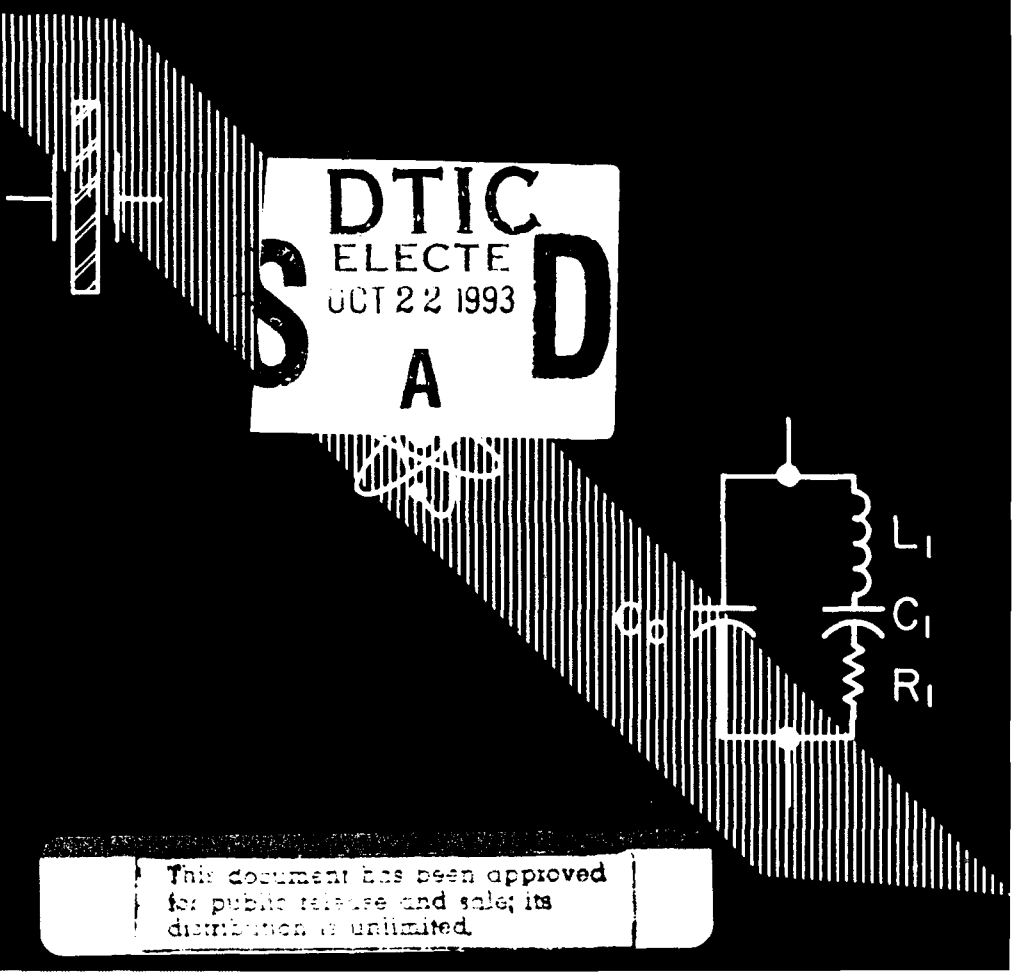
AD-A272 017

Proceedings of the 44th Annual

Symposium on Frequency Control

1990



DTIC
ELECTE
OCT 22 1993
S A D

This document has been approved for public release and sale; its distribution is unlimited.

ADDITIONAL NOTE TO DTIC for CITATIONS for the Proceedings of the Annual Frequency Control Symposia

Suggested Descriptors:

accelerometers
aging
Allan variance
atomic clocks
atomic frequency standards
atomic resonators
cesium
cesium standards
clocks
crystals
etching
filters
frequency control
frequency standards
global positioning system
GPS
hydrogen masers
noise(electrical and electromagnetic)
oscillators
phase noise
quartz
quartz crystals
quartz oscillators
quartz resonators
radiation effects
resonators
rubidium
rubidium standards
sensors
stability
surface acoustic wave devices
synthesizers
time
timekeeping
time transfer
transducers
vibration effects

Abstract: Main headings in the Table(s) of Contents are good to use for Abstract data.

Proceedings of the 44th Annual

Symposium on Frequency Control

1990

Cosponsored by the

U.S. Army
Electronics Technology and
Devices Laboratory

and

The Institute of Electrical and
Electronics Engineers, Inc.

Ultrasonics, Ferroelectrics and
Frequency Control Society

IEEE Catalog No. 90CH2818-3

Library of Congress No. 87-654207

23-25 May 1990

Stouffer Harborplace Hotel
Baltimore, Maryland

Abstracting is permitted with the credit to the source. Libraries are permitted to photocopy beyond the limits of US 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, 21 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for commercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved. Copyright © 1990 by the Institute of Electrical and Electronics Engineers, Inc.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SYMPOSIUM CHAIRMEN

General Chairman	David W. Allan
Technical Program Chairman	Thomas E. Parker
Local Arrangements Chairman	Michael Driscoll
Editorial Chairman	John R. Vig
Publicity Chairman	Raymond L. Filler

TECHNICAL PROGRAM COMMITTEE VICE CHAIRMEN

Marvin Frerking	Rockwell International
Lute Maleki	Jet Propulsion Laboratory
Joel Martin	Oklahoma State University

TECHNICAL PROGRAM COMMITTEE

David Allan, National Institute of Standards & Technology
Arthur Ballato, U.S. Army LABCOM
Martin Bloch, Frequency Electronics, Inc.
Jean-Simon Boulanger, National Research Council of Canada
Jan Brown, Fischer Controls International, Inc.
Mary Chiu, Johns Hopkins University, APL
Leonard Cutler, Hewlett-Packard Company
Michael Driscoll, Westinghouse Systems Center
Errol EerNisse, Quartztronics, Inc.
Raymond Filler, U.S. Army LABCOM
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, National Institute of Standards & Technology
James Pon-Wei Hou, Quartz Frequency Tech., Ltd., Taiwan
Charles Jensik, Piezo Crystal Company
Tom Jerse, Hewlett Packard Company
Gary Johnson, Sawyer Research Products, Inc.
Robert Kinsman, Motorola, Inc.
William Klepczynski, U.S. Naval Observatory
Jack Kusters, Hewlett Packard Company
Theodore Lukaszek, U.S. Army LABCOM
Lute Maleki, Jet Propulsion Laboratory
Donald Malocha, University of Central Florida
Joel Martin, Oklahoma State University
Thrygve Meeker, Consultant
Yuzo Nakagawa, Toyo Communication Equipment Co., Japan
Thomas Parker, Raytheon Research Division
Victor Reinhardt, Hughes Aircraft Company
Gerald Roberts, General Electric Company
Vincent Rosati, U.S. Army LABCOM
Stanley Schodowski, U.S. Army LABCOM
Tadashi Shiosaki, Kyoto University, Japan
Robert Smythe, Piezo Technology, Inc.
Samuel Stein, Ball Corporation
Dan Stevens, AT&T Bell Laboratories
Richard Sydnor, Jet Propulsion Laboratory
John Vig, U.S. Army LABCOM
Rolf Weglein, Hughes Aircraft Company
Joseph White, Naval Research Laboratory
Gernot Winkler, U.S. Naval Observatory
Nicholas Yannoni, U.S. Air Force/RADC

TECHNICAL SESSION CHAIRMEN

MEASUREMENT OF ENVIRONMENTAL SENSITIVITIES OF PRECISION OSCILLATORS
(Panel Discussion)
Helmut Hellwig, Air Force Office of Scientific Research

MODE SHAPES AND ACCELERATION SENSITIVITY
Raymond Filler, US Army LABCOM

STATISTICAL PROCESS CONTROL
Thrygve Meeker, Consultant

RESONATOR PROCESSING
Roger Ward, Quartztronics, Inc.

DISTRIBUTION AND MEASUREMENT
Gernot Winkler, US Naval Observatory

SAW DEVICES AND FILTERS
Donald Malocha, University of Central Florida

ATOMIC FREQUENCY STANDARDS I
Richard Sydnor, Jet Propulsion Laboratory

CRYSTAL OSCILLATORS AND SYNTHESIZERS
Marvin Frerking, Rockwell International

MODE SHAPES AND ACCELERATION SENSITIVITY
Roger Williamson, STC Technology, Ltd.

ATOMIC FREQUENCY STANDARDS II
Mary Chiu, Johns Hopkins University, APL

POSTER SESSION
Dan Stevens, AT&T Bell Labs

THEORY AND DESIGN OF PIEZO RESONATORS
Robert Smytne, Piezo Technology, Inc.

UHF AND MICROWAVE SOURCES
Michael Driscoll, Westinghouse Systems Center

NOBEL SESSION
David Wineland, National Institute of Standards and Technology

THERMAL HYSTERESIS
William Hanson, Piezo Crystal Company

SENSORS AND DEVICES
Charles Adams, Hewlett-Packard Company

TABLE OF CONTENTS

PLENARY SESSION PAGE

Award Presentations 1

Nobel Session 2

Nobel Lectures:

Experiments with Separated Oscillatory Fields and Hydrogen Masers 3
N. Ramsey, Harvard University

Experiments with an Isolated Subatomic Particle at Rest 12
H. Dehmelt, University of Washington

ATOMIC FREQUENCY STANDARDS

Interference Fringes from Single-Cavity Excitation of an Atomic Beam 34
A. DeMarchi, University of Ancona, Italy, & R.E. Drullinger & J.H. Shirley, NIST

Effects of Neutron Fluence on the Operating Characteristics of Diode Lasers Used in Atomic Frequency Standards 39
R.P. Frueholz, J.C. Camparo & S.B. Delcamp, The Aerospace Corporation & C.E. Barnes, California Institute of Technology

Frequency Locking of Laser Diodes Using an Optically Pumped Cesium Beam Tube 44
K. Hisadome & M. Kihara, NTT Transmission Systems Laboratories, Japan

Zero-Crossing Technique for Clock-Transition Detection in a Rb Frequency Standard 53
A. Stern & M. Golosovsky, Time & Frequency Ltd., Israel

Development of a Digitally Controlled Rubidium Atomic Oscillator 59
N. Ishihara, S. Fukugawa, S. Kodama & T. Makabe, NEC Corporation, Japan

Performance of SAO Model VLG-12 Advanced Hydrogen Masers 66
E.M. Mattison & R.F.C. Vessot, Smithsonian Astrophysical Observatory

Optimization of Hydrogen Consumption in Hydrogen Masers 71
B. Jaduszliwer & Y.C. Chan, The Aerospace Corporation

Optically Pumped Primary Frequency Standards **(Invited)** 76
R.E. Drullinger, NIST

Linear Ion Trap Based Atomic Frequency Standard 82
J.D. Prestage, G.J. Dick & L. Maleki, California Institute of Technology

Improved Performance of the Superconducting Cavity Maser at Short Measuring Times 89
R.T. Wang & G.J. Dick, California Institute of Technology

Fine-Structure Artifact of the Velocity Distribution of Cs Beam Tubes as Measured by the Pulsed Microwave Power Technique 94
W.A. Johnson, S.K. Karuza & F.J. Voit, The Aerospace Corporation

Total Process Control and its Implementation in the Manufacturing of Cesium Beam Tubes 98
E.R. Straka, R.W. Sarrica & L. Hansen, Hewlett-Packard Company

FREQUENCY AND TIME DISTRIBUTION AND MEASUREMENT[1]

New Inexpensive Frequency Calibration Service from NIST **(Invited)** 107
D.W. Allan, D.D. Davis, J. Levine, M.A. Weiss, N. Hironaka & D. Okayama, NIST

1. A paper belonging in this section was received after the Proceedings pages had already been numbered. It can be found at page 630.

A Redundant Timing Source for Digital Telecommunication Network Synchronization . . . 117
S.M. Bass, P. Vlitas & R.M. Garvey, Frequency & Time Systems

Frequency and Time Stability of GPS and GLONASS Clocks . . . 127
P. Daly & I.D. Kitching, The University of Leeds, UK and D.W. Allan & T.K. Peppler, NIST

A Novel Approach to the Simultaneous Measurement of Phase and Amplitude Noise of Oscillators . . . 140
K.W. Wan, J. Austin & E. Vilar, Portsmouth Polytechnic Institute, England

Time & Frequency System for Satellites Ground Stations . . . 145
E. Peled, M. Zelitzki, A. Nemesh & A. Stern, Time & Frequency Ltd., Israel

A Rubidium Frequency Standard and a GPS Receiver: A Remotely Steered Clock System with Good Short-Term and Long-Term Stability . . . 151
D.W. Allan & J. Levine, NIST

Radar Stability Self-Test Using Bulk Acoustic Wave Delay Lines . . . 161
P.A. Sorrell, Westinghouse Electric Corporation

THEORY, DESIGN AND PROCESSING OF RESONATORS, FILTERS, SAW DEVICES AND SENSORS

Thermal Hysteresis in Quartz Resonators—A Review . . . 165
J.A. Kusters, Hewlett-Packard Company & J.R. Vig, US Army LABCOM

Thermal Hysteresis in Quartz Crystal Resonators & Oscillators . . . 176
R.L. Filler, US Army LABCOM

Orientational Dependence of "True" SC-Cuts . . . 185
E.P. EerNisse, Quartztronics, Inc. & J.A. Kusters, Hewlett-Packard Company

Determination of the Optimum Orientation of an SC-Cut Resonator Using a Pulsed Laser . . . 193
P.E. Morley, STC Components & R.J. Williamson, STC Technology, England

Contamination Control Methods—A Review of Recent Progress . . . 201
J.R. Vig, US Army LABCOM

Resistivity, Adhesive Strength, and Residual Stress Measurements of Thin Film Metallizations on Single Crystal Quartz . . . 207
P.T. Vianco, W.R. Conley & J.K.G. Panitz, Sandia National Laboratories

A Parametric Study of the Variables Involved in Quartz Growth . . . 216
G.R. Johnson, R.A. Irvine & J.W. Foise, Sawyer Research Products, Inc.

A Study of the Electrodiffusion Process in Quartz . . . 222
D.W. Hart, J. Frank, D. Smith, J.J. Martin, Oklahoma State University & J.G. Gualtieri, US Army LABCOM

Nonuniformities in the Air-Sweeping of Quartz . . . 228
J.Gualtieri, J. Kosinski & R.A. Murray, US Army LABCOM

Plasma Etching of Monocrystal Quartz . . . 238
L. Spassov, E. Yossifov & D. Yankov, Bulgarian Academy of Sciences, Bulgaria & H.-U. Poll & S. Schreiter, Technical University Karl-Marx-Stadt, DDR

Automatic Frequency Control in Chemical Etching of Quartz Crystal Blanks . . . 246
F. Sauerland, Transat Corporation

A Comparison of Laser-Induced Twinning of Quartz Based on Characteristics and Supplier . . . 251
K.L. Blisnuk, J.G. Gualtieri & R.A. Murray, US Army LABCOM

A Quartz Standard Automatic Frequency Adjustment System . . . 259
M. Mourey, S. Galliou, J.P. Valentin & R.J. Besson, École Nationale Superieure de Mecanique et des Micro-techniques, France

Fabrication Considerations for a Coupled-Triple Resonator Crystal . . . 267
G.E. Roberts, A.G. Staples & S. Toliver, Ericsson GE Mobile Communications, Inc.

Usage of a Quartz Resonator as a Piezosorption Sensor for Mercury Vapours in Air . . . 277
L. Spassov, D. Yankov, D. Mihailov & R. Zaharieva, Bulgarian Academy of Science, Bulgaria & A.N. Mogilevski & A.D. Mayorov, Academy of Science of USSR, USSR

Statistical Process Control for Frequency Control **(Invited)** . . . 285
B. Thomas, Rockwell International Corporation

A Synopsis of Quality Involvement/Improvement Programs and the Ramifications on Our Industry . . . 294
C. Jensik, R. Zellers & R. Lackey, Piezo Crystal Company

Raman-Nath Diffraction by Means of Thickness-Shear Oscillating AT-Cut Quartz Crystal Resonator . . . 298
S. Yamagata, Hokkaido University of Education, Japan

Simulation Networks for Piezoelectric Bimorph Actuators . . . 309
J.G. Smits, Boston University & A. Ballato, US Army LABCOM

Narrowband SAW Filters for IF Applications **(Invited)** . . . 316
B. Horine, Sawtek, Inc.

Novel Types of Guided SAW Reflectors for Micro-Resonators . . . 323
G. Golan, G. Griffel, A. Seidman & N. Croitoru, Tel-Aviv University, Israel

FBAR Filters at GHz Frequencies . . . 332
C. Vale, J. Rosenbaum, S. Horitz, S. Krishnaswamy & R. Moore, Westinghouse Electric Corporation

New Designs for Resonators and Filters Using Lithium Tantalate . . . 337
J. Détaint, J. Schwartzel, H. Carru, R. Lefévre & C. Joly, Centre National d'Etudes de Télécommunications &
B. Capelle, Y. Zheng & A. Zarka, L.M.C. Université, France

Linear Phase Filtering . . . 349
B. D'Albaret, L. Bidart & G. Sillioc, Q.K. Piezoéléctronique, France

Two-Dimensional Analysis of Thickness-Shear and Flexural Vibrations in Rectangular AT-Cut Quartz Plates Using a One-Dimensional Finite Element Method . . . 358
H. Sekimoto & Y. Watanabe, Tokyo Metropolitan University & M. Nakazawa, Shinsyu University, Japan

Precise Determination of VHF Quartz Crystal Resonator Equivalent Parameters . . . 363
Y. Oomura & Y. Watanabe, Tokyo Metropolitan University, Japan

A New Low Frequency Thermally Compensated Contour Mode Resonator . . . 367
C. Bourgeois, Swiss Center for Electronics and Microtechnology, Inc., Switzerland

Trapped-Energy Piezoelectric Resonators with Elliptical Ring Electrodes . . . 372
K. Nakamura, Tohoku University, R. Yasuike & K. Hirama, Toyo Communication Equipment Company &
H. Shimizu, The University of Electro-Communications, Japan

Variational Analysis of New Shape Length Extensional Mode Quartz Crystal Resonator Taking Account of Lateral Motion . . . 378
H. Kawashima & M. Nakazato, Seiko Electronic Components, Ltd., Japan

Simulation of Noise Processes in Thickness Shear Resonators Caused by Multilayer Adsorption and Desorption of Surface Molecules . . . 387
Y.K. Yong, Rutgers University

Dynamics of Piezoelectric Laminae Under a Bias . . . 394
M.C. Dökmeci, Istanbul Teknik Universitesi, Istanbul

Piezoelectric Beams and Angular Rate Sensors . . . 406
J. Söderkvist, University of Uppsala, Sweden

QUARTZ RESONATOR MODE SHAPES AND ACCELERATION SENSITIVITY

Mode Shape Analysis Techniques Using Synchroton X-Ray Topography **(Invited)** . . . 416
B. Capelle, A. Zarka, Y. Zheng, Universités P. & M. Curie, & J. Détaint & J. Schwartzel, Centre National d'Études des Télécommunications, France

Non-Destructive Means for Determining Mode Shape in Thickness-Shear Resonators **(Invited)** . . . 424
R.J. Williamson, STC Technology Ltd., England

Adjustment of Resonator G-Sensitivity by Circuit Means . . . 437
R.C. Smythe & W.H. Horton, Piezo Technology Inc.

Electronic Desensitization of Resonators to Accelerations . . . 444
A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan & R. McGowan, US Army LABCOM

On the Influence of a Fabrication Imperfection on the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports . . . 452
Y.S. Zhou & H.F. Tiersten, Rensselaer Polytechnic Institute

An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators with Rectangular Supports . . . 461
H.F. Tiersten & Y.S. Zhou, Rensselaer Polytechnic Institute

Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes . . . 468
P.C.Y. Lee & X. Guo, Princeton University

Performance Characteristics of a 10 MHz VCO with a Dual BVA Resonator . . . 474
M. Calhoun, P. Kuhnle & R. Sydnor, California Institute of Technology

A New Factor Affecting the Acceleration Sensitivity of the Resonance Frequency of Quartz Crystal Resonators . . . 478
W.P. Hanson, L.C. Heishman, Piezo Crystal Company & T.R. Meeker, Consultant

Measurements of Acceleration-Induced Phase Noise in Surface Acoustic Wave Devices . . . 488
J. Kosinski, A. Ballato & T. Lukaszek, US Army LABCOM

Phase Noise Measurements on Vibrating Crystal Filters . . . 493
R.L. Clark, Pennsylvania State University

CRYSTAL OSCILLATORS, SYNTHESIZERS, UHF AND MICROWAVE SOURCES

Spectral Purity of Direct Digital Frequency Synthesizers **(Invited)** . . . 498
V.F. Kroupa, Czechoslovak Academy of Sciences, Czechoslovakia

An Exact Spectral Analysis of a Number Controlled Oscillator Based Synthesizer . . . 511
J.F. Garvey & D. Babitch, California Microwave, Inc.

Design Techniques for Achieving State-of-the-Art Oscillator Performance **(Invited)** . . . 522
G.K. Montress & T.E. Parker, Raytheon Research Division

Universal, Computer Facilitated, Steady State Oscillator Analysis Theory . . . 536
B. Parzen, Consulting Engineer

High Spectral Purity X-Band Source . . . 542
F.L. Walls & C.M. Felton, NIST & T.D. Martin, Gravity Research Institute

Low Phase Noise Dielectric Resonator Oscillator . . . 549
R. Jones & V. Estrick, Hughes Aircraft Company

High-Order Harmonic Mixing with GaAs Schottky Diodes . . . 555
M. Prevedelli, Universita di Firenze, Italy & F.L. Walls & S.P. Beaton, NIST

A Multiple Modulator Fractional Divider . . . 559
B. Miller & B. Conley, Hewlett Packard Company

A Novel Design of a 1.8 GHz Input Odd Ratio Frequency Divider . . . 569
D. Korn, Naval Research Laboratory & C. Deierling, Select Electronics

Ultra-Low Noise Microwave Phase Stabilizer Using Sapphire Ring Resonator . . . 577
G.J. Dick, J. Saunders & T. Tucker, California Institute of Technology

An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using a Coupling Quartz Crystal Resonator . . . 585
H. Kawashima, Seiko Electronic Components, Ltd., Japan

A B.V.A. Quartz Crystal Oscillator for Severe Environments . . . 593
R.J. Besson & M. Mourey, École Nationale Supérieure de Mécanique et des Microtechniques, France

Factors Influencing Stability in the Microprocessor-Compensated Crystal Oscillator . . . 597
A. Benjaminson, General Technical Services, Inc.

An Update on the TMXO . . . 615
D.A. Symonds & D. Allen, Piezo Technology Inc.

Superimposing Low-Phase-Noise, Low-Drift Instrumentation Techniques on RF Design . . . 622
C.M. Felton, NIST

LATE SUBMISSION ON "FREQUENCY AND TIME DISTRIBUTION AND MEASUREMENT SESSION" (see page iv.)

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) 630
S.R. Stein, Ball Communication Systems Division & J. Evans, Defense Communications Engineering Center

AUTHOR INDEX 636

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL 637

PROCEEDINGS AVAILABILITY INFORMATION 640

1990 AWARD WINNERS

The Cady Award

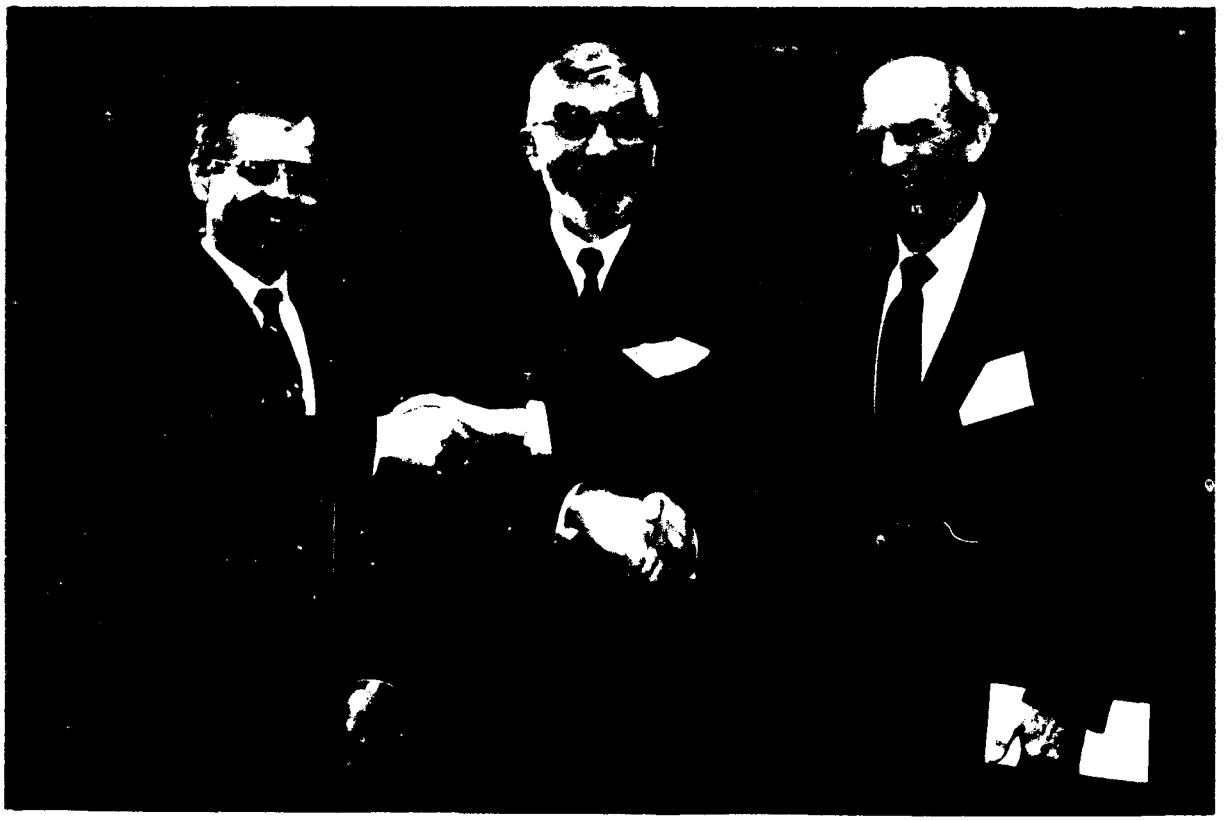
The Cady Award was presented to John R. Vig, U. S. Army Electronics Technology and Devices Laboratory, "for outstanding contributions to the development of improved quartz crystals and processing techniques, significantly advancing the field of precision frequency control and timing." The award was presented by John A. Kusters, Hewlett Packard Co.

The Raby Award

The Rabi Award was presented to Claude Audoin "for original contributions to the theoretical and experimental foundations of microwave frequency standards and their metrology." The award was presented by Jacques Vanier, National Research Council, Canada.

The Sawyer Award

The Sawyer Award was presented to William H. Horton, Piezo Technology, Inc., "For technical and industrial leadership in the development and manufacture of quartz crystals, filters, and oscillators." The award was presented by Gary R. Johnson, Sawyer Research Products, Inc.



John R. Vig, Cady Award winner; Claude Audoin, Rabi Award winner; and William H. Horton, Sawyer Award winner; after the award presentations.

NOBEL SESSION

The 44th Annual Symposium on Frequency Control featured a special session honoring the recipients of the 1989 Nobel Prize in physics: Professors Norman F. Ramsey of Harvard University, Hans G. Dehmelt of the University of Washington, and Wolfgang Paul of the University of Bonn. Two of the three laureates presented papers. (Prof. Paul was unable to attend the Symposium.) Prof. Ramsey gave an historically rooted account of experiments with separated oscillatory fields and hydrogen masers, and Prof. Dehmelt provided a visionary discussion of experiments with an isolated subatomic particle at rest. This session was a highlight of the Symposium. It offered a collegial interaction among those who built the foundations for today's atomic clocks and those who continue the research toward tomorrow's exotic quantum electronic frequency and time standards, which will probably rely on single particles at rest, in free space, unperturbed by collisions or fields.

Professors Norman F. Ramsey and Hans G. Dehmelt at the Nobel Session held in their honor.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

EXPERIMENTS WITH SEPARATED OSCILLATORY FIELDS AND HYDROGEN MASERS

Norman F.Ramsey

Lyman Laboratory of Physics, Harvard University
Cambridge, Massachusetts, USA

Abstract

Descriptions are given of the methods of separated and successive oscillatory fields and of the atomic hydrogen maser. Their applications to precision spectroscopy of atoms and molecules are discussed along with specific examples of fundamental measurements best made with such devices. The applications of these methods to atomic clocks are discussed. A discussion is also given of fundamental experiments which require highly stable clocks.

The Method of Successive Oscillatory Fields

In the summer of 1937 following two years at Cambridge University, I went to Columbia University to work with I. I. Rabi. After I had been there only a few months, Rabi invented[1-4] the molecular beam magnetic resonance method so I had the great good fortune to be the only graduate student to work with Rabi and his colleagues[1-2] on one of the first two experiments to develop and utilize magnetic resonance spectroscopy, for which Rabi received the 1944 Nobel Prize in Physics.

By 1949, I had moved to Harvard University and was looking for a way to make more accurate measurements than were possible with the Rabi method and in so doing I invented the method of separated oscillatory fields.[3-6]. In this method the single oscillatory magnetic field in the center of the Rabi device is replaced by two oscillatory fields, one at the entrance and one at the exit of the space in which the properties of the atoms or molecules are studied. As I will discuss, the separated oscillatory fields method has many advantages over the single oscillatory field method and in subsequent years it has been extended to many experiments beyond those of molecular beam magnetic resonance.

Let me now review the successive oscillatory field method, particularly in its original and easiest to explain application -- the measurement of nuclear magnetic moments. The extension to more general cases is then straightforward.

The method was initially an improvement on Rabi's resonance method for measuring nuclear magnetic moments, whose principles are illustrated schematically in Figure 1. Consider a classical nucleus with spin angular momentum $\hbar\mathbf{J}$ and magnetic moment $\mu = (\mu/J)\mathbf{J}$. Then in a static magnetic field $\mathbf{H}_0 = H_0\,\mathbf{k}$, the nucleus, due to the torque on the nuclear angular momentum, will precess like a top about $\mathbf{H}_0$ with the Larmor frequency ν_0 and angular frequency ω_0 given by

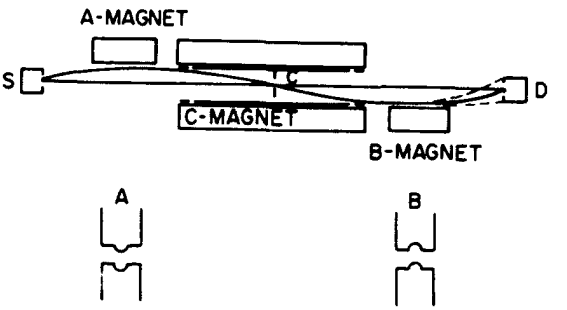


Figure 1. -- Schematic diagram of a molecular beam magnetic resonance apparatus. A typical molecule which can be detected emerges from the source, is deflected by the inhomogeneous magnetic field A. passes through the collimator and is deflected to the detector by the inhomogeneous magnetic field B. If, however, the oscillatory field in the C region induces a change in the molecular state, the B magnet will provide a different deflection and the beam will follow the dashed lines with a corresponding reduction in detected intensity. In the Rabi method, the oscillatory field is applied uniformly throughout the C region as indicated by the long rf lines F, whereas in the separated oscillatory field method the rf is applied only in the regions E and G.

$$\omega_0 = 2\pi\nu_0 = \frac{\mu H_0}{\hbar I} \qquad (1)$$

as shown in Figure 2. Consider an additional magnetic field

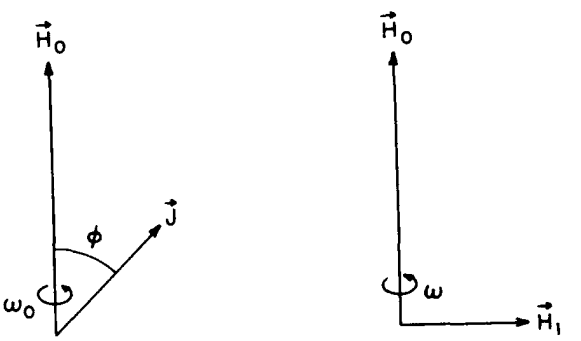


Figure 2. -- Precession of the nuclear angular momentum **J** (left) and the rotating magnetic field $\mathbf{H}_1$ (right) in the Rabi method.

CH2818-3/90/0000-003 $1.00 © 1990 IEEE

$\mathbf{H}_1$ perpendicular to H_0 and rotating about it with angular frequency ω. Then, if at any time $\mathbf{H}_1$ is perpendicular to the plane of $\mathbf{H}_0$ and $\mathbf{J}$, it will remain perpendicular to it provided $\omega = \omega_0$. In that case, in a coordinate system rotating with $\mathbf{H}_1$, $\mathbf{J}$ will precess about $\mathbf{H}_1$ and the angle ϕ will continuously change in a fashion analogous to the motion of a "sleeping top"; the change of orientation can be detected by allowing the molecular beam containing the magnetic moments to pass through inhomogeneous fields as in Figure 1. If ω is not equal to ω_0, $\mathbf{H}_1$ will not remain perpendicular to $\mathbf{J}$; so ϕ will increase for a short while and then decrease, leading to no net change. In this fashion the Larmor precession frequency ω_0 can be detected by measuring the oscillator frequency ω at which there is maximum reorientation of the angular momentum and hence a maximum change in beam intensity for an apparatus as in Figure 1. This procedure is the basis of the Rabi molecular beam resonance method.

The separated oscillatory field method in this application is much the same except that the rotating field $\mathbf{H}_1$ seen by the nucleus is applied initially for a short time τ, the amplitude of $\mathbf{H}_1$ is then reduced to zero for a relatively long time T and then increased to $\mathbf{H}_1$ for a time τ, with phase coherency being preserved for the oscillating fields as shown in Figure 3. This can be done, for example, in the molecular-

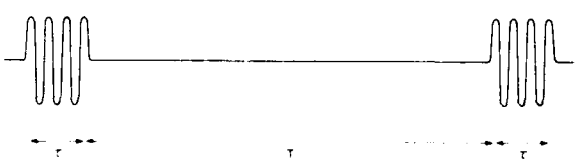


Figure 3. --Two separated oscillatory fields, each acting for a time τ, with zero amplitude oscillating field acting for time T. Phase coherency is preserved between the two oscillatory fields so it is as if the oscillation continued, but with zero amplitude for time T.

beam apparatus of Figure 1 in which the molecules first pass through a rotating field region, then a region with no rotating field and finally a region with a second rotating field driven phase coherently by the same oscillator.

If the nuclear spin angular momentum is initially parallel to the fixed field (so that ϕ is equal to zero initially) it is possible to select the magnitude of the rotating field so that ϕ is 90^0 or $\pi/2$ radians at the end of the first oscillating region. While in the region with no oscillating field, the magnetic moment simply precesses with the Larmor frequency appropriate to the magnetic field in that region. When the magnetic moment enters the second oscillating field region there is again a torque acting to change ϕ. If the frequency of the rotating field is exactly the same as the mean Larmor frequency in the intermediate region there is no relative phase shift between the angular momentum and the rotating field.

Consequently, if the magnitude of the second rotating field and the length of time of its application are equal to those of the first region, the second rotating field has just the same effect as the first one -- that is, it increases ϕ by another $\pi/2$, making $\phi = \pi$, corresponding to a complete reversal of the direction of the angular momentum. On the other hand, if the field and the Larmor frequencies are slightly different, so that the relative phase angle between the rotating field vector and the precessing angular momentum is changed by π while the system is passing through the intermediate region, the second oscillating field has just the opposite effect to the first one; the result is that ϕ is returned to zero. If the Larmor frequency and the rotating field frequency differ by an amount such that the relative phase shift in the intermediate region is exactly an integral multiple of 2π, ϕ will again be left at π just as at exact resonance.

In other words if all molecules had the same velocity, the transition probability would be periodic as in Figure 4.

