration Sensitivity of the Plano-Convex Resonator Due to its Thickness Asymmetry," Proc. 45th ASFC, May 1991, pp. 289-297.

[29] Y. S. Zhou and H. F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Stiffened by Identical Top and Bottom Quartz Cover Plates Supported by Clips," Proc. 45th ASFC, May 1991, pp. 298-308.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

MARKET SURVEY OF ACCELERATION-INSENSITIVE SAW OSCILLATORS

J. Himmel, R. McGowan, J. Kosinski and T. Lukaszek

US ARMY LABCOM, Electronics Technology & Devices Laboratory
Ft. Monmouth, NJ 07703-5601

ABSTRACT

SAW oscillators serve as key components in many military systems, such as IFF equipment on helicopters and tanks, guidance and surveillance systems on UAV's, smart munitions, etc. The timing for military computer systems and high speed bus systems may also be derived from SAW oscillators. The ever-increasing performance demands of these systems combined with the harsh vibration environments in which they operate has led to required acceleration sensitivity of 10^{-12} /g for state-of-the-art military applications.

In order to determine how closely commercially available SAW stabilized oscillators approach this requirement, a market survey of both fixed frequency and voltage controlled oscillators (VCO) is being conducted. The results of the complete market survey will lead to a greater understanding of commercial vendors capabilities to satisfy the hostile vibration environment of today's battlefield. The purpose of this paper is to provide a preliminary report on devices tested to date.

INTRODUCTION

The harsh operational environments experienced by modern military weapon systems produce a set of demanding system requirements far in excess of commercial requirements. Combinations of environmental effects can degrade weapon system performance at times when maximum performance is needed. Typical mobile military system platforms including fixed wing, rotary wing and wheeled or tracked vehicles produce severe vibration environments for on board electronics. The severe vibration environment not only introduces mechanical failures into equipment due to stress and fatigue, but also introduces degraded electronic performance which is less widely understood.

Among these electronic performance characteristics, the acceleration sensitivity of crystal oscillators has for some time been recognized as one of the limiting factors in the ultimate performance of military systems. The improved system performance which can be achieved by reducing the vibration sensitivity of crystal oscillators can often be dramatic. A 12 dB improvement in oscillator phase noise under vibration will double the range of a radar system for a constant target size, or improve the target detection capability by a factor of 16 for constant range.

Oscillators tested thus far have exhibited acceleration sensitivities ranging from 10^{-9}/g to 10^{-7}/g. The range of responses exhibited by the oscillators was as myriad as the number of oscillators tested and several unexpected results were observed in the test data.

TEST SYSTEM AND PROCEDURE

In order to test each SAW oscillator, a computer-controlled shake table system is employed, as illustrated in the block diagram, Figure 1, and as seen in the photograph of the test setup, Figure 2. The test fixture used to evaluate each of the oscillators is comprised of two major components, namely, a mounting bracket and a circuit board, as pictured in Figure 3. The mounting bracket, of aluminum construction, is bolted to the shake table and provides a stable platform to mount the circuit board. Each circuit board was fabricated on 0.060" thick pc board, having 2.0 oz. of laminated copper on both sides, in order to insure minimum variation in the test fixture

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

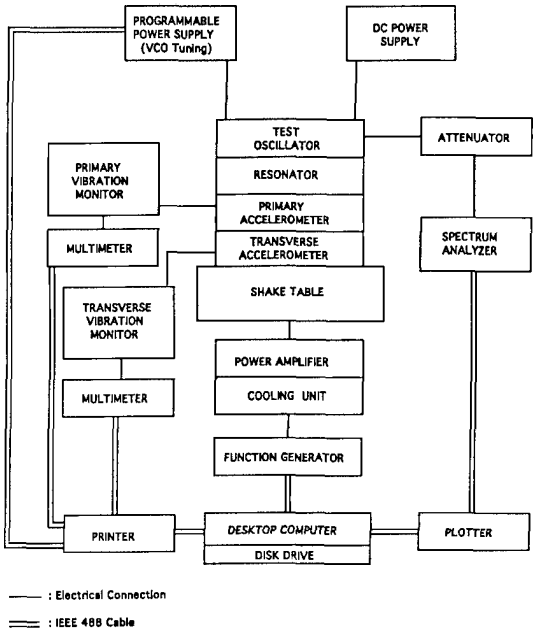


FIGURE 1 BLOCK DIAGRAM OF SHAKE TABLE TEST SYSTEM

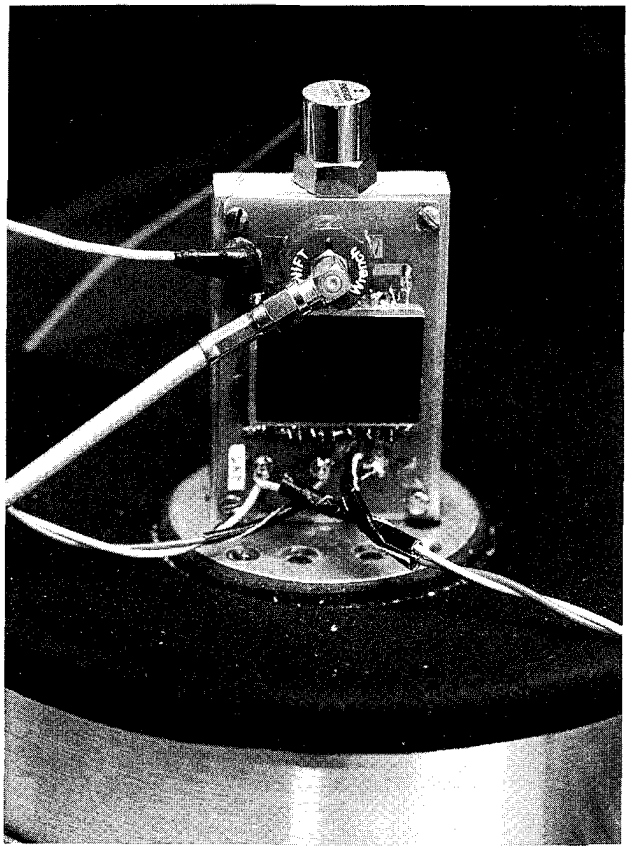


FIGURE 3 PHOTOGRAPH OF TEST FIXTURE

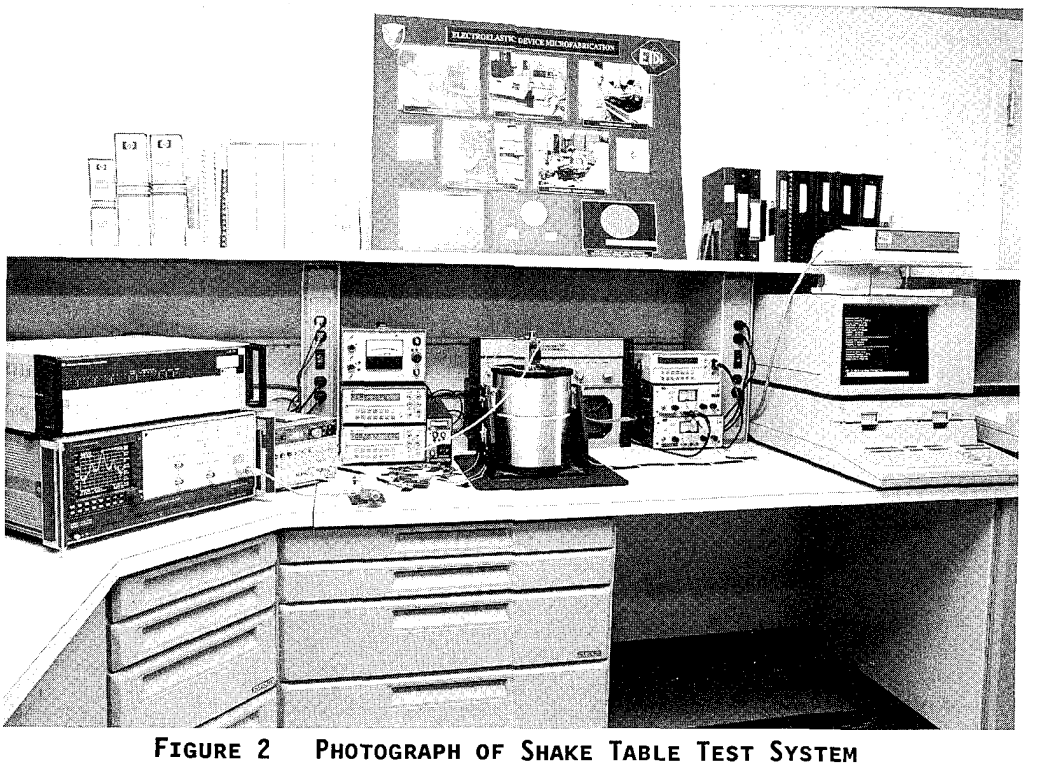


FIGURE 2 PHOTOGRAPH OF SHAKE TABLE TEST SYSTEM

configuration. The boards, whose dimensions were 4.8 cm x 7.3 cm, were used to mount the SAW oscillator and any necessary external circuitry to the mounting bracket.