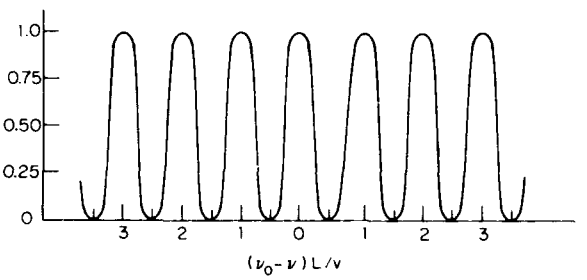


Figure 4. -- Transition probability as a function of the frequency $\nu = \omega/2\pi$ that would be observed in a separated oscillatory field experiment if all the molecules in the beam had a single velocity.

However, in a molecular beam resonance experiment one can easily distinguish between exact resonance and the other cases. In the case of exact resonance, the condition for no change in the relative phase of the rotating field and of the precessing angular momentum is independent of the molecular velocity. In the other cases, however, the condition for integral multiple of 2π relative phase shift is velocity dependent, because a slower molecule is in the intermediate region longer and so experiences a greater shift than a faster molecule. Consequently, for the non-resonance peaks, the reorientations of most molecules are incomplete so the magnitudes of the non-resonance peaks are smaller than at exact resonance and one expects a resonance curve similar to that shown in Figure 5, in which the transition probability for a particle of spin 1/2 is plotted as a function of frequency.

Although the above description of the method is primarily in terms of classical spins and magnetic moments, the method applies to any quantum mechanical system for which a transition can be induced between two energy states W_i and W_f which are differently focussed. The resonance frequency ω_0 is then given by

$$\omega_0 = (W_i - W_f)/\hbar \qquad (2)$$

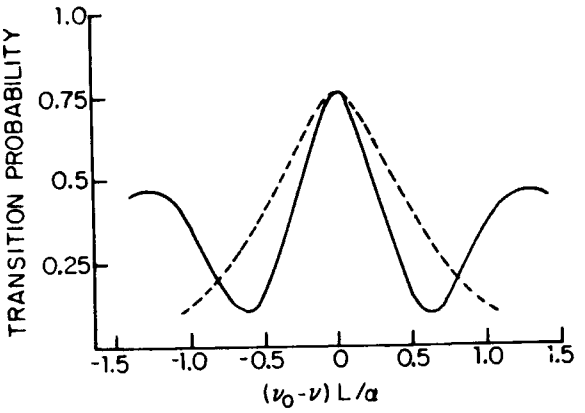


Figure 5. -- When the molecules have a Maxwellian velocity distribution, the transition probability is as shown by the full line for optimum rotating field amplitude. (L is the distance between oscillating field regions, a is the most probable molecular velocity and ν is the oscillatory frequency = $\omega./2\pi$) The dashed line represents the transition probability with the single oscillating field method when the total duration is the same as the time between separated oscillatory field pulses.

and one expects a resonance curve similar to that shown in Figure 5, in which the transition probability for a particle of spin 1/2 is plotted as a function of frequency.

From a quantum-mechanical point of view, the oscillating character of the transition probability in Figures 4 and 5 is the result of the cross term in the calculation of the transition probability from probability amplitudes. Let C_{iif} be the probability amplitude for the nucleus to pass through the first oscillatory field region with the initial state i unchanged but for there to be a transition to state f in the final field, whereas C_{iff} is the amplitude for the alternative path with the transition to the final state f being in the first field with no change in the second. The cross term $C_{iif}C^*_{iff}$ produces an interference pattern and gives the narrow oscillatory pattern of the transition probability shown in the curves of Figures 4 and 5. Alternatively the pattern can in part be interpreted as resulting from the Fourier spectrum of an oscillating field which is on for a time τ off for T and on again for τ , as in Figure 3,. However, the Fourier interpretation is not fully valid since with finite rotations of $\mathbf{J}$, the problem is a non-linear one. Furthermore, the Fourier interpretation obscures some of the key advantages of the separated oscillatory field method. I have calculated the quantum mechanical transition probabilities[3,5,7,8] and these calculations provide the basis for Figure 5.

The separated oscillatory field method has a number of advantages including the following:

(1) The resonance peaks are only 0.6 as broad as the corresponding ones with the single oscillatory field method. The narrowing is somewhat analogous to the peaks in a two slit optical interference pattern being narrower than the central diffraction peak of a single wide slit whose width is equal to the separation of the two slits.

(2) The sharpness of the resonance is not reduced by non-uniformities of the constant field since both from the qualitative description and from the theoretical quantum analysis, it is only the space average value of the energies along the path that enter Eq. (2) and are important.

(3) The method is more effective and often essential at very high frequencies where the wave length of the radiation used may be comparable to or smaller than the length of the region in which the energy levels are studied.

(4) Provided there is no unintended phase shift between the two oscillatory fields, first order Doppler shifts and widths are eliminated.

(5) The method can be applied to study energy levels in a region into which an oscillating field can not be introduced; for example, the Larmor precession frequency of neutrons can be measured while they are inside a magnetized iron block.

(6) The lines can be narrowed by reducing the amplitude of the rotating field below the optimum. The narrowing is the result of the low amplitude favoring slower than average molecules.

(7) If the atomic state being studied decays spontaneously, the separated oscillatory field method permits the observation of narrower resonances than those anticipated from the lifetime and the Heisenberg uncertainty principle provided the two separated oscillatory fields are sufficiently far apart; only states that survive sufficiently long to reach the second oscillatory field can contribute to the resonance. This method, for example, has been used by Lundeen and others[9] in precise studies of the Lamb shift.

The advantages of the separated oscillatory field method have led to its extensive use in molecular and atomic beam spectroscopy. One of the best known is in atomic cesium standards of frequency and time which will be discussed later.

Although in most respects, the separated oscillatory field method offers advantages over a single oscillatory field, there are sometimes disadvantages. In studying complicated overlapping spectra the subsidiary maxima of Figure 5 can cause confusion. Furthermore, it is sometimes difficult at the required frequency to obtain sufficient oscillatory field strengths with two short oscillatory fields, whereas adequate field strength may be achieved with a weaker, longer oscillatory field. Therefore for most molecular beam resonance experiments, it is best to have both separated oscillatory fields and a single long oscillatory field available so the most suitable method under the circumstances can be used.

As in any high precision experiment, care must be exercised with the separated oscillatory field method to avoid obtaining misleading results. Ordinarily these potential distortions are more easily understood and eliminated with the separated oscillatory field method than are their counterparts in most other high-precision spectroscopy. Nevertheless, the effects are important and require care in high-precision measurements. I have discussed the various effects in detail elsewhere[3,7,8,10] but I will briefly summarize them here.

Variations in the amplitudes of the oscillating fields from their optimum values may markedly change the shape of the resonance, including the replacement of a maximum transition probability by a minimum. However, symmetry about the exact resonance frequency is preserved, so no measurement error need be introduced by such amplitude variations.[7,8]

Variations of the magnitude of the fixed field between, but not in, the oscillatory field regions do not ordinarily distort a molecular beam resonance provided the average transition frequency (Bohr frequency) between the two fields equals the values of the transition frequencies in each of the two oscillatory field regions alone. If this condition is not met, there can be some shift in the resonance frequency.[7,8]

If, in addition to the two energy levels between which transitions are studied, there are other energy levels partially excited by the oscillatory field, there will be a pulling of the resonance frequency as in any spectroscopic study and as analyzed in detail in the literature.[3,7,8]

Even in the case when only two energy levels are involved, the application of additional rotating magnetic fields at frequencies other than the resonance frequency will produce a net shift in the observed resonance frequency, as discussed elsewhere.[3,7,8] A particularly important special case is the effect identified by Bloch and Siegert,[11] which occurs when oscillatory rather than rotating magnetic fields are used. Since an oscillatory field can be decomposed into two oppositely rotating fields, the counter-rotating field component automatically acts as such an extraneous rotating field. Another example of an extraneously introduced rotating field is that which results from the motion of an atom through a field H_0 whose direction varies in the region traversed. The theory of the effects of additional rotating fields at arbitrary frequencies has been developed by Ramsey[7,8,10,12] Winter,[10] Shirley,[13] Code,[12] and Greene.[14]

Unintended relative phase shifts between the two oscillatory field regions will produce a shift in the observed resonance frequency.[13,14,15] This is the most common source of possible error, and care must be taken to avoid it either by eliminating such a phase shift or by determining the shift -- say by measurements with the molecular beam passing through the apparatus first in one direction and then in the opposite direction.

A number of extensions to the separated oscillatory field method have been made since its original introduction:

(1) It is often convenient to introduce phase shifts deliberately to modify the resonance shape.[15] As discussed above, unintended phase shifts can cause distortions of the observed resonance, but some distortions are useful. Thus, if the change in transition probability is observed when the relative phase is shifted from $+\pi/2$ to $-\pi/2$ one sees a dispersion curve shape[15]. A resonance with dispersion shape provides greater sensitivity for detecting small shifts in the resonance frequency.

(2) For most purposes the highest precision can be obtained with just two oscillatory fields separated by the maximum time, but in some cases it is better to use more than two separated oscillatory fields.[4] I have calculated the theoretical resonance shapes[7] with two, three, four and infinitely many oscillatory fields. The infinitely many oscillatory field case, of course, by definition becomes the same as the single long oscillatory field if the total length of the transition region is kept the same and the infinitely many oscillatory fields fill in the transition region continuously. For many purposes this is the best way to think of the single oscillatory field method, and this point of view makes it apparent that the single oscillatory field method is subjected to complicated versions of all the distortions discussed in the previous section. It is noteworthy that, as the number of oscillatory field regions is increased for the same total length of apparatus, the resonance width is broadened; the narrowest resonance is obtained with just two oscillatory fields separated the maximum distance apart. Despite this advantage, there are valid circumstances for using more than two oscillatory fields. With three oscillatory fields the first and largest side lobe is suppressed, which may help in resolving two nearby resonances; for a larger number of oscillatory fields additional side lobes are suppressed, and in the limiting case of a single oscillatory field there are no side lobes. Another reason for using a large number of successive pulses can be the impossibility of obtaining sufficient power in a single pulse to induce adequate transition probability with a small number of pulses.

(3) The earliest use of the separated oscillatory field method involved two oscillatory fields separated in space, but it was early realized that the method with modest modifications could be generalized to a method of successive oscillatory fields with the separation being in time, say by the use of coherent pulses.[16]

(4) If more than two successive oscillatory fields are utilized it is not necessary to the success of the method that they be equally space in time;[4] the only requirement is that the oscillating fields be coherent -- as is the case if the oscillatory fields are all derived from a single continuously running oscillator. In particular, the separation of the pulses can even be random,[16] as in the case of the large box hydrogen maser[17] discussed later. The atoms being stimulated to emit move randomly into and out of the cavities with oscillatory fields and spend the intermediate time in the large container with no such fields.

(5) The full generalization of the successive oscillatory field method is excitation by one or more oscillatory fields that vary arbitrarily with time in both amplitude and phase.[7,8]

(6) V. F. Ezhov and his colleagues,[6,18] in a neutron-beam experiment, used an inhomogeneous static field in the region of each oscillatory field region such that initially when the oscillatory field is applied conditions are far from resonance. Then, when the resonance condition is slowly approached, the magnetic moment that was originally aligned parallel to $\mathbf{H}_0$ will adiabatically follow the effective magnetic field on a coordinate system rotating with $\mathbf{H}_1$ until at the end of the first oscillatory field region the moment is parallel to $\mathbf{H}_1$. This arrangement has the theoretical advantage that the maximum transition probability can be unity even with a velocity distribution, but the method may be less well adapted to the study of complicated spectra.

(7) I emphasized earlier that one of the principal sources of error in the separated oscillatory field method is that which arises form uncertainty in the exact value of the relative phase shift in the two oscillatory fields. Jarvis, *et al.* [19] have pointed out that this problem can be overcome with a slight loss in resolution by driving the two cavities at slightly different frequencies so that there is a continual change in the relative phase. In this case the observed resonance pattern will change continuously from absorption to dispersion shape. The envelope of these patterns, however, can be observed and the position of the maximum of the envelope is unaffected by relative phase shifts. Since the envelope is about twice the width of a specific resonance there is some loss of resolution

in this method, but in certain cases this loss may be outweighed by the freedom from phase-shift errors.

(8) The method has been extended to electric as well as magnetic transitions and to optical laser frequencies as well as radio- and microwave-frequencies. The application of the separated oscillatory field method to optical frequencies requires considerable modifications because of the short wave lengths, as pointed out by Blaklanov, Dubetsky and Chebotsev[20]. Successful applications of the separated oscillatory field method to lasers have been made by Bergquist,[21] Lee,[21] Hall,[21] Salour,[22] Cohen-Tannoudji,[22] Bordé,[23] Hansch,[24] Chebotayev[25] and many others.[25]

(9) The method has been extended to neutron beams and to neutrons stored for long times in totally reflecting bottles.

(10) In a recent beautiful experiment, S. Chu and his associates.[26] have successfully used the principle of separated oscillatory fields with a fountain of atoms that rises up slowly, passes through an oscillating field region, falls under gravity and passes again through the same oscillatory field region. This fountain experiment was attempted many years ago by J. R. Zacharias and his associates,[3] but it was unsuccessful because of the inadequate number of very slow atoms. Chu and his collaborators used laser cooling[27,28,29] to slow the atoms to a low velocity and obtained a beautifully narrow separated oscillatory fields resonance pattern.

The Atomic Hydrogen Maser

The atomic hydrogen maser grew out of my attempts to obtain even greater accuracy in atomic beam experiments. By the Heisenberg uncertainty principle (or by the Fourier transform), the width of a resonance in a molecular beam experiment cannot be less than approximately the reciprocal of the time the atom is in the resonance region of the apparatus. For atoms moving through a 1 m long resonance region at 100 m/s this means that the resonance width is about 100 Hz wide. To decrease this width and hence increase the precision of the measurements required an increase in this time. To increase the time by drastically lengthening the apparatus or selecting slower molecules would decrease the already marginal beam intensity or greatly increase the cost of the apparatus. I therefore decided to plan an atomic beam in which the atoms, after passing through the first oscillatory field would enter a storage box with suitably coated walls where they would bounce around for a period of time and then emerge to pass through the second oscillatory field. My Ph.D. student, Daniel Kleppner,[30] undertook the construction of this device as his thesis project. The original configuration required only a few wall collisions and was called a broken atomic beam resonance experiment. Initially the beam was cesium and the wall coating was teflon. The experiment[30] was a partial success in that a separated oscillatory field pattern for an atomic hyperfine transition was obtained, but it was weak and disappeared after a few wall collisions. The results improved markedly when a paraffin was used for the wall coating and a hyperfine resonance was eventually obtained after 190 collisions giving a resonance width of 100 Hz, but with the resonance frequency shifted by 150 Hz.

To do much better than this, we decided we would have to use an atom with a lower mass and a lower electric polarizability to reduce the wall interactions. Atomic hydrogen appeared ideal for this purpose, but atomic hydrogen is notoriously difficult to detect. We, therefore, calculated the possibility of detecting the transitions through their effects on the electromagnetic radiations. Townes[31] had a few years earlier made the first successful maser (acronym for microwave amplifier by stimulated emission of radiation) but no one had previously made a maser based on a magnetic dipole moment or on a frequency as low as that of an atomic hyperfine transition. We concluded, however, that if the resonance could be made narrow enough by multiple wall collisions, we should be able to obtain maser oscillations. The apparatus was designed and constructed by Goldenberg, Kleppner and myself[32] and after a few failures we obtained maser oscillations at the atomic hydrogen hyperfine frequency. Both the proton and the electron have spin angular momenta **I** and **J** as well as magnetic moments. The atomic hyperfine transitions are those for which there is a change of the relative orientation of these two magnetic moments between the initial and final states in Eq. (2). We studied H atoms in the $1^2S_{1/2}$ ground electronic state and mostly observed the transitions ($F=1, m=0 \rightarrow F=0, m=0$) where F is the quantum number of the total angular momentum $\mathbf{F} = \mathbf{I} + \mathbf{J}$ and m is the associated magnetic quantum number.

The principles of an atomic hydrogen maser are shown schematically in Figure 6. An intense electrical discharge in

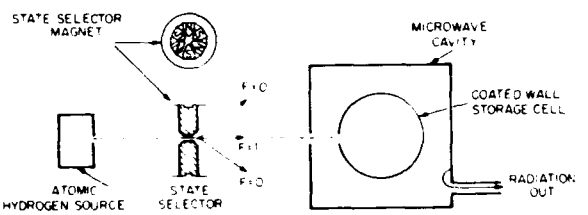


Figure 6. -- Schematic diagram of atomic hydrogen maser. Only the paths of the m=0 atoms are shown since the m=1 atoms are not involved in the Dm=0 transitions studied.

the source converts commercially available molecular hydrogen (H_2) into atomic hydrogen (H). The atoms emerge from the source into a region that is evacuated to 10^{-6} torr and enter a state selecting magnet which has three north poles alternating in a circle with three south poles. By symmetry, the magnetic field is zero on the axis and increases in magnitude away from the axis. Since the energy of a hydrogen atom in the F=1 m=0 state increases with energy and since mechanical systems are accelerated toward lower potential energy, an atom in F=1 state that is slightly off axis will be accelerated toward the axis, i.e. the F=1 state will be focussed onto the small aperture of the 15cm diameter teflon coated storage cell whereas the F=0 state is defocussed. As a result, if the atomic beam flows steadily, the storage bottle in equilibrium will contain more high energy F=1 atoms than low. If these atoms are exposed to microwave radiation at the hyperfine frequency, more atoms are stimulated to go from the higher energy state to the lower one than in the opposite direction. Energy is then released from the atoms and makes the microwave radiation stronger. Thus the device is an amplifier or maser. If the storage cell is placed inside a tuned cavity, an oscillation at the resonance frequency will increase in magnitude until an equilibrium value is reached. At this

level the oscillation will continue indefinitely, with the energy to maintain the oscillation coming from the continuing supply of hydrogen atoms in the high energy hyperfine state. The device then becomes a free running maser oscillator at the atomic hyperfine frequency.

The atomic hydrogen maser oscillator has unprecedented high stability due to a combination of favorable features. The atoms typically reside in the storage cell for 10 seconds, which is much longer than in an atomic beam resonance apparatus so the resonance line is much narrower. The atoms are stored at low pressure so they are relatively free and unperturbed while radiating. The first order Doppler shift is removed, since the atoms are exposed to a standing wave and since the average velocity is extremely low for atoms stored for 10 seconds. Masers have very low noise levels, especially when the amplifying elements are isolated atoms. Over periods of several hours the hydrogen maser stability is better than 1×10^{-15}.

The major disadvantage of the hydrogen maser is that the atoms collide with the walls at intervals, changing slightly the hyperfine frequency and giving rise to wall shifts of 1×10^{-11}. However, the wall shifts are highly stable and can be experimentally determined by measurements utilizing storage bottles of two different diameters or with a deformable bulb whose surface to volume ratio can be altered. As in all precision measurements, care must be taken in adjusting and tuning the hydrogen maser to avoid misleading results. These limitations and precautions are discussed in a series of publications by various authors.[32,33,34] The designs of hydrogen masers have been modified in many ways either for special purposes or for increased stability and reliability. For example different hyperfine transitions have been used and masers have been operated in relatively strong magnetic fields. A hydrogen maser has also been operated[17] with a storage bottle that is much larger than the wave length of the stimulating radiation by confining the microwave power to two small cavities so that it functions as a separated oscillatory field device. The atoms that are stimulated to emit radiation move randomly into and out of two oscillatory field cavities and spend the intermediate time in the large container where there is no oscillatory field. Due to the larger size of the storage box there are longer storage times and less frequent wall collisions, so the resonances are narrower and the wall shifts are smaller than for a normal hydrogen maser.

Precision Spectroscopy

Now that I have discussed extensively the principles of the separated oscillatory field method and of the atomic maser, I shall give some illustrations as to their value. One major category of applications is to precision spectroscopy, especially at radio and microwave frequencies. Another category of applications is to atomic clocks and frequency standards.

It is difficult to summarize the spectroscopic applications since there are so many of them. Many beautiful experiments have been done by a large number of scientists in different countries. I shall, therefore use just a few illustrations from experiments in which I have been personally involved.

My graduate students have made precision measurements of the radiofrequency spectra of different molecules in various rotational states. For each of these states more than seven different molecular properties can be inferred and thus the variations of the properties with changes in the rotational and vibrational quantum numbers can be determined. These properties include nuclear and rotational magnetic moments, nuclear quadrupole interactions, nuclear spin-spin magnetic interactions, spin rotational interactions, etc. I shall illustrate the accuracy and significance of the measurements with a single example. With both D_2 and LiD we have accurately measured[35,36] the deuteron quadrupole interaction eqQ where e is the proton electric charge, q is the gradient of the molecular electric field at the deuteron and Q is the deuteron quadrupole moment which measures the shape of the deuteron and in particular its departure from spherical symmetry. These measurements were made with a high resolution molecular beam apparatus based on the method of separated oscillatory fields. We found for eqQ the value $+225{,}044 \pm 20$ Hz in D_2 and $+34{,}213 \pm 33$ Hz in LiD. Since q has been calculated[37,38] for each of these quite different molecules, two independent values of Q can be calculated. The results agree to within 1.5% which confirms the validity of the difficult calculation; with it we find $Q = 2.9 \times 10^{-27}$ cm^2.

In an experiment with collaborators[39] at the Institut Laue-Langevin at Grenoble, France, we have used the separated oscillatory field method with a beam of slow neutrons to make an accurate measurement of the neutron magnetic moment and found[37,40] it to be $-1.91304275 \pm 0.00000045$ nuclear magnetons. In a somewhat different experiment with neutrons moving so slowly that they can be bottled for more than 80 s in a suitable storage vessel, we have used the method of successive oscillatory fields with the two coherent radiofrequency pulses being separated in time rather than space. In this manner and as a fundamental test of time reversal symmetry, we have recently set a very low upper limit for the neutron electric dipole moment by finding[41] its value to be $(-3 \pm 5) \times 10^{-26}$ e cm.

The atomic hydrogen maser gives very accurate data on the microwave spectrum of the ground electronic state of the hydrogen atom. The hyperfine frequency $\Delta\nu$ for atomic hydrogen has been measured in our laboratory and in a number of other laboratories. The best value[42,43] is

$$\Delta\nu_H = 1{,}420{,}405{,}751.7667 \pm 0.0009 \text{ Hz}$$

This value agrees with present quantum electromagnetic theory[44] to within the accuracy of the theoretical calculation and can be used to obtain information on the proton structure. Similarly accurate values have been found for atomic deuterium and tritium and the dependence[45] of these results on the strengths of externally applied electric fields have been measured. With a modified form of the hydrogen maser designed to operate at high magnetic fields, the ratio of magnetic moment of the electron to that of the proton is found[40,46] to be -658.210688 ± 0.000006. Incidentally when this result is combined with the beautiful electron measurements from Professor Dehmelt's labortory[40,47,48] we obtain the best values for the free proton magnetic moment in both Bohr and nuclear magnetons.

Atomic Clocks

In the past 50 years there has been a major revolution in time keeping, with accuracy and reproducibility of the best clocks at the end of that period being approximately a million times those at the beginning. This revolution in time keeping and frequency control is due to atomic clocks.

Any clock or frequency standard depends on some regular periodic motion such as the pendulum of the grandfather's clock. In the case of atomic clocks the periodic motion is internal to the atoms and is usually that associated with an atomic hyperfine structure as discussed in the section on atomic hydrogen maser.

In the most widely used atomic clocks, the atom whose internal frequency provides the periodicity is cesium and the usual method of observing it is with a separated oscillatory field magnetic resonance apparatus as in Figure 1. The first commercial cesium beam clock was developed in 1955 by a group led by J. R. Zacharias[4] and in the same year L. Essen and V. L. Parry[4] constructed and operated the first cesium beam apparatus that was extensively used as an actual frequency standard. Subsequently many scientists and engineers throughout the world contributed to the development of atomic clocks, as discussed in greater detail elsewhere.[4]

Cesium atomic clocks now have an accuracy and stability of about 10^{-13} which was so far superior to all previous clocks that in 1967 the internationally adopted definition of the second was changed from one based on motion of the earth around the sun to 9,192,631,770 periods of the cesium atom.

For many purposes even greater stability is required over shorter time intervals. When such stability is needed the hydrogen maser is frequently used with a stability of 10^{-15} over periods of several hours.

Atomic clocks based on the above principles have for a number of years provided clocks of the greatest stability and accuracy and these are sufficiently great that further improvements might seem to be neither desirable nor feasible. But as we shall see in our final section, there are applications that already push atomic clocks to their limits and there are many current developments with great promise for the future. These include improvements to the existing devices, use of higher frequency, use of lasers, electromagnetic traps for storing both ions and atoms, laser cooling, etc.

Applications for Accurate Clocks

Accurate atomic clocks are used for so many different purposes that a list of them all is tediously long so I shall here just briefly mention a few that push clock technology to its limit.

In radio astronomy one looks with a parabolic reflector at the radio waves coming from a star just as in optical astronomy one looks with an optical telescope at the light waves coming from a star. Unfortunately, in radio astronomy the wavelength of the radiation is about a million times longer than the wavelength of light. The resolution of the normal radio telescope is therefore about a million times worse since the resolution of a telescope depends on the ratio of the wave length to the telescope aperture. However, if there are two radio telescopes on opposite sides of the earth looking at the same star and if the radio waves entering each are matched in time, it is equivalent to a single telescope whose aperture is the distance between the two telescopes and the resolution of such a combination exceeds that of even the largest single optical telescope. However, to do such precise matching in time each of the two radio telescopes needs a highly stable clock, usually an atomic hydrogen maser.

One of the exciting discoveries in radio astronomy has been the discovery of pulsars, that emit their radiation in short periodic pulses. Precision clocks have been needed to measure the pulsar periods and the changes in the periods with time; these changes sometimes occur smoothly and sometimes abruptly. Of particular interest from the point of view of time measurements, are the millisecond pulsars which have remarkable constancy of period, rivaling the stability of the best atomic clocks. In fact one of these pulsars is so stable that it may eventually be suitable as a standard of time over long periods.[49] Another millisecond pulsar is part of a rapidly rotating binary star that is slowly changing its period of rotation.[49] This slow change in rotation can be attributed to the loss of energy by the radiation of gravity waves -- the first experimental evidence for the existence of gravity waves.

Time and frequency can now be measured so accurately that wherever possible other fundamental measurements are reduced to time or frequency measurements. Thus the unit of length by international agreement has recently been defined as the distance light will travel in a specified time and voltage has just been represented in terms of frequency measurements.

Accurate clocks have provided important tests of both the special and general theories of relativity. In one experiment, a hydrogen maser was shot in a rocket to a 6,000 mile altitude and its periodic rate changed with speed and altitude just as expected by the special and general theories of relativity. In other experiments, observers have measured the delays predicted by relativity for radio waves passing near the sun.

Precision clocks make possible an entirely new and more accurate navigational system, the global positioning system or GPS. A number of satellites containing accurate clocks transmit signals at specific times so any observer receiving and analyzing the signals from four such satellites can determine his position to within ten yards and the correct time within one hundredth of a millionth of a second (10^{-8} s).

A particularly fascinating navigation feat dependent on accurate clocks was the recent and highly successful tour of the Voyager spacecraft to Neptune. The success of this mission depended upon the ground controllers having accurate knowledge of the position of the Voyager. This was accomplished by having three large radio telescopes at different locations on the earth, each of which transmitted a coded signal to Voyager which in turn transmitted the signals back to the telescopes. The distances from each telescope to Voyager could be determined from the elapsed times and thus Voyager could be located. To achieve the required timing accuracy, two hydrogen masers were located at each telescope. Due to the rotation of the earth in the eight hours required for the electromagnetic wave to travel from the earth to Voyager and back again at the speed of light, the telescope transmitting the signal in some cases had to be different from the one receiving; this placed an additional stringent requirement on the clocks. Thus, highly stable clocks were essential for the spectacular success of the Voyager mission to Neptune.

References

1. I. I. Rabi, J. R. Zacharias, S. Millman and P. Kusch, Phys. Rev. 53, 318 (1938) and 55, 526 (1939).

2. J. M. B. Kellogg, I. I. Rabi, N. F. Ramsey and J. R. Zacharias, Phys. Rev. 55, 729 (1939); 56, 728 (1939) and 57, 677 (1940).

3. N. F. Ramsey, Molecular Beams, Oxford Press (1956 and 1985).

4. N. F. Ramsey, History of Atomic Clocks, Journal of Research of NBS 88, 301 (1983). This paper contains an extensive list of references.

5. N. F. Ramsey, Phys. Rev. 76, 996 (1949) and 78, 695 (1950).

6. N. F. Ramsey, Physics Today 33 (7), 25 (July 1980).

7. N. F. Ramsey, Phys. Rev. 109, 822 (1958).

8. N. F. Ramsey, Jour. Phys. et Radium 19, 809 (1958).

9. S. R. Lundeen, P. E. Jessop and F. M. Pipkin, Phys. Rev. Lett. 34, 377 and 1368 (1975).

10. N. F. Ramsey, Phys. Rev. 100, 1191 (1955).

11. F. Bloch and A. Siegert, Phys. Rev. 57, 522 (1940).

12. R. F. Code and N. F. Ramsey, Phys. Rev. A4, 1945 (1971).

13. J. H. Shirley, J. Appl. Phys. 34, 783 (1963).

14. G. Greene, Phys. Rev. A18, 1057 (1970).

15. N. F. Ramsey and H. B. Silsbee, Phys. Rev. 84, 506 (1951).

16. N. F. Ramsey, Rev. Sci. Inst. 28, 57 (1957).

17. E. Uzgiris and N. F. Ramsey, Phys. Rev. A1, 429 (1970).

18. V. F. Ezhov, S. N. Ivanov, I. M. Lobashov, V. A. Nazarenko, G. D. Porsev, A. P. Serebrov and R. R. Toldaev, Sov. Phys. - JETP 24, 39 (1976).

19. S. Jarvis, D. J. Wineland and H. Hellwig, J. Appl. Phys. 48, 5336 (1977).

20. Y. V. Blaklanov, B. V. Dubetsky and V. B. Chebotsev, Appl. Phys. 9, 171 (1976).

21. J. C. Bergquist, S. A. Lee and J. L. Hall, Phys. Rev. Let. 38, 159 (1977) and Laser Spectroscopy III, 142 (1978).

22. M. M. Salour, C. Cohen-Tannoudji, Phys. Rev. Lett. 38, 757 (1977); Laser Spectroscopy III, 149 (1978), Appl. Phys. 15, 119 (1978) and Phys. Rev. A17, 614 (1978).

23. C. J. Bordé., C. R. Acad. Sci. Paris 284B, 101 (1977).

24. T. W. Hansch, Laser Spectroscopy III, 149 (1978).

25. V. P. Chebotayev, A. V. Shishayev, B. Y. Yurshin, L. S. Vasilenko, N. M. Dyuba and M. I. Skortsov, Appl. Phys. 15, 43, 219 and 319 (1987).

26. M. Kasevich, E. Riis, S. Chu and R. S. DeVoe, Phys. Rev. Lett.63, 612 (1989).

27. D. Wineland and H. Dehmelt, Bull. Am. Phys. Soc. 18, 1521 (1973) and 20, 60, 61, 637 (1975).

28. T. W. Hansch and A. L. Schawlow Opt. Commun. 13, 68 (1975) and review by U. S. Letokhow, Comments on Atomic and Molecular Physics 6, 119 (1977).

29. D. J. Wineland and W. M. Itano, Physics Today 40, (6) 34 (June 1987).

30. D. Kleppner, N. F. Ramsey and P. Fjelstadt, Phys. Rev. Lett. 1, 232 (1958).

31. J. P. Gordon, H. Z. Geiger and C. H. Townes, Phys. Rev. 95, 282 (1954) and 99, 1264 (1955).

32. H. M. Goldenberg, D. Kleppner and N. F. Ramsey, Phys. Rev. 8, 361 (1960) and Phys. Rev. 126, 603 (1962).

33. D. Kleppner, H. C. Berg, S. B. Crampton, N. F. Ramsey, R. F. C. Vessot, H. E. Peters and J. Vanier, Phys. Rev. 138, A972 (1965).

34. J. M. V. A. Koelman, S. B. Crampton, H. T. C. Luiten and B. J. Verhaar, Phys. A38, 3535 (1988). This paper contains an extended series of references to other papers on hydrogen maser limitations, principles and practices.

35. R. F. Code and N. F. Ramsey, Phys. Rev. A4, 1945 (1971).

36. R. R. Freeman, A. R. Jacobson, D. W. Johnson and N. F. Ramsey, Jour. Chem. Phys. 63, 2597 (1975).

37. R. V. Reid and M. L. Vaida, Phys. Rev. A7, 1841 (1943).

38. K. K. Docken and R. R. Freeman, J. Chem. Phys. 61, 4217 (1974).

39. G. L. Green, N. F. Ramsey, W. Mampe, J. M. Pendlebury, K. Smith, W. B. Dress, P. D. Miller and P. Perrin, Phys. Rev. D20, 2139 (1979).

40. E. R. Cohen and B. Taylor, Rev. Mod. Phys. 59, 1121 (1987).

41. K. F. Smith, N. Crampin, J. M. Pendlebury, D. J. Richardson, D. Shiers, K. Green, A. I. Kilvington, J. Moir, H. B. Prosper, D. Thompson, N. F. Ramsey, B. R. Heckel, S. K. Lamoreaux, P. Ageron, W. Mampe and A. Steyerl, Phys. Lett. B136, 327 (1984) and B234, 191 (1990).

42. H. Hellwig, R. F. Vessot, M. Levine, P. W. Zitzewitz, D. W. Allan and D. T. Glaze, IEEE Trans. Instruments and Measurements IM-19, 200 (1970).

43. L. Essen, M. J. Donaldson, M. J. Bangham and E. G. Hope, Nature 229, 110 (1971).

44. G. L. Baldwin and D. R. Yennie, Phys. Rev. D37, 498 (1988).

45. P. C. Gibbons and N. F. Ramsey, Phys. Rev. A5, 73 (1972).

46. P. F. Winkler, D. Kleppner, T. Myint and F. G. Walther, Phys. Rev. A5, 83 (1972) and E. Cohen and B. Taylor, Phys. Lett. B204 (April 1988)..

47. R. S. vanDyck, P. B. Schwinberg and H. Dehmelt, Atomic Physics 9, 53 (1984) (World Scientific, Singapore).

48. R. S. vanDyck, F. L. Moore, D. L. Farnum and .P. B. Schwinberg, Bull. Am. Phys. Soc. 31, 244 (1986) and Atomic Physics 9, 75 (1984) (World Scientific, Singapore)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

EXPERIMENTS WITH AN ISOLATED SUBATOMIC PARTICLE AT REST

Nobel Lecture, December 8, 1989

by

HANS DEHMELT

Department of Physics, University of Washington, Seattle, WA 98195, USA

"You know, it would be sufficient to really understand the electron." Albert Einstein

The 5th century B.C. Philosopher's Democritus' smallest conceivable in-divisible entity, the a-tomon (the un-cuttable), is a most powerful but not an immutable concept. By 1930 it had already metamorphosed twice: from something similar to a molecule, say a slippery atomon of water, to Mendeleyev's chemist's atom and later to electron and to proton, both particles originally assumed to be of small but finite size. With the rise of Dirac's theory of the electron in the late twenties their size shrunk to mathematically zero. Everybody "knew" then that electron and proton were indivisible Dirac point particles with radius R = 0 and gyromagnetic ratio g = 2.00. The first hint of cuttability or at least compositeness of the proton came from Stern's 1933 measurement of proton magnetism in a Stern-Gerlach molecular beam apparatus. However this was not realized at the time. He found for its normalized dimensionless gyromagnetic ratio not g = 2 but

$$g = (\mu/A)(2M/q) \approx 5,$$

where μ, A, M, q are respectively magnetic moment, angular momentum, mass and charge of the particle. For comparison the obviously composite $^4He^+$ ion, also with spin ½, according to the above formula has the |g| value 14700, much larger than the Dirac value 2. Also, along with this large |g| value went a very finite radius of this atomic ion of $\approx 3 \times 10^{-11}$ m. And indeed, with Hofstadter's high energy electron scattering experiments in the fifties the proton radius grew again to $R = 0.86 \times 10^{-15}$ m, roughly in proportion to the excess value of 3 in g. Similar later work at still

CH2818-3/90/0000-012 $1.00 © 1990 IEEE

higher energies found 3 quarks inside the "indivisible" proton. Today everybody "knows" the *electron* is an indivisible atomon, a Dirac point particle with radius R = 0 and g = 2.00.... But is it? Like the proton, it could be a composite object. History may well repeat itself once more. This puts a very high premium on precise measurements of the g factor of the electron.

GEONIUM SPECTROSCOPY

The metastable pseudo-atom geonium (Van Dyck et al. 1978 and 1986) has been expressly synthesized for studies of the electron g factor under optimal conditions. It consists of an individual electron permanently confined in an ultrahigh vacuum Penning trap at 4K. The trap employs a homogeneous magnetic field B_0 = 5T and a weak electric quadrupole field. The latter is produced by hyperbolic electrodes, a positive ring and two negative caps spaced $2Z_0$ = 8 mm apart, see Fig. 1. The potential, with A a constant, is given by

$$\phi(xyz) = A(x^2 + y^2 - 2z^2),$$

with an axial potential well depth

$$D = e[\phi(000) - \phi(00Z_0)] = 2eAZ_0^2 = 5eV.$$

The trapping is mostly magnetic. The large magnetic field dominates the motion in the geonium atom. The energy levels of this atom shown in Figure 2 reflect the cyclotron motion, at frequency $\nu_c = eB_0/2\pi m$ = 141 GHz, the spin precession, at $\nu_s \approx \nu_c$, the anomaly or g-2 frequency $\nu_a = \nu_s - \nu_c$ = 164 MHz, the axial oscillation, at ν_z = 60 MHz, and the magnetron or drift motion at frequency ν_m = 13 kHz. The electron is continuously monitored by exciting the ν_z-oscillation and detecting via radio the 10^8-fold enhanced spontaneous 60 Mhz emission. A corresponding signal appears in Figure 3. Doppler side-band cooling has made continuous confinement in the trap center of an electron for 10 months (Gabrielse et al. 1985) possible. This process makes the electron absorb rf photons deficient in energy and supply the balance from energy stored in the electron motion to be cooled. The corresponding shrinking of the radius of the magnetron motion is displayed in Figure 4. Extended into the optical region, the cooling scheme is most convincingly demonstrated in Figure 5. The

transitions of primary interest at ν_c, ν_a, ν_m are much more difficult to detect than the ν_z oscillation. Nevertheless the task may be accomplished by means of the continuous Stern-Gerlach effect (Dehmelt 1988a), in which the geonium atom itself is made to work as a 10^8-fold amplifier. In the scheme a single ν_a-photon of only ≈ 1 μeV energy gates the absorption of ≈ 100 eV of rf power at ν_z. The continuous effect uses an inhomogeneous magnetic field in a similar way as the classic one. However, the field takes now the form of a very weak Lawrence cyclotron trap or magnetic bottle shown in Figure 6. The bottle adds a minute monitoring well only

$$D_m = (m + n + \tfrac{1}{2})0.1\mu eV$$

deep to the axial well of large electrostatic depth D = 5 eV, with m, n respectively denoting spin and cyclotron quantum numbers. Thus jumps in m or n show up as jumps in ν_z,

$$\nu_z = \nu_{z0} + (m + n + \tfrac{1}{2})\delta,$$

with $\delta = 1.2$Hz in our experiments, and ν_{z0} the axial frequency of the electron without a magnetic bottle. Random jumps in m, n occur, when spin or cyclotron resonances are excited. Figure 6A shows an early example of a series of such jumps in m or spin flips. For the spin spontaneous transitions are totally negligible. Standard text books discuss transitions between two sharp levels induced by a broad electromagnetic spectrum $\rho(\nu)$: The transition rate from either level is the same and is proportional to the spectral power density $\rho(\nu_s)$ of the radiation field at the transition frequency ν_s. Ergo, the average dwell times in either level are the same, compare Fig. 6A. In the geonium experiments the frequency of the weak rf field is sharp, but the spin resonance is broadened and has a shape $G_s(\nu)$. One may convince oneself that moving the sharp frequency of the rf field upwards over the broad spin resonance should produce the same results as moving a broad rf field of spectral shape $\rho(\nu) \propto G_s(\nu)$ downwards over a sharp spin resonance: The rate of all spin flips or jumps in m in either direction counted in the experiment is proportional to $G_s(\nu)$. To obtain the plot of $G_s(\nu)$ in Fig. 7 the frequency of the rf field was increased in small steps, and at each step spin flips were counted for a fixed period of about ½ hour. From our ν_s, ν_c data for electron and positron (Van Dyck et al. 1987) we have determined

$\frac{1}{2}g^{exp} = \nu_s/\nu_c = 1.001\ 159\ 652\ 188(4),$

the same for particle and anti-particle. The error in the difference of their g-factors is only half as large. Heroic quantum electro-dynamical calculations (Kinoshita 1988) have now yielded for the shift of the g factor of a point electron associated with turning on its interaction with the electromagnetic radiation field

$\frac{1}{2}(g^{point} - 2) = \frac{1}{2}\Delta g^{KINOSHITA} = 0.001\ 159\ 652\ 133(29).$

In the calculations $\Delta g^{KINOSHITA}$ is expressed as a power series in α/π. Kinoshita has critically evaluated the experimental α input data on which he must rely. He warns that the error in his above result, which is dominated by the error in α, may be underestimated. Muonic, hadronic and other small contributions to g amount to less than about 4×10^{-12} and have been included in the shift. Kinoshita's result may be used to correct the experimental g value and find

$g = g^{exp} - \Delta g^{KINOSHITA} = 2 + 11(6)\times10^{-11}.$

ELECTRON RADIUS R?

Extrapolation from known to unknown phenomena is a time-honored approach in all the sciences. Thus from known g, and R values of other near-Dirac particles and our *measured* g value of the electron I attempt to extrapolate a value for its radius. Stimulated by theoretical work of Brodsky & Drell (1980), I (1989a) have plotted $|g - 2|$ vs $R/\lambda\!\!\!^{-}_C$ in Figure 8 for the helium3 nucleus, triton, proton, and electron. Here $\lambda\!\!\!^{-}_C$ is the Compton wavelength of the respective particle. The plausible relation given by Brodsky and Drell (1980) for the simplest composite theoretical model of the electron,

$|g - 2| = R/\lambda\!\!\!^{-}_C$, or

$|g - g_{DIRAC}| = (R - R_{DIRAC})/\lambda\!\!\!^{-}_C$

fits the admittedly sparse data surprisingly well. Even for such a very different spin ½

structure as the atomic ion $^4He^+$ composed of an α-particle and an electron the data point does not fall too far off the full line. Intersection in Figure 8 of this line with the line $|g-2| = 1.1x10^{-10}$ for the Seattle g data yields for the electron the extrapolated point shown and with $\lambda\!\!\!^{-}_C = 0.39x10^{-10}$ cm an electron radius

$R \approx 10^{-20}$ cm.

The row of X's reflects the data range defined by the uncertainty in the Seattle g data and the upper limit $R < 10^{-17}$ cm determined in high energy collision experiments. It appears that this combination of current data is not in harmony with electron structure models assuming special symmetries that predict the quadratic relation $|g-2| \approx (R/\lambda\!\!\!^{-})^2$ shown by the dashed line. This favors the linear relation used in the above extrapolation of R for the electron. Thus, the electron may have *size and structure!*

If one feels that the excess g value $11(6)x10^{-11}$ measured is not significant because of its large relative error then, the value $R \approx 10^{-20}$ cm given here still constitutes an important new upper limit. Changing the point of view, the close agreement of g^{point} with g^{exp} provides the most stringent experimental test of the fundamental theory of Quantum Electrodynamics in which $R = 0$ is assumed. Furthermore the near-identity of the g values measured for electron and positron in Seattle constitutes the most severe test of the CPT theorem or mirror symmetry of a *charged* particle pair.

LEMAÎTRE'S "L'ATOME PRIMITIF" REVISITED - A SPECULATION

Beginning 1974 Salam and others have proposed composite electron and quark models (Lyons 1983). On the strength of these proposals and with an eye on Figure 8, I view the electron as the third approximation of a Dirac particle, d_3 for short, and as composed of three fourth-approximation Dirac or d_4 particles. The situation is taken to be quite similar to that previously encountered in the triton and proton subatomic particles, respectively assumed to be of type d_1 and d_2. In more detail, three d_4 subquarks of huge mass $m' \approx 10^{10}\, m_e$ in a deep square well make up the electron in this working hypothesis. However, their mass $3m'$ is almost completely compensated by strong binding to yield a total relativistic mass equal to the observed mass m_e of the electron. Figure 8 may even suggest a more speculative extrapolation: The e-constituents, in the infinite regression $N \rightarrow \infty$ proposed in Figure 9, have ever more

massive, ever smaller sub-sub-.... constituents d_N. However, these higher order subquarks are realized only up to the "cosmon" with N = C, the most massive particle ever to appear in this universe*. At the beginning of the universe, a lone bound cosmon-anticosmon pair or life time-broadened cosmonium atom state of near-zero total relativistic mass/energy was created from Vilenkin's (1984) metastable "nothing" state of zero relativistic energy in a spontaneous quantum jump of cosmic rarity. Similar, though much more frequent, quantum jumps that have recently been observed in a trapped Ba^+ ion are shown in Figure 10. In this case the system also jumps spontaneously from a state (ion in metastable $D_{5/2}$ level plus no photon) to a new state (ion in $S_{1/2}$ ground level plus photon) of the same total energy. The "cosmonium atom" introduced here is merely a modernized version** of Lemaître's (1950) "l'atome primitif" or world-atom whose explosive radioactive decay created the universe. At the beginning of the world the short-lived cosmonium atom decayed into an early gravitation-dominated standard big bang state that eventually developed into a state, in which again rest mass energy, kinetic and gravitational potential energy add up to *zero* (see formula 8 of Jordan 1937). The electron is a much more complex particle than the cosmon. It is composed of 3^{C-3} cosmon-like d_C's, but only two particles of this type formed the cosmonium world-atom from which sprang the universe. In closing, I should like to cite a line from *William Blake*,

"To see a world in a grain of sand ---"

and allude to a possible parallel

-- to see worlds in an electron --

Brodsky, S. J., and Drell, S. D. (1980) "Anomalous Magnetic Moment and Limits on Fermion Substructure," Phys. Rev. D 22, 2236.

Dehmelt, H. (1983) "Stored Ion Spectroscopy", in Advances in Laser spectroscopy, F. T. Arecchi, F. Strumia & H. Walther, Eds., Plenum, New York.

Dehmelt, H. (1988a) "Single Atomic Particle Forever Floating at Rest in Free Space: New Value for Electron Radius," Physica Scripta T22, 102.

Dehmelt, H. (1988b) "New Continuous Stern Gerlach Effect and a Hint of 'the' Elementary Particle," Z. Phys. D 10, 127-134.

Dehmelt, H. (1989a) "Geonium Spectra * Electron Radius * Cosmon" in *High Energy Spin Physics,* 8th International Symposium, K. Heller, Ed. (AIP Conference Proceedings No. 187, New York) p. 319.

Dehmelt, H. (1989b) "Triton,..electron,..cosmon..: An infinite regression?, Proc. Natl. Acad. Sci. USA 86, 8618-8619.

Dehmelt, H. (1990) "Less is more: Experiments with an Individual Atomic Particle at Rest in Free Space," Am. J. Phys. 58, 17.

Gabrielse, G., Dehmelt, H., and Kells, W. (1985) "Observation of a Relativistic, Bistable Hysteresis in the Cyclotron Motion of a Single electron" Phys. Rev. Letters 54, 537.

Jordan, P., (1937) "Die physikalischen Weltkonstanten" Naturwissenshaften 25, 513.

Kinoshita, T. (1988) "Fine-Structure Constant Derived from Quantum Electrodynamics," Metrologia 25, 233.

Lemaître, G. (1950) *THE PRIMEVAL ATOM* (Van Nostrand, New York) p. 77.

Lyons, L. (1983) "An Introduction to the Possible Substructure of Quarks and Leptons," Progress in Particle and Nuclear Physics 10, 227, see references cited herein.

Nagourney, W., Sandberg, J., and Dehmelt, H. (1986) "Shelved optical electron amplifier: Observation of quantum jumps." Phys. Rev. Letters 56, 2797.

Van Dyck, Jr., R. S., Ekstrom, P., and Dehmelt, H. (1977) "Precise Measurement of Axial, Magnetron, and Spin-Cyclotron Beat Frequencies on an Isolated 1-meV Electron", Phys. Rev. Lett. 38, 310

Van Dyck, Jr., R. S., Schwinberg, P. B. & Dehmelt, H. G. (1978) "Electron Magnetic Moment from Geonium Spectra," in New Frontiers in High Energy Physics (Eds. B. Kursunoglu, A. Perlmutter, and L. Scott), Plenum New York.

Van Dyck, Jr., R. S., Schwinberg, P. B. & Dehmelt, H. G. (1986) "Electron Magnetic Moment from Geonium Spectra: Early Experiments and Background Concepts," Phys. Rev. D 34, 722.

Van Dyck, Jr., R. S., Schwinberg, P. B. & Dehmelt, H. G. (1987) "New High Precision Comparison of Electron/Positron g-Factors," Phys. Rev. Letters 59, 26.

Vilenkin, A. (1984) "Quantum Creation of Universe," Phys. Rev. D 30, 509-515.

Wineland, D., Ekstrom, P. and Dehmelt, H. (1973) "Monoelectron Oscillator," Phys. Rev. Lett. 31, 1297.

*A single "monotheistic" entity terminating a *finite* sequence of preons, pre-preons, pre-pre-preons.... has been briefly discusssed in "Preon Model" by J. C. Pati, A. Salam and J. Strathdee, Nucl. Phys. B185, 416 (1981).

**This is by no means the first modernization attempt. M. Goldhaber has kindly brought it to my attention that he had introduced a different "cosmon" already in 1956 in his paper "Speculations on Cosmogony," SCIENCE 124, 218.

FIGURE CAPTIONS

Fig. 1. Penning trap. The simplest motion of an electron in the trap is along its symmetry axis, along a magnetic field line. Each time it comes too close to one of the negatively charged caps it turns around. The resulting harmonic oscillation took place at about 60 Mhz in our trap. Reproduced from (Dehmelt 1983) with permission, copyright Plenum Press.

Fig. 2. Energy levels of geonium. Each of the cyclotron levels labeled n is split first by the spin - magnetic field interaction. The resulting sublevels are further split into the oscillator levels and finally the manifold of magnetron levels extending downwards. Reproduced from (Van Dyck et al. 1978) with permission, copyright Plenum Press.

Fig. 3. Rf signal produced by trapped electron. When the electron is driven by an axial rf field, it emits a 60 MHz signal, which was picked up by a radio receiver. The signal shown was for a very strong drive and an initially injected bunch of 7 electrons. One electron after the other was randomly "boiled" out of the trap until finally only a single one is left. By somewhat reducing the drive power, this last electron could be observed indefinitely. Reproduced from (Wineland et al. 1973) with permission, copyright American Institute of Physics.

Fig. 4. Side-band "cooling" of the magnetron motion at ν_m. By driving the axial motion not on resonance at ν_z but on the lower Doppler side-band at ν_z-ν_m, it is possible to force the metastable magnetron motion to provide the energy balance $h\nu_m$, and thereby expand the magnetron orbit radius. Conversely, an axial drive at $\nu_z+\nu_m$ shrinks the radius. The roles of upper and lower side-bands are reversed here from the case of a particle in a well where the energy increases with amplitude because the magnetron motion is metastable and the total energy of this motion *decreases* with radius. Reproduced from (Van Dyck et al. 1978) with permission, copyright Plenum Press.

Fig. 5. Visible blue (charged) barium atom Astrid at rest in center of Paul trap photographed in natural color. The photograph strikingly demonstrates the close localization, $<< 1\ \mu$m, attainable with geonium cooling techniques. Stray light from

the lasers focussed on the ion also illuminates the ring electrode of the tiny rf trap of about 1 mm internal diameter. Reproduced from (Dehmelt 1988) with permission, copyright the Royal Swedish Academy of Sciences.

Fig. 6. Weak magnetic bottle for continuous Stern-Gerlach effect. When in the lowest cyclotron and magnetron level the electron forms a 1 μm long wave packet, 30 nm in diameter, which may oscillate undistorted in the axial electric potential well. The inhomogeneous field of the auxiliary magnetic bottle produces a minute spin-dependent restoring force that causes the axial frequency ν_z for spin $\uparrow$ and $\downarrow$ to differ by a small but detectable value. Reproduced from (Dehmelt 1988a) with permission, copyright Springer Verlag.

Fig. 6A. Spin flips recorded by means of the continuous Stern-Gerlach effect. The random jumps in the base line indicate jumps in m at a rate of about 1/minute when the spin resonance is excited. The upwards spikes or "cyclotron grass" are explained by expected rapid random thermal excitation and spontaneous decay of cyclotron levels with an average value $<n> \approx 1.2$. Adapted from (Van Dyck et al. 1977) with permission, copyright American Institute of Physics.

Fig. 7. Plot of electron spin resonance in geonium near 141 GHz. A magnetic radiofrequency field causes random jumps in the spin quantum number. As the frequency of the exciting field is stepped through the resonance in small increments, the number of spin flips occurring in a fixed observation period of about ½ hour are counted and then plotted vs frequency. (Actually the 141 GHz field flipping the spin is produced by the cyclotron motion of the electron through an inhomogeneous magnetic rf field at $\nu_s - \nu_c = 164$ MHz.) Reproduced from (Van Dyck et al. 1987) with permission, copyright American Institute of Physics.

Fig. 8. Plot of $|g-2|$ values, with radiative shifts removed, vs reduced rms radius $R/\lambda\!\!\!^{-}_C$ for near-Dirac particles. The full line $|g-2| = R/\lambda\!\!\!^{-}_C$ predicted by the simplest theoretical model provides a surprisingly good fit to the data points for proton, triton and helium3 nucleus. It may be used to obtain a new radius value for the *physical* electron from its intersection with the line $|g-2| = 1.1\times10^{-10}$ representing the Seattle electron g data. The data are much less well fitted by the relation $|g-2| = (R/\lambda\!\!\!^{-}_C)^2$, which is shown for comparison in the dashed line. The atomic ion $^4He^+$ is definitely

not a near-Dirac particle, but even its data point does not fall too far off the full line. Adapted from (Dehmelt 1990) with permission, copyright American Institute of Physics.

Fig. 9. Triton model of near-Dirac particles. Reproduced from (Dehmelt 1989b) with permission, copyright the National Academy of Sciences of the USA.

Fig. 10. Spontaneous decay of Ba^+ ion in metastable $D_{5/2}$-level. Illuminating the ion with a laser tuned close to its resonance line produces strong resonance fluorescence and an easily detectable photon count of 1600 photons/sec. When later an auxiliary, weak Ba^+ spectral lamp is turned on the ion is randomly transported into the metastable $D_{5/2}$ level of 30 sec lifetime and becomes invisible. After dwelling in this shelving level for 30 sec on the average, it drops down to the $S_{1/2}$ ground state *spontaneously* and becomes visible again. This cycle then repeats. Reproduced from (Nagourney et al. 1986) with permission, copyright American Institute of Physics.

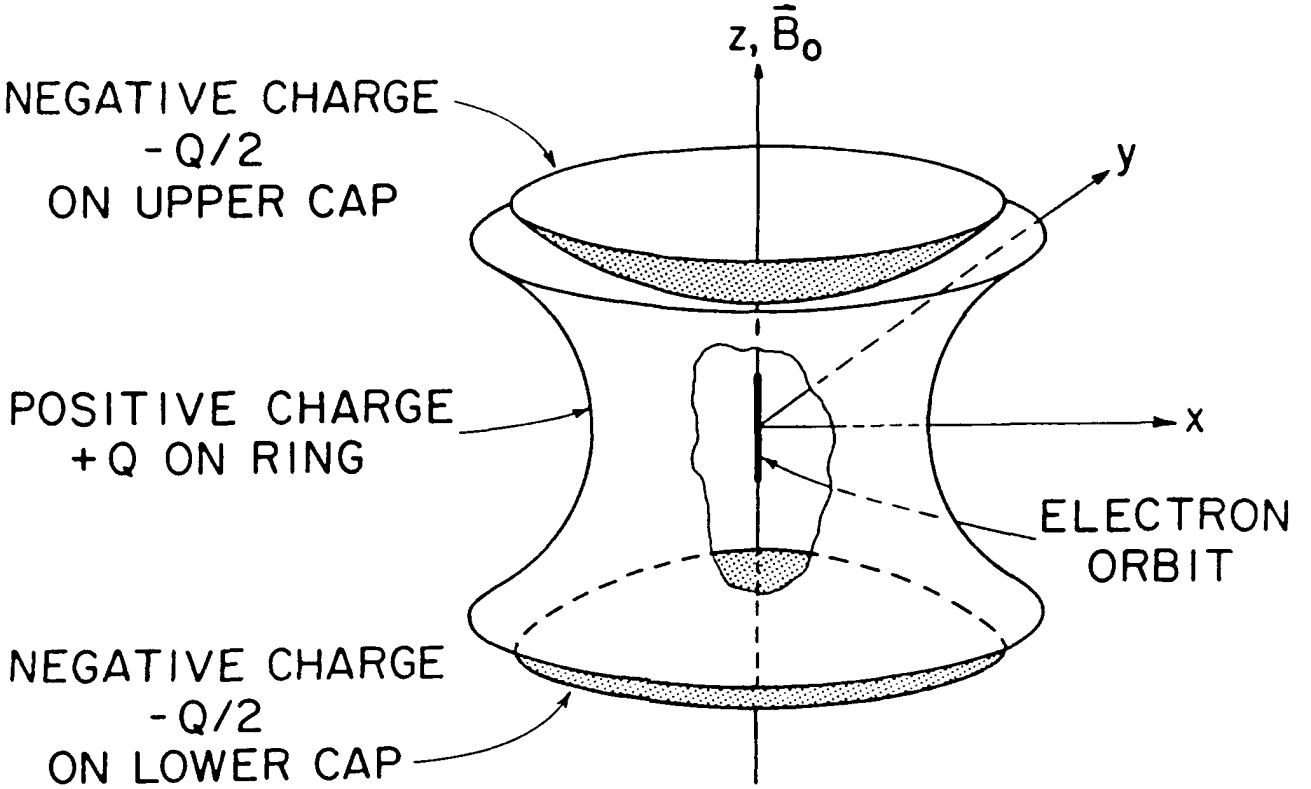

z, $\vec{B}_o$
NEGATIVE CHARGE
-Q/2
ON UPPER CAP
y
POSITIVE CHARGE
+Q ON RING
x
ELECTRON
ORBIT
NEGATIVE CHARGE
-Q/2
ON LOWER CAP

FIGURE 1

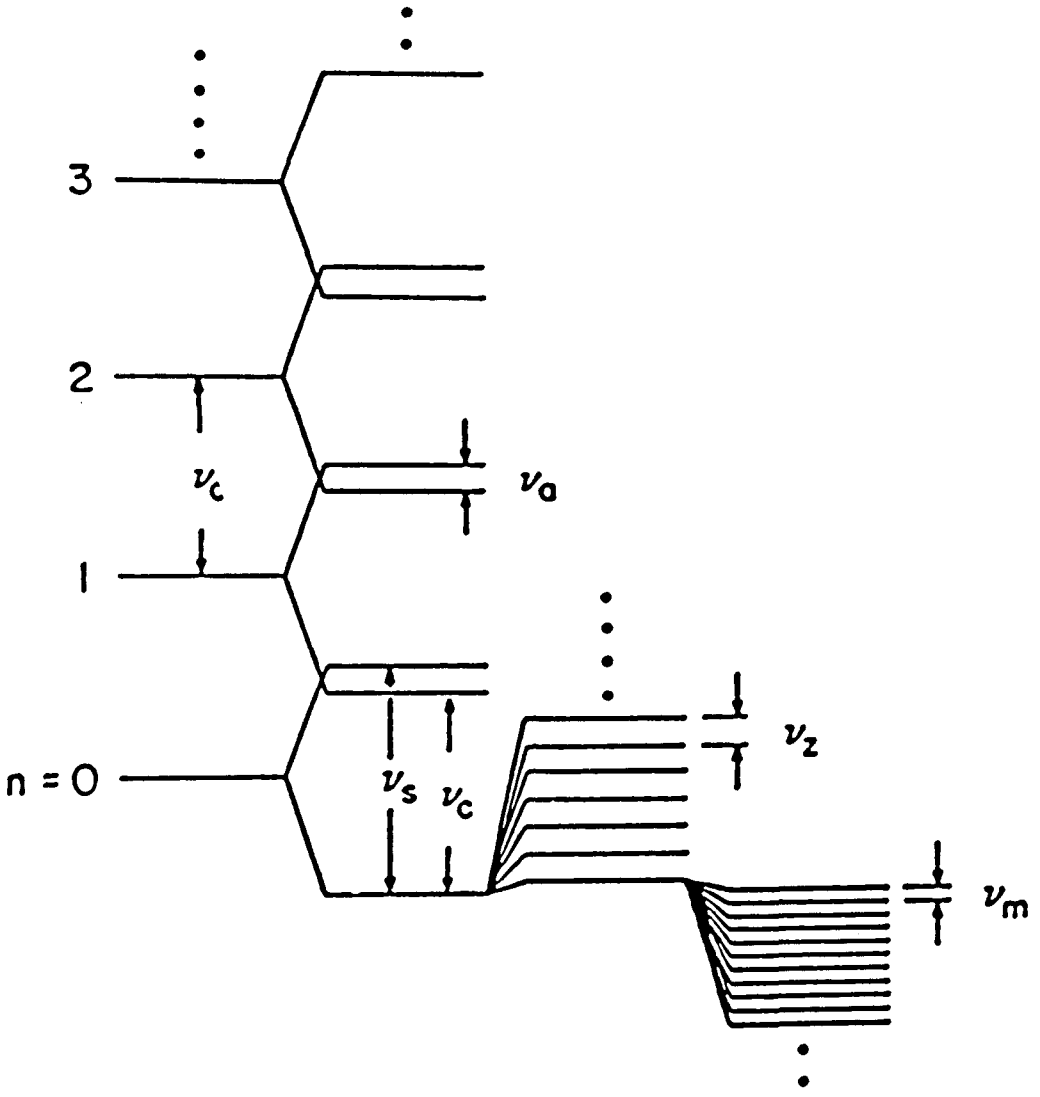


FIGURE 2

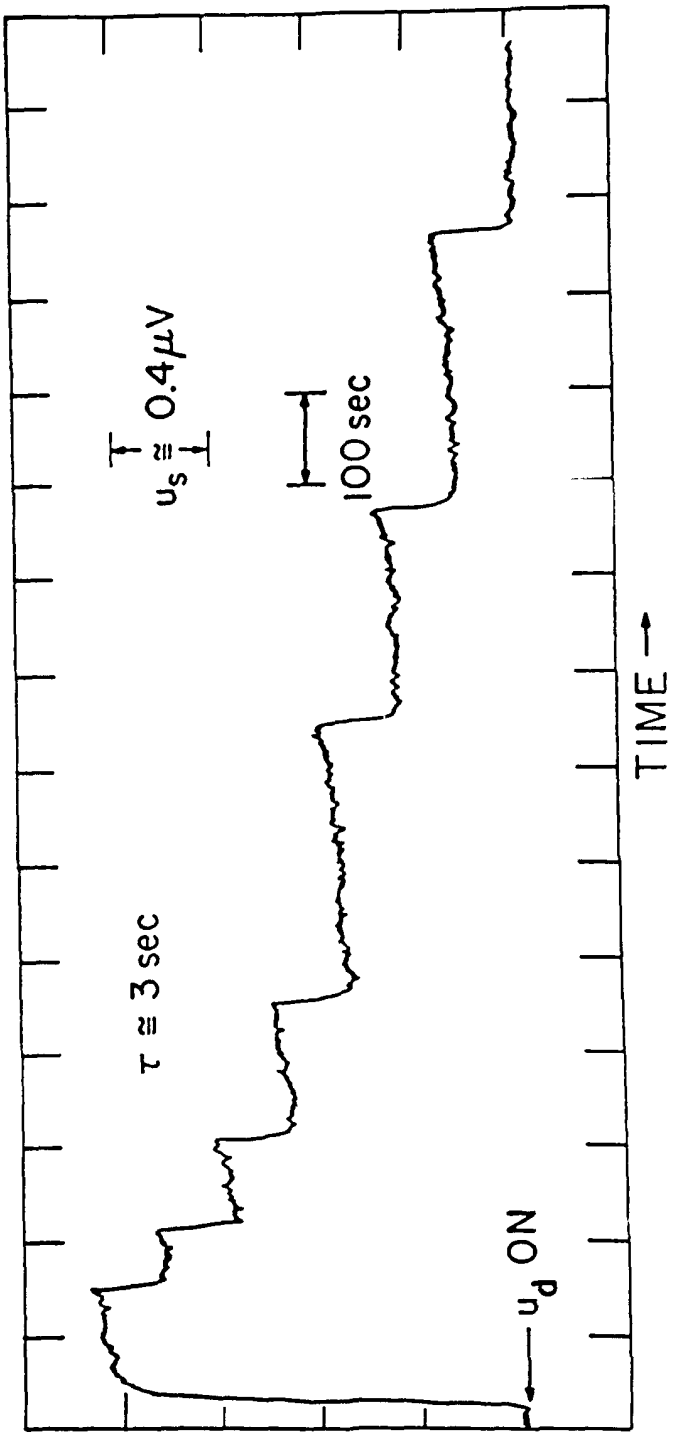

$u_s \cong 0.4\mu V$
100 sec
TIME →
$\tau \cong 3$ sec
u_d ON
← RE-EMITTED SIGNAL

FIGURE 3

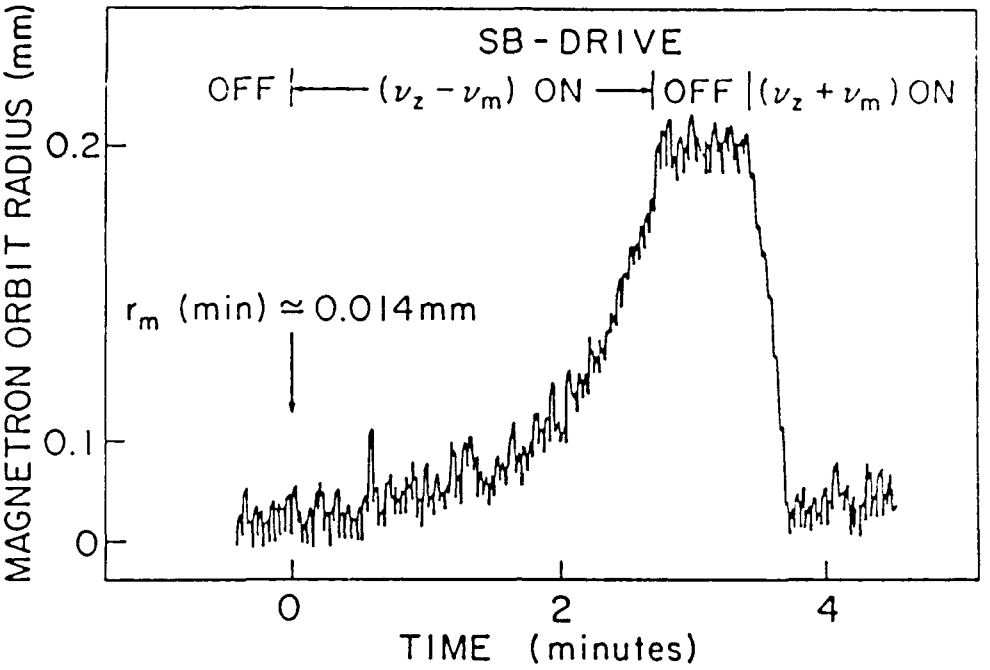

SB - DRIVE
OFF
$(\nu_z - \nu_m)$ ON
OFF
$(\nu_z + \nu_m)$ ON
MAGNETRON ORBIT RADIUS (mm)
0.2
0.1
0
r_m (min) ≅ 0.014mm
0
2
4
TIME (minutes)

FIGURE 4

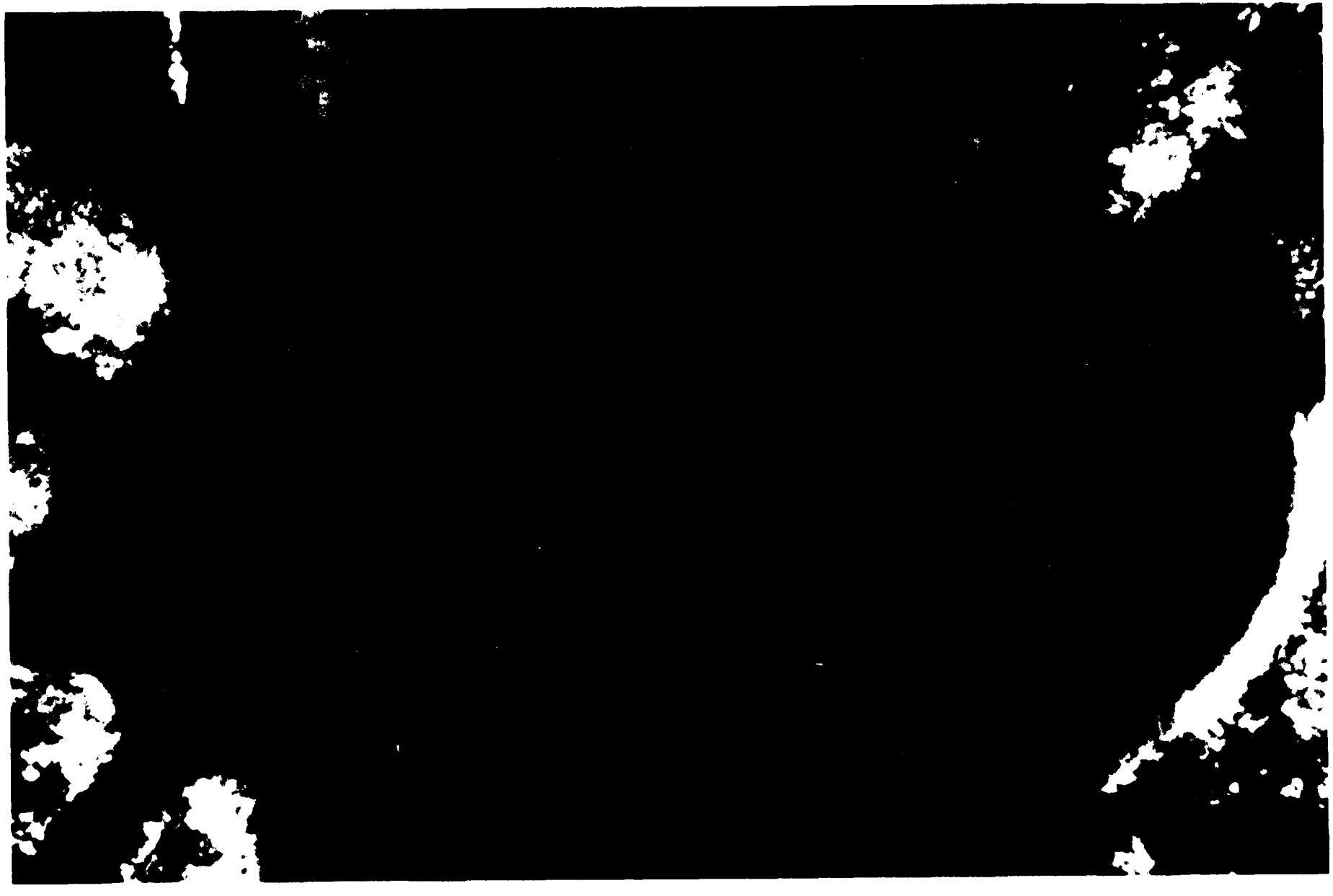


FIGURE 5. VISIBLE CHARGED BLUE BARIUM ATOM ASTRID AT REST IN CENTER OF 1 mm DIAMETER PAUL TRAP PHOTOGRAPHED IN NATURAL COLOR (NAGOURNEY, SANDBERG & D 1986)