After the device under test is powered up and the spectrum analyzer is calibrated, the device is subjected to a sinusoidal acceleration of 2 g's. Two accelerometers are mounted on the fixture. One is used in a feedback loop to maintain a peak acceleration level of 2 g's along the desired test axis while the other is used to measure any unwanted transverse acceleration. The vibration frequency is swept from 90 Hz to 9990 Hz. If the device is a VCO, tuning voltage is also varied. At each vibration frequency (and each tuning voltage, where applicable), measurements are taken of the power levels and frequencies of the carrier, the first upper sideband and the first lower sideband. The acceleration sensitivity is calculated by the computer using the average of the first sideband power levels, the vibration frequency, the peak acceleration along the test axis and the carrier frequency.

Acceleration sensitivity was measured along three orthogonal directions using the axial convention indicated in Figure 4. The three components of the acceleration sensitivity vector Γ are denoted by Γ_x, Γ_y and Γ_z.

For SAW oscillators, the first sideband-to-carrier power level ratio $\mathcal{L}_v^1(f)$ often exceeds -26 dBc, resulting in a modulation index ß which is greater than 0.1. In order to calculate the acceleration sensitivity under these conditions, the "moderate" index approximation is utilized as explained in detail by Kosinski [1]. The procedure is as follows:

(1) The first sideband-to-carrier power level ratio is measured in dBc using the spectrum analyzer.

(2) The data is converted to decimal format using

$$[J_1(ß)/J_0(ß)] = 10^{[\mathcal{L}_v^1(f)/20]} \qquad (1)$$

(3) The modulation index is determined by using the "moderate" modulation index approximation:

$$ß = \frac{-2 + (4 + 8\cdot[J_1(ß)/J_0(ß)]^2)^{1/2}}{[J_1(ß)/J_0(ß)]} \qquad (2)$$

(4) The magnitude of the ith component of the acceleration sensitivity vector is given by

$$\Gamma_i = \frac{ß\cdot f_v}{|\mathbf{A}|\cdot f_0} \qquad (3)$$

where f_v is the vibration frequency, $\mathbf{A}$ is the peak acceleration vector and f_0 is the unperturbed output carrier frequency of the device.

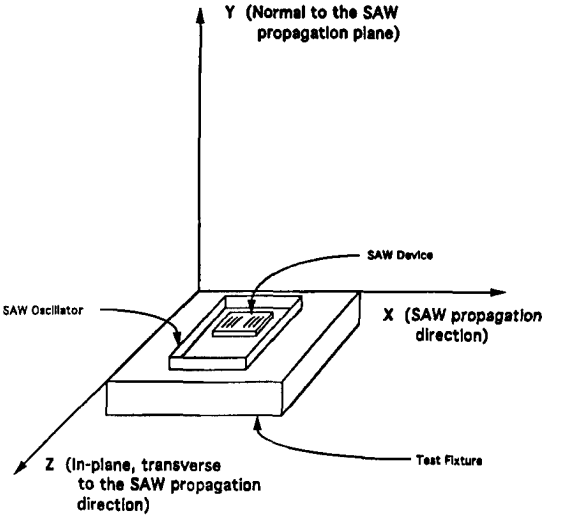


FIGURE 4 AXIAL CONVENTION FOR VIBRATION STUDY

TEST VALIDITY AND VERIFICATION

In the majority of the SAW oscillators, the component of acceleration sensitivity in the Y-direction, $|\Gamma_y|$, was an order of magnitude higher than the other components of the acceleration sensitivity vector. Due to the construction of the test fixture, this presented a problem in the form of unwanted transverse acceleration when measuring $|\Gamma_x|$ and $|\Gamma_z|$. As explained by Kosinski [1], when $|\Gamma_y|$ is substantially greater than $|\Gamma_x|$ and $|\Gamma_z|$, a small transverse motion in the test fixture can

cause a greater effect on the oscillator than the substantially larger motion along the desired axis. In order to assess the extent to which the data were affected, the values of $|\Gamma_y|$ were used to calculate the transverse motion induced power levels for the x-axis and z-axis tests. The transverse motion sideband levels were plotted together with and compared to the measured sideband levels in order to determine the validity range for the x-axis and z-axis data. Examples of such plots are illustrated in Figure 5. Whenever the transverse motion induced sideband levels were equal to or greater than the measured sideband levels, the data can only be interpreted as representing the upper bound for the acceleration vector component in question. A superior test fixture is under construction in order to reduce the problem of transverse acceleration.

Most of the tests resulted in mechanical resonances in the part of the test system consisting of the SAW resonator, the SAW oscillator, the test fixture and the shake table. The vibration frequencies at which the resonances occurred varied from one device to another and from one mounting to another.

Limitations in the measurement system have been discussed by Kosinski [2]. Errors in measuring the vibration frequency during acceleration contribute to the error in the calculation of Γ via the following equation:

$$\frac{\delta\Gamma}{\Gamma} = \frac{\epsilon}{f_v} , \qquad (4)$$

where ϵ is the error in f_v.

Errors ϵ in measuring either the peak acceleration or the carrier frequency contribute

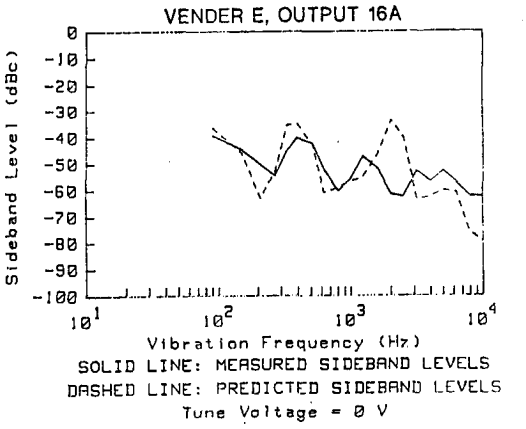
$$\frac{\delta\Gamma}{\Gamma} = \frac{-\epsilon}{\phi + \epsilon} , \qquad (5)$$

where ϕ is the measurand.

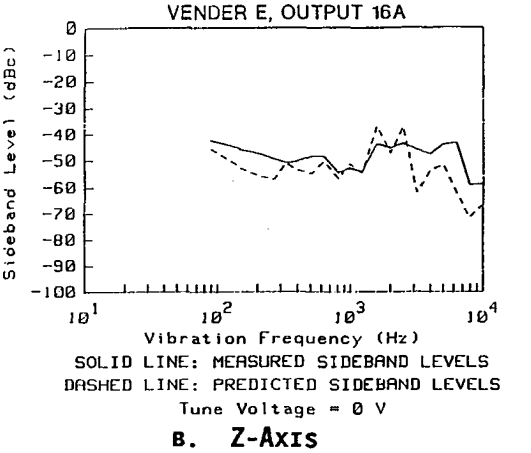
Errors in measuring the sideband levels contribute

$$\frac{\delta\Gamma}{\Gamma} = 10^{(\epsilon/20)} - 1 . \qquad (6)$$

where ϵ is the error in measuring the sideband-to-carrier power level ratio in dBc. The dominant source of error is the sideband power ratio measurement, which introduces an error in $|\Gamma|$ ranging from +19% to -16% [2].



A. X-Axis



B. Z-Axis

FIGURE 5 ANALYSIS OF THE TRANSVERSE ACCELERATION

TEST RESULTS & CONCLUSIONS

Typical test results for each type of oscillator evaluated in the survey are presented in Figures 6-13. The 2 dimensional plots in Figures 6-9 are representative of the four types of fixed frequency SAW oscillators tested. The four devices had the following characteristics:

- Vender A - 500 MHz SAW oscillator with a sinusoidal output,
- Vender B1 - 225 MHz SAW oscillator with a complementary ECL output,
- Vender B2 - 357 MHz SAW oscillator with a complementary ECL output,
- Vender E - 250 MHz SAW oscillator with two pairs of differential ECL outputs.

The 3 dimensional plots seen in Figures 10-13 show data from the four types of voltage controlled SAW oscillators (VCSO) examined in the survey. The device descriptions are as follows:

- Vender B3 - VCSO with two sinusoidal outputs, one at 701 MHz, and the other at 1403 MHz,
- Vender C1 - VCSO with a 240 MHz sinusoidal output,
- Vender C2 - VCSO with a 400 MHz sinusoidal output,
- Vender D - VCSO with a 622 MHz sinusoidal output.

The information provided by the shake table test results is much more comprehensive than the information which is usually available from industry. For example, for a single-frequency SAW oscillator, acceleration sensitivity is often given at only one vibration frequency. However, the two dimensional plots displayed in this report indicate that acceleration sensitivity may vary significantly with vibration frequency. For VCO's, industry often provides acceleration sensitivity at only one tuning voltage. While Figure 11 indicates that for some VCO's Γ may be relatively constant with tuning voltage, it turns out to be unwise to assume that all VCO's follow the same behavior. For example, Figure 12 displays a large dependence of $|\Gamma_x|$ and $|\Gamma_z|$ on tuning voltage at the lower vibration frequencies.