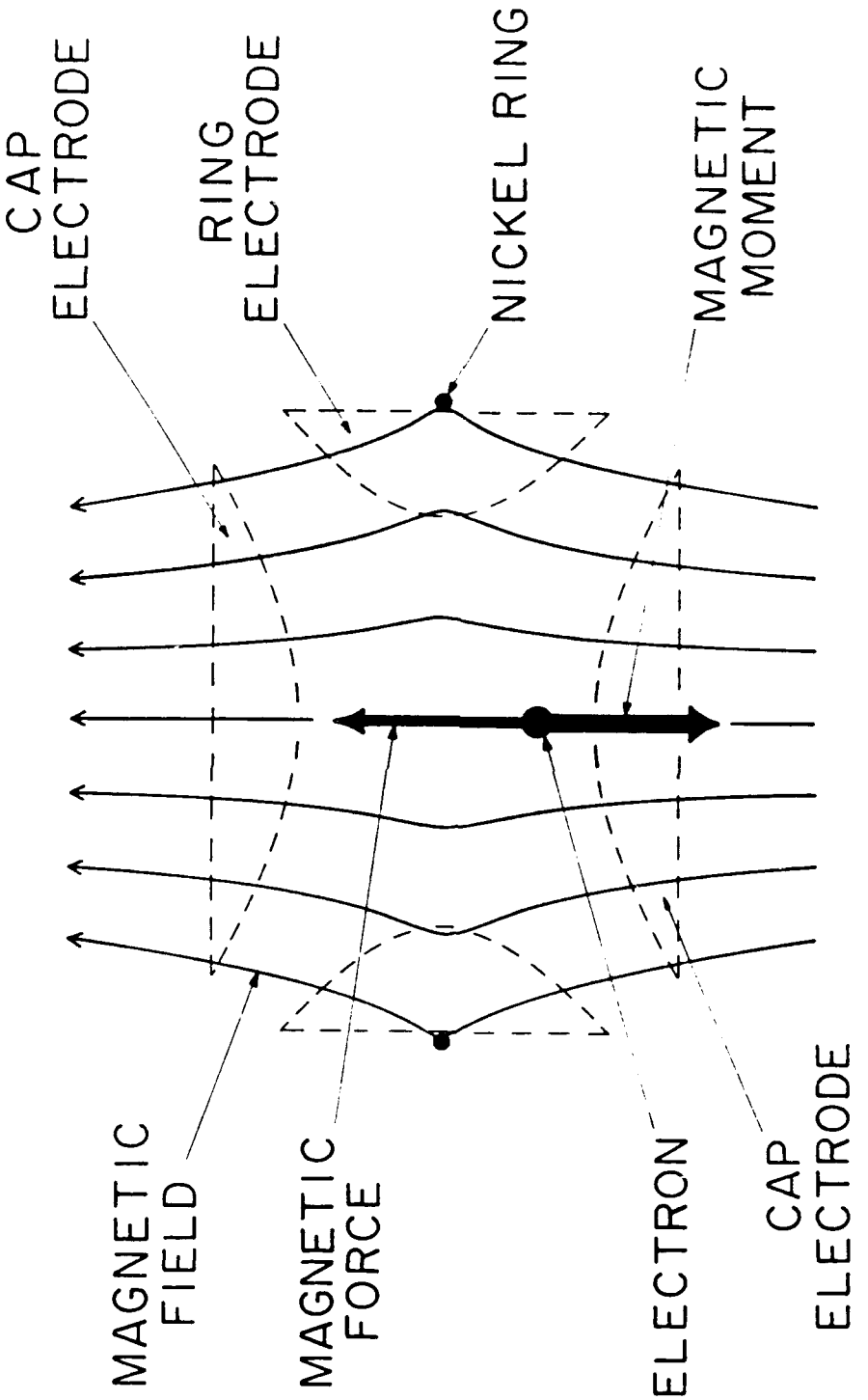

CAP ELECTRODE
RING ELECTRODE
NICKEL RING
MAGNETIC MOMENT
MAGNETIC FIELD
MAGNETIC FORCE
ELECTRON
CAP ELECTRODE

FIGURE 6

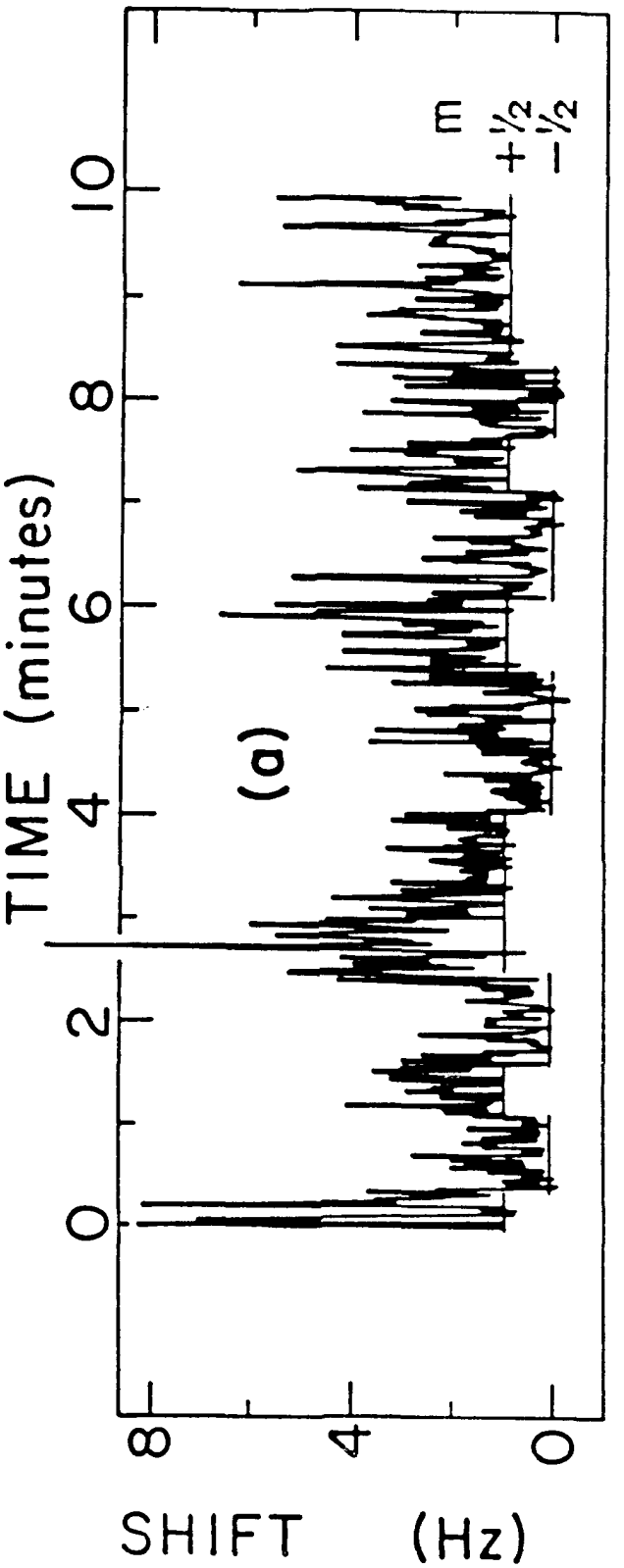


FIGURE 6A

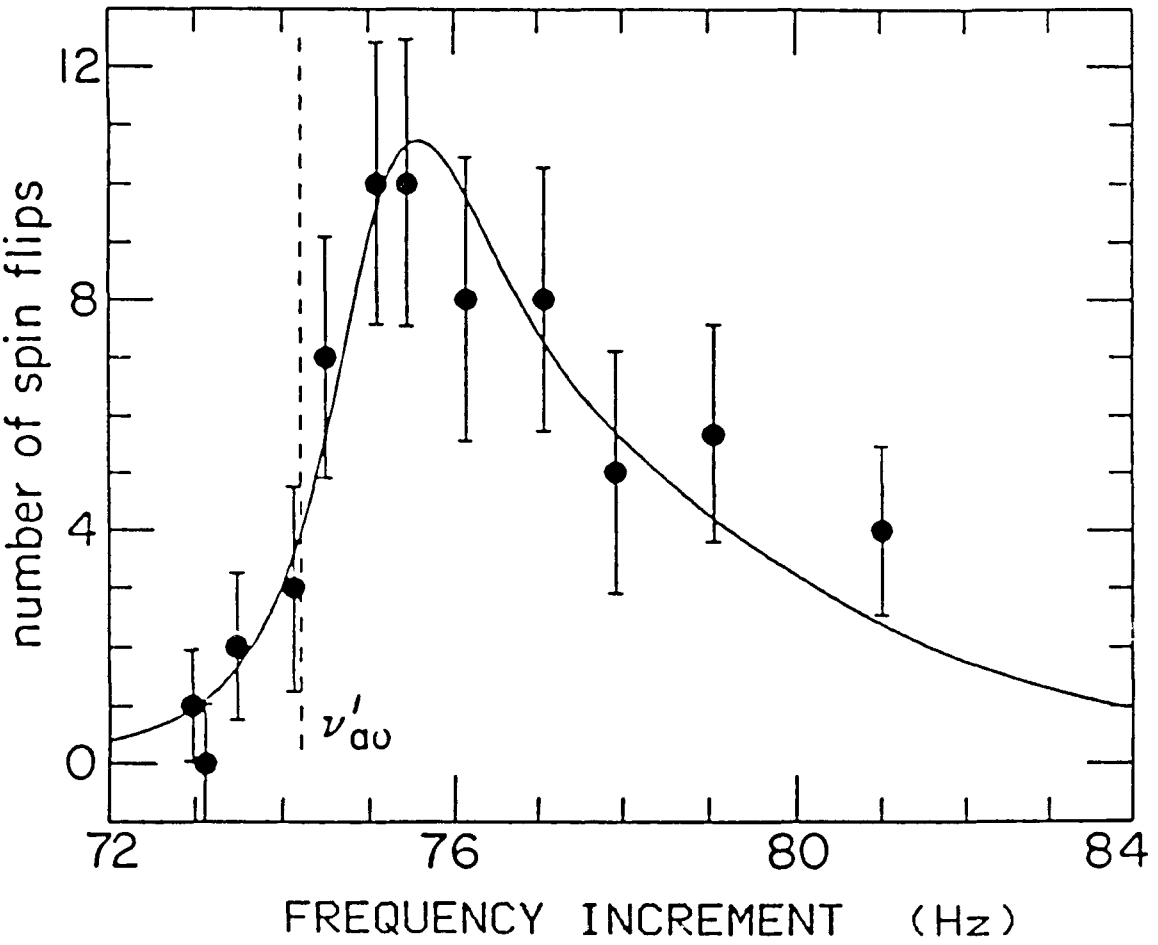

number of spin flips
12
8
4
0
72
76
80
84
FREQUENCY INCREMENT (Hz)
ν'_{ao}

FIGURE 7

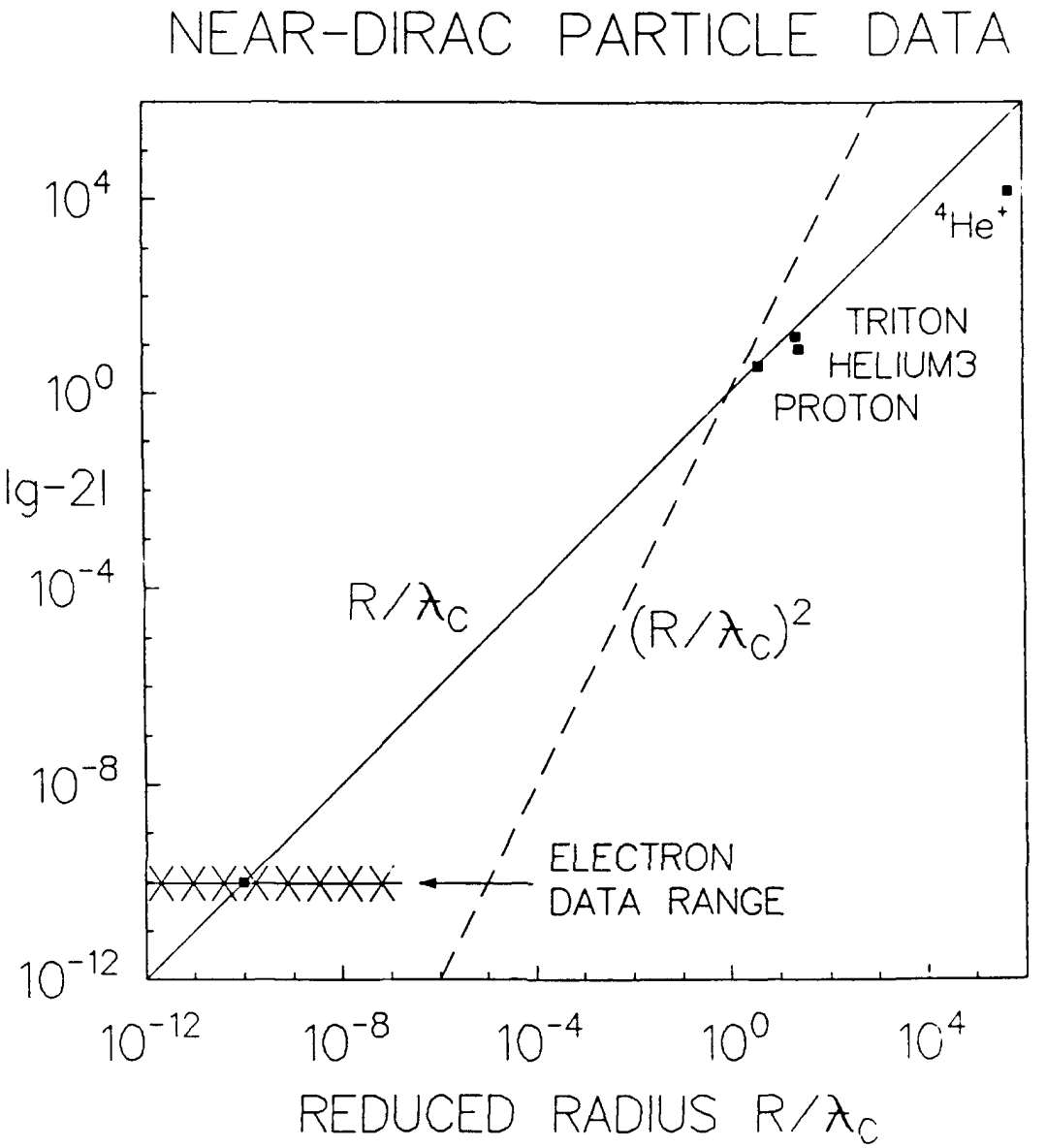


FIGURE 8/DEHMELT/DECEMBER 27, 1989

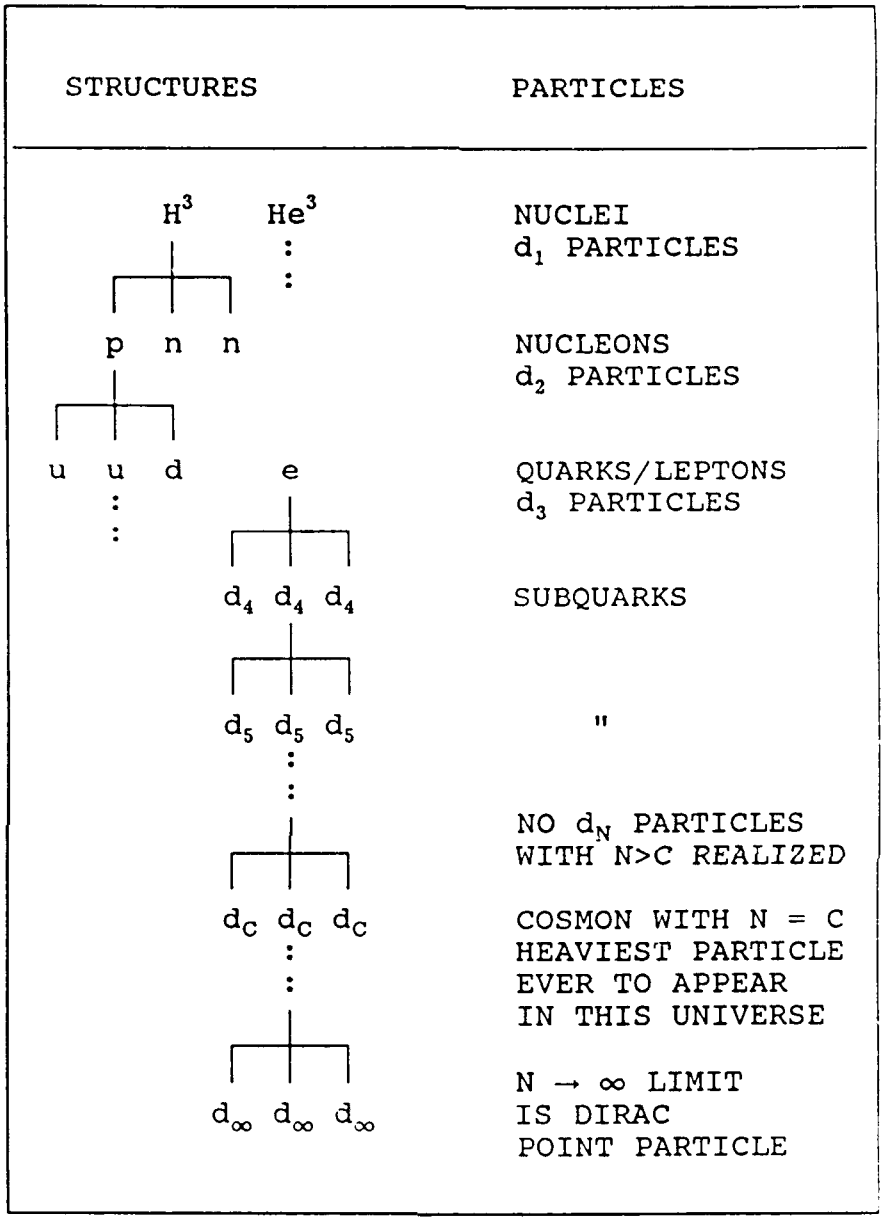

STRUCTURES
PARTICLES
H3
He3
NUCLEI
d1 PARTICLES
p n n
NUCLEONS
d2 PARTICLES
u u d
e
QUARKS/LEPTONS
d3 PARTICLES
d4 d4 d4
SUBQUARKS
d5 d5 d5
"
NO dN PARTICLES
WITH N>C REALIZED
dC dC dC
COSMON WITH N = C
HEAVIEST PARTICLE
EVER TO APPEAR
IN THIS UNIVERSE
d∞ d∞ d∞
N → ∞ LIMIT
IS DIRAC
POINT PARTICLE

FIGURE 9

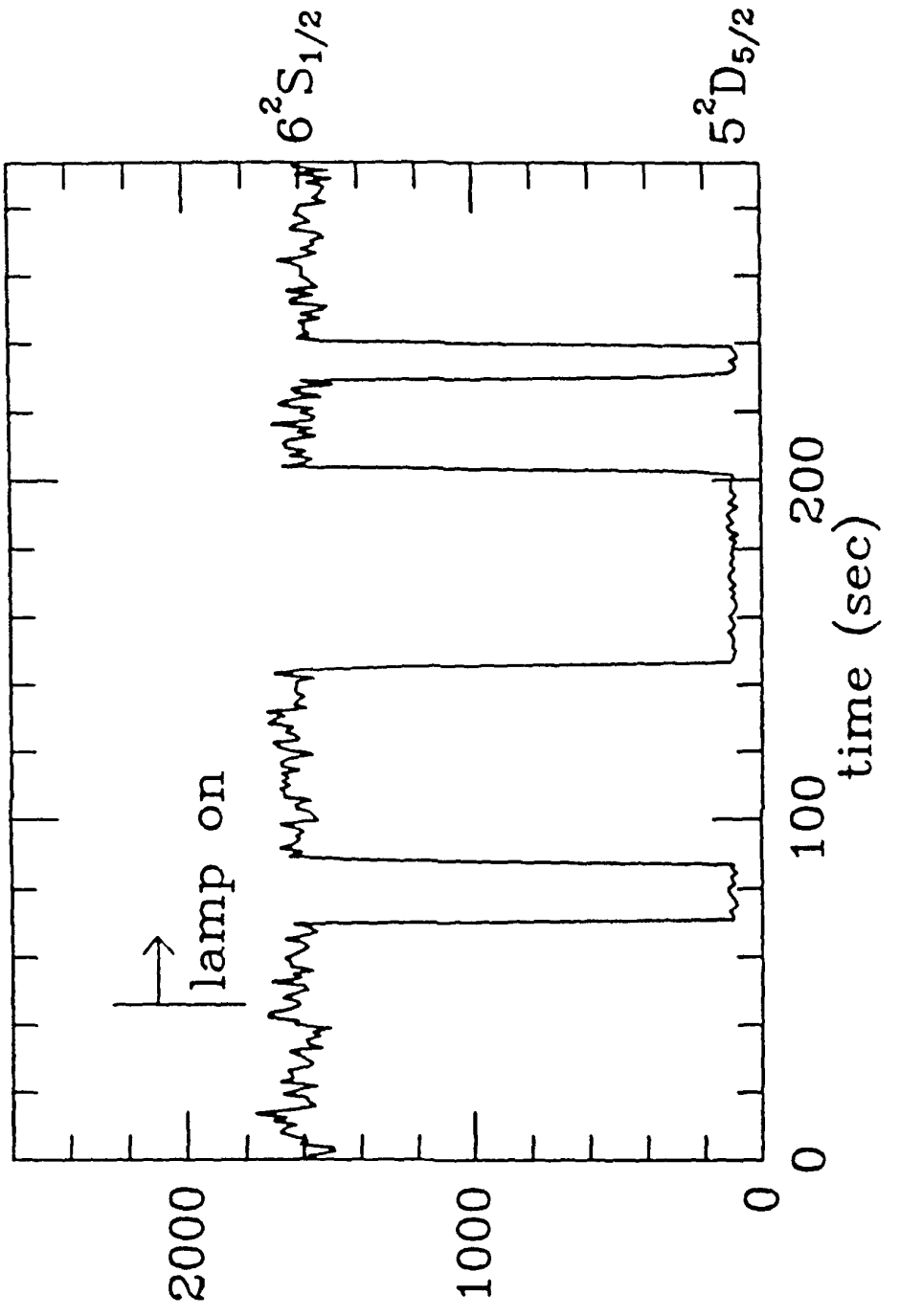

$6^2S_{1/2}$
$5^2D_{5/2}$
lamp on
2000
1000
0
0
100
200
time (sec)
fluorescence photon counts/sec

FIGURE 10

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

INTERFERENCE FRINGES FROM SINGLE-CAVITY EXCITATION OF AN ATOMIC BEAM*

A. DeMarchi,
University of Ancona, Ancona, Italy
and
R. E. Drullinger and J. H. Shirley
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway
Boulder, Colorado 80303

Abstract

A cyclindrical cavity operated in the TE_{013} mode was used for excitation of the hyperfine transition in an optically pumped cesium beam spectrometer. In the configuration we used the atoms see the rf H-field reverse its direction twice. The observed lineshapes show an interference structure similar to Ramsey interference. Theoretically derived lineshapes are in good agreement with the observations. A comparison is made between these lineshapes and corresponding Ramsey lineshapes. The effects of phase variations within the cavity are also discussed briefly.

Introduction

A prototype atomic beam device was set up to test optical pumping as a means of state preparation and detection. Since the magnetic field in this device is longitudinal, a suitable Ramsey-style microwave excitation structure was not readily available. Instead, a simple cylindrical cavity was substituted. This cavity was operated in the TE_{013} mode with the atomic beam passing along the cylindrical axis. The microwave magnetic field amplitude seen by the atoms then has the form of three half periods of a sine wave (see Fig. 1). We expected this form of excitation would exhibit line narrowing similar to that achieved with Ramsey excitation.

Experiment

The prototype beam tube was made as follows. The vacuum chamber was assembled from commercial vacuum components and pumped with a turbo molecular pump. The cesium oven was a piece of 3/8 inch copper tube with a 3 mm graphite aperture. This type of oven does not last long because of the absorption of cesium by graphite, but it is simple to make and delivers a fairly well collimated beam. The laser beams enter and exit the chamber through near normal incidence, high quality anti-reflection coated windows epoxied to the tube. The fluorescence collection optics are identical to those developed for NIST-7 [1]. The C-field coil is wound on an aluminum cylinder with a diameter of about 20 cm. There is a single layer of magnetic shielding with no end caps. The separation of the optical pumping region from the detection region is 25 cm.

The cavity and its mode pattern are shown schematically in Fig. 1. The cavity is made of 2 inch copper plumbing pipe with brass end plugs. The latter have a 5 mm axial hole for the atomic beam passage and a $\lambda/4$ mode filter to discriminate against the degenerate TM and lower modes. The cavity is fed in the center by a small loop in the end of the 0.087 inch coaxial feed line. There is no provision for coupling adjustment. The

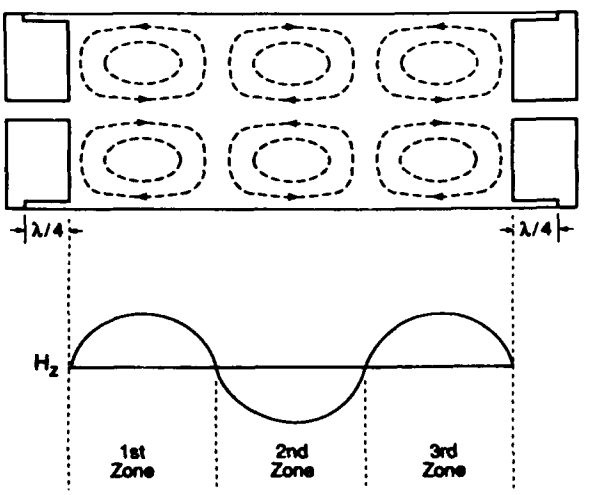


Figure 1. Schematic of TE_{013} mode showing longitudinal magnetic field amplitude in the three excitation zones.

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

resonant frequency was tuned by adjusting the position of the end plugs before they were fixed in place. The final tuning is done by temperature adjustment. The cavity has a loaded Q of about 20 000.

The experiments were done with a single diode laser narrowed by optical feed-back [2] and locked to a saturated absorption feature in a separate cesium cell. A second optical frequency was synthesized from the laser by an acousto-optic modulator. This allowed us to pump on the F = 4 →F' = 4 transition and detect on the F = 4 → F' = 5 cycling transition.

Representative experimental lineshapes for the clock resonance, (F = 3, m = 0 to F = 4, m = 0), are shown in Figs. 2a and 3a together with theoretical lineshapes Figs. 2b and 3b for the same conditions.

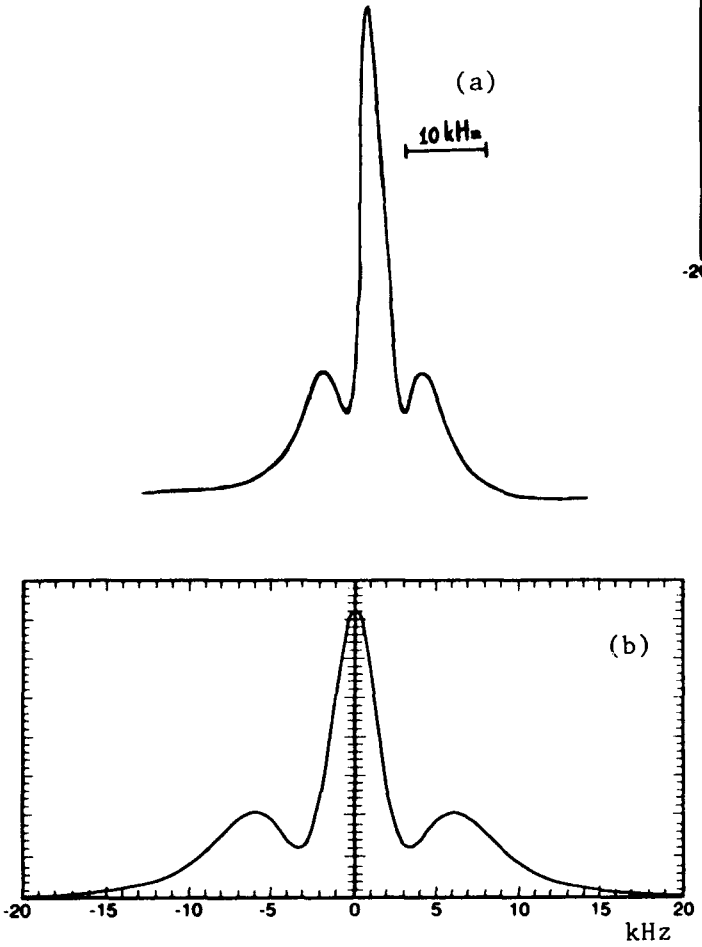


Figure 2. Experimental(a) and theoretical(b) lineshapes obtained at optimum power.

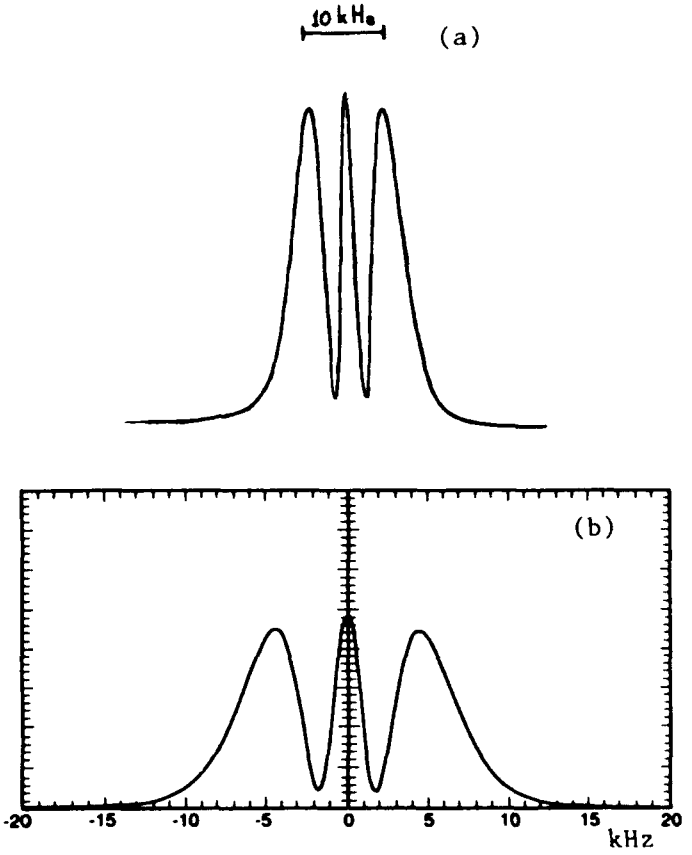


Figure 3. Experimental(a) and theoretical(b) lineshapes obtained at 6 dB below optimum power.

Theory

In the notation of Ramsey [3, Chap. V.3], the evolution of the probability amplitudes of the two hyperfine states is given by the time-dependent Schrodinger equation

$$i(d/dt)C_p(t) - bg(t)e^{i\omega t}C_q(t),$$
$$i(d/dt)C_q(t) - bg(t)e^{-i\omega t}C_p(t) + \omega_o C_q(t), \quad (1)$$

with the initial conditions $C_p(0) = 1$ and $C_q(0) = 0$. We have chosen the energy of the initial state p to be zero. The rotating field approximation has been made. The Rabi frequency 2b is proportional to the microwave magnetic field amplitude, ω is the microwave angular frequency, and ω_0 is the atomic resonance frequency. The function g(t) represents the time dependence of the microwave field amplitude. For the TE_{013} mode g(t) is shown in Fig. 1

and is given by $(\pi/2)\sin(\pi t/\tau)$, where τ is the transit time through one of the three zones, and g is normalized so that $\int_0^\tau g(t)dt = \tau$.

Weak excitation

The theory is most easily developed and understood in the limit of weak excitation. To first order in b, C_p remains equal to 1. The probability amplitude for excitation after an atom has traversed n zones is then

$$C_q = -ibe^{-i\omega_0\tau}G_n(\lambda) \qquad (2)$$

where

$$G_n(\lambda) = \int_0^{n\tau} g(t)e^{i\lambda t}dt = -\frac{\pi^2}{2}\tau\frac{(1-e^{in\lambda\tau}\cos n\pi)}{(\lambda^2\tau^2-\pi^2)} \qquad (3)$$

and $\lambda = \omega_0-\omega$ is the detuning from resonance of the microwave field. G_n is well-behaved at $\lambda\tau = \pm\pi$ since both numerator and denominator vanish together for integer n. If we consider the function g(t) to be zero outside the range $0<t<n\tau$, then the integration limits in (3) can be extended to plus and minus infinity and $G_n(\lambda)$ becomes the Fourier transform of g(t). The transition probability becomes

$$P_n(\lambda) = b^2|G_n|^2 = \pi^4b^2\tau^2\frac{\cos^2(n\lambda\tau/2)}{(\lambda^2\tau^2-\pi^2)^2} \qquad (4)$$

for n odd. For n even, replace $\cos(n\lambda\tau/2)$ by $\sin(n\lambda\tau/2)$. The result (4) is shown in Fig. 4 for n=3. With the aid of some trigonometric identities we can factor the transition probability for one zone of excitation from (4):

$$P_n = |F_n|^2P_1. \qquad (5)$$

For n = 3 the remaining factor,

$$F_3 = 2\cos\lambda\tau-1 = e^{i\lambda\tau}-1+e^{-i\lambda\tau}, \qquad (6)$$

is the sum of three exponentials. These exponentials relate the transition amplitude in the second and third zones to that in the first zone by phases that correspond to the difference in phase evolution between the field and the state q. These phases interfere to either destroy or enhance the basic probability P_1. Hence, we refer to F_3 as the interference factor. For $\lambda\tau = \pi$, $F_3 = -3$ and P_3 is enhanced nine times over P_1. This enhancement is shown by the strong side peaks in Fig. 4. The interference factor also narrows the central peak and introduces additional zeros.

Note that if no excitation took place in the second zone, we would have a form of Ramsey excitation with the drift time $T=\tau$. The second term in (6) would then be missing and (5) would reduce to

$$P_R = (2\cos\lambda\tau)^2P_1, \qquad (7)$$

the usual expression for weak two-zone Ramsey excitation [3, Chap. V.4]. This lineshape, also shown in Fig. 4, has a central peak 1.4 times broader than the central peak for three zone excitation. It also has prominent side peaks, but they are not enhanced above the central peak.

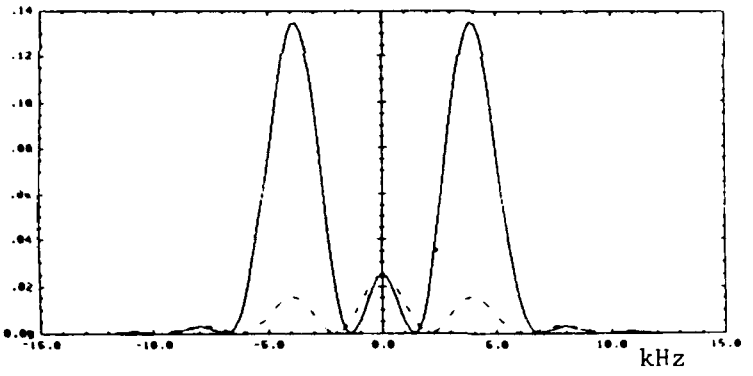


Figure 4. Comparison of lineshapes from three-zone excitation (solid line) with Ramsey excitation (dashed line) at low power with single velocity atoms.