Figure 14 summarizes the results of all the SAW oscillators tested. Figures 14a, 14b and 14c show the magnitudes of the various components for all of the outputs of the units tested. Figure 15 shows the magnitude of the total acceleration sensitivity vector $|\Gamma|$ for each device output measured, i.e., it is the root-sum-squared of the three orthogonal components. Table 1 serves as a key of the various venders and oscillator types represented in Figure 14. Typical values of acceleration sensitivity components range from 10^{-10} /g to slightly above 4.0×10^{-8} /g. Notice the similarity of $|\Gamma_y|$ and $|\Gamma|$. As mentioned earlier, this illustrates that $|\Gamma_y|$ is usually the most dominant component of acceleration sensitivity. $|\Gamma|$ ranges from 1.4×10^{-9} /g to 4.4×10^{-8} /g. Clearly, the Army's requirement of 10^{-12} /g is three orders of magnitude better than the best oscillator tested to date in the survey.

Table 2 provides the acceleration sensitivity result summary for each output of the 16 devices tested. The data presented in this table includes the device number where **A, B, C** and **D** after the device number denote the different outputs of the device. Column 2 identifies the vender of a particular device. The next three columns provide the magnitude of the acceleration sensitivity for the three vector components. Column 6 gives the magnitude of the total acceleration sensitivity vector.

The relatively small data sample represented in this market survey shows a large degree of variability, not only from manufacturer to manufacturer and device to device as one might expect, but also from one output to another for devices with more than one output. For example, in one device with two outputs, there was a difference of a factor of four in the acceleration sensitivity between the two outputs. In the continuation of this study, combined with the measurements of acceleration sensitivity of the oscillators, there will be an empirical study of the physical construction of the oscillators. Notably, during the evaluation of the physical construction of the oscillator, emphasis will be placed on the aspect ratio of the SAW device, the mounting of the SAW, type of components used (i.e., discrete & chip versus thin film & microstrip), etc. The combination of the acceleration sensitivity data with the observations of the physical construction of the devices should provide an excellent tool for the design and development of a state-of-the-art vibration insensitive SAW based oscillators.

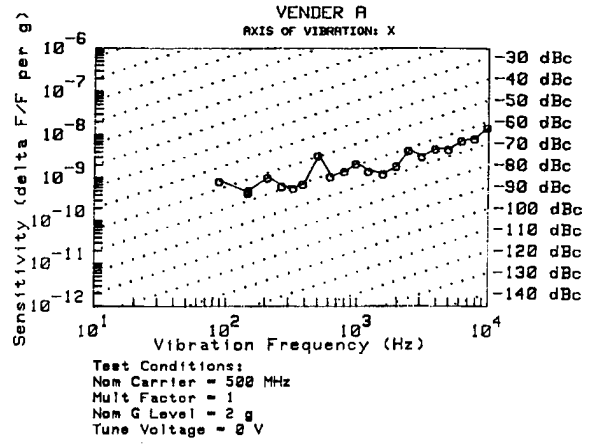
ACKNOWLEDGEMENT

The authors would like to thank the CECOM Center for EW/RSTA for its support in this effort.

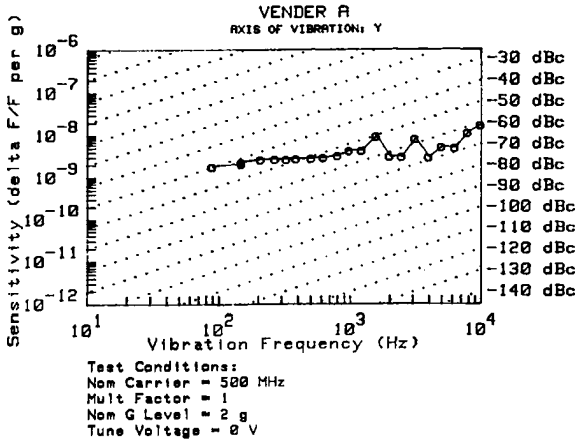
REFERENCES

[1] J. Kosinski and A. Ballato, "Advances in Acceleration Sensitivity Measurement and Modeling", elsewhere in these Proceedings.

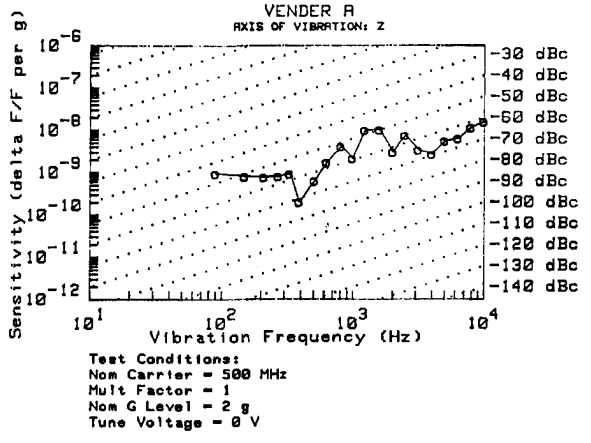
[2] J. Kosinski, A. Ballato, and T. Lukaszek, "Measurements of Acceleration-Induced Phase Noise in Surface Acoustic Wave Devices", Proc. 44th Annual Frequency Control Symposium, pp. 488-492, May 1990



A. X-Axis

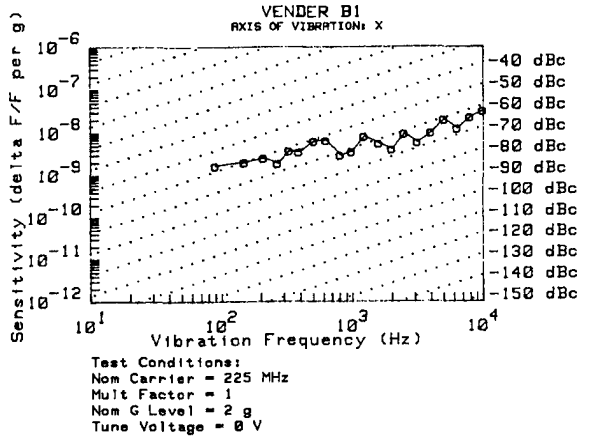


B. Y-Axis

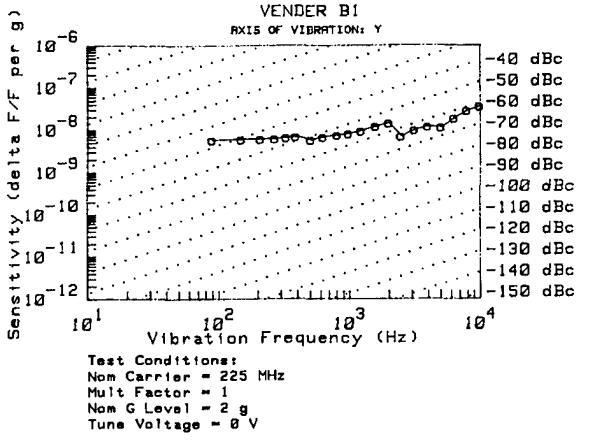


C. Z-Axis

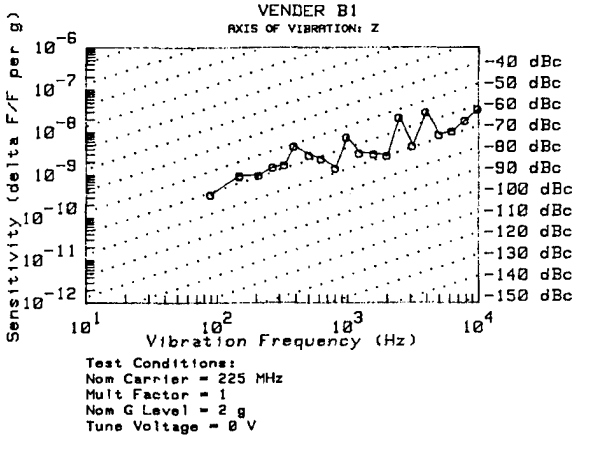
FIGURE 6 ACCELERATION SENSITIVITY FOR VENDER A IN THREE ORTHOGONAL DIRECTIONS



A. X-Axis

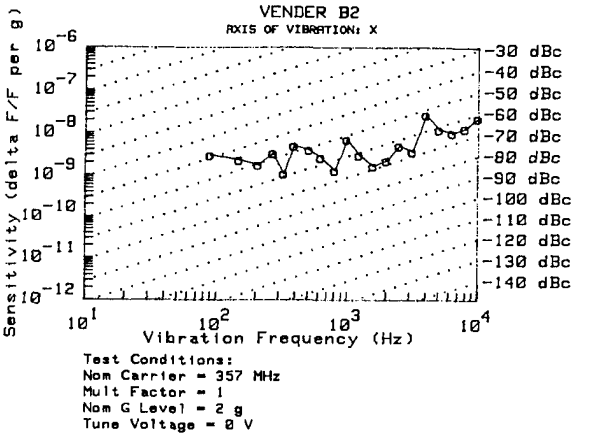


B. Y-Axis

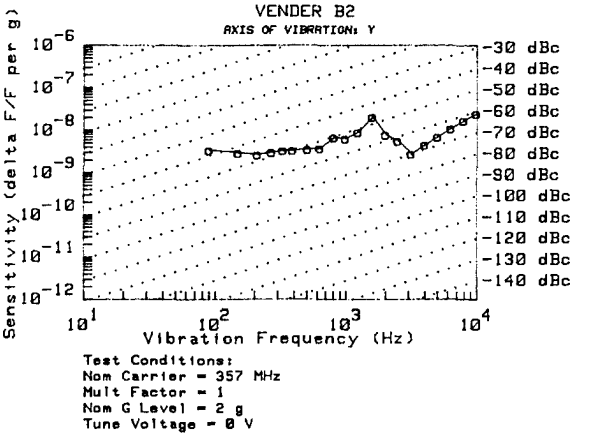


C. Z-Axis

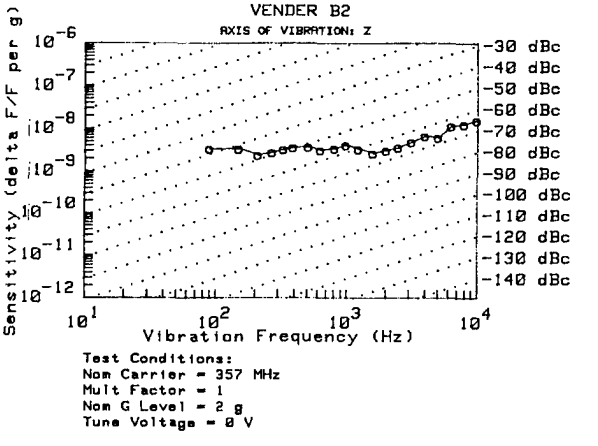
FIGURE 7 ACCELERATION SENSITIVITY FOR VENDER B1 IN THREE ORTHOGONAL DIRECTIONS



A. X-Axis

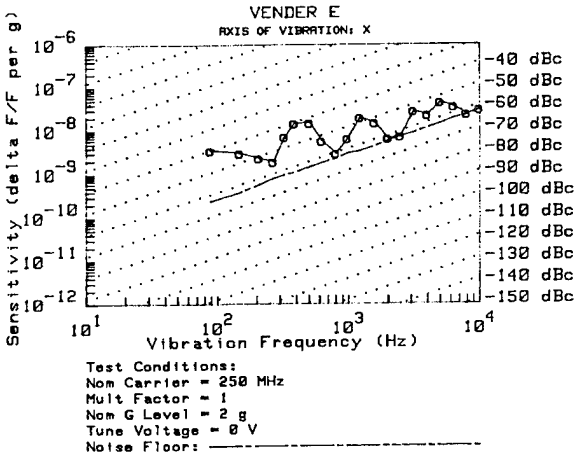


B. Y-Axis

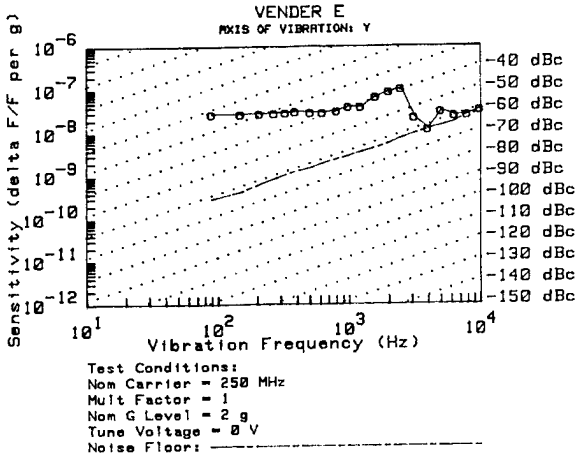


C. Z-Axis

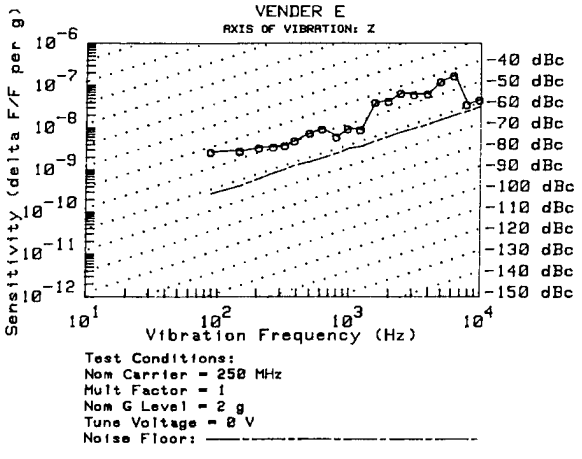
FIGURE 8 ACCELERATION SENSITIVITY FOR VENDER B2 IN THREE ORTHOGONAL DIRECTIONS



A. X-Axis

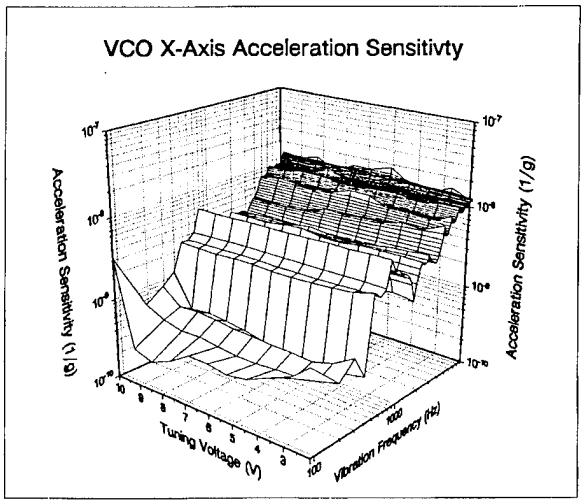


B. Y-Axis

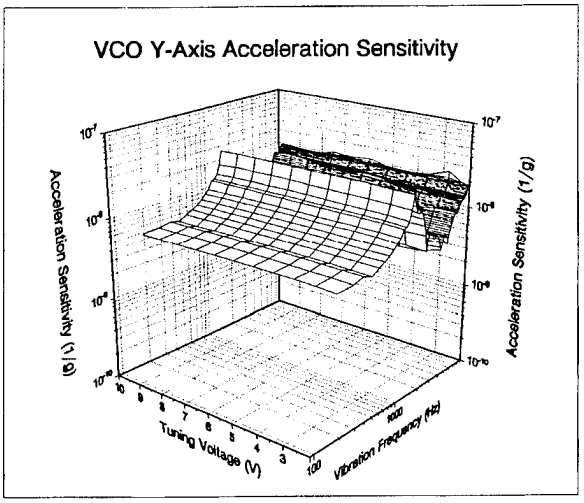


C. Z-Axis

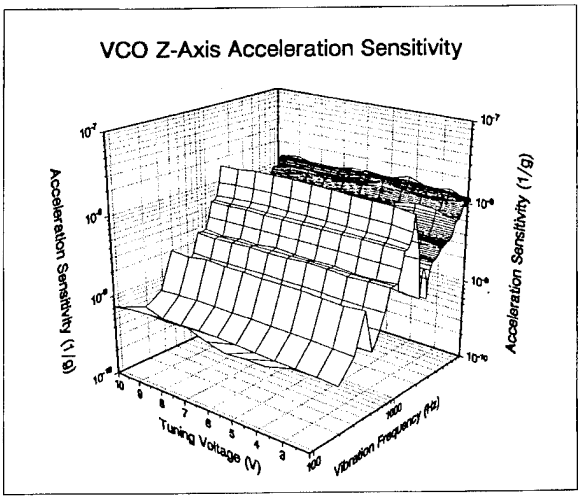
FIGURE 9 ACCELERATION SENSITIVITY FOR VENDER E IN THREE ORTHOGONAL DIRECTIONS



A. X-Axis

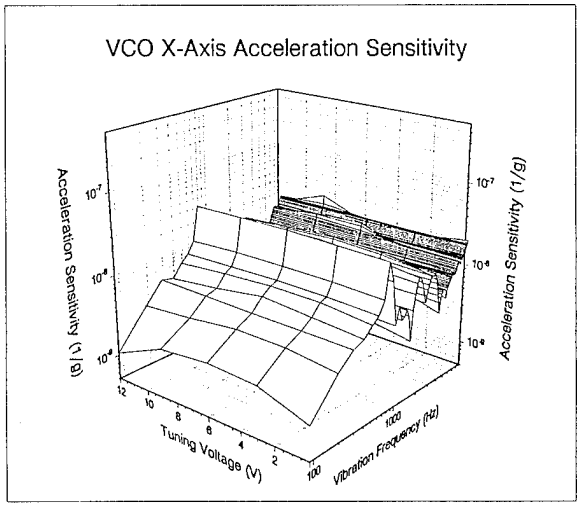


B. Y-Axis

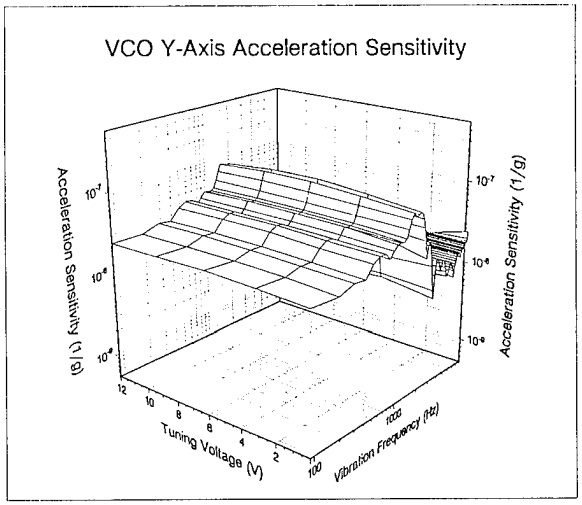


C. Z-Axis

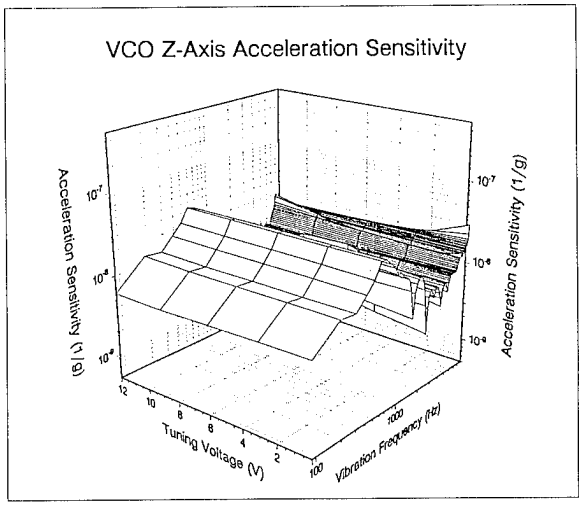