The results (4) and (5) also apply to values of n larger than 3 corresponding to excitation by a TE_{01n} mode. The interference factor F_n becomes a polynomial in $\cos\lambda\tau$ related to the Tschebyscheff polynomials. The lineshape has very strong, narrow side peaks for single-velocity atoms. The central peak has a width $\Delta\nu_n = k_n/2n\tau$ where $k_1 = 2.38$, $k_3 = 1.04$, and k_n approaches unity for n large. A corresponding Ramsey excitation would include excitation by only the first and last zones with a drift time between excitation zones $T = (n-2)\tau$. For such a Ramsey excitation

(7) holds with $\lambda\tau$ replaced by $(n-1)\lambda\tau/2$ in the argument of the cosine. The central peak has a width $\Delta\nu_n = q_n/2(n-1)\tau$ where $q_3 = .98$ and q_n approaches unity for n larger. Hence for n large the multi-zone excitation and corresponding Ramsey excitation yield lineshapes with similar widths.

Strong excitation

If $\psi(t)$ represents the state vector whose components are $C_p(t)$ and $C_q(t)$, then the time evolution of $\psi(t)$ can be expressed by a 2 x 2 evolution matrix $U(t,t_0)$ such that

$$\psi(t) = U(t,t_0)\ \psi(t_0). \qquad (8)$$

For three-zone excitation the evolution can be broken down into the product of evolutions across each zone:

$$\psi(3\tau) = U(3\tau,2\tau)\ U(2\tau,\tau)\ U(\tau,0)\ \psi(0). \qquad (9)$$

In terms of the solution of (1) at $t = \tau$ we have the evolution matrix

$$U(\tau,0) = \begin{pmatrix} C_p(\tau) & -C_q(\tau)^* e^{-i\omega_0\tau} \\ C_q(\tau) & C_p(\tau)^* e^{-i\omega_0\tau} \end{pmatrix}. \qquad (10)$$

$U(2\tau,\tau)$ and $U(3\tau,2\tau)$ are the same as $U(\tau,0)$ except that $C_q(\tau)$ is replaced by $-C_q(\tau)e^{-i\omega\tau}$ and $C_q(\tau)e^{2i\omega\tau}$ respectively. Thus, knowing the solution of (1) for the first excitation zone allows us to easily find the solution for several zones by matrix multiplication. The results for the transition probability are

$$P_1 = |C_q(\tau)|^2 \quad \text{for one zone,}$$

$$P_2 = 4y^2P_1 \quad \text{for two zones,}$$

$$P_3 = (1-4y^2)^2P_1 \quad \text{for three zones,} \qquad (11)$$

where $y = \mathrm{Im}[C_p(\tau)\exp(i\lambda\tau/2)]$.

The one-zone transition probability P_1 is again a factor in the multi-zone transition probability. In the weak excitation limit $C_p(\tau)$ is unity so that $y = \sin(\lambda\tau/2)$ in agreement with (6). For stronger excitation $C_p(\tau)$ decreases, making the side peaks less prominent.

For the actual sine-wave form of g(t) the Schrodinger equation (1) was integrated numerically across one zone to find $\psi(\tau)$. The transition probability for 1, 2, and 3 zones was then found from the relations (11). A sample result is plotted in Fig. 5, for optimum power. Saturation reduces the value of y in (11) allowing the central peak to broaden and reducing the side peaks. Also shown in Fig. 5 is the corresponding Ramsey lineshape for two zones of excitation. The central Ramsey peak is now only 0.6 times as wide as for three zone excitation. Saturation does not affect the interference factor for Ramsey excitation.

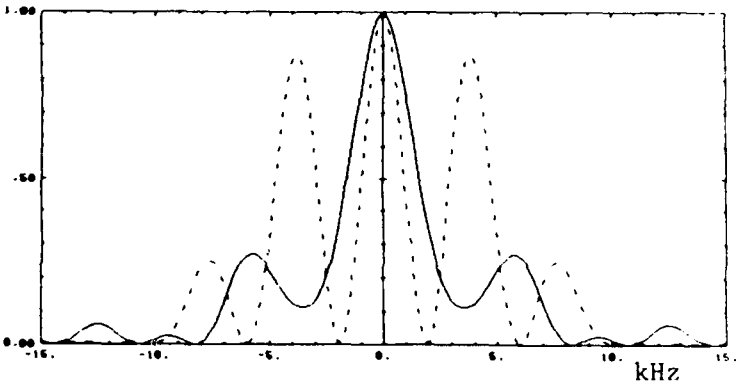


Figure 5. Comparison of lineshapes from three-zone excitation (solid line) with Ramsey excitations (dashed line) at optimum power with single velocity atoms.

Velocity Average

In the experimental situation τ is not fixed, but has a broad distribution of values corresponding to the velocity distribution of atoms in the beam. Since the position of the side lobes, but not the central peak, depends on τ, the side lobes are greatly reduced and broadened by averaging over τ values. A thermal velocity distribution weighted by 1/v for detection by a cycling transition [4] was used to average the numerical results. Figure 2b shows the resulting calculated lineshape for optimum excitation for comparison with the experimental curve in Fig. 2a. At weaker excitation levels, the central peak shrinks and narrows faster than the side lobes as shown in Figs. 3a and 3b. At still lower excitation the central peak becomes only half as high as the side lobes.

Theoretical predictions are in agreement with experimental observations.

Cavity Phase Variations

Spatial phase variations within the cavity can lead to frequency biases, in analogy with the end-to-end phase shift in Ramsey cavities. Only phase variations associated with modes of a symmetry different from the desired one can produce a bias [5]. In our cavity the closest such modes are TE_{012} and TE_{014}. However, since they resonate about 1.1 GHz away from the TE_{013} mode and have very high Q, the resulting frequency bias would be very small even if the cavity feed excited them, which it is not designed to do. An experimental search was made for other modes using probes within the cavity. The only resonances seen in our cavity other than the desired one were those of the TE_{011} and TE_{015} modes, which have a symmetry that will not produce frequency shifts. Furthermore, the degenerate TM_{113} mode and the lower TE_{11n} modes were not observed, indicating that the mode filter works well. In conclusion, we feel that spatial phase variations can be analyzed and shown not to be a problem in this type of cavity.

References

[1] R. E. Drullinger, Jon Shirley, D. J. Glaze, L. W. Hollberg and A. DeMarchi, "Progress Toward an Optically Pumped Cesium Beam Frequency Standard" in Proc. 40th Annual Symposium on Frequency Control, pp. 428-431, 1986.

[2] B. Dahmani, L. Hollberg, and R. Drullinger, "Frequency Stabilization of Semiconductor Lasers by Resonant Optical Feedback", Optics Letters, vol. 12, 876-878, 1987.

[3] N. F. Ramsey, Molecular Beams, Oxford University Press, London, 1956.

[4] G. Avila, E. de Clercq, M. de Labachelerie, and P. Cerez, "Microwave Ramsey Resonances from a Laser Diode Optically Pumped Cesium Beam Resonator," IEEE Trans. Inst. Meas. vol. IM-34, 139-143, 1985.

[5] A. de Marchi, O. Francescangeli, G.P. Bava, "Feasibility of End-to-End Phase Shift Correction from the Outside of Sealed Cesium Beam Tubes," CPEM 1990 (to be published in IEEE Trans. Inst. Meas.).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

EFFECTS OF NEUTRON FLUENCE ON THE OPERATING CHARACTERISTICS OF DIODE LASERS USED IN ATOMIC FREQUENCY STANDARDS

R. P. Frueholz, J. C. Camparo, and S. B. Delcamp, Chemistry and Physics Laboratory, The Aerospace Corporation, P. O. Box 92957, Los Angeles, CA 90009-2957

C. E. Barnes, California Institute of Technology, Jet Propulsion Laboratory, 4800 Oak Grove Dr. Pasadena, CA 91109

Abstract: One of the next major advances in rubidium and cesium atomic clock technology will center on the use of diode lasers for optical pumping. The atomic clocks used on board satellites have the potential to interact with various forms of radiation that are not present in the laboratory environment, and the effects of this radiation on the laser's operating characteristics relevant to clock applications are not well known.

The present paper describes an ongoing experiment to study the effects of neutron fluence on the operating characteristics of Mitsubishi Transverse Junction Stripe (TJS) AlGaAs diode lasers. Different models of the TJS diode laser produce optical radiation in both the 780 and 850 nm range, appropriate for optical pumping in rubidium and cesium atomic clocks, respectively. In this phase, a set of TJS diode lasers has been exposed to a neutron fluence of 2×10^{12} n/cm^2, and four laser characteristics were examined after each exposure. The laser's light output versus injection current and single-mode linewidth versus output power both influence the efficiency of optical pumping and hence the atomic clock's signal-to-noise ratio. We have also measured the laser's single-mode wavelength versus injection current (laser tuning). Since the diode laser must remain tuned to the appropriate atomic transition, any degradation in the ability to tune the laser will impact atomic clock reliability. Finally, the diode laser's gain curve has been studied at several injection currents below threshold. This diode laser characteristic is taken as an indicator of the neutron damage mechanisms in the laser's semiconductor material. Changes in these characteristics due to the neutron exposure are reported.

Introduction

The next generation of rubidium (Rb) and cesium (Cs) atomic clocks will employ diode lasers for optical pumping to improve their frequency stabilities. It is anticipated that the introduction of laser optical pumping in each of these devices will improve their frequency stabilities by factors of nearly 100 over current designs [1,2]. For space applications, the laser pumped clocks will have to function in the presence of various forms of radiation. In particular, the effects of radiation on diode laser function must be investigated. In this paper, we discuss experiments in progress in the laboratories of The Aerospace Corporation to determine the effects of neutron irradiation on the operating characteristics of diode lasers relevant to clock function. The emphasis on characteristics pertinent to atomic clock operation distinguishes this study from previous investigations of neutron effects on diode laser operation [3-5].

Atomic clock operation requires a population imbalance to be established between the two hyperfine levels whose energy difference defines the clock's internal frequency. Optical pumping is a very efficient means of generating the desired population imbalance. The AlGaAs diode lasers [6] are ideal sources of optical pumping radiation. They are compact, solid state devices whose size is attractive for use in the satellite environment. Additionally, they have a number of spectral properties that are well suited to optical pumping in Rb and Cs atomic clocks. The intensities of light they emit are adequate for efficient optical pumping in atomic standards. By varying the mole fraction of Al in these devices, lasers with wavelengths that may be tuned to either Rb or Cs optical resonances are available. Finally, the spectral width of the laser light is sufficiently narrow to allow effective optical pumping in the gas-cell environment of the Rb standard, as well as the atomic-beam environment of the Cs atomic clock.

Prior to the present studies, the bulk of information concerning the effects of neutron exposure on diode laser operation dealt with the radiation-induced modifications of the diode laser's output power versus injection current curve [3-5]. In the present study, four tests are used to characterize a laser's performance before and after exposure to neutrons. First, the laser's output power versus injection current curve is measured. This characteristic affects the laser optical pumping rate, and hence the signal-to-noise ratio and frequency stability of the atomic clock. This curve also identifies the laser's threshold current, the minimum current at which laser emission takes place. Then the laser's single-mode linewidth versus inverse output power curve is obtained. Again this characteristic influences the efficiency of optical pumping and hence the atomic clock's signal-to-noise ratio. The laser's single-mode wavelength versus injection current curve (laser tuning) is measured. As the injection current is increased, the internal temperature of the lasing media also increases. The increasing temperature causes the

CH2818-3/90/0000-039 $1.00 © 1990 IEEE

lasing wavelength to shift to longer values due to changing of the media's refractive index. As the diode laser must remain tuned to the appropriate atomic transition, any degradation in the ability to tune the laser will impact the clock's reliability. Finally, the diode laser's gain versus wavelength curve is obtained at several injection currents below threshold. Though this characteristic is not as closely related to atomic clock performance as the others, it is perhaps a better indicator of the neutron damage mechanisms in the laser's semiconductor material. The range of tests encompassed in this study should provide a fuller picture of the effects of neutron exposure on the diode laser characteristics relevant to atomic standard operation.

Experimental Procedure

Experimentation is being performed on six Mitsubishi TJS ML 3101 diode lasers, labeled A through F. The lasers are being characterized at operating temperatures of both 15 and 30°C. Prior to irradiation, the lasers were characterized using the previously mentioned four tests. We found that the lasers divided themselves into two groups. Lasers C and D displayed relatively low threshold currents, approximately 15 mA, while lasers A, B, E, and F showed higher threshold currents, all very near 35 mA. Typically, this type of laser diode has a threshold current near 20 mA, with a maximum of approximately 40 mA. After initial characterization, the lasers were exposed to a neutron fluence of 2×10^{12} n/cm^2 (E > 1 MeV) at room temperature at the Sandia Pulsed Reactor (SPR III). The fluence represents the average value of the measurements of three sulfur dosimeters with appropriate corrections applied to convert the dosimeter values to fluences for neutrons with energies greater than 1 MeV.

Upon return to our laboratory, the lasers' optical properties were remeasured. To ensure the reproducibility of our spectral measurements, a control laser, unexposed to neutrons, was also recharacterized. Its spectral properties were required to remain constant prior to proceeding with the characterization of the exposed lasers. In this paper, we report the results of this single neutron exposure. We anticipate continuing the study, increasing the exposure to higher neutron fluences.

Results

Prior to exposure, all of the diode lasers displayed spectral characteristics consistent with normally behaving devices. After exposure, the two low threshold lasers showed no discernible changes in operating characteristics. In contrast, the four high threshold lasers showed marked changes in their performances. These effects will be reviewed in the following paragraphs.

In Fig. 1, typical output power versus injection current curves are displayed for a low threshold laser (D) and a high threshold laser (E) before and after neutron exposure. No effect of neutron exposure is observed for the low threshold laser. In contrast, laser E shows a slight increase in its threshold current upon neutron exposure. Also, after exposure, this laser can produce no more than 2.4 mW of optical power. The reduced output power could degrade the frequency stability of a clock employing this laser. The behaviors of the two low threshold lasers were consistent, as were the behaviors of the four high threshold lasers. Linewidth versus inverse power curves are presented in Fig. 2 for lasers D and E before and after neutron exposure. The low threshold lasers show no measurable changes due to this exposure. In contrast, the high threshold lasers show increases in their linewidths upon exposure. Again, the potential for changes in linewidth upon neutron exposure would have to be taken into account when considering the application of these lasers in atomic clocks.

Wavelength versus injection current curves for a low threshold laser (D) and a high threshold-laser (E) before and after neutron exposure are displayed in Fig. 3. The apparent changes in laser D's tuning curve upon neutron exposure fall within our normal range of tuning curve reproducibility. Consequently, we cannot state that the radiation exposure had any effect on this laser's tuning. However, laser E displayed a significant change in its tuning characteristics. It is apparent that wavelengths accessible prior to irradiation are no longer attainable after exposure. This could be a serious limitation to the use of this laser in an atomic clock. Again, the two low threshold lasers displayed similar lacks of sensitivity to neutron exposure. Two of the four high threshold lasers were unaffected by the radiation, while the other two showed the effects just discussed.

Review of the diode laser gain curves may give some indication of the origins of the neutron-induced effects. One aspect of the information supplied in these curves is summarized on the graphs shown on Fig. 4. Laser C, representing the low threshold lasers, is unaffected by the neutron exposure. Laser E, a high threshold laser, shows a consistent shift of its gain curve peak to higher energies after neutron exposure. This is somewhat surprising, since the increased threshold currents observed after exposure would indicate increased active region temperature during lasing. Consequently, a shift to lower energies (longer wavelength emission) after neutron exposures would have been expected.

The lack of sensitivity of the low threshold lasers' output power versus injection current curves to this level of neutron exposure is consistent with the findings of Barnes [4]. In the present study, we find that this behavior extends to a range of optical characteristics relevant to optical pumping as would be performed in an atomic clock.

Conclusions

Application of the diode lasers of the type used in this study has the potential to significantly improve the frequency stabilities of the Rb and Cs standards. It is apparent though, that neutron exposure can lead to tuning curve modifications, increased spectral linewidths, and reduced output powers. Furthermore,

these results should extend to AlGaAs diode lasers in general. All of these effects can result in degraded atomic standard performance or potential failure. This is not to say, though, that these devices cannot be used in atomic standards subject to a neutron environment. Rather, if the ultimate operational environment has the potential for neutron exposure, care must be taken in use of the standards. As a first step to reducing a standard's potential sensitivity to neutron exposure, diode lasers with thresholds below typical values should be employed. Additionally, it would be wise to have a laser wavelength control system with sufficient sophistication to correct any small wavelength shifts that might occur upon exposure. To this point, we have addressed the effects of neutron exposure on the operating characteristics of diode lasers at a phenomenological level. With further exposures at increasing fluences and additional analysis, we hope to obtain a more fundamental understanding. Results of this more complete study will be forthcoming.

References

[1] J. C. Camparo and R. P. Frueholz, "Fundamental Stability Limits for the Diode-Laser-Pumped Atomic Frequency Standard," *J. Appl. Phys.*, vol. 59, pp. 3313-3317, May 1986.

[2] R. E. Drullinger, J. H. Shirley, D. J. Glaze, and L. Hollberg, "An Optically Pumped Primary Frequency Standard," in *Proceedings of the Fourth Symposium on Frequency Standards and Metrology*, 1989, pp. 116-119.

[3] C. E. Barnes, "Increased Radiation Hardness of GaAs Lasers at High Current Densities," *J. Appl. Phys.*, vol. 45, pp. 3485-3489, August 1974.

[4] C. E. Barnes, "Neutron Damage Effects in Laser Diodes," in *Proceedings of SPIE*, vol. 328, 1982, pp. 88-95.

[5] C. E. Barnes, "The Effect of Neutron Irradiation on the High Temperature Operation of Injection Laser Diodes," in *Proceedings of SPIE*, vol. 506, 1984, pp. 21-223.

[6] J. C. Camparo, "The Diode Laser in Atomic Physics," *Contemp. Phys.*, vol. 26, pp. 44-477, September/October 1985.

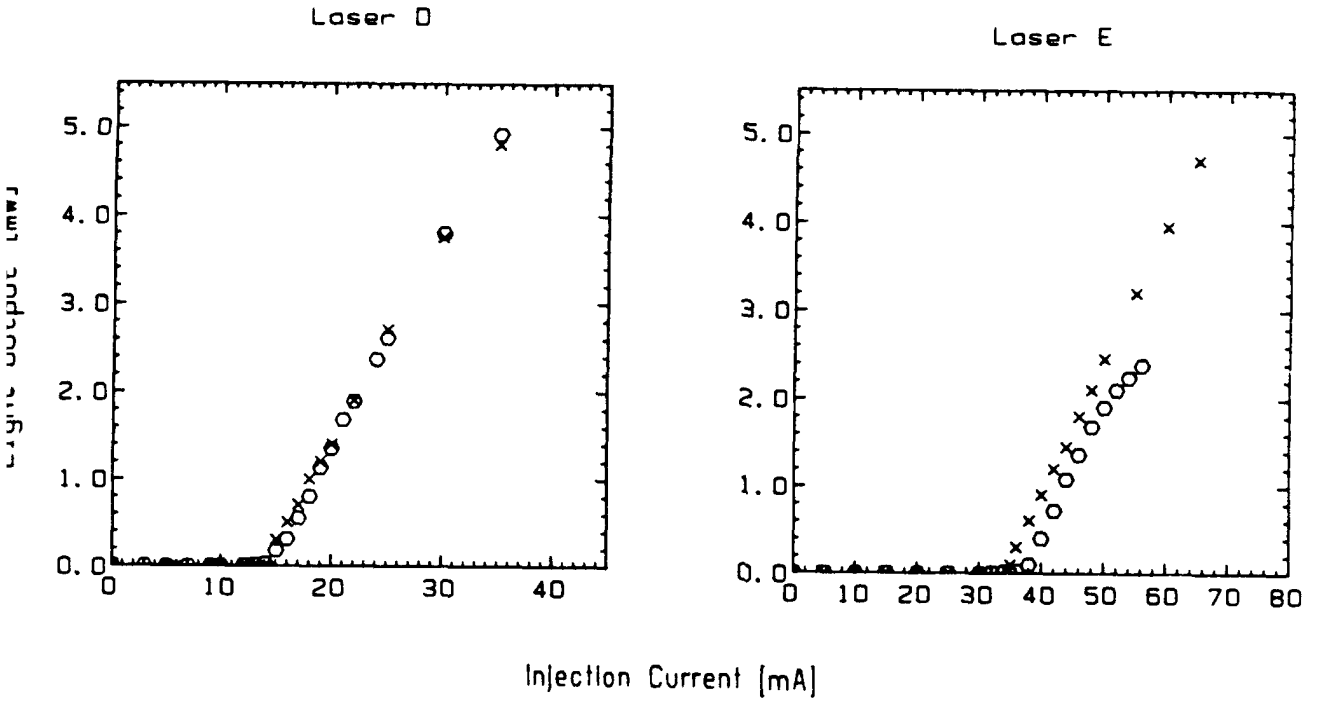


Fig. 1. Output power versus injection current for a low threshold laser (D) and a high threshold laser (E) before (X) and after (○) neutron exposure.

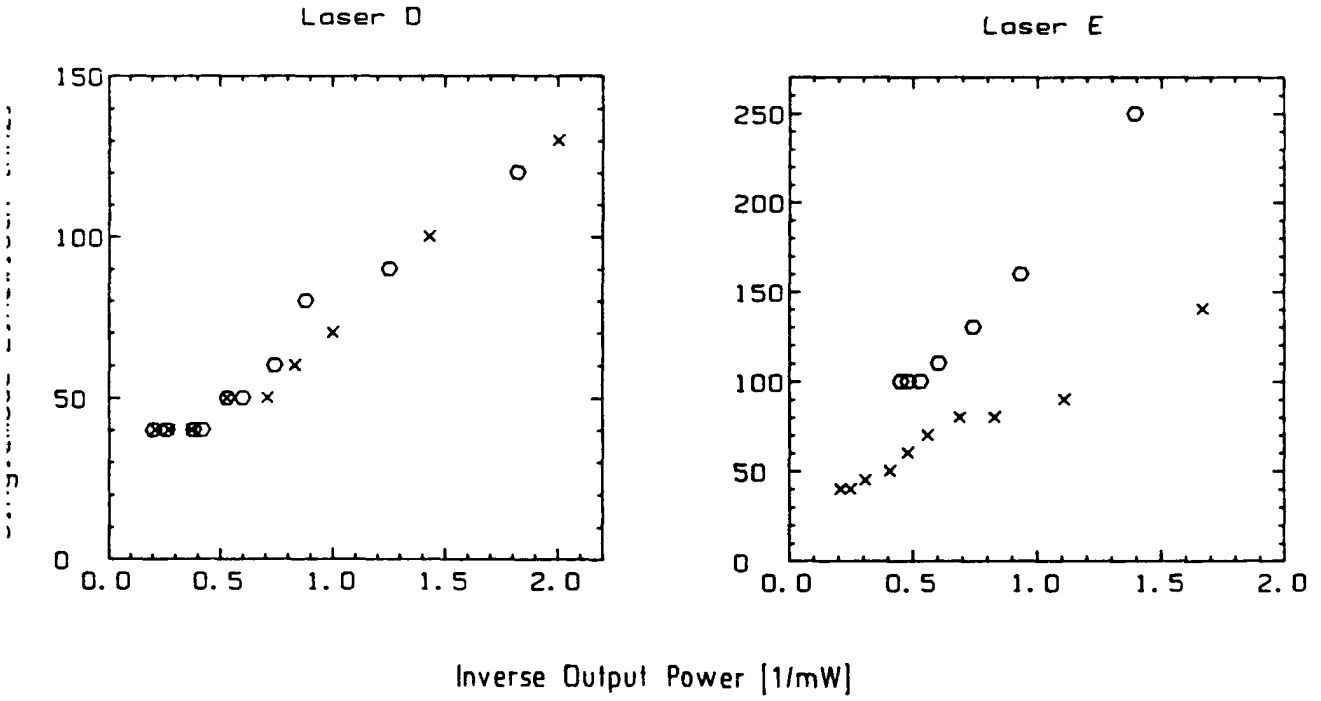


Fig. 2. Linewidth versus inverse power for a low threshold laser (D) and a high threshold laser (E) before (X) and after (○) neutron exposure.

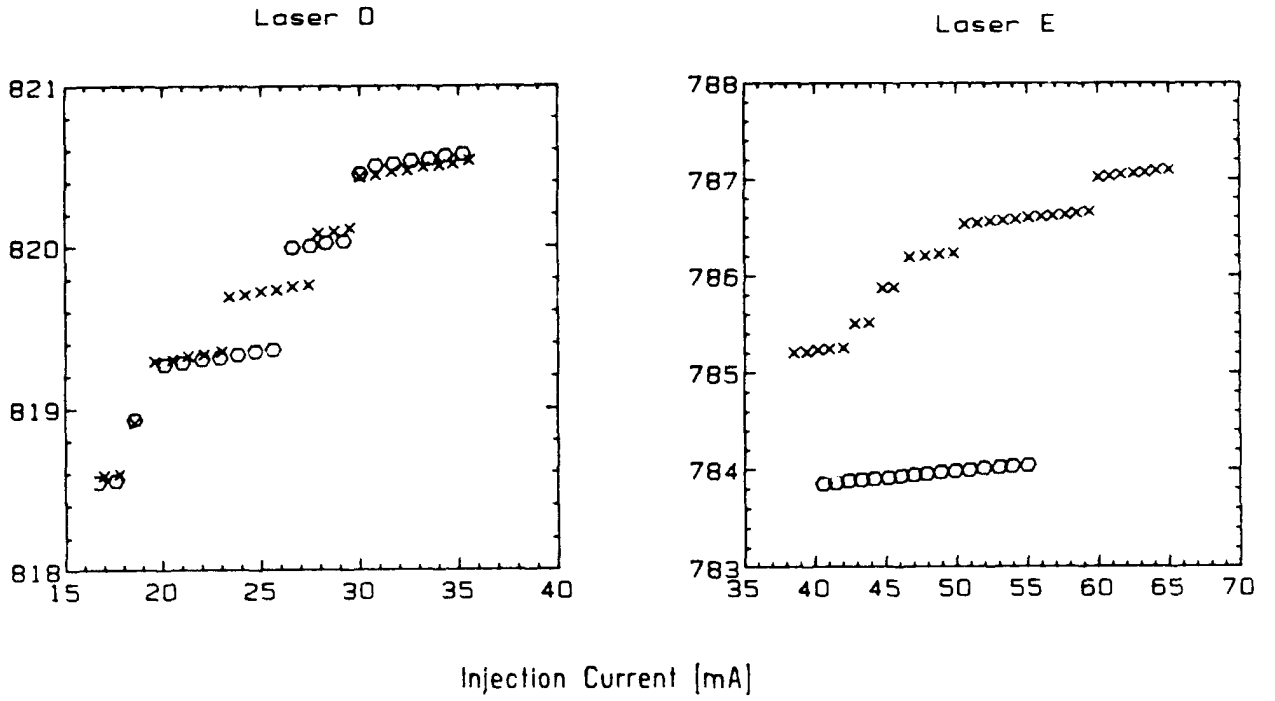


Fig. 3. Wavelength versus injection current for a low threshold laser (D) and a high threshold laser (E) before (X) and after (○) neutron exposure.

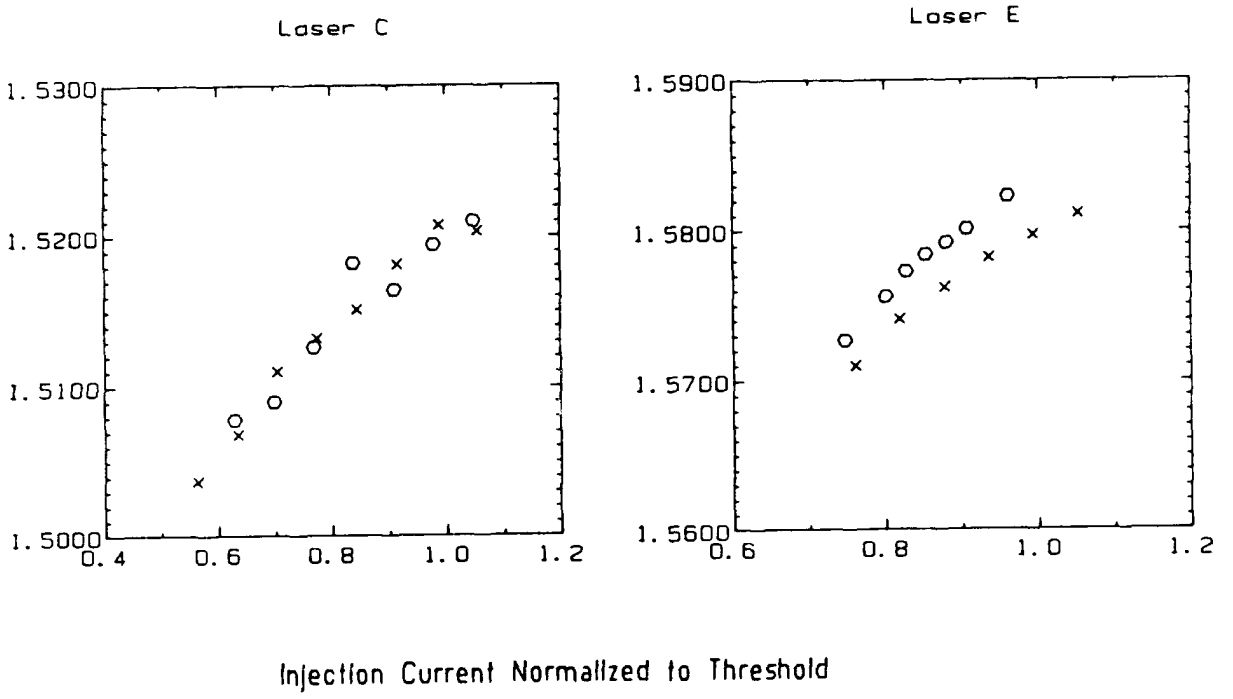


Fig. 4. Energy of gain curve for a low threshold laser (C) and a high threshold laser (E) before (X) and after (○) neutron exposure.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY LOCKING OF LASER DIODES USING AN OPTICALLY PUMPED CESIUM BEAM TUBE

Kenji HISADOME and Masami KIHARA
NTT Transmission Systems Laboratories
1-2356 Take, Yokosuka, Kanagawa, 238-03 Japan

ABSTRACT

A new optically pumped cesium beam tube has been designed. This tube is 680 mm long, and has two cesium beam ovens and four atom-laser interaction zones. The frequency of a laser diode is locked to the fluorescence from the cesium beam. The frequency fluctuation of this frequency-locked laser diode and the short-term frequency stability of the optically pumped cesium beam frequency standard using this laser diode are theoretically estimated according to the measured beam current values. Also, the directivity of the cesium beam tube is measured to determine the proper operational condition. When oven temperature is 130 °C, a practical lifetime of about 4 years (1 g Cs metal enclosed) and a good short-term frequency stability of $\sigma_{yO}(\tau) = 3\times10^{-12}/\sqrt{\tau}$ can be obtained.

1. INTRODUCTION

Cesium beam frequency standards generate an accurate and stable frequency synchronized to the "clock transition" frequency (9.192631770 GHz) of cesium atoms using deflection magnets and the Ramsey resonance method[1]. Their frequency accuracy and long-term frequency stability are better than rubidium gas cell frequency standards and oven controlled crystal oscillators, and they are smaller than hydrogen maser frequency standards. They are used in several fields: navigation, communication, time keeping, instrumentation, etc.

Optically pumped cesium beam frequency standards use lasers instead of conventional deflection magnets for atomic state preparation and microwave resonance detection. Their frequency accuracy, long-term frequency stability, and short-term frequency stability are expected to be better than conventional cesium beam frequency standards. With recent advances in laser diode technology, several laboratories are in the process of developing them[2-15].

NTT is developing a small optically pumped cesium beam frequency standard as a clock frequency source for NTT's synchronous digital communication networks[16]. A new optically pumped cesium beam tube is created as a quantum resonator for this optically pumped cesium standard. This paper reports a scheme and preliminary results of the new cesium bean tube. Short-term frequency stability and the proper operational condition for the optically pumped cesium standard using this tube are also discussed.

2. CESIUM BEAM TUBE

Scheme

The scheme of the optically pumped cesium beam tube is shown in figure 1. It is 680 mm long and has two cesium beam ovens and four atom-laser interaction zones. With these ovens and interaction zones, it is possible to simultaneously operate two counter-propagating beams[17], to cancel out a cavity phase shift by summing up the resonance signals obtained from the two beams(Fig. 2). A more detailed explanation is given in the appendix.

Windows and mirrors

The cesium beam tube has four pairs of windows for laser beams and four spheroidal mirrors for fluorescence detection. These windows are made of anti-reflection coated fused silica and their reflectivity at 850 nm is less than 0.4

CH2818-3/90/0000-044 $1.00 © 1990 IEEE

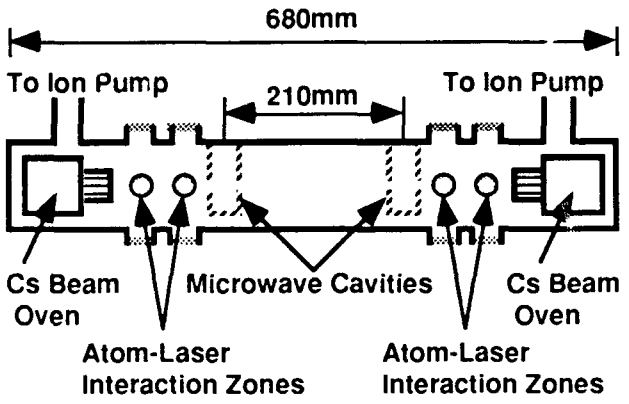


Figure 1. Optically Pumped Cs Beam Tube.

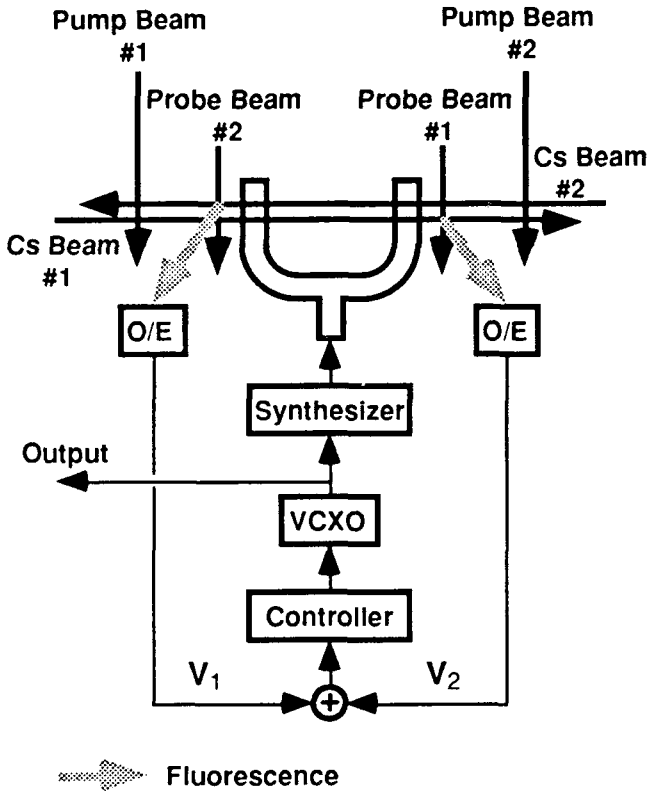


Figure 2. Cancellation of Cavity Phase Shift by Two Stimultaneous Counter Propagating Cs Beams.

%. The laser beam and cesium beam, running perpendicular to the paper, cross each other at one focus of the mirror in figure 3. Fluorescence from the cesium beam is focused by the mirror onto a silicon photodiode located at the other focus.

Ovens

The ovens have beam collimators made of copper. The temperatures of the beam collimators are measured by thermistor sensors and their fluctuations are controlled to less than 0.1 °C. The beam collimator is composed of channels each of whose lengths are 17 mm and whose cross section areas are 0.03 mm^2. The directivity of the beam collimator is shown in section 5.

Material

The energy levels of atoms generally change due to the Zeeman effect caused by environmental magnetic fields. Therefore the material for the beam tube must be nonmagnetic to avoid unexpected Zeeman frequency offset. Moreover, gas emission from the material must be kept to minimum because the mean free path of the cesium atoms in the cesium beam tube must be much longer than the beam tube length. Consequently, aluminium alloy is employed to satisfy the above conditions.

Vacuum

Two ion vacuum pumps are used for an evacuation speed of 10 liters per second are used. The vacuum level in the beam tube can be held at about 10^{-8} Torr. The mean free path at this vacuum level is on the order of 10^2 m.

3. OBSERVATION OF Cs - D2 LINE

An absorption line from the ground state $|6^2S\,1/2\rangle$ to the excited state $|6^2P3/2\rangle$, the Cs - D2 line (resonant wavelength in vacuum is 852.35 nm), is used for the optical pumping. These states have the hyperfine structures shown in figure 4 due to nuclear spin interaction. In this figure, F and F' are the total angular momentum quantum number in the ground state $|6^2S\,1/2\rangle$ and the excited state

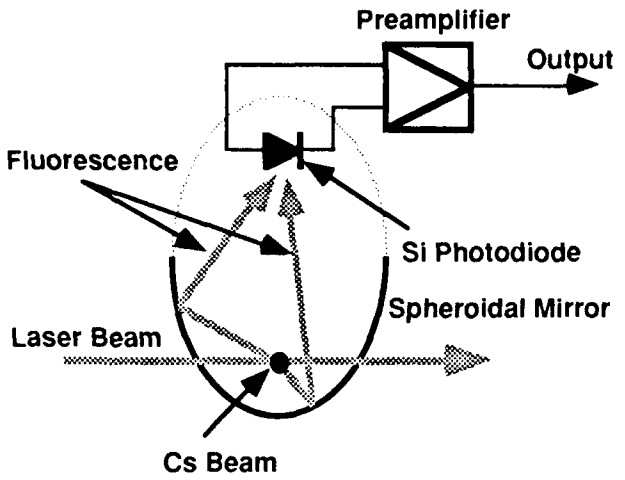


Figure 3. Spheroidal Mirror for Fluorescence Detection.

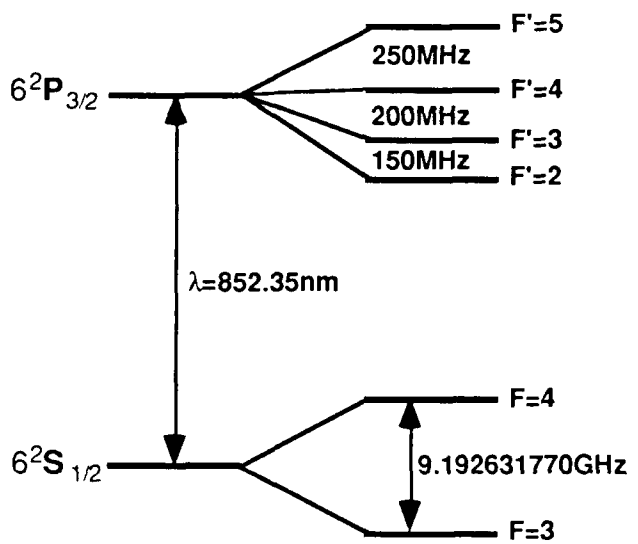


Figure 4. Hyperfine Structures in Cs-D_2 Line.

$|6^2P3/2\rangle$ respectively. According to the selection rule, there are six allowed transitions (F = 3 → F' = 2, 3, 4 and F = 4 → F'= 3, 4, 5) in the Cs - D_2 line.

Fluorescence corresponding to the Cs - D_2 line is observed with the cesium beam tube and temperature-stabilized laser diode module. Experimental conditions are as follows.

- Temperature fluctuation of LD module; $< 10^{-2}$ K.
- Input optical power; -4 dBm.
- Oven temperature; 150 °C.
- Sweeping range of laser diode injection current; 0.1 μA.

The fluorescence from the cesium beam is focused onto the silicon photodiode whose detection area is 33 mm^2, and its photocurrent is amplified by a low noise FET operational amplifier.

The observation results are shown in figure 5(a). Three transition lines (F = 4 → F'= 3, 4, 5) can be found for 0.1 μA sweeping range of injection current (900 MHz sweeping range of laser frequency). Peak value for cycling transition line (F = 4 → F'= 5) is 400 nA (in atomic state preparation zone) and 20 nA (in microwave resonance detection zone) respectively. Full width at half-maximum of the transition line is 90 MHz.

A frequency discrimination signal is used to lock the frequency of the laser diode to one of the transition lines. To obtain the discrimination signal, rectangular modulation current with an amplitude of 25 μA and a modulation frequency of 10 kHz is added to the injection current of the laser diode. The frequency of the laser diode is shifted ±45 MHz with the modulation current. The synchronously detected result of the fluorescence signal is shown in figure 5(b). The three zero-crossing points correspond to the three peaks of the fluorescence signal in figure 5(a).

4. ESTIMATION OF SHORT-TERM FREQUENCY STABILITY

Based on the experimental results obtained in the previous section, the short-term frequency stability of our optically pumped cesium beam frequency standard is theoretically estimated, considering both the frequency and power fluctuation of the laser diode. The estimated short-term frequency stability is ultimate, because 150 °C in the previous experimental conditions is the highest oven tem-

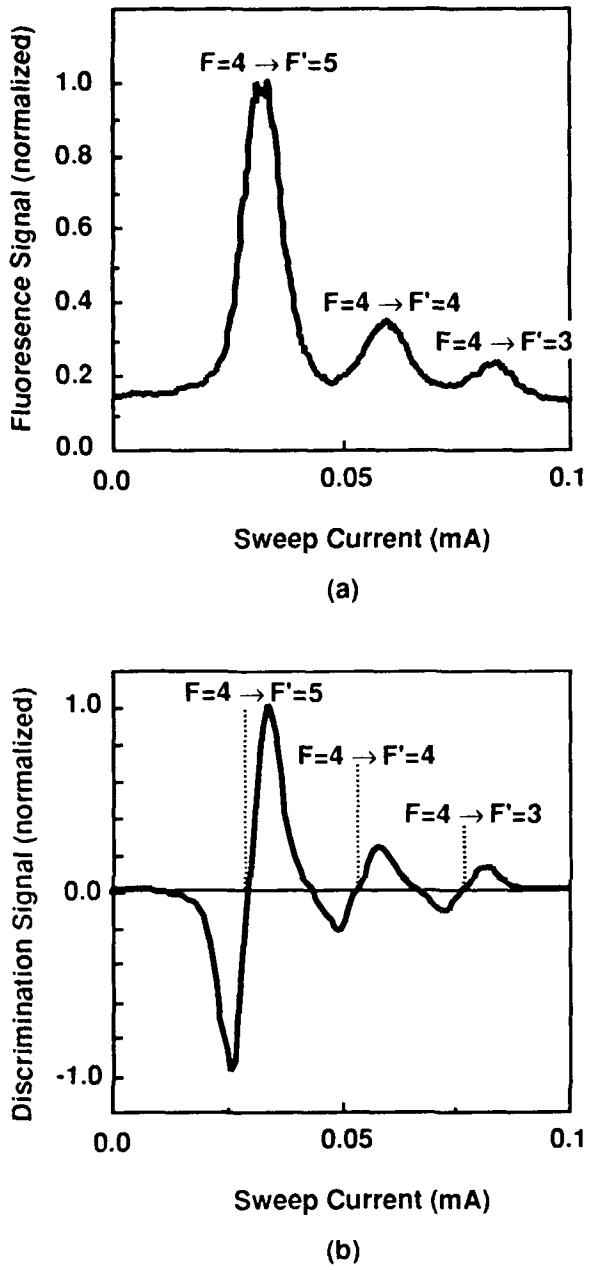


Figure 5. Fluoresecence Signal (a) and Frequency Discrimination Signal (b).

perature allowed for aluminum alloy used as cesium beam tube material.

A block diagram of the optically pumped cesium standard analyzed in this section is shown in figure 6. In this figure, laser diode frequency is stabilized by fluorescence from the interaction zone close by the cesium beam oven.

Laser diode

When modulation current $\pm\Delta I$ with a frequency of f_L, is added to the injection current, the LD frequency is given by

$$\nu_L(t) = \nu_{LO}(t) \pm \Delta\nu_L$$

$$\nu_{LO}(t) = \nu_{LF} - K_L I_C(t) + \delta\nu_L(t) \qquad (1)$$

with $\Delta\nu_L = K_L \Delta I$

where

ν_{LF}: free-running LD mean frequency
$I_C(t)$: control current
$\delta\nu_L(t)$: free-running LD frequency fluctuation
K_L : constant.

The fluorescence for cycling transition can be approximated by Lorentzian shape. When $\Delta\nu_L$ is equal to one half the transition linewidth, the fluorescent signal is approximated by

$$V_L(t) \cong \frac{1}{2}\{1 + \frac{\nu_r - \nu_{LO}(t)}{W_L/2}\}\alpha\{P_O + \delta P(t)\} + V_n(t) \qquad (2)$$

where

P_O: mean power
$\delta P(t)$: power fluctuation
$V_n(t)$: noise in the control system (beam shot noise, atom-photon conversion noise, stray light shot noise, etc.[18])
W_L: transition linewidth
ν_r: resonance frequency
α: coefficient dependent on collection efficiency of the mirror, quantum efficiency of the photodiode, beam intensity, etc.

The control current is given by

$$I_C(t) \equiv h_L(t) \otimes \{\frac{1}{2}\alpha P_O \frac{\nu_r - \nu_{LO}(t)}{W_L/2} + \frac{1}{2}\alpha \cdot \delta P(f_L;t) + V_n(f_L;t)\} \tag{3}$$

where

$h_L(t)$; integrator impulse response

$\delta P(f;t)$; f component of $\delta P(t)$

$V_n(f;t)$; f component of $V_n(t)$.

In equation (3), $\otimes$ represents the convolution integral.

From equations (1) and (3), the two-sample variance of the frequency fluctuation of the stabilized laser diode is given by

$$\sigma_{yL}^2(\tau) = \frac{1}{2Q_L^2}\{\frac{S_{\delta P}(f_L)}{P_C^2} + \frac{S_{Vn}(f_L)}{(\alpha P_O/2)^2}\}\frac{1}{\tau} \tag{4}$$

with

$$Q_L = \nu_r / W_L$$

where

$S_{\delta P}(f)$: power spectral density of $\delta P(t)$

$S_{Vn}(f)$: power spectral density of $V_n(t)$.

In this equation, τ is an average time of a frequency.

Optically pumped Cs standard

The microwave frequency, synthesized from the VCXO frequency and modulated by a rectangular signal whose frequency is f_O, is given by

$$\begin{aligned} \nu_{M0}(t) &= M\nu_O(t) \\ \nu_M(t) &= \nu_{M0}(t) \pm \Delta\nu_O \\ \nu_O(t) &= \nu_{OF} - K_O V_C(t) + \delta\nu_O(t) \end{aligned} \tag{5}$$

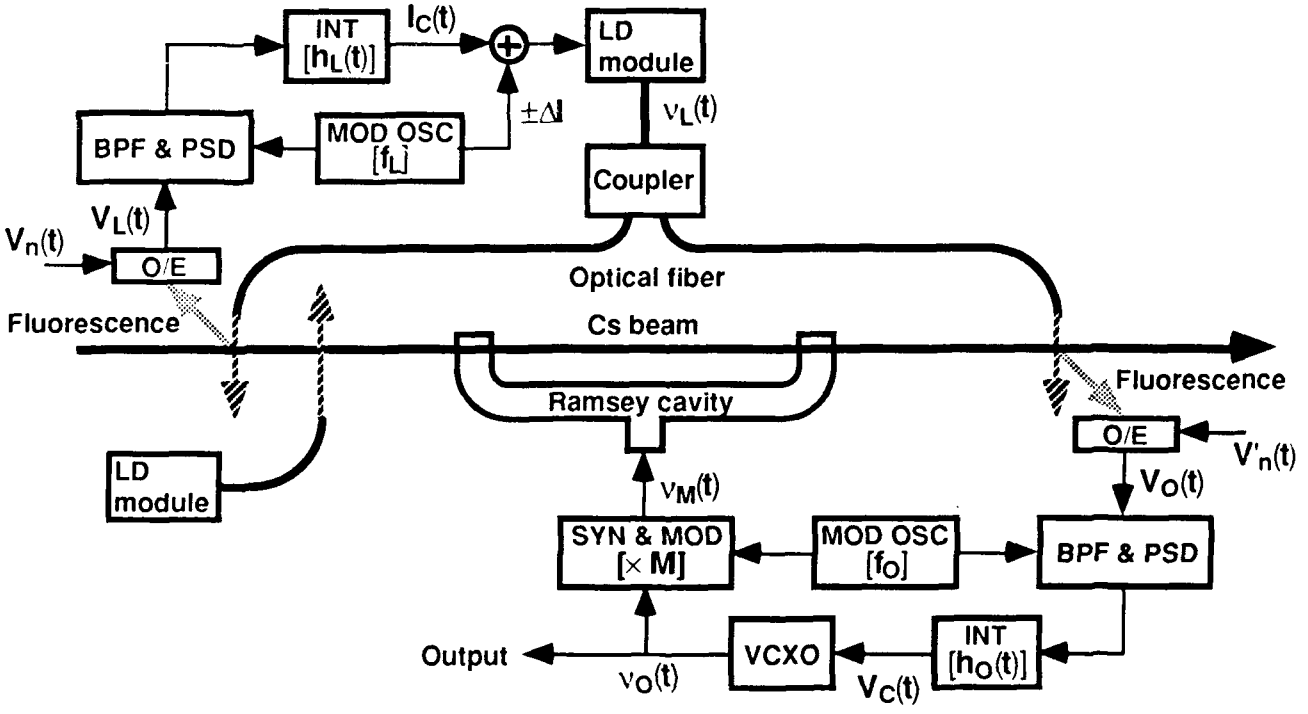


Figure 6. Optically Pumped Cs Standard Analyzed in Sec. 4.

BPF: band pass filter, PSD: phase sensitive detector
INT: integrator, MOD OSC: modulation oscillator
SYN: synthesizer, MOD: modulator

where

ν_{OF}: free-running VCXO mean frequency
$\delta\nu_O(t)$: free-running VCXO frequency fluctuation
K_O: constant
$V_C(t)$: control voltage
M: multiplication rate of the microwave synthesizer.

Microwave transition probability can be approximated by

$$\cos^2\left[\frac{\pi\{\nu_{CL}-\nu_M(t)\}}{2W_O}\right]$$

where

W_O: Ramsey resonance linewidth
ν_{CL}: "clock transition" frequency.

When $\Delta\nu_O = W_O/2$, the fluorescent signal is approximated by

$$V_O(t) = \frac{1}{2}\left[1\pm\frac{\pi\{\nu_{CL}-\nu_{MO}(t)\}}{W_O}\right]\times\frac{1}{2}\left\{1+\frac{\delta\nu_L(t)}{W_L/2}\right\}$$
$$\times\alpha'\{P_O+\delta P(t)\}+V'_n(t) \quad (6)$$

where

$V'_n(t)$: noise in the control system.
α': coefficient.

The control voltage is given by

$$V_C(t) = h_O(t)\otimes\left[\frac{1}{4}\alpha' P_O\frac{\pi\{\nu_{CL}-\nu_M(t)\}}{W_O}\right.$$
$$\left.+\frac{1}{4}\alpha' P_O\frac{\delta\nu_L(f_O;t)}{W_L/2}+\frac{1}{4}\alpha'\cdot\delta P(f_O;t)+V_n(f_O;t)\right] \quad (7)$$

where

$h_O(t)$; integrator impulse response
$\delta\nu_L(f;t)$; f component of $\delta\nu_L(t)$.

From equations (5) and (7), the two-sample variance of the frequency fluctuation of the optically pumped cesium standard is given by

$$\sigma_y^2(\tau) = \frac{1}{8\pi^2 Q_O^2}\left\{\frac{S_{\delta P}(f_O)}{P_O^2}+Q_L^2 S_{yL}(f_O)+\frac{4S_{Vn}(f_O)}{(\alpha' P_O/2)^2}\right\}\frac{1}{\tau} \quad (8)$$

with

$$Q_O = M\nu_{CL}/W_O.$$

In this equation, the first term reflects the power fluctuation of the laser diode, the second the frequency fluctuation of the laser diode, and the third the noise in the control system.

$\sigma_{yL}(\tau=1s)$ and $\sigma_{yO}(\tau=1s)$ as calculated for $S_{\delta P}(f)/P_O^2$ are shown in figure 7. The parameters for the calculations are shown in table 1. These parameters are obtained from the measured results in the previous section. In figure 7, the ultimate values of the square root of the two-sample variance of the frequency fluctuation of the frequency-locked laser diode and the optically pumped cesium standard using the laser diode are $\sigma_{yL}(\tau) = 4.5\times10^{-13}/\sqrt{\tau}$ and $\sigma_{yO}(\tau) = 9\times10^{-13}/\sqrt{\tau}$, respectively, when $S_{\delta P}(f)/P_O^2<10^{-11}$/Hz. This $S_{\delta P}(f)/P_O^2$ value can be easily realized.

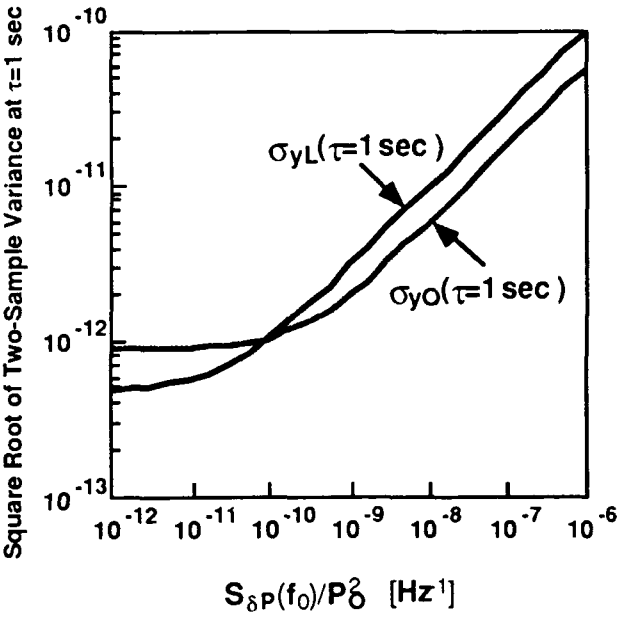


Figure 7. Frequency Fluctuations of Laser Diode and Optically Pumped Cs Standard.

5. PROPER OPERATIONAL CONDITION

In general, the consumption of cesium in a cesium beam oven and saturation of the carbon getters degrade the lifetime of cesium beam tubes. Fogging of the windows and mirrors by cesium atoms can also degrade the lifetime of a cesium beam tube when optical pumping is used. Off-axis atoms causes these degradations. Thus, it is necessary to measure the directivity of the beam collimator to determine the proper operational condition for cesium beam tubes.

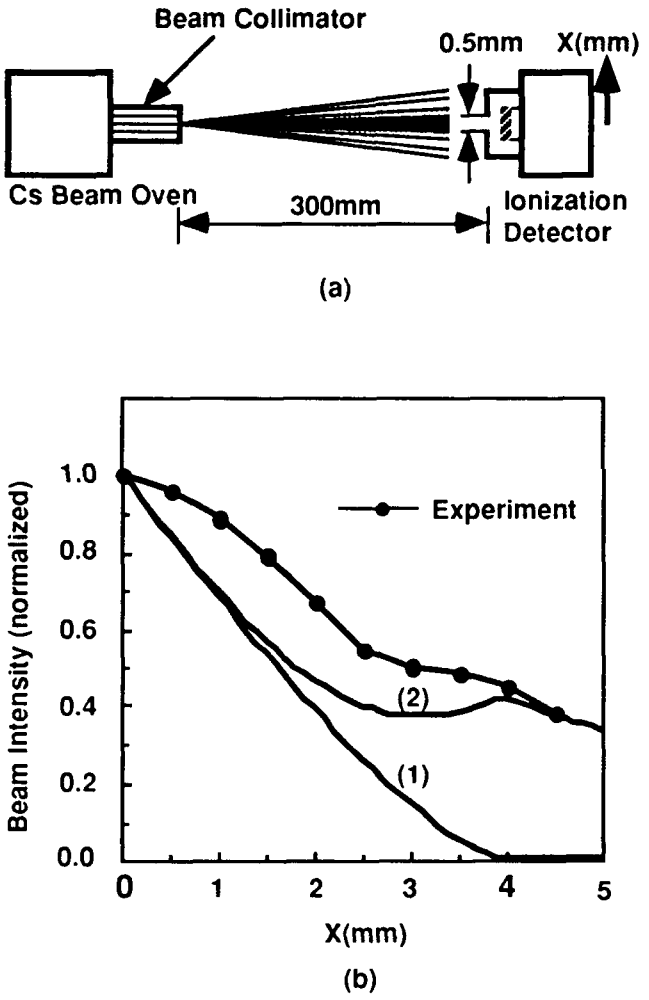


Figure 8. Measurement Scheme of Beam Collimator Directicity (a) and Measured Beam Pattern (b).

Directivity of beam collimator

The directivity of an experimental beam collimator similar to the beam collimator in the cesium beam tube is measured. The measurement scheme is shown in figure 8(a). An ionization detector whose detection slit is 0.5 mm × 5 mm is placed far 300 mm from the beam collimator and is moved perpendicular to the beam axis.

The measured beam pattern is shown in figure 8(b). Theoretical patterns(19) are shown by lines (1) and (2) in this figure. Line (1) ignores re-emission from the walls of the collimator and line (2) considers it. The measured beam pattern is affected by re-emission.

Transmission probability (inverse of peaking factor, κ) indicates the directivity of the beam collimator. The κ^{-1} obtained from the theoretical beam pattern (2) is 0.018, and the κ^{-1} obtained from the measured beam pattern is 0.023.

Proper operational condition

In the following calculation, it is assumed that only cesium consumption limits the lifetime of the cesium beam tube. This consumption can be calculated by the measured transmission probability and the overall cross section area (0.36 mm^2) of the beam collimator. The relation between lifetime (1 g Cs enclosed), short-term frequency stability and oven temperature is shown in figure 9. The lifetime and short-term frequency stability of NTT's present cesium standard are also shown in this figure.

The proper operational condition can be determined by this graph. When oven temperature is 130 °C, a lifetime of about 4 years per gram, and a short-term frequency stability of $\sigma_{yO}(\tau) = 3\times10^{-12}/\sqrt{\tau}$ can be obtained. The lifetime is longer than that of NTT's present cesium standard, and the short-term frequency stability is less than one tenth that of NTT's present cesium standard.

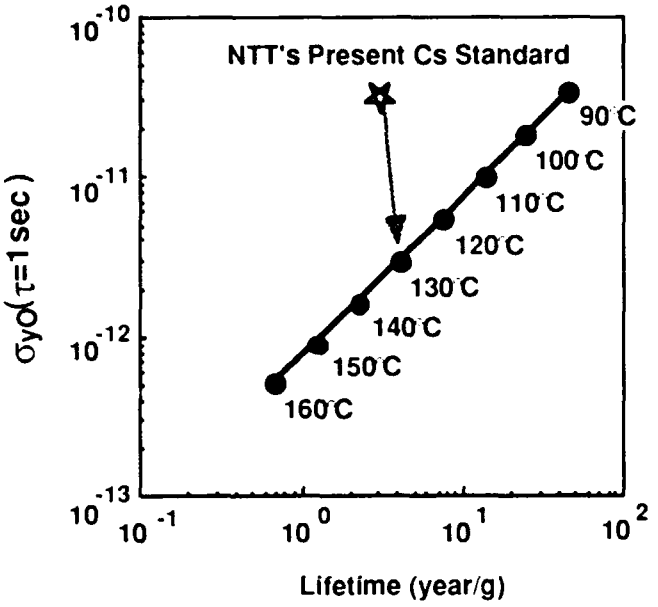


Figure 9. Lifetime, Frequency Stability and Oven Temperature of Optically Pumped Cs Beam Tube.

6. SUMMARY

A new optically pumped cesium beam tube has been created. This cesium beam tube is 680 mm long, and has two cesium beam ovens and four atom-laser interaction zones. Measured fluorescent current is 20 nA in the microwave resonance detection zone, when oven temperature is 150 °C and input optical power is -4 dBm.

Short-term frequency stability of our new optically pumped cesium beam frequency standard is theoretically estimated according to the measured values. The ultimate value of the square root of the two-sample variance of the frequency fluctuation is $\sigma_{yO}(\tau) = 9\times10^{-13}/\sqrt{\tau}$.

Using the measured directivity of the beam collimator, the proper operational condition for the cesium beam tube is determined. A practical lifetime of about 4 years (1 g Cs metal enclosed) and a good short-term frequency stability, $\sigma_{yO}(\tau) = 3\times10^{-12}/\sqrt{\tau}$ can be obtained when the temperature is 130 °C.

ACKNOWLEDGMENTS

The authors wish to thank Dr. T. Miki, Executive Manager; Dr. M. Koyama, Executive Manager; and Mr. K. Kikuchi, Research Group Leader of the NTT Transmission Systems Laboratories, for their encouragement and suggestions.

REFERENCES

1. J. Vanier and C. Audoin, *The quantum physics of atomic frequency standards*, Vol. 2, Chapter 5, Adam Hilger, Bristol, 1989.

2. S. Ohshima, Y. Nakadan, T. Ikegami, Y. Koga, R. Drullinger, and L. Hollberg, "Characteristics of an optically pumped Cs frequency standard at the NRLM", *IEEE Trans. Instrum. Meas.*, IM38, 2, pp. 533-536, 1989.

3. S. Ohshima, Y. Nakadan, T. Ikegami, and Y. Koga, "A beam reversal experiment for the estimation of microwave power shifts in an optically pumped Cs beam frequency standard", *IEEE Trans. Instrum. Meas.*, IM38, 6, pp. 1100-1103, 1989.

4. V. Gandelier, V. Giordano, A. Hamel, G. Theobald, P. Cerez, and C. Audoin, "Frequency stability of an optically pumped cesium beam frequency standard", *Appl. Phys.*, B49, pp. 365-370, 1989.

5. V. Giordano, A. Hamel, P. Petit, G. Theobald, N. Dimarcq, P. Cerez, and C. Audoin, "New design for an efficient optically pumped cesium beam tube", *Proc. 43rd Annual Symposium on Frequency Control*, pp. 130-134, 1989.

6. L. L. Lewis, "Low noise laser for optically pumped cesium beam standards", *Proc. 43rd Annual Symposium on Frequency Control*, pp. 151-157, 1989.

7. E. De Clercq, A. Clairon, B. Dahmani, A. Gerard, and P. Aynie, "Design of an optically pumped Cs laboratory frequency standard", in *Frequency standards and metrology*, ed. A. De Marchi, pp. 120-125, 1989.

8. W. Yiqiu, X. Linzhen, W. Depei, and Y. Shutong,

"Preliminary results and some considerations on optically pumped cesium beam frequency standard", in *Frequency standards and metrology*, ed. A. De Marchi, pp. 126-131, Springer, Berlin, 1989.

9. P. Thomann, H. Schweda, and G. Busca, "Preliminary results on small optically pumped cesium resonator", in *Frequency standards and metrology*, ed. A. De Marchi, pp. 392-394, Springer, Berlin, 1989.

10. D. H. Yang and Y. Q. Wang, "Detection by cycling transition in an optically pumped cesium beam frequency standard", *Optics Communications*, 73, 4, pp. 285-288, 1 989.

11. J. Umezu, B. Komiyama, H. Saitho, and Y. Ohta, "Experiments on optically pumped cesium beam frequency standard", *Proc. 2nd European Frequency and Time Forum*, p. 521, 1988.

12. T. McClelland, I. Pascaru, J. Zacharski, N. H. Tran, and M Meirs, "An optically pumped cesium beam frequency standard for military applications", *Proc. 41st Annual Symposium on Frequency Control*, pp. 59-65, 1987.

13. B. Dahmani, L. Hollberg, and R. E. Drullinger, "Frequency stabilization of semiconductor lasers by resonant optical feedback", *Optics Letters*, 12, 11, pp. 876-878, 1987.

14. R. E. Drullinger, J. Shirley, D. J. Glaze, L. W. Hollberg, and A. DeMarchi, "Toward an optically pumped cesium beam frequency standard", *Proc. 41st Annual Symposium on Frequency Control*, pp. 428-431, 1986.

15. A. Derbyshire, R. E. Drullinger, M. Feldman, D. J. Glaze, D. Hilliard, D. A. Howe, L. L. Lewis, J. H. Shirley, I. Pascaru, and D. Stanciulescu, "Optically pumped small cesium beam standards; A status report", *Proc. 40th Annual Symposium on Frequency Control*, pp. 18-21, 1985.

16. M. Kihara and K. Hisadome, "Design of high performance, portable, optically pumped cesium beam standards", *CPEM'88 Digest*, 248-249, 1988.

17. L. L. Lewis, M. Feldman, and J. C. Bergquist, "Impact of lasers on primary frequency standards and precision spectroscopy", *J. Phys.*, 42, C8, pp. 271-281, 1981.

18. V. Giordano, V. Candelier, A. Hamel, C. Audoin, G. Theobald, and P. Cerez, "Noise in the optical detection of atoms in a beam", *Optics Communications*, 67, 4, pp. 287-292, 1988.

19. J. A. Giordmaine and T. C. Wang, "Molecular beam formulation by long parallel tubes", *J. Appl. Phys.*, 31, 3, pp. 463-471, 1960.

APPENDIX

In figure 2, the two resonance signals, V_1 and V_2 are given by

$$V_1 = V_{10}\cos^2\left[\frac{\pi(\nu_{CL}-\nu_M+\Delta\nu_{CPS})}{2W_O}\right]$$

$$V_2 = V_{20}\cos^2\left[\frac{\pi(\nu_{CL}-\nu_M-\Delta\nu_{CPS})}{2W_O}\right]$$

where $\Delta\nu_{CPS}$ is cavity phase shift.

V_1 and V_2 are summed up and synchronously detected. The detected signal, V_D is given by

$$V_D = K_D\left(\nu_{CL}-\nu_M+\frac{V_{10}-V_{20}}{V_{10}+V_{20}}\Delta\nu_{CPS}\right)$$

where K_D is a constant.

When $V_{10} \equiv V_{20}$, the cavity phase shift $\Delta\nu_{CPS}$ can be cancelled out.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ZERO-CROSSING TECHNIQUE FOR CLOCK-TRANSITION DETECTION IN A Rb FREQUENCY STANDARD

A. STERN M. GOLOSOVSKY

T.F.L. Time & Frequncy Ltd.
Holon, 58117, Israel

Abstract

We propose a new technique of detection of the clock-transition, in a Rb frequency standard, by means of magnetic field modulation. In this technique the microwave is held constant and magnetic field is rapidly swept across the zero value. This zero crossing produces a transient in the transmitted light intensity. The amplitude of the transient is strongly dependent on microwave frequency and shows picks corresponding to the Rb hyperfine transitions. The pick which corresponds to the clock transition can be utilized to lock the servo-loop in a rubidium frequency standard.
Zero crossing detection technique has the obvious advantage that it has a low sensitivity to a stray magnetic field since the signal is generated when the total magnetic field is nearly zero.

Introduction

This paper describes the experimental results and interpretation of effects that are produced by the reversal of the magnetic "C" field in a rubidium atomic frequency standard. We have first encountered these phenomena when alternating the "C" field in order to cancel the disturbance of external field, by means of averaging [1].

In the experiment we sweep the magnetic field through zero, while (a) transmitting light through the Rb^{87} vapor cell and (b), injecting a microwave signal at the clock-transition-frequency into the microwave cavity, where the vapor cell is located. We observe positive and negative transient signals in the transmitted light intensity.
We attribute these signals to non-adiabatic processes that occurs between the Zeeman sublevels of the hyperfine structure of the rubidium atom.

Related but different effects were observed before.

First, the very well known Majorama effect, [2], deals with non-adiabatic transitions that occurs between the Zeeman sublevels of a Cs beam as it passes through zero magnetic field. However, no microwave or pumping light are present.

H.G. Dehmelt, [3], and other authors, [4,5], had investigated the change in the transmission of a polarized light in alkali metals vapor under a reversal of a magnetic field. (No microwave is present). These experimentsa are closely related to the Hanle effect [6], that deals with the change of the intensity and polarization of a scattered light from a vapor, which is irradiated by polarized light and is exposed to a very small magnetic field.

More recently, Camparo & Frueholz, [7], have studied the effect of adiabatic rapid passage of the microwave frequency through resonance. In this case the magnetic field is kept constant and the transmitted light intensity is being recorded, showing positive signals.

In our experiment we observe transients only when a microwave EM field at the clock transition frequency is present. When plotting the transient amplitude versus the microwave frequency we obtain a sort of Ramesy pattern with center line and side lobes. The center line is picked at a frequency which corresponds to the zero field clock transition frequency.

One can make use of this line for the detection of the clock transition and locking the servo-loop in a rubidium frequency standard.

In the following, we describe the experimental set-up, and report the results. We discuss the results and describe a model for their interpretation. Finally, we suggest a double modulation technique for the detection of the clock transition signal produced by the zero-crossing.
A preliminary description of the results had been published before [1].

Experimental

For the experimental set-up we have used our ordinary commercial "Physics Package" which is installed in the Rubidium Frequency Standard model TF-4000A.

CH2818-3/90/0000-053 $1.00 © 1990 IEEE

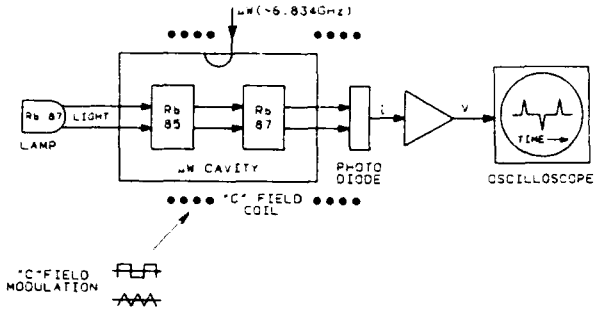


Fig. 1: Experimental set-up; a schematic drawing.

Schematic drawing is shown in fig. 1.
The light from a Rb^{87} lamp is being filtered by a Rb^{85} filter-cell and is transmitted through a Rb^{87} resonance-cell. Both cells are located inside a rectangular microwave cavity which resonates in a pseudo TE101 mode. The cavity is being operated at the clock-transition frequency that corresponds to a magnetic "C" field around 200 mgauss.
The magnetic field is induced by a pair of Helmholtz coils, to produce a highly homogeneous field. The "C" field current is modulated, around zero, in a square-wave or a triangle-wave modulation.
The cavity and coils are magnetically shielded by a double mu-metal shield. The residual magnetic field in the cavity is estimated to be around 1 mgauss.

The transmitted light is detected by a photo-diode whose current is preamplified and fed into an oscilloscope or a chart recorder. This detection-recording chain was verified to have a flat response in the frequency range of interest (for exact measurements of rise and decay times we have used the oscilloscope).

Results

Fig. 2 shows the transients that were recorded with a square-wave modulation and, figs 3-7 exhibit transients that result from a triangle wave modulation. "C" field modulation are depicted by the upper graphs whereas the transient signals are depicted by the lower graph in each figure. Horizontal time scale ranges and "C" field rates of change (slopes) are given in each figure.

Fig. 2 and fig. 3 exhibit a similar behavior, where sharp positive signals appear on each zero-crossing and a smaller but wider positive signals appear on each second zero-crossing.
Figs 3 to 7 exhibit the development of the signals as we decrease the slope, dH/dt.
The negative signals grow, and the positive signals decrease and become wider.

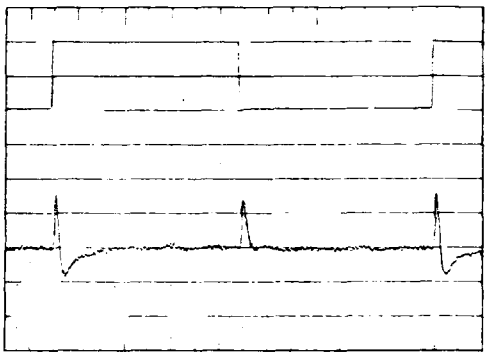


Fig. 2: Transmitted light, square-wave modulation. Total time (horizontal scale) is 50mS.