FIGURE 10 ACCELERATION SENSITIVITY FOR VENDER B3 IN THREE ORTHOGONAL DIRECTIONS



A. X-AXIS

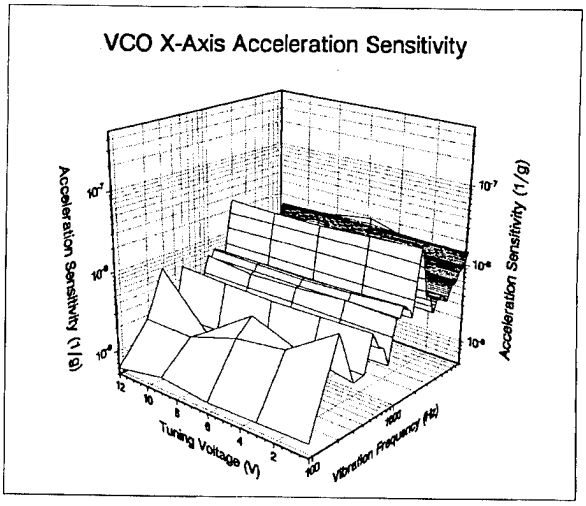


B. Y-AXIS

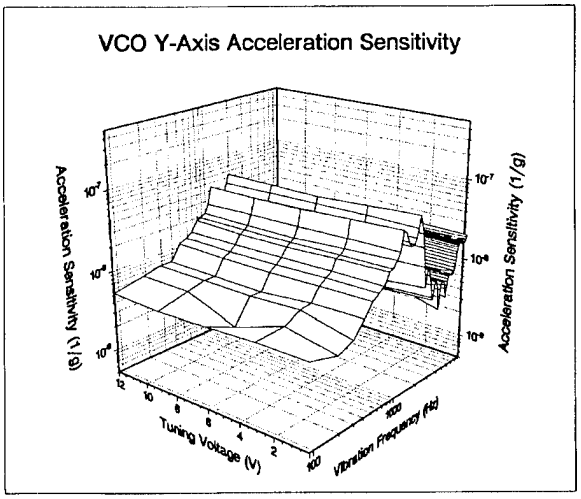


C. Z-AXIS

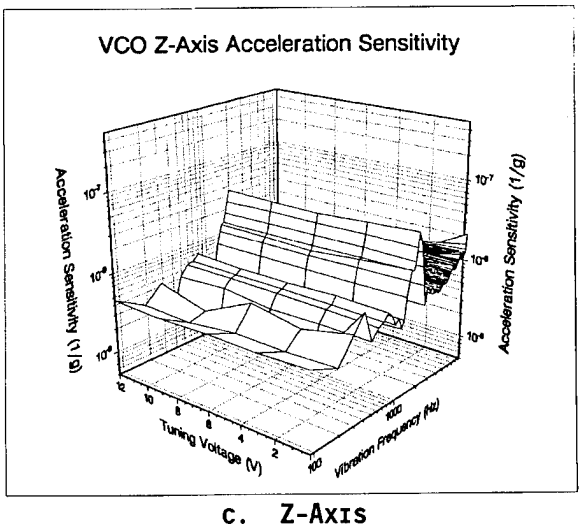
FIGURE 11 ACCELERATION SENSITIVITY FOR VENDER C1 IN THREE ORTHOGONAL DIRECTIONS



A. X-AXIS

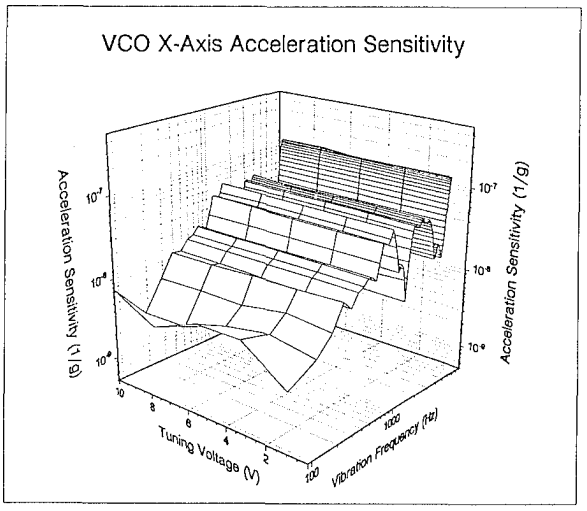


B. Y-AXIS

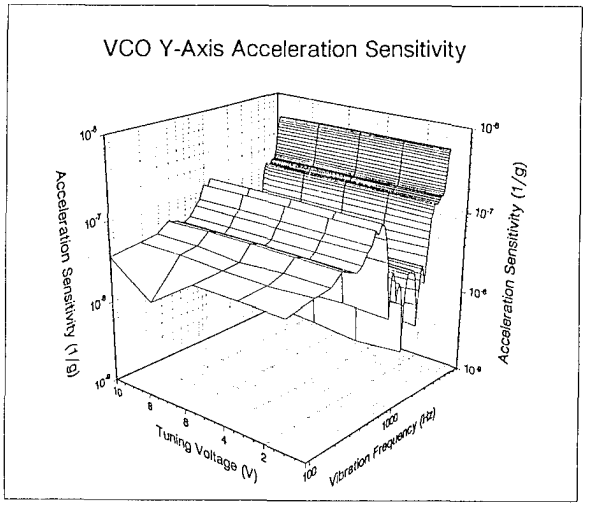


C. Z-AXIS

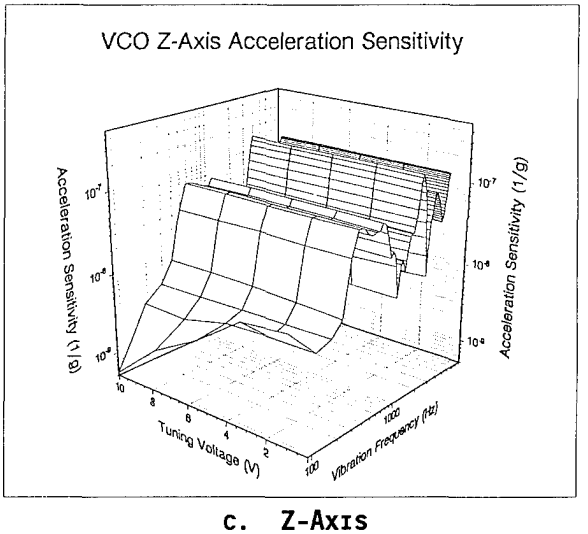
FIGURE 12 ACCELERATION SENSITIVITY FOR VENDER C2 IN THREE ORTHOGONAL DIRECTIONS



A. X-AXIS

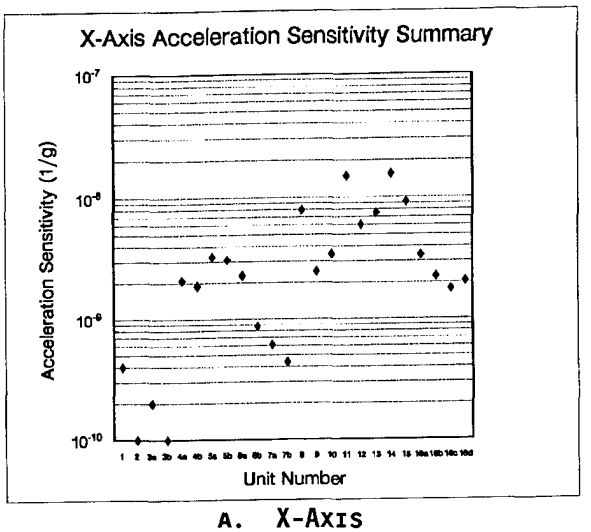


B. Y-AXIS

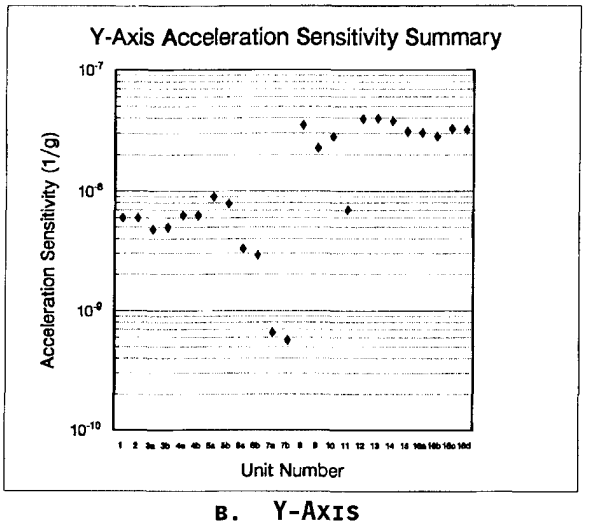


C. Z-AXIS

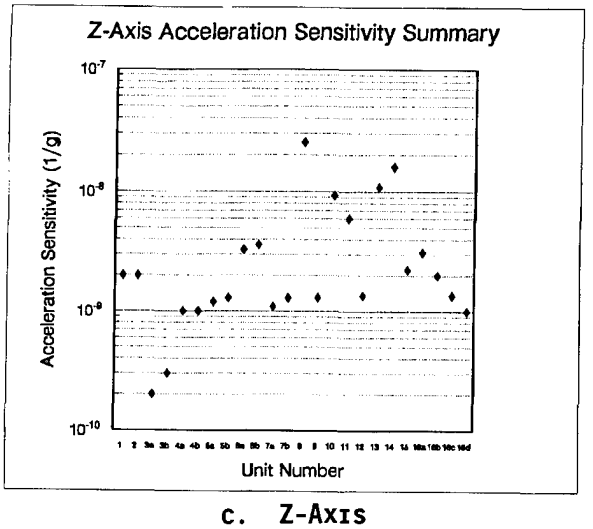
FIGURE 13 ACCELERATION SENSITIVITY FOR VENDER D IN THREE ORTHOGONAL DIRECTIONS



A. X-AXIS



B. Y-AXIS



C. Z-AXIS

FIGURE 14 ACCELERATION SENSITIVITY SUMMARIES FOR THE THREE VECTOR COMPONENTS

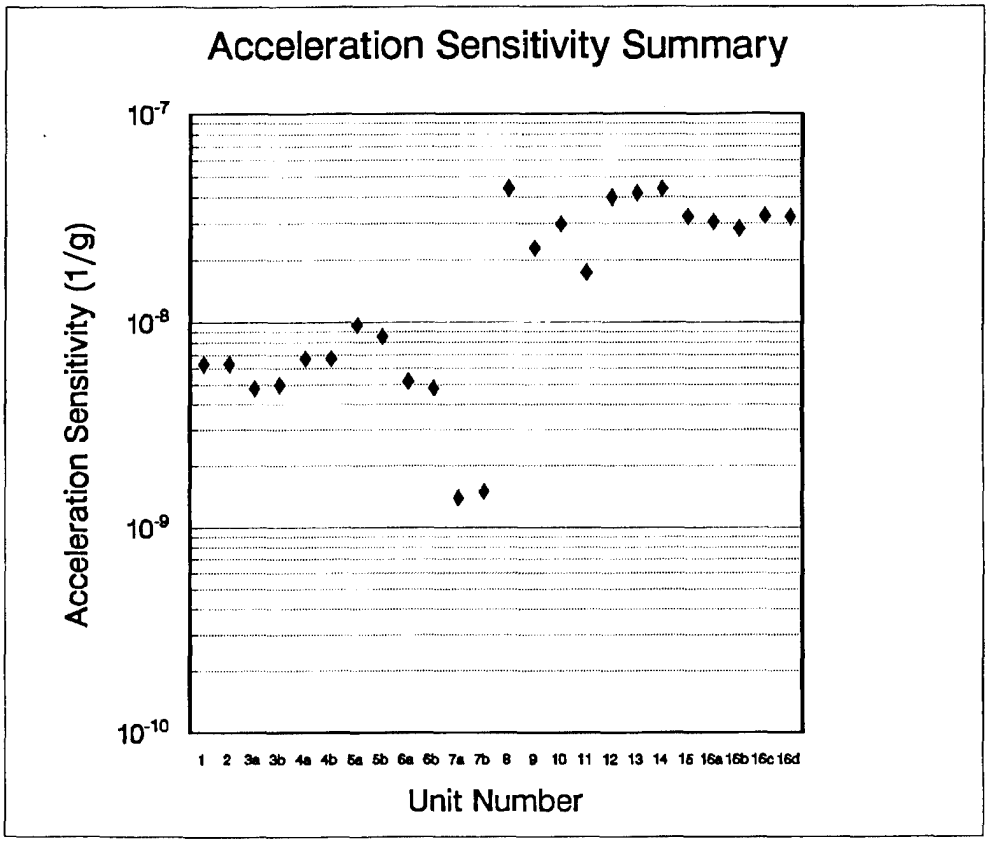


FIGURE 15 SUMMARY OF THE TOTAL ACCELERATION SENSITIVITY VECTOR

TABLE 1 VENDER KEY

UNIT #	VENDER	OSCILLATOR DESCRIPTION
1-2	A	FIXED FREQUENCY; 500 MHz
3A	B	VCSO; 701 MHz; 1x OUTPUT
3B	B	VCSO; 1403 MHz; 2x OUTPUT