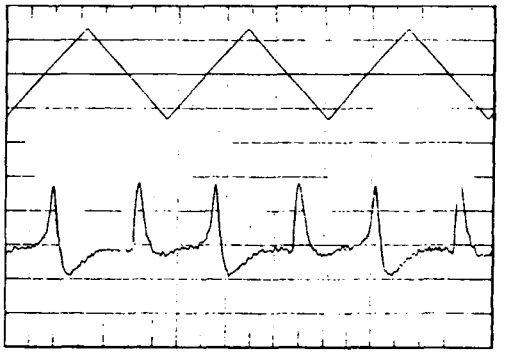


Fig. 3: Transmitted light, triangle-wave modulation. Total time is 30mS, slope is 50gs/sec.

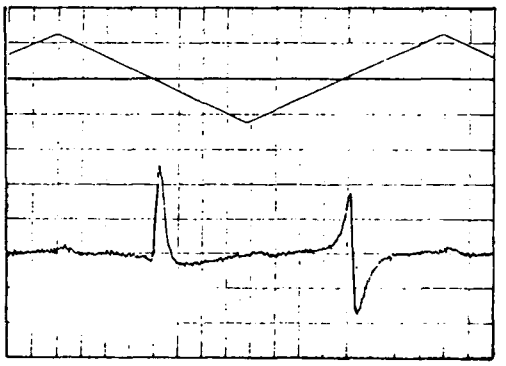


Fig. 4: Transmitted light, triangle-wave modulation. Total time is 50mS, slope is 40gs/sec.

In fig. 8 we plot the positive signals amplitude and width as a function of the slope dH/dt.
We see that the amplitude increases linearly from zero and reaches a constant

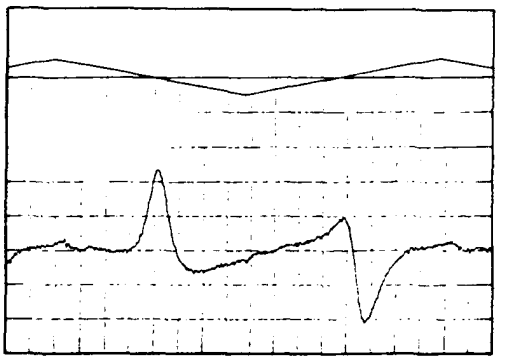


Fig. 5: Transmitted light, triangle-wave modulation. Total time is 50mS, slope is 10gs/sec.

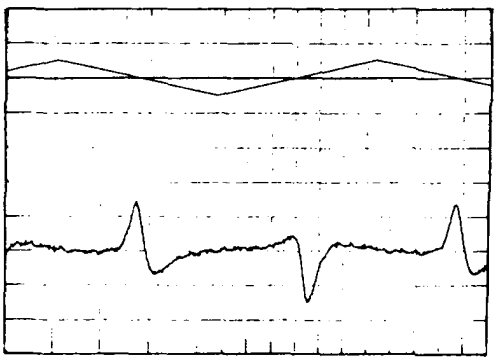


Fig. 6: Transmitted light, triangle-wave modulation. Total time is 150mS, slope is 5gs/sec.

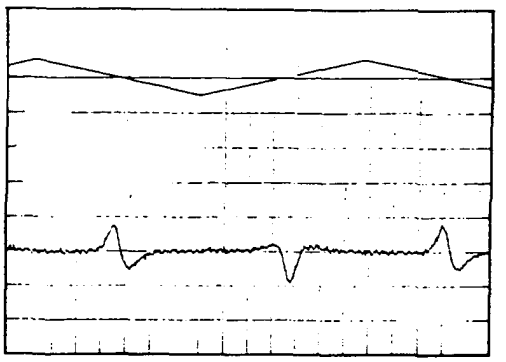


Fig. 7: Transmitted light, triangle-wave modulation. Total time is 150mS, slope is 2gs/sec.

value around 8 gauss/sec. On the other hand, the width is constant for slopes less than 8 gauss/sec and decreases sharply for larger slopes.

A similar behavior is observed for the rise-time and for the decay-time, (of the positive signal) but is not shown here.

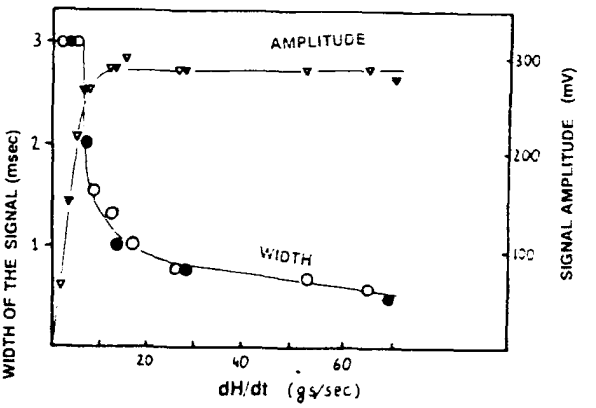


Fig. 8: Dependence of the amplitude and width of transient signal on the slope of a triangle wave

- ● ▼ -The frequency of triangle wave is 48Hz, the amplitude is changed
- ○ ▽ -The amplitude of triangle wave is 0.35GS the frequency is changed

Figures 9 and 10 exhibit the transient signal intensity versus the microwave frequency. The intensity is detected in the following technique: "C" field is square-wave modulated. This give rise to transient signals at double the modulation frequency. These are being detected by a Lock-in-Amp tuned to the 2nd harmonics. Then the microwave frequency is slowly scanned.

Fig. 9 exhibits a wide scan of about 0.5MHz centered around the zero-field, clock-transition frequency. We observe symmetrical spectrum of 7 lines. These lines are attributed to the various σ and π transitions as indicated in the figure. The center line is the narrowest one and exhibits side-lobes which might indicate some kind of interference phenomenon.

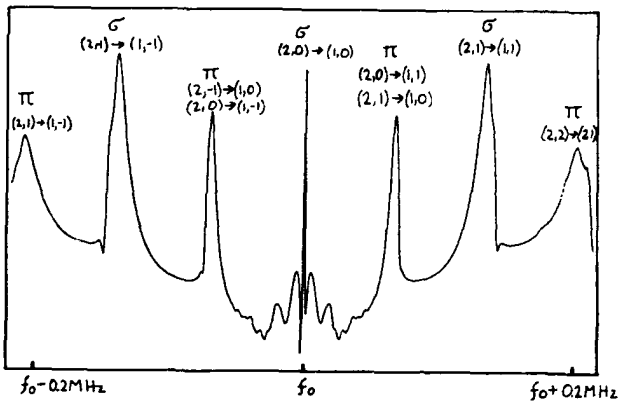


Fig. 9: Zero-Crossing signal versus microwave frequency, a wide scan. The "C" field is square wave modulated at 220Hz and signal is detected by a Lock-in-Amp tuned to the 2nd harmonics.

Fig. 10 depicts an enlargement of the center line together with a plot of a line obtained in the traditional way; i.e., the "C" field is kept constant around 300mgauss, the microwave frequency is amplitude modulated at 444Hz and signal is detected by a Lock-in-Amp tuned to the 2nd harmonics. We see that the field modulated line is narrower than the frequency modulated line and that it is peaked at frequency that is associated with a zero "C" field.

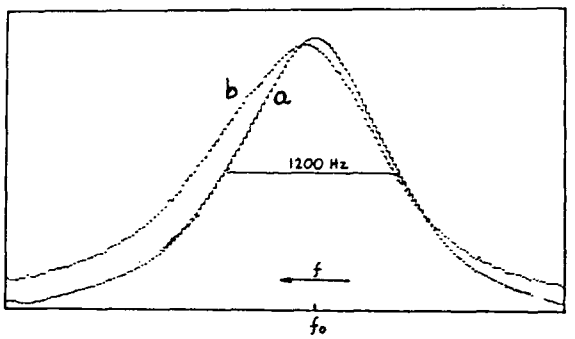


Fig. 10: a. Zero-Crossing signal versus microwave frequency, an enlargement of the clock transition line. "C" field is square-wave modulated at 444Hz and signal is detected by a Lock-in-Amp tuned to the 2nd harmonics.
b. Constant "C" field line. Frequency is amplitude modulated at 444Hz and signal is detected via a Lock-in-Amp tuned to 888Hz, "C" field is set to 300mgauss.

Discussion

We concentrate mostly on the observations that are related to the center line (fig. 9) and present a preliminary model to explain some of the features. A more comprehensive analysis shall be published elsewhere.

The transient signals that are observed in figures 2 to 7 are associated with the zero-crossing of the magnetic field. However, as seen from fig. 8 we have two regions of interest: (a) a rapid zero-passage region, where dH/dt>8 gauss/sec and (b) a slow-zero passage, where dH/dt<8 gauss/sec.
The first region is attributed to non-adiabatic dephasing phenomena, in the following model.

The adiabatic condition states that the relative field variation must be smaller then the Larmor precession frequency, i.e.

$$(1/H)(dH/dt) < |\gamma| \cdot H \qquad (1)$$

where $|\gamma| = 2\mu_s/(2I+1)$ is the gyromagnetic-ratio, μ_s is the Bohr magneton and I is the nuclear spin.

For a given slope, dH/dt, the condition (1) breaks down for $H < H_A$, where H_A is defined by the equation,

$$1/H_A(dH/dt) = |\gamma| H_A \qquad (2)$$

We define a non-adiabatic region in the H-t plane by $H < H_A$ and $t < \tau_A$ (see fig.11) where

$$\tau_A = H_A/(dH/dt) \qquad (3)$$

combining (2) and (3) we obtain

$$1/\tau_A = J|\gamma| \cdot \sqrt{dH/dt} \qquad (4)$$

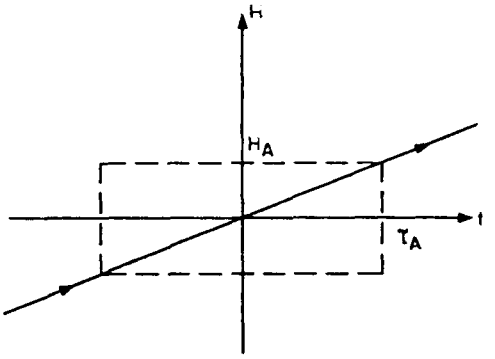


Fig.11: Illustrating the non-adiabatic region, in the H-t plane.

In fig. 12 we redraw the data of fig. 8. 2/(width) is plotted versus the square-root of dH/dt. In the fast passage region (dH/dt> 8gauss/sec), we obtain a straight line with slope of $0.7\times10^3 (\text{sec}\cdot\text{gauss})^{-1/2}$. This is the same order of magnitude as $J|\gamma| = 2.1\times10^3 (\text{sec}\cdot\text{gauss})^{-1/2}$.

Thus, we see that equation (4) roughly predicts the experimental data.
The experimental line in fig. 12, however, does not cross zero, as predicted by equation (4). This might happen due to residual magnetic fields and hysteresis of magnetic materials that are not included in the model.

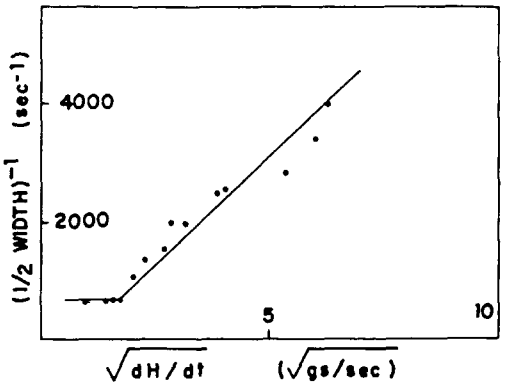


Fig. 12: Signal width dependence on dH/dt

The mechanism for the generation of the transient signals in the fast zero passage region is as follows. Under a steady state conditions (constant "C" field) we have a steady light absorption which is proportional to the number of Rb^{87} atoms (N_1) in the ground state (F=1, m_f=0) of the hyperfine structure. As the system goes through the non-adiabatic region the Zeeman sublevels are no longer defined, the Rb atoms loose coherence with the microwave radiation at 6.834.. GHz, and T_2, the dephasing time, goes to zero. This, has the same effect as turning off the microwave for a period of $2\cdot\tau_A$. This, in turn, decreases N_1 and the light absorption decreases. Thus, we obtain positive signals with width of $2\tau_A$. Negative signals, on the other hand, are attributed to a constructive in-phasing phenomena whereas several transitions between the Zeeman sublevels combine in phase to increase N_1. This is also indicated by the side-lobes around the central line in fig. 9.

In the slow passage region, i.e., dH/dt< 8gauss/sec, the width of the signal and T_2 are dominated by thermal relaxation processes, and therefore are not dependent on dH/dt as exhibited in fig. 8.
Finally, a note concerning the non-central lines in fig. 8. These are attributed to the various π and σ transitions between the (F=2, $m_{f,2}$) and (F=1, $m_{f,1}$), $\Delta m_f = 0, \pm 1$ as indicated in the figure.
(The existence of the π lines is evidence of a microwave magnetic field perpendicular to the "C" field. This is due to the field line shape of our TE101 cavity).

The non-central lines cannot be related to a zero magnetic field. They are generated by a square-wave modulation where the "C" field is being modulated between (+) and (-) approximately 200mgauss.
Now, consider, for instance the σ transition (2,1) (1,1).
Its frequency dependence on H is,

$$f = f_o + b\cdot H \quad b \approx 1.4\text{MHz/gauss} \qquad (5)$$

and its width is approximately 7KHz, or 5mgauss in terms of magnetic field.
As the field changes between (+) and (-), it rapidly departs the (+) resonance and enters the (-) resonance. The spins cannot follow adiabatically this field reversal, hence we obtain a non-adiabatic signal as explained before. However, in this case it is dependent on the magnetic field (the amplitude of the square-wave) via equation (5).

Detection of the clock transition

As seen from fig. 10 magnetic field zero-crossing generates a line which is centered around the zero-field transition frequency. This line is narrower than the conventional line and thus can be used to lock the servo-loop in a rubidium frequency standard. According to our interpretation this line is generated by a constructive and destructive interference between the Zeeman sublevels of the hyperfine structure. Thus, we expect a stronger signal. The short-term- stability is a function of the signal-to-noise ratio and the line Q, or the linewidth. The noise, however, have not been measured so far in our experiment.
Assuming, no degradation in the noise figure, we can expect a better short-term-stability.

Fig. 13 presents a scheme for the detection of the signal generated by zero-crossing and for locking the servo-loop. The "C" field is being modulated at 400Hz and the microwave frequency is modulated at lower frequency of 100Hz. The output signal from the photo-diode is first detected by a 2nd harmonic detector, to produce a dc signal whose line shape is shown in fig. 10. Then, the frequency modulation at 100Hz is detected by a 1st harmonic detector to produce the error curve (the discriminator pattern) in the conventional scheme.

The main advantage of this technique is the canceling of the sensitivity to external magnetic field, since the signal is generated when the total magnetic field is nearly zero.

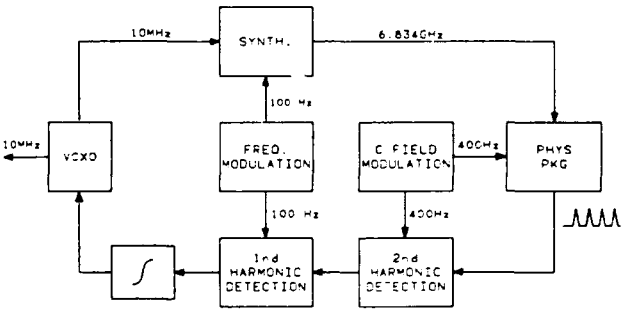


Fig. 13: A suggested scheme for the detection of the zero-crossing signal and for locking the servo-loop in a rubidium-frequency-standard.

Summary

When the magnetic field of an optically pumped Rb vapor crosses zero or reverse sign, in the presence of a microwave EM field positive and negative transient signals are generated in the transmitted light intensity. We have associated these signals to a non-adiabatic dephasing of the Zeeman sublevels. We have suggested a scheme for the application of the zero-crossing signals in a rubidium frequency standard.

However the work described in this paper is not completed and additional study of the basic phenomena as well as its application is planned.

Acknowledgement

The authors acknowledge Y. Barash for a most helpful and enlightening discussions.

References

[1] A. Stern, M. Golosovski, Y. Elbaz, A. Hertz and A. Lepek, "TF-4000A, TFL's High Performance Ruggedized Rubidium Frequency Standard". Proc. 43rd Ann. Symp. on Frequency Control, 1989, pp.124.

[2] J. Vanier and C. Audoin, The Quantum Physics of Atomic Frequency Standard. Bristol and Philadelphia, Adam Hilger, 1989, ch4, pp 512-531.

[3] H. G. Dehmelt, Phys. Rev. 105, 1487, (1957).

[4] W. B. Hawkins, Alignment of Sodium Atoms. Phys. Rev. 98, 478, (1955).

[5] J. C. Lehmann and C. Cohen Tannoudji, CR. Acad. Sci.(Paris) 258, 4463 (1964).

[6] W. Hanle, Z. Phys. 30, 93 (1924).

[7] J. C. Camparo and R. P. Frueholz, Phys. Rev. A 30, 803, (1984).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DEVELOPMENT OF A DIGITALLY CONTROLLED RUBIDIUM ATOMIC OSCILLATOR

Naoki Ishihara, Shin-ichi Fukugawa
Shigenori Kodama, Takayoshi Makabe

NEC Corp. Transmission Div.
1753 Shimonumabe, Nakahara-ku
Kawasaki, Kanagawa, 211 Japan

Abstract

Rubidium atomic oscillators (RbOSCs) are widely applied in many systems. In most cases, they are used as "independent frequency reference sources". Recently, however, it has also become important to use them under "mutually synchronized" conditions.

The authors have developed a digitally controlled rubidium oscillator (DCRO), which can synchronize another frequency source. The $\pm 3X10^{-8}$ wide frequency control range with good linearity and $1X10^{-12}$ fine control step are highlighted.

In order to achieve the wide frequency range with the fine step, a synthesizer was developed which has three stage phase locked loop oscillators with variable frequency dividers. Data for controlling the frequency dividers, written in a 3Mbit ROM, was calculated using a computer.

Along with $\sigma y(\tau)=6X10^{-12}/\tau^{1/2}$ fine frequency stability, a more than 20 year long life and compactness were realized, which can easily satisfy telecommunications and broadcasting system requirements.

Introduction

RbOSCs are becoming widely used as compact and practical frequency sources. In most cases, they are used as independent frequency standards. Recently, their reliability has been markedly improved and their size has been reduced as well. Therefore, it is foreseen that RbOSC will be used in a new concept equipment; such as "High quality synchronized network" or "Highly accurate frequency sources".

Figure 1 shows the permissible time interval error for communication systems. A Cesium Frequency Standard (CsOSC) is necessary to meet the necessary less than $1 X10^{-11}$ long term frequency stability. Also, the phase locked loop Crystal Oscillator (XLOSC) was available to meet the requirements for short term frequency stability, such as the range between 1 second and 1 million seconds. However, if the synchronization should be interrupted for some reason, a serious network quality deterioration would occur, since the XLOSCs do not have sufficient self running frequency stability. Therefore, if a compact, practical and frequency controllable RbOSC

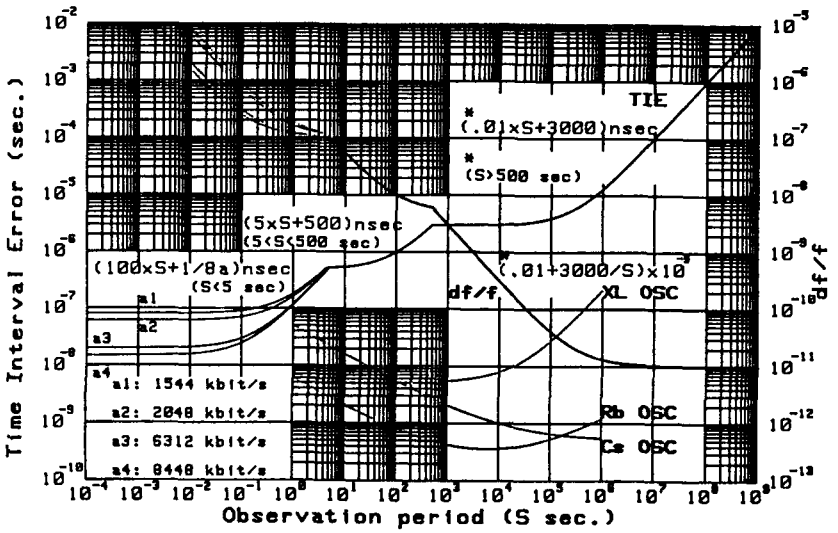


Fig. 1. Time interval error (CCITT G.811)

CH2818-3/90/0000-059 $1.00 © 1990 IEEE

could be realized, the network system quality would be rapidly improved. Implementing these factors resulting a high quality synchronized network.

The authors would like to propose a new concept. Previously, RbOSC had "Excellent Frequency Stability" and XLOSC had "Wide Frequency Control Range","Long Life Time", and "Compact Size". The authors' target was to realize a new frequency source, which has both RbOSC advantages and XLOSC advantages.

Basic approach

Design target

The main design targets for the DCRO are:

1. Digital frequency control
2. $1X10^{-12}$ frequency control step
3. $\pm 3X10^{-8}$ frequency control range

Digital frequency control function is useful to realize lengthy time constant, in consideration of the DCRO application fields.

Next, consider frequency control accuracy. XLOSC and RbOSC, which have a flicker floor on the order of 10^{-12}, have a practically significant frequency control accuracy of approximately 1×10^{-12}.

Along with this feature, in consideration of the frequency control range, the XLOSC stability is on the of the order of 10^{-8}. For an oscillator of this type to make a frequency variable with a digital input, a $\pm 3.3 \times 10^{-8}$ frequency control range is considered appropriate, since the oscillator matches beneficially with on LSI universally applicable at 8 bits or 16 bits. (see Table 1).

TARGET PERFORMANCES	BACKGROUND ITEMS
DIGITAL FREQUENCY CONTROL	• To Realize Lengthy time constant
FREQUENCY CONTROL STEP (1×10^{-12})	• Flicker floor for XLOSC & RbOSC
FREQUENCY CONTROL RANGE ($\pm 3 \times 10^{-8} \leq 1 \times 10^{-12} \times 2^{16}$)	• XLOSC Stability • Interface to digital IC

Table 1. DCRO design target

Frequency control method

There are two possible methods to make a RbOSC change its output frequency.

Figure 2 shows blockdiagrams for these frequency control methods. One is to vary the rubidium atoms resonant frequency. The other is to control the servo loop in the RbOSC.

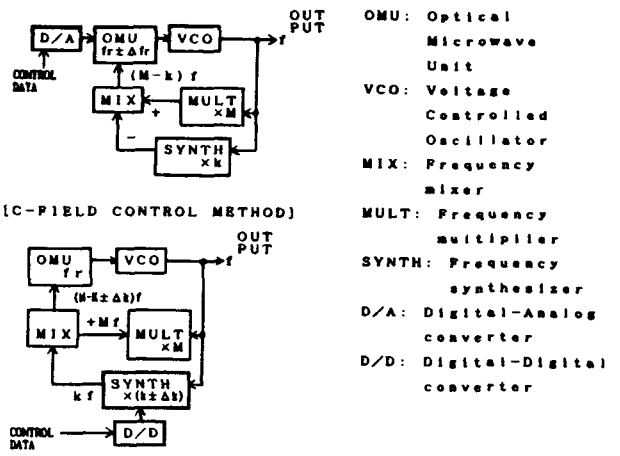


Fig. 2. Blockdiagrams for frequency control methods

To make an atomic resonant frequency vary, one method generally used is the C-field added to a rubidium resonant cell, with its intensity and resonant frequency changed externally. This is called the C-field control method.

To control the servo loop in RbOSC, the simplest method is to control the synthetic ratio for frequency synthesizers in the RbOSC. This is called the frequency synthesizer control method.

The C-field control method has a simple circuit configuration. Nevertheless, it has the following two disadvantages:

First, the RbOSC output frequency varies in proportion to the square of a C-field strength. To provide a wider variable frequency range, the nonlinearity of an output frequency will become problematical.

Second, it is necessary to apply a large current in the C-field beforehand, to provide a wider variable frequency range. This involves the problem that, especially if a small cavity type were employed, it would disturb the C-field uniformity and the resonant frequency changes excessively, in relation to the level fluctuation for an input micro-wave signal for the cell in the cavity. As a result, this large current disturbs the RbOSC frequency stability.

Therefore, the authors adopted the frequency synthesizer method.

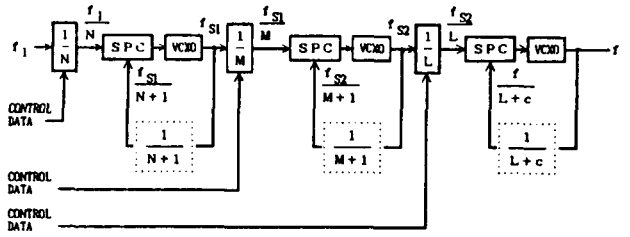
Synthesizer configuration

Another approach which could be chosen would be to select an appropriate synthesizer configuration. The frequency synthesizer control method makes the frequency vary effectively. Nevertheless, it should be noted that the frequency synthesizer control method has a small loop in the

synthesizer and a large loop in the RbOSC, which are controlled separately. It is necessary, therefore, to materialize a frequency synthesizer responsive at a speed high enough not to affect the RbOSC loop. The frequency synthesizer control method also has a circuit configuration which is more complicated than that for the C-field control method. The most important problem to be solved is to develop a synthesizer unit with wide frequency control range, good linearity, fine frequency control steps and fast response time in a simple circuitry.

A phase-locked loop oscillator could be effectively used to form a simple circuitry. It has only one frequency divider and permits response speed to rise easily, while still allowing simple circuitry configuration. If it is desired to realize a wide frequency control range with fine steps, only a one stage synthesizer and a two stage synthesizer would not enough to realize the required "Lock range", "Response time" and "Control error". The authors have designed a frequency synthesizer to cope with this problem, in which three stages of sampling phase-locked loop oscillator are connected in series to materialize the desired performances.

Fig. 3. Synthesizer blockdiagram



FREQUENCY SYNTHESIS RATIO

$$k=\frac{f}{f_1}=\frac{f_{S1}}{f_1}\cdot\frac{f_{S2}}{f_{S1}}\cdot\frac{f}{f_{S2}}=\left(\frac{N+1}{N}\right)\left(\frac{M+1}{M}\right)\left(\frac{L+C}{L}\right)\cdots(1)$$

RESPONSE TIME

$$t=\frac{1}{f_C}=\frac{1}{\alpha A}\fallingdotseq 1\,ms\cdots(2)$$

M, N, L: Division ratio
c: Integral number
f_c: Cut off frequency [Hz]
A: VCXO Sensitivity [Hz/V]
α: Sampling Phase Comparator Sensitivity [V/rad]

Figure 3 is a blockdiagram of the three-stage phase-locked loop oscillator. For control inputs, N, M and L are determined as frequency division ratios for the phase-locked loop oscillator. The circuit is designed so as to allow the three phase-locked loop oscillator to be synchronized in series. In this case, Frequency Synthesizing Ratio k may be expressed as Eq. (1). The response speed for this phase-locked loop oscillator, is dependent upon Voltage Controlled Oscillator (VCO) Sensitivity A and Sampling Phase Comparator Sensitivity α expressed as Eq. (2). It is determined, based on the lowest response speed in the three stages. Recently, a response speed of approximately 1 ms was obtained as a theoretical value. Compared with the 0.5 thru 1 second time constant for the loop in the RbOSC as a whole, a satisfactorily high speed characteristic was available.

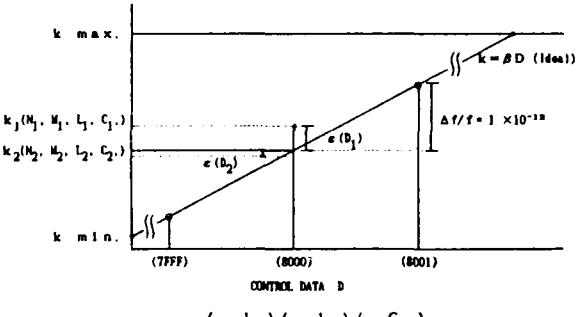
Division ratio calculation

Once a circuit configuration has been determined, the frequency may be made variable when values N, M and L can be determined. The DCRO configuration is such that 16-bit data are used for control inputs, so that the frequency will be continuously variable, with values N, M and L recorded in their memories. In order to linearly move a frequency synthesizer output, N, M and L are generally not continuous and have random values.

Table 2 shows a control data example. The control data specifies the RbOSC output frequency, and data 8000 corresponds to the f_0 frequency. Data 0000 corresponds to the minimum frequency and Data FFFF corresponds to the maximum frequency for the RbOSC. Concerning the f_0 area, for example, if data 7FFF is given, the frequency synthesizer is offset about 0.007Hz from the center frequency and the RbOSC output is controlled to -1×10^{-12}. This division ratio was calculated by computer to obtain an optimum combination of L, M and N.

CONTROL DATA (HEX)		0000	7FFF	8000	8001	FFFF
DIVISION RATIO	L	583	566	550	550	582
	M	3126	3268	3212	3148	3109
	N	2902	3146	3024	3083	3056
SYNTHESIZER FREQUENCY (Hz)		5,312,723.962	5,312,500.007	5,312,500.000	5,312,499.994	5,312,276.049
RbOSC FREQ.(Δf/f)		-3.28×10^{-6}	-1.02×10^{-12}	0	$+0.88\times10^{-12}$	$+3.28\times10^{-6}$

Table 2. A control data example



SYNTHESIS RATIO $k_1=\left(1+\frac{1}{N_1}\right)\left(1+\frac{1}{M_1}\right)\left(1+\frac{C_1}{L_1}\right)$

CONTROL ERROR $\varepsilon(D_1)=k_1-\beta D_1$

Fig. 4. Algorithm to obtain N, M, L & C

In order to appropriately choose the N, M and L values, the authors have adopted the following method.

Figure 4 shows how to obtain the division ratio. Initially, the range for synthesis ratio k, division ratio N, M, L and C are set. Next a combination of N, M, L and C is selected and implemented and synthesis ratio k is obtained. If the k value is within the k min. and k max. range, it is compared to the ideal value. The difference is defined to control error $\varepsilon(D)$. For example, after $\varepsilon(D1)$ is stored, another combination of N, M, L and C is given and $\varepsilon(D2)$ are obtained. If the absolute value of $\varepsilon(D2)$ is smaller than that of $\varepsilon(D1)$, $\varepsilon(D2)$ is stored. Repeating this calculation, optimum values in a combination of N, M, L are obtained.

Stabilizing resonant frequency

The frequency synthesizer must cover the frequency control range and the dispersion range for rubidium gas cell resonant frequency. Buffer gas is sealed in the rubidium gas cell, to reduce the Doppler width. Conversely, however, the atomic resonant frequency varies with a dispersion of sealed gas pressure. To mass-produce the RbOSC, therefore, the rubidium gas cell must have a resonant frequency as constant as possible.

All rubidium resonant cells are operated under oven controlled temperature conditions. Therefore, if it is desired to stabilize the resonant frequency for the gas cell, it is effective to stabilize the number of molecules in the buffer gas. To produce a rubidium resonant cell, the authors have developed the following new methods.

1. Controlling the sealing pressure in the cell, according to the manufacture stage ambient temperature.

2. Stabilizing cell temperature with an air cooling system when chipping off the gas cell.

Thus, the rubidium resonant frequency has been stabilized to a σ =2Hz dispersion level.

DCRO construction

Photo 1 shows a rear view of the DCRO prototype. It is separated into the synthesizer unit and the RbOSC unit.

Photo 2 shows the synthesizer unit for the DCRO. The three-stage phase-locked loop oscillator is found at the left side of this photo. The upper right area shows an input port for control data. This synthesizer employs five EPROMs. The division ratio for the frequency synthesizer is written in

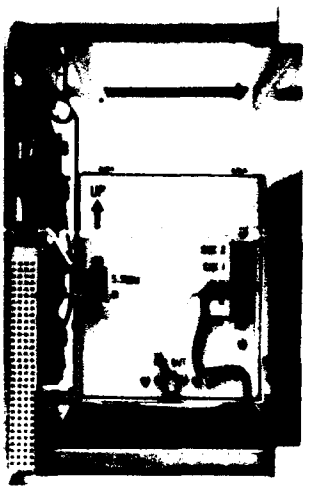


Photo 1 DCRO Rear view

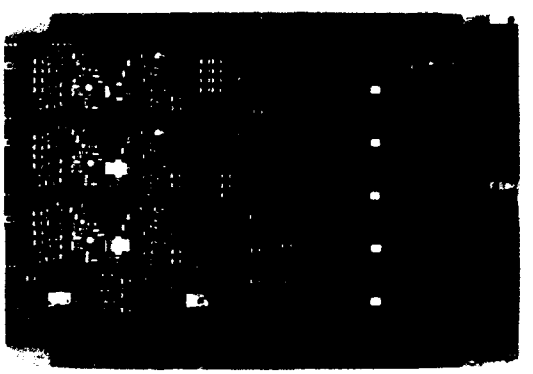


Photo 2 Synthesizer unit

Photo 3 Rubidium oscillator unit

the ROMs. The total EPROM capacity is 2.5Mbits.

Photo 3 shows the RbOSC unit. It is also possible to use it as an independent frequency standard.

Improvement in lifetime

To achieve the reliability at a level identical with the level of a XLOSC, the authors have also prolonged the life of both rubidium lamps and rubidium resonant cells. Rubidium cells have their operating life reduced, possibly due to the following factors:

1. A residual amount of metal has decreased, due to rubidium metal reaction with glass.

2. Q has dropped, due to rubidium metal dispersion.

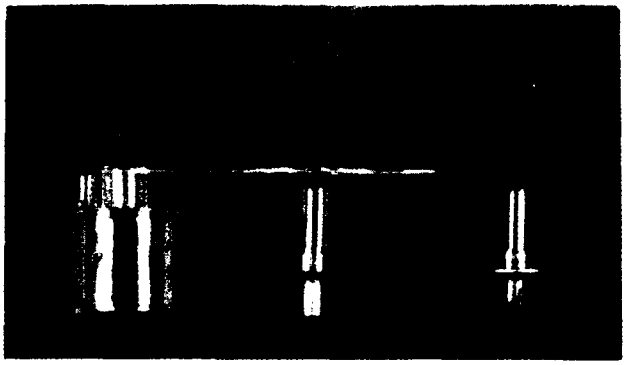


Photo 4 Rubidium resonant cell and rubidium lamps

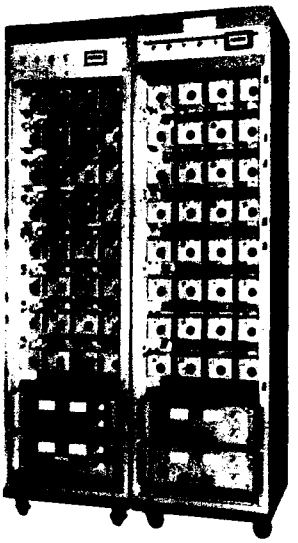


Photo 5 Rubidium lamp evaluation equipment

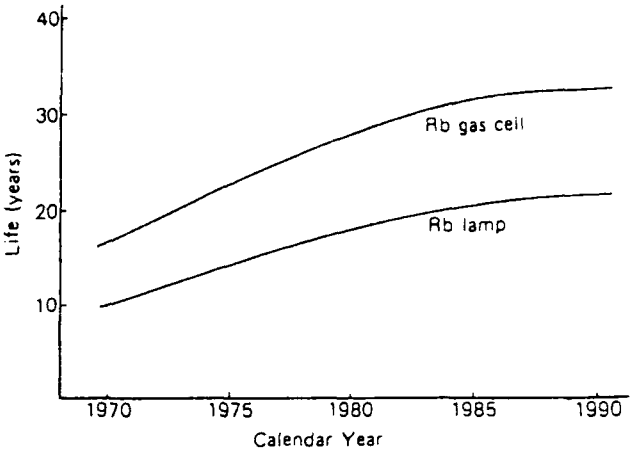


Fig. 5. Improvement in optical device life

The authors have noted that the speed of a reaction between rubidium metal and glass and the diffusion rate vary according to the surface condition of glass. Thus, consideration was paid to

1. A thermal design to prevent Rb metal from diffusing.

2. An improvement in glass material and surface conditions.

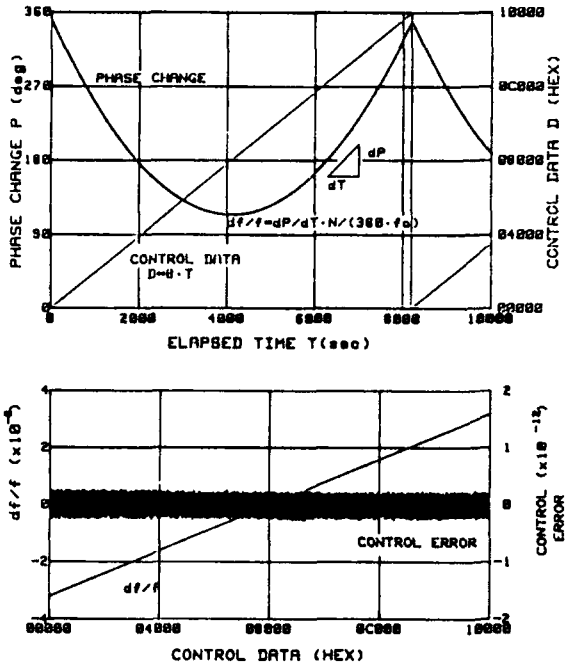
The authors have designed and evaluated the rubidium lamps and cells. Photo 4 shows the Rb cells. No.1 is the Rb resonant cell. No.2 is the Rb lamp (used in the DCRO). No.3 is the same kind as No.2, and was tested for more than 20 years in NEC.