4-5	B	FIXED FREQUENCY; 225 MHz COMPLEMENTARY ECL OUTPUT
6-7	B	FIXED FREQUENCY; 357 MHz COMPLEMENTARY ECL OUTPUT
8-9	C	VCSO; 240 MHz
10-11	C	VCSO; 400 MHz
12-15	D	VCSO; 622 MHz
16	E	FIXED FREQUENCY; 250 MHz TWO PAIRS OF DIFFERENTIAL ECL OUTPUTS

TABLE 2 SUMMARY OF ACCELERATION SENSITIVITY

DUT	VENDER	$\lvert\Gamma_x\rvert$	$\lvert\Gamma_y\rvert$	$\lvert\Gamma_z\rvert$	$\lvert\Gamma\rvert$
1	A	≤ 2.00	0.40	6.00	≤ 6.30
2	A	≤ 2.00	≤ 1.00	6.00	≤ 6.40
3A	B3	0.20	0.20*	4.80	4.80
3B	B3	0.30	0.10*	5.00	5.00
4A	B1	2.10	1.00	6.30	6.70
4B	B1	1.90	1.00	6.30	6.70
5A	B1	3.30	1.20	9.00	9.70
5B	B1	3.10	1.30	7.90	8.60
6A	B2	3.30	2.30	3.30	5.20
6B	B2	3.60	0.87	2.90	4.80
7A	B2	1.10	0.61	0.66	1.40
7B	B2	1.30	0.44	0.57	1.50
8	C1	≤ 25.4	≤ 7.97	35.2	≤ 44.1
9	C1	1.31	≤ 2.48	22.7	≤ 22.9
10	C2	≤ 9.31	≤ 3.42*	≤ 28.1*	≤ 29.8
11	C2	≤ 5.99*	≤ 14.9*	≤ 7.00*	≤ 17.5
12	D	≤ 6.02	≤ 1.36*	39.5	≤ 40.0
13	D	≤ 7.51	≤ 10.9	≤ 39.8*	≤ 41.9
14	D	≤ 15.5	≤ 16.0	≤ 37.9*	≤ 44.0
15	D	≤ 9.25	2.23	31.0	≤ 32.4
16A	E	≤ 3.37	3.12	30.2	30.5
16B	E	≤ 2.26	≤ 2.01	28.3	28.5
16C	E	≤ 1.77	≤ 1.36	32.6	32.7
16D	E	≤ 2.05	≤ 1.00	32.2	32.3

* Varied substantially with tuning voltage.

1992 IEEE FREQUENCY CONTROL SYMPOSIUM

INTERNATIONAL STANDARDIZATION IN THE GLOBAL MARKETPLACE

D. Canon Bradley

EG&G Mound Applied Technologies, Inc.
P. O. Box 3000, Miamisburg, OH 45343-3000

Abstract: This paper will address the role of standardization in a changing global market. The European community will economically unite in January of 1993. The International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC) are writing documents setting qualification procedures to establish acceptable uniform quality systems within companies throughout the world. A review and comparison of ISO 9000 and IECQ documents will be made and a status report of their implementation will be given as they relate to the Frequency Control Industry.

I. INTRODUCTION

The International Standardization Organization (ISO) and the International Electrotechnical Commission (IEC) are two of the better known and widely accepted standardization bodies. Although both organizations have their Central Offices in Geneva, Switzerland, the work of writing and adopting the documents is done by professional technical personnel from the countries having membership in either the ISO and/or the IEC organizations. The Central Offices follow the development of all documents seeing that the work is done in agreement with established procedures and publish and maintain the documents after approval of the majority of the participating countries is obtained.

The ISO was organized in London, England, in 1946 and currently has a membership of 90 countries. The ISO's scope of responsibilities includes all technologies. Although ISO has 2,651 technical bodies it does not have active Technical Committees (TC) for all technologies.

The International Electrotechnical Commission's Central Office was organized in St. Louis, Missouri, USA, in 1904 and presently has a membership of 43 countries. The IEC is structured very similar to ISO, having approximately 2,000 technical bodies in the form of 83 Technical Committees, 110 Sub Committees (SC) and many Working Groups (WG) under most sub-committees. In some cases both organizations have committees addressing the same technology. When this occurs a Joint Committee is established having a membership from both organizations to coordinate the documents being written.

Each member country has a National Committee that serves as the coordinating body for their country to the international organizations. Within each country's National Committee, there is a Technical Advisor (TA) for each technology (TC) of interest. The Technical Advisor for the United States National Committee (USNC) for the Piezoelectric and Dielectric Devices for the frequency control and selection industry currently is Charles Adams. He also serves as the chief U. S. delegate at the IEC TC-49 committee meetings.

The purpose of having an international

0-7803-0476-4/92 $3.00 © 1992 IEEE

standardization system is to enhance trade between countries. Thus, the objective of the International Electrotechnical Quality (IECQ) assessment arm is to facilitate international trade in electronic components of assessed quality in accordance with the principle of reciprocity. Electronic components qualified by IEC will be equally acceptable within the participating countries of the system without discrimination. This objective can only be accomplished if the system of testing is accepted by all member countries. This will result in IECQ certified components, made and handled by approved manufacturers and distributors, being used with confidence by buyers worldwide.

The ISO 9000 document is very similar in principle and differs only in that it is less definitive as to the mechanics of the system. An example of one difference is the IECQ system stipulates that one agency per country be assigned the responsibility of designated Inspectorate. One of the responsibilities of the Inspectorate is the administration of the IECQ within the country to insure consistency of application. The ISO 9000 document only defines the responsibilities of the Inspectorate without designating that only one agency will be allowed in each country; thus any number of agencies within a country can self-appoint inspectorates.

Both the ISO 9000 and the IECQ systems are defined by several groups of documents. Within the USA, the United States National Committee (USNC) has adopted the IECQ documents which are listed as part of the NECQ document system. The documents are:

QC001001 - 1986 Basic Rules of the IEC Quality Assessment System for Electronic Components (NECQ)

QC001002 - 1986 Rules of Procedure of the IEC Quality Assessment System for Electronic Components (IECQ)

QC001003 - 1988 Guidance Documents

QC001004 - Specifications List (published February, June and October)

QC001005 - Qualified Products List (published June and December)

Similarly the American Society of Quality Control has incorporated the ISO 9000 documents into their system of National Standards, consisting of:

ISO 9001: Quality Systems - Model for Quality Assurance in Design/Development Production, Installation and Servicing ANSI/ASQC Q91-1987

ISO 9002: Quality Systems - Model for Quality Assurance in Production and Installation ANSI/ASQC Q92-1987

ISO 9003: Quality Systems - Model for Quality Assurance in Final Inspection and Test ANSI/ASQC Q93-1987

ISO 9004: Quality Management and Quality System Elements - Guidelines ANSI/ASQC Q94-1987

Two other documents that are very informative of how the system should function within the company, and the company within the national/international system, are:

ISO/IEC Guide 25: 1990 General requirements for the competence of calibration and testing laboratories.

IEC Guide 102: 1989 Electronic Components. Specification structures

for quality assessment
(Qualification Approval
and Capability Approval)

The ISO and IEC qualification systems will have an impact on the manufacturer, distributor and the testing laboratory. The impact is minor for those agencies and/or companies that are currently complying with the more demanding existing qualification requirements. To comply with the IECQ system, manufacturers must establish a quality system, documented production processes and make independent testing arrangements that meet IECQ requirements and the applicable requirements of ISO 9001 or 9002. The Chief Inspector or Quality Manager will be independently approved by the National Supervising Inspectorate (NSI). The NSI will perform regular component test and conformance inspections.

The manufacturer must provide a certificate of conformity of the product to the requirements of the specifications. These specifications must be under the control of the Chief Inspector or Quality Manager. This will be a change for companies that have specification control responsibility in departments other than the Quality department.

It is the responsibility of the distributor to maintain the traceable records provided to him by the manufacturer and the identity of certified components will be maintained for at least two years. The distributor will issue a certificate of conformity declaring that the storage and handling of the components are in accordance with the systems requirements and are issued under the direct control of the distributor's Chief Inspector or Approved Signatory. If components exceed the specified shelf life, they must be retested by an approval test facility before they can be shipped in compliance with IECQ.

The staff of the testing laboratory must be competent in the fields of technology associated with the component being tested and be adequately equipped with the necessary equipment required to perform the test prescribed in the IECQ specifications. The laboratory is also responsible for maintaining a credible, traceable calibration system to support the integrity of their test results.

There are two required stages of approval in the IECQ system for the manufacturer, and a choice of two types of approval in the second stage of approval. The first stage is being granted approval by NSI for the facilities Quality system to meet all the requirements of IECQ and ISO 9001 and ISO 9002. The second stage is to be granted either a product qualification approval (QA) or a capability approval (CA). A product qualification approval is the approval granted to a manufacturer for an individual component or range of components which meet the assessment requirements given in a Detail Specification which is acceptable within the system. QA is granted on the acceptance of the test report which is representative of normal production and of the design and range of values. A capability approval is the approval granted to a manufacturer whose capability for manufacturing processes and quality control methods, including relevant design aspects, covering a specific component technology, fulfills the requirements of the relevant generic specification. A CA is granted on the acceptance of test results of specified Capability Qualifying Components (CQCs) which must undergo the same processes and inspection procedures that are applicable to normal production.

At the writing of this paper, there are 25 countries (See Appendix I) subscribing to the IECQ system and 29 components covered by IECQ specifications (See Appendix II) including frequency control devices.

The IECQ documents for quartz crystal

resonators are presently being published and the documents on oscillators, filters, and piezoelectric ceramics is under discussion.

The companies that have an interest in complying with the IECQ or ISO-9000 qualification systems are encouraged to contact their country's NSI and individuals active in the process to obtain the most recent and accurate information on compliance. Appendix III lists individuals in the USA active in this qualification system.

ACKNOWLEDGEMENTS & REFERENCES

Robert Mills, ECCB
Charles Zegers, Secretary USNC/IEC
Jack Kinn, VP Engineering EIA
Charles Adams, TA to USNC for IEC TC-49
ISO 9000
IECQ Documents
IECQ A Guide to The Worldwide Electronic Component Certification System
Electronic Business, January 27, 1992

APPENDIX I

Member Countries of IECQ and Their Inspectorates

Australia*	
Belgium	CEB
China	CEPREI
Czechoslowakia	EZU
Denmark	EC
France	SNQ
Germany	VDE
Hungary	MEEI
India	STQC
Israel	SII
Italy	IMQ
Japan	RCJ
Korea (Republic of)	KAITECH
Netherlands *	
Norway	EC
Poland*	
Singapore	RCJ
Spain *	
South Africa	SABS
Sweden	EC
Switzerland	ASE/SEV
United Kingdom	DQA/TS(COMP),MOD
USA	UL
Russian Federation	Russian Federation State Committee for Standards
Yugoslavia*	

* Non Certifying Countries

APPENDIX II

Component Types, Covered by IECQ (June 1991)

1. Attenuators, fibre optic
2. Branching devices, fibre optic
3. Cables, radio frequency (including data cables)
4. Capacitors, fixed
5. Connectors, low frequency, for printed boards
6. Connectors, low frequency, rectangular
7. Connectors, for optical fibres and cables
8. Connectors, radio frequency
9. Inductors and transformers
10. Discrete semiconductor devices (diodes, transistors, and thyristors)
11. Fibre optic mechanical splices and accessories.
12. Filters, ceramic
13. Integrated circuits, film and hybrid film*
14. Integrated circuits, monolithic
15. Potentiometers
16. Printed boards*
17. Quartz crystal units
18. Relays, electromechanical, all-or-nothing
19. Resistors, fixed
20. Resistor networks *
21. Resonators, ceramic

22. Semiconductor opto-electronic and liquid crystal devices
23. Surge protective devices
24. Switches, electro-mechanical
25. Switches, keyboard
26. Thermistors, directly heated, positive step function temperature coefficient
27. Tubes, electronic (including cathrode-ray tubes)
28 Tubes, generator
29 Varistors

* These component types are covered by CAPABILITY APPROVAL.

APPENDIX III

For Additional Information contact:

Mr. Jack Kinn
VP Engineering
EIA
2001 Pennsylvania Avenue, NW
Washington, DC 20006-1813
Phone: 202-457-4900
FAX: 202-457-4985

Mr. Harvey Berman
Underwriters Laboratories
1285 Walt Whitman Road
Melville, NY 11747
Phone: 516-271-6200, extension 229

APPENDIX IV

IECQ/ISO Acronyms (May 1992)

ANSI	American National Standards Insititute
ASQC	American Society of Quality Control
CA	Capability Approval
CECC	European Commission for Component Certification
CENELEC	European Commission for ElectrotechnicalStandardization
CQC	Capability Qualifying Component
ECCB	Electronic Components Certification Board, Inc. manages NECQ
EIA	International Electrotechnical Commission
IEC	International Electrotechnical Commission
IECQ	International Electrotechnical Commission's quality assessment arm
NECQ	National Electronic Components Quality Assessment System
NSI	National Supervising Inspectorate (same as SSI)
NSO	National Standards Organization
OEM	Original Equipment Manufacturer
QA	Qualification Approval
QAS	Quality Assessment System
SC	Sub Committee of TC
SMB	Systems Management Board
SSI	Systems Supervising Inspectorate (same as NSI)
SSO	Systems Standard Organization
SPC	Statistical Process Control
TC	Technical Committee of IEC
TQM	Total Quality Management
UL	Underwriters Laboratories
USNC	United States National Committee
WG	Working Group of a SC or a TC

AUTHOR INDEX

Abdrafikov, S.N. 679
Abramson, I.V. 442,499
Aguasca, A. 492
Allan, D.W. 300,782
Ascarrunz, F.G. 54
Asseev, P.I. 607
Audoin, C. 86,763
Avramov, I.D. 391

Bahadur, H. 657
Ballato, A. 610,838
Barnes, J. 782
Basevich, A.B. 123
Bastien, F. 597
Bauer, S.M. 603
Bava, E. 81
Behagi, A. 207,415
Berends, R.W. 70
Besson, R. 327,798
Bezdelkin, V.V. 707
Bierlein, J.D. 732
Blair, D.G. 167
Bogdanov, P.P. 306
Bootnik, M. 108
Boulanger, J.-S. 6
Bowen, C. 648
Bradley, D.C. 861
Breakiron, L.A. 816
Brendel, R. 485,794
Broquetas, A. 492
Buckingham, M.J. 167
Buissson, X. 696

Camparo, J.C. 104
Candet, E. 561
Capelle, B. 553,639,696
Cavin, M. 384
Chantry, P.J. 114
Chu, D.K.T. 732
Clark, R.F. 76
Clayton, L.D. 582
Cordara, F. 782
Cutler, L.S. 127,188,763,816

Del Casale, A. 265
DeMarchi, A. 81,265
Detaint, J. 553,639,696,744
Detoma, E. 108
Dick, G.J. 58,172,176,225,349
Diener, W.A. 86,225
Dikidzhi, A.N. 499
Djian, F. 485
Douglas, R.J. 6
Drakhlis, B.G. 505
Driscoll, M.M. 157,193,334,763
Drullinger, R.E. 104,183

Eernisse, E.P. 582,763
Eisenhauer, N. 384
El Habti, A. 597
Endres, T.J. 213

Filippov, S.B. 603
Filler, R.L. 231,470,763

Gagnepain, J.J. 284,327,807
Galysh, I. 317
Garvey, R.M. 763,782
Gevorkyan, A.G. 123,306
Giffard, R.P. 127,188
Goiffon, A. 744
Gouzhva, Y.G. 306
Granveaud, M. 807
Greenhall, C.A. 262
Groslambert, J. 201,284
Grouzinenko, V.B. 607,707
Gualtieri, J.G. 724
Guo, W. 151

Hall, J.L. 44
Hammell, R.L. 225
Handel, P.H. 327
Hanson, W. 782
Hardin, R. 238
Hellwig, H. 762,816
Himmel, J. 849
Hofman, W. 683
Houck, G. 648
Howe, D.A. 300
Hruska, C.K. 690
Hunsperger, R.G. 732

Ibanez, A. 744
Ionov, B.P. 425

Johnson, J.L. 143
Joly, C. 639

Kabanovich, I.V. 673
Karlquist, R.K. 134
Kasdin, N.J. 274
Kawashima, H. 434,525
Kinnison, J.D. 798
Kinsman, R. 782
Kirk, A. 225
Kirkpatrick, J.B. 213
Koehler, A.P. 798
Kolodieva, S.V. 667
Koppang, P.A. 92
Kosinski, J. 356,724,838,849
Kosykh, A.V. 425
Kramer, G. 39
Kress, D.W. 356
Kuhnle, P.F. 225
Kusters, J.A. 143,782

Landis, G.P. 317
Lantz, E. 284
Larionov, I.M. 713
Lasewicz, P.T. 465
Lee, D. 32
Lee, P.C.Y. 537,626
Levy, B. 108
Liang, X-P. 163
Liberman, I. 114
Lin, S.T. 724
Lipphardt, B. 39
Logan Jr., R.T. 310,420
Lowe, J.P. 183
Lu, Y. 724
Luiten, A.N. 167
Lukaszek, T. 610,849
Lutes, G.F. 310

MacMillan, C.C. 92
Madej, A.A. 76
Magerl, G. 76
Maleki, L. 58,70,420
Malocha, D.C. 384
Mann, A.G. 167
Marianneau, G. 485
Mateescu, I. 561
Maurer, R.H. 798
McAvoy, B.R. 114
McGowan, R. 849
Medvedev, A.V. 713
Meier, H. 378
Merrell, T.D. 193
Mizan, M. 409
Montress, G.K. 340,356
Moriizumi, T. 610
Murray, M. 724

Nakazawa, M. 532,610
Nelson, C.W. 64

Nelson, L.M. 831
Newnham, R.E. 513
Niiyama, H. 610
Norton, J.R. 465

Olivier, M. 201
Owings, H.B. 92

Parker, T.E. 340,356
Parzen, B. 453
Pedrotto, G. 108
Peters, H.E. 92
Petrosian, I.G. 505
Philippot, E. 639,744
Phillips, W.D. 56
Plana, J.M. 492
Pogrebnyak, A.P. 679
Pond, P.A. 64
Prestage, J.D. 58

Riley, W.J. 763,807
Robert, E. 485
Rolston, S.L. 56
Romanov, L.N. 667
Rubiola, E. 265
Ruderman, W. 724
Rueger, L.J. 465
Rukhlenko, A.S. 371
Russer, P. 378

Sankey, J.D. 76
Santiago, D.G. 176
Schwartzel, J. 553,639,696,744
Seidel, D.J. 70
Sekimoto, H. 532
Semenov, B.N. 603
Siemsen, K.J. 76
Smythe, R.C. 763,782
Ssakharov, S.A. 713
Stern, A. 108
Su, W. 231
Sunaga, K. 434
Suter, J.J. 798
Sydnor, R.L. 225,794
Szekely, C. 104

Takeuchi, M. 610
Tavella, P. 81
Telle, H.R. 27
Tiersten, H.F. 614
Tjoelker, R.L. 58
Toki, M. 448
Toudic, Y. 639,696
Tovstik, P.E. 603
Tsuzuki, Y. 448
Turner, S.D. 207,415

Vanier, J. 807
Vasiliev, A.M. 425
Vernotte, F. 284
Vianco, P.T. 753
Vig, J.R. 470,763,798,816
Volkov, A.A. 505
Vorohkovsky, Y.L. 505,603
Walls, F.L. 2,64,183,290,
327,782,807,831
Walls, W.F. 257
Walter, T. 274
Wang, J. 172,537
Wang, R.T. 58
Watanabe, Y. 532
Weglein, R.D. 763
Weinert, R.W. 157
Weiss, C.O. 39
Weiss, M.A. 300
Whitford, B.G. 76
Wiedemann, W. 794
Wilber, W.D. 724
Williams, A. 70
Winkler, G.M.R. 816
Wischert, W. 201
Wong, N.C. 32

Yang, J.S. 626
Yankowski, M. 238
Yong, Y-K. 567

Zaki, K.A. 163
Zarka, A. 553,639,696,744
Zhang, Z. 567
Zheng, Y. 553,639,696
Zholnerov, V.S. 123
Zhou, W. 270
Zhou, Y.S. 614
Zhu, M. 44
Zomp, J.M. 114

IEEE 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	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.

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