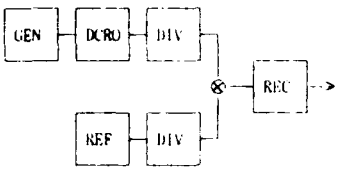
Photo 5 shows Rb lamp evaluation equipment. 32 lamps are tested for all together in one rack.

Rubidium lamp reliability was discussed by NEC personnel in this symposium in 1974. Prior to making that report, efforts were continuously made to improve the rubidium lamp and resonant cell. NEC has conducted a long-term aging test for over 20 years on the rubidium lamp and have confirmed that the emitted light is stable. Concerning the rubidium gas cell, moreover, long-term aging tests have been conducted. Results have confirmed that Q and rubidium absorption are stable. Figure 5 shows the improvement in optical device life. More than 20 year life time can be expected for the rubidium lamp and rubidium resonant cell.

Test data

Figure 6 shows frequency control characteristics. The upper characteristics line indicated the RbOSC output, determined by phase change vs. time lapse with the DCRO control data input





GEN : DATA GENERATOR
DIV : FREQUENCY DIVIDER (1/1024)
REC : RECORDER
REF : FREQUENCY REFERENCE

[MEASURING SYSTEM BLOCKDIAGRAM]

Fig. 6. Frequency control characteristic

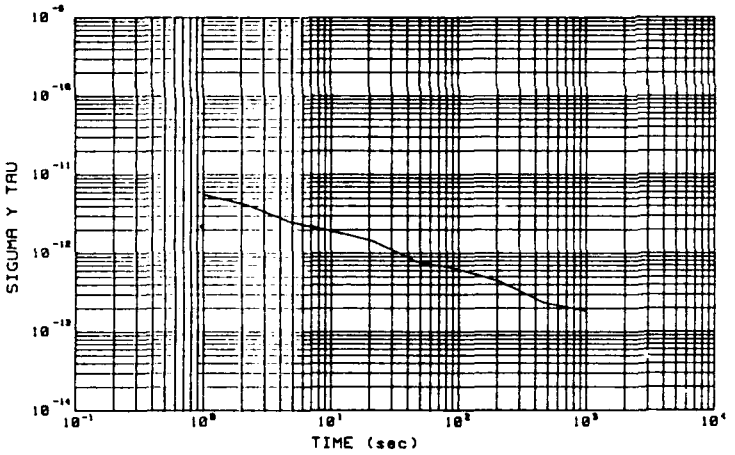


Fig. 7. DCRO short term frequency stability

ITEM	PERFORMANCE	OBSERVATION
FREQ. CONTROL	Range $\pm 3\times10^{-8}$ Step 1×10^{-12}	16Bit data control
SHORT TERM FREQ. STABILITY	$6\times10^{-12}/\tau^{1/2}$	1～1000sec.
FREQ. TEMPERATURE CHARACTARISTICS	2×10^{-12}/deg.	0～50℃
LIFE TIME	>20 years	
DIMENSIONS	110mm height 89mm width 123mm length 2kg weight	WITHOUT CONTROL UNIT

Table 3. DCRO performance

forced to change from "0000" to "FFFF". The lower characteristics line is converted into $\Delta f/f$ and HEX data. It may be concluded, from this, that the frequency is varying smoothly and linearly within a wide range of frequencies.

Figure 7 shows typical characteristic short-term frequency stability. In this DCRO, no filter cell was employed, to make its configuration simple. Nevertheless, a favorable short-term frequency stability of $\sigma y(\tau)=6X10^{-12}/\tau^{1/2}$ was achieved.

The main DCRO performances are shown Table 3. Excellent frequency control characteristics, frequency stability, long life and compactness were realized.

Conclusion

The authors have proposed a new kind of oscillator DCRO, which includes both RbOSC advantages and XLOSC advantages. The DCRC development was successful. Results involved excellent frequency control characteristics, excellent frequency stability, long life time and compact size. The authors expect that such DCROs will be used as new frequency standards and will expand their application fields in the future.

Acknowledgment

The authors are deeply indebted to many persons for their encouragement and assistance in the process of research on the digitally controlled rubidium atomic oscillator. They wish to express their thanks to Messrs. A. Tsukui and Y. Kobayashi.

References

(1) N. Ishihara and H. Oyamada, et al., "Development of a Compact, Highly Reliable Rb-3100 Rubidium Atomic Oscillator," NEC Res. & Develop., No. 86, pp. 37-44, July 1987.

(2) H. Oyamada, et al., "A Consideration of Rubidium Lamp Stability for Rubidium Frequency Standard," Proc. 28th Annual Symp. on Frequency Control, pp. 340-343, 1974.

(3) H. Uchida, et al., "Automatic Frequency Controlled Rubidium Frequency Standard," Proc. 26th Annual Symp. on Frequency Control, pp. 211-215, 1972.

(4) T. Yuuki et al., "Rubidium Frequency Standard," NEC Res. & Develop., pp. 42-56, October 1973.

(5) M. Arditi and T. R. Carver, "Pressure, Light, and Temperature Shifts in Optical Detection of 0-0 Hyperfine Resonance of Alkali Metals," Phys. Rev., 124, 3, pp. 800-809, November 1, 1961.

(6) R. C. Brewer, "High Intensity Low Noise Rubidium Light Source," R. S. I., 32, 12, pp. 1356-1358, December 1961.

(7) V. B. Gerard, "Laboratory Alkali Metal Vapor Lamps for Optical Pumping Experiments," J. S. I., 39, pp. 217-218, 1962.

(8) T. Yuuki, et al., "Rubidium Frequency Standard," NEC Res. Develop., 31, pp. 42-56, October 1973.

(9) W. E. Bell, et al., "Alkali Metal Vapor Spectral Lamps," R. S. I., 32. 6, pp. 688-692, June 1961.

(10) H. Uchida, Y. Sato, H. Kumamoto and H. Oyamada, "Automatic Frequency Controlled Rubidium Frequency Standard," Proc. of the 26th Annual Symp. on Frequency Control, pp. 211-215, 1972.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PERFORMANCE OF SAO MODEL VLG-12 ADVANCED HYDROGEN MASERS

Edward M. Mattison and Robert F.C. Vessot

Smithsonian Astrophysical Observatory
Cambridge, Massachusetts 02138

Abstract

The SAO model VLG-12 hydrogen maser incorporates several technological advances, including a single-state atomic selection system, improved bulb collimation, a metal-sealed vacuum system, sorption vacuum pumps, a digital frequency synthesizer and a digital monitoring and control system. We describe the performance of two VLG-12 masers that have operated at the U.S. Naval Observatory since October 1989. We discuss the effectiveness of the single-state selection system and of the storage bulb coatings. The masers' initial frequency drift rates of several parts in 10^{15} per day decreased by factors of 2 to 4 over a period of 200 days.

Introduction

The model VLG-12 hydrogen maser, produced by the Smithsonian Astrophysical Observatory, incorporates several technological improvements designed to provide increased frequency stability and operating reliability. Two VLG-12 masers (designated P24 and P25) have been in operation at the United States Naval Observatory (USNO) since October, 1989. Here we describe novel features of the VLG-12 maser and measurements of the masers' performance.

Features of VLG-12 masers

A major innovation in the VLG-12 maser is the inclusion of a single-state hydrogen beam selection system[1]. In the traditional state-selection scheme, a hexapole magnet immediately following the maser's hydrogen dissociator removes from the atomic hydrogen beam atoms in the ($F = 0$, $m_F = 0$) and ($F = 1$, $m_F = -1$) hyperfine states (referred to as the "a" and "b"states), leaving ($F = 1$, $m_F = 0$ and 1) state atoms (states "c" and "d") to enter the maser's interaction region. The d-state atoms do not contribute to maser oscillation, but broaden the atomic linewidth by spin-exchange relaxation, thus increasing cavity pulling[2] of the maser oscillation frequency. The presence of d-state atoms also creates the possibility of frequency shifts that depend upon hydrogen density and magnetic field inhomogeneities.

In the VLG-12's single-state selection system the first hexapole magnet is followed by an adiabatic fast passage (AFP) state inversion mechanism[3]that transfers d-state atoms into the b state, leaving the c-state atoms unchanged. A second hexapole magnet then removes the b-state atoms, leaving a beam of atoms almost entirely in the desired c state. By eliminating unnecessary spin-exchange relaxation, the single-state selection system reduces the atomic linewidth and thus the effect of cavity frequency pulling. In addition, by removing atoms in other than the c state, the system substantially reduces frequency shifts caused by magnetic field inhomogeneities.

The storage bulb's entrance aperture is equipped with multiple Teflon tubes that increase the bulb's atomic storage time and decrease the maser's linewidth, thus reducing the effect of cavity pulling. The increased long-term stability resulting from the narrow linewidth is important in frequency standards to be used for timekeeping.

Operational reliability is improved through the use of a vacuum system employing all metal seals, thus avoiding low-level leaks and outgassing that can result from elastomer seals. The hydrogen dissociator bulb is located within the vacuum envelope and is sealed with indium to eliminate the need for either glass-to-metal or elastomer seals. Hydrogen is removed by sorption vacuum pumps that require no operating power. A small ion pump removes residual gasses produced from outgassing of the maser's components.

The VLG-12's rf receiver system is equipped with a computer-controllable digital frequency synthesizer[4] with a frequency resolution of 7 parts in 10^{18}. This high resolution allows the maser's output

CH2818-3/90/0000-066 $1.00 © 1990 IEEE

frequency to be closely matched to that of another oscillator or to a calculated time scale. A digital control and monitoring system[5] permits external computer control of the maser's hydrogen beam flux, cavity-tuning varactor diode voltage, and synthesizer frequency. The system monitors 32 channels of maser operating parameters and stores up to 1024 sets of automatically recorded data in non-volatile memory that is accessible by an external computer.

Density-independent relaxation rate and state selection effectiveness

The density-independent relaxation rate γ_t and the quality parameter q are performance measures of the maser's atom storage mechanism and of its state selection system. γ_t is the total atomic relaxation rate due to mechanisms independent of the hydrogen density in the storage bulb – that is, other than spin exchange relaxation. Assuming that for each of these mechanisms the longitudinal and transverse relaxation rates are equal, $\gamma_1 = \gamma_2$, γ_t can be written as

$$\gamma_t = \gamma_b + \gamma_r + \gamma_m + \gamma' \qquad (1)$$

Here γ_b is the escape rate of atoms through the bulb's entrance collimator, γ_r is the rate of loss of hydrogen atoms due to recombination on the bulb's storage surface, γ_m is the spin relaxation rate due to magnetic field inhomogeneities in the storage region, and γ' represents any other density-independent relaxation process for which $\gamma_1 = \gamma_2$. Generally, the first three terms on the right side of Eq. 1 are dominant.

The quality parameter q is given by

$$q = \frac{\sigma_{se}\,\bar{v}\,\hbar}{8\pi\mu_o^2}\,\frac{\gamma_t}{\gamma_b+\gamma_r}\,\frac{V_c}{\eta V_b Q_c}\left(\frac{I_{tot}}{I}\right) \qquad (2)$$

Here σ_{se} is the hydrogen-hydrogen spin-exchange cross section, $\bar{v}$ is the average relative speed of hydrogen atoms in the storage bulb, $\hbar$ is Planck's constant, μ_o is the Bohr magneton, V_c and V_b are the volumes of the resonant cavity and storage bulb, Q_c is the loaded cavity Q, and η is the magnetic filling factor[6]. I is the flux of c-state atoms entering the storage bulb, while I_{tot} is the total flux of atoms in all hyperfine states entering the bulb. Because q is proportional to (I_{tot}/I), it is a useful measure of the effectiveness of the state-selection system. For properly operating traditional (one-magnet) selection systems, $(I_{tot}/I) \cong 2$, while for single-state selection, $(I_{tot}/I) \cong 1$; thus in a perfectly operating system, we expect the ratio $q_{off}/q_{on} = 2$, where q_{off} and q_{on} are the values of q with the AFP system turned off and on, respectively. If $\gamma_m + \gamma'$ is small, the ratio $[\gamma_t/(\gamma_b+\gamma_r)] \cong 1$. Inserting the values of atomic parameters and typical VLG-12 maser dimensions into Eq. 2 gives

$$q = 0.024\,\frac{\gamma_t}{\gamma_b+\gamma_r}\left(\frac{I_{tot}}{I}\right) \qquad (3)$$

γ_t and q are determined by measuring Q_l, the maser's line Q, as a function of rf output power, and fitting a quadratic form to $(Q_l)^{-1}$ as a function of the power radiated by the atomic beam[7]. The values of γ_t and q were measured for masers P24 and P25 before the masers were transported to USNO, and again for P25 in May, 1990. The results are shown in the table below.

Maser	γ_t (sec^{-1})	q_{on}	q_{off}	q_{off}/q_{on}
P24 8/89	.77	.034	.062	1.83
P25 9/89	.78	.048	n/a	1.76*
P25 5/90	1.16	.061	n/a	n/a

P25's state-selection magnets have longitudinal magnetic field components that result in state inversion even in the absense of the AFP rf field. As a consequence, P25's single-state selection system operates at all times, and a measurement of q_{off} is difficult. The ratio q_{off}/q_{on} indicated by * was measured earlier, using special techniques. The parameters for P24 were not measured in May 1990 because the maser was in use by USNO and could not be disturbed.

The γ_t relaxation rates for P24 and P25 in August and September, 1989, were both approximately 0.77 sec^{-1}, corresponding to an equivalent line Q (in the absense of spin exchange) of approximately 5.8×10^9. Table 1 shows that for maser P25, γ_t increased by 49% between September 1989 and May 1990, and q_{on} increased by 27%. Increases in maser storage bulb relaxation rates over time, due presumably to deterioration or contamination of the Teflon storage surface, have been observed previously in our laboratory[8] and by other workers. The improved storage bulb collimation used in the VLG-12 masers makes recombination a larger fraction of the total relaxation rate, so a change in the bulb's surface quaility becomes more evident. All else being

constant, an increase in the recombination rate γ_r would make the ratio $\gamma_l/(\gamma_b+\gamma_r)$ closer to unity (smaller), and would thus decrease the value of q. The increase in P25's q, and part of the increase in γ_t, may be due to magnetic field inhomogeneities within the storage bulb. Subsequent to the measurements reported here the maser was degaussed, with a resulting improvement in its performance. Remeasurements of γ_t and q will be made in the future.

The values of q_{on} and q_{off} for maser P24 allow us to estimate the effectiveness of the state selection system. Assuming $[\gamma_l/(\gamma_b+\gamma_r)] \equiv 1$, Eq. 3 and the data of Table 1 give $(I_{tot}/I)_{on} = 1.42$, and $(I_{tot}/I)_{off} = 2.58$. If I, the flux of c-state atoms, is unchanged by the operation of the AFP system, these values imply that the single-state selection system removes 73% of non–c-state atoms from the beam. If, as is likely, $[\gamma_l/(\gamma_b+\gamma_r)] > 1$ due to magnetic field inhomogeneities, the percentage of undesired atoms removed is higher.

Sensitivity to External Magnetic Fields

The magnetic shields that surround the maser's microwave cavity attenuate changes in the external magnetic field, reducing their effects within the atomic interaction region. The residual field that penetrates the shields can shift the maser's frequency primarily through two effects: (i) the quadratic variation of the $\Delta m_F = 0$ hyperfine transition energy with magnetic field, which can be calculated from a measurement of the change in the Zeeman frequency for $\Delta m_F = \pm 1$ transitions, and (ii) $\Delta F = 0$, $\Delta m_F = \pm 1$ hyperfine transitions by oscillating atoms moving through transverse dc magnetic field gradients.[9] The motional shift is proportional to $(\rho_d-\rho_b)$, the Zeeman population difference in the storage bulb, as well as to the product of the radial dc magnetic field amplitude and the radial rf field amplitude, averaged over the storage bulb[10]. Thus if $(\rho_d-\rho_b)$ is minimized, as by single-state selection, motional shifts due to residual field gradients can be reduced.

The effectiveness of the single-state selection system in achieving this reduction is indicated in Fig. 1. The variation in maser frequency with external field was measured using a Helmholtz coil array surrounding the maser. Measurements were made at high and low hydrogen beam flux, with the internal maser field parallel to and opposite to the earth's field direction, and with the AFP system on and off. Shown in Fig. 1 are the predicted sensitivities calculated from the measured variation in Zeeman frequency with applied field. With the AFP system turned off, the measured sensitivities vary widely from the predicted values, while with the AFP system operating, the measured sensitivities are consistent with the calculations, indicating that single-state selection significantly reduces the effect of motional frequency shifts.

Long-term frequency performance

The frequencies of masers P24 and P25, relative to reference maser P18, are shown in Fig. 2 over a span of approximately 200 days. The abscissa is given in modified Julian date [MJD]; for reference, MJD 47900 = 9 January 1990. The frequencies were calculated from hourly simultaneous phase measurements relative to a common crystal oscillator; isolated erroneous points due to the measurement system, and frequency offsets introduced when the masers were tuned, have been removed from the data.

Both masers exhibit long-term frequency drifts that decreased over time. At the beginning of the observation period P24's average drift rate of fractional frequency relative to P18 was approximately 8 parts in 10^{15} per day, and at the end of the period it was approximately 2 parts in 10^{15} per day. The corresponding rates for P25 were 1.6×10^{-14}/day and 7×10^{-15}/day. The drift rate of P18 relative to the time scale of the Bureau International des Poids et Mesures (BIPM) over the same time period was approximately 5×10^{-16}/day at the start and 2×10^{-16}/day at the end; thus P24's drift relative to BIPM was roughly 10% less than the drift measured relative to P18.

Qualitatively similar behavior has been observed in many other masers. Masers P18 and P19, for example, which have been in operation since 1983, exhibited drift rates of several parts in 10^{15} per day soon after they were built; currently their drift rates are a few parts in 10^{16} per day. All drifts are toward higher frequencies. Such behavior has been ascribed to shrinkage of the joints between the resonant cavity's cylinder and endplates, and to relaxation of tensile stress in the cavity's inner silver coating[11]. It is not clear why P25 has a drift rate more than twice that of P24's. One difference between the masers is that P25's cavity was chilled to –25°C prior to assembly in the maser, in order to remove tensile stress in its silver coating. The

resulting compressive coating stress, or an unknown change in the cylinder's Cervit[12] material, may have resulted in the larger drift rate, although relaxation of compressive coating stress is expected to decrease, rather than increase, the maser's frequency.

Conclusions

The single-state selection system used in VLG-12 masers operates properly, removing at least 73% of unwanted atoms from the hydrogen beam and substantially eliminating frequency shifts due to motion of hydrogen atoms through magnetic field gradients. The long-term behavior of the storage bulbs' surface coatings, which appear to be deteriorating with time, will be monitored through measurements of q and γ_t. The masers' drift rates, which are consistent with the performance of other masers, will also be monitored, by comparison with the USNO time scale. Because the long-term frequency drifts are quite consistent, the masers' frequencies are highly predictable, making them suitable for use in timekeeping.

Acknowledgements

For their unstinting help in obtaining data and operating the masers, we are grateful to Dr. Gernot M.R. Winkler, Paul Wheeler, Anthony Kubik, and Dr. Mihran Miranian of the United States Naval Observatory.

References

1. E.M. Mattison and R.F.C. Vessot, "Single-state selection system for hydrogen masers." *Proc. 19th Annual Precise Time and Time Interval Applications and Planning Meeting*, p. 107 (1987)
2. D. Kleppner, H.M. Goldenberg, and N.F. Ramsey "Theory of the hydrogen maser." *Phys. Rev.* **126**, 603 (1962).
3. C. Audoin, M. Desaintfuscien, P. Petit, and J.-P. Schermann, "Design of a double focalization in a hydrogen maser." *IEEE Trans. Instrum. Meas.* **IM-17**, 351 (1968).
4. E.M. Mattison and L.M. Coyle, "Phase noise in direct digital synthesizers." *Proc. 42nd Annual Frequency Control Symposium*, p. 352 (1988).
5. L.M. Coyle and E.M. Mattison, "A microcomputer-based adaptive data logging system." *Proc. 35th International Instrumentation Symposium*, p. 285 (1989).
6. 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." *Phys. Rev A* **138**, 972 (1965).
7. E.M. Mattison, R.F.C. Vessot and W. Shen, "Single-state selection system for hydrogen masers." *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, **UFFC-34**, 622 (1987).
8. E.M. Mattison, R.F.C. Vessot, C. Bain, S. Wasserman, and G. Whitesides, "Surface interaction of atomic hydrogen with Teflon." *Proc. 41st Annual Frequency Control Symposium*, p. 95 (1987).
9. S.B. Crampton and H.T.M. Wang, "Duration of hydrogen-atom spin-exchange collision." *Phys. Rev. A* **12**, 1305 (1975).
10. S.B. Crampton and H.T.M. Wang, "Density-dependent shifts of hydrogen maser standards." *Proc. 28th Annual Frequency Control Symposium*, p. 355 (1974).
11. E.M. Mattison and R.F.C. Vessot, "Time and temperature stability of silver-coated ceramics for hydrogen maser resonant cavities." *Proc. 20th Annual Precise Time and Time Interval* Applications and Planning Meeting, p. 313 (1988).
12. Cervit is a low-expansion glass-ceramic manufactured by Owens-Illinois, Inc.

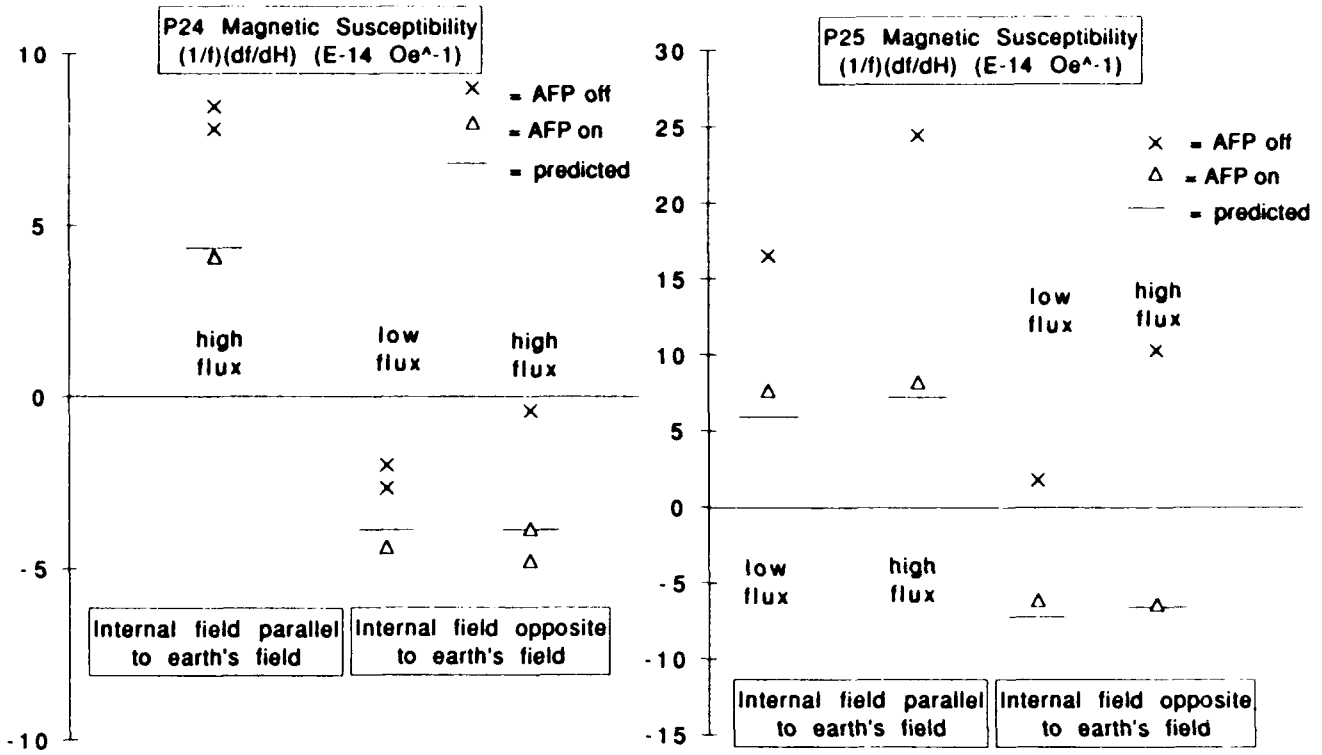


Figure 1. P24 and P25 Frequency Variation with External Magnetic Field

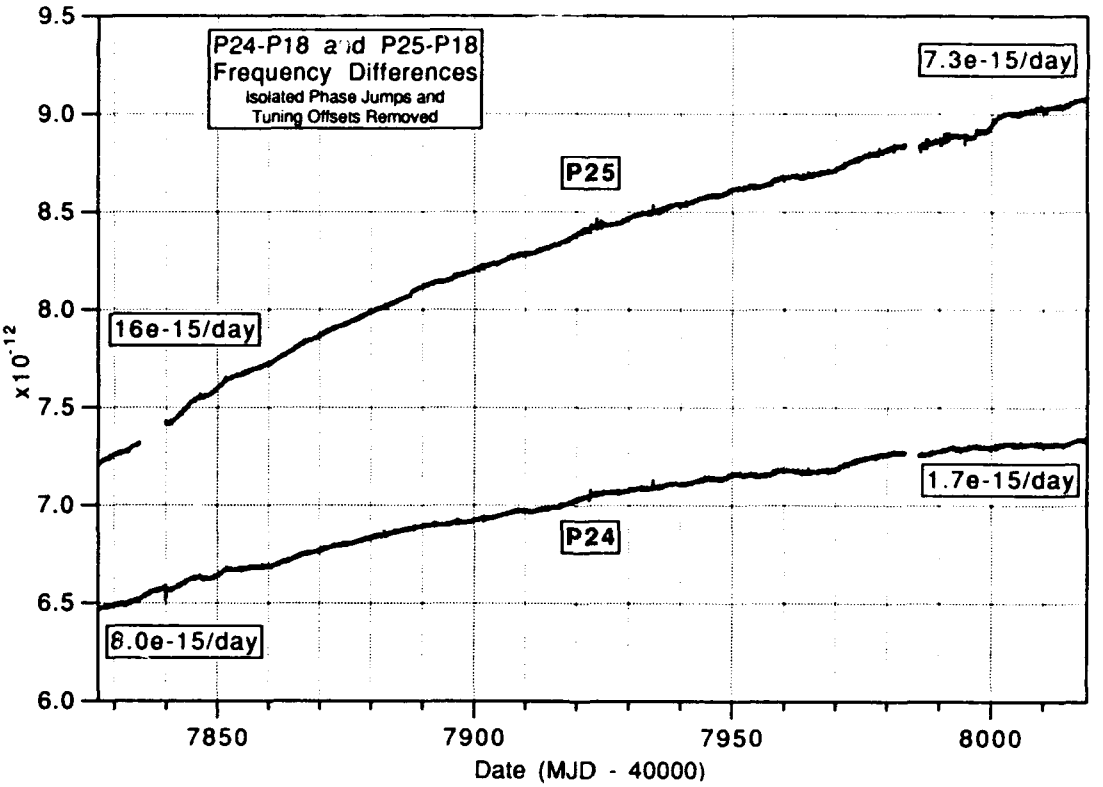


Figure 2. P24 and P25 Frequencies vs. Maser P18

OPTIMIZATION OF HYDROGEN CONSUMPTION IN HYDROGEN MASERS

Bernardo Jaduszliwer and Yat C. Chan

Chemistry and Physics Laboratory, The Aerospace Corporation,
P. O. Box 92957, Los Angeles, CA 90009

Abstract: Hydrogen consumption is an important reliability issue for space-qualified hydrogen masers. An excess of atoms in the upper state of the maser transition must be continuously fed into the maser bulb. Atomic hydrogen is produced in an rf discharge dissociator, and state selection is accomplished by a multipole magnet which focuses atoms in the desired state at the maser's bulb entrance orifice. The focusing properties of these magnets depend strongly on atomic speed. The quality of the match between the speed distribution of the atoms leaving the dissociator and the velocity-dependent transmission of the state selector plays an important role in determining the maser's hydrogen consumption budget. We have determined speed distributions of hydrogen atoms effusing from an rf discharge dissociator, and found that they are significantly narrower than Maxwellians. We have also performed realistic calculations of the focusing properties of hexapole magnet atomic state selectors, and explored the conditions under which good matches between atomic speed distribution and state selector transmission can be achieved.

Introduction

The possibility of using hydrogen masers as frequency standards on board spacecraft requires careful consideration of long-term reliability issues. Since many of the likely failure modes of a maser involve either the atomic hydrogen source system or vacuum pump problems due to the hydrogen load, the maser's reliability will be enhanced by efficient use of its hydrogen supply.

To operate the maser, an excess of hydrogen atoms in the $F=1, M=0$ hyperfine state must be continuously fed [1] into the maser bulb. An rf discharge dissociator breaks hydrogen molecules into atoms, and the state selector (typically a hexapole or quadrupole magnet) focuses those atoms having $F=1, M=0$ at the maser's bulb entrance orifice and defocuses those having $F=0, M=0$, thus creating the population inversion required for maser operation. Clearly, the first requirement for efficient hydrogen use is that a large fraction of the hydrogen molecules flowing into the dissociator exit as atoms. But additionally there is a subtler requirement, caused by the fact that the focusing properties of the state-selecting magnet depend on the atomic velocities; since faster atoms will be deflected less than slower ones during their magnetic field traverse, the "focal length" of the magnet will be longer for those faster atoms. Thus, only atoms within a fairly narrow range of velocities will be focused at the bulb entrance orifice by a given state selector design. If there is a mismatch between the velocity distribution of the atoms coming out of the dissociator and the velocity-dependent transmission of the state selector, the hydrogen use efficiency of the maser could be seriously impaired.

Relatively little is known of the velocity distribution of atoms effusing out of rf discharge dissociators. In many cases the tacit assumption is made that the atoms will be in thermal equilibrium with the dissociator wall, but that is not necessarily the case. The threshold energy for molecular dissociation by collision with electrons in the discharge plasma is about 8.5 eV [2]. Since the molecular binding energy is only about 4.7 eV, each atom carries away approximately 2 eV of excess kinetic energy. Depending on dissociator geometry and gas density, the hydrogen atoms may, or may not, undergo enough bulk and wall collisions to thermalize fully. Velocity distributions of atoms exiting dissociators followed by cryogenically cooled thermal accommodators have been measured [3,4] but these results are not applicable to our problem, since such accommodators would not be used in space-qualified masers.

We have determined the velocity distributions of hydrogen atoms effusing out of an rf discharge dissociator having a geometry and operating parameters resembling those of a maser dissociator. We have also performed realistic calculations of the velocity-dependent transmission of hexapole-magnet state selectors, investigated designs yielding good matches between atomic velocity distribution and magnet transmission, and explored the consequences of possible mismatches. This paper discusses the relevant techniques and presents some results.

Velocity Distributions

Atomic hydrogen velocity distributions have been determined using a magnetic deflection technique described in detail elsewhere [5]. Fig. 1 shows the experimental arrangement. Hydrogen gas is fed through a temperature-controlled Pd-Ag leak [6] into a cylindrical double-walled Pyrex bulb, 15 cm long and 1.9 cm in internal diameter. Compressed-air flow between the walls provides cooling, and rf power is inductively coupled to the discharge by an external 25-turn coil. The hydrogen beam exits the dissociator through a 0.1-cm-long, 0.025-cm-wide slit, is collimated by a second

CH2818-3/90/0000-071 $1.00 © 1990 IEEE

slit, 0.025 cm wide, set at $d = 63.7$ cm away from the source slit, and then travels between the polepieces of an $L = 11.4$-cm-long electromagnet configured in the "two-wire" geometry described by Rabi et al. [7]. After traversing a $D = 71.3$-cm drift space, the beam is detected by a quadrupole mass analyzer. Our dissociator operates over a wide range of rf power levels and hydrogen pressures, as shown in Fig. 2; dissociation fractions of up to 80% can be achieved.

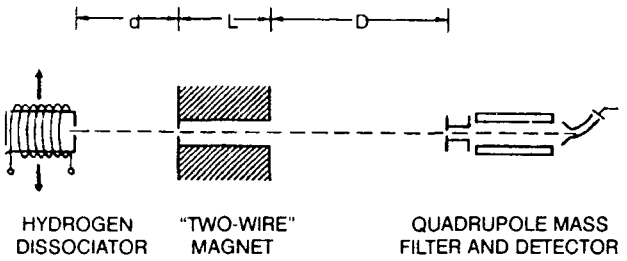


Fig. 1. Schematic view of the experimental arrangement. The hydrogen dissociator can be displaced transversely.

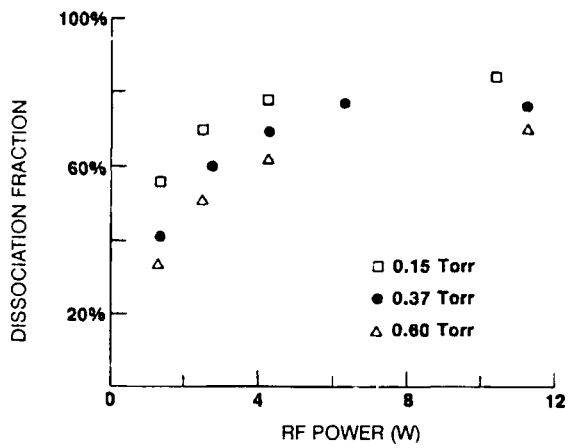


Fig. 2. Atomic hydrogen beam fraction vs rf discharge power. Total dissociator bulb pressures are indicated.

Molecular hydrogen, having no magnetic dipole moment, will travel through the magnet without deflection, but hydrogen atoms will be deflected by the inhomogeneous magnetic field in opposite directions depending on the sign of their magnetic moments. The dissociator is attached to the rest of the apparatus by flexible vacuum bellows, and can be displaced transversely by micrometer screws; in this way, the angular distribution of atoms deflected by the field can be measured. We have shown previously [5] that if $f_0(X)$ is the distribution of detected atoms as a function of source slit position at zero magnetic field and $g(V)$ the speed distribution of the atoms leaving the source slit, then the distribution of detected atoms when the magnetic field is turned on, $f(X)$, will be given by

$$f(X) = \sum_{F,M} \int_0^\infty f_0(X - K_{FM}/V^2)\, g(V) dV \qquad (1)$$

where

$$K_{FM} = \frac{0.492}{a} \frac{\mu}{m} BLd \left(1 + \frac{d}{L + d}\right) \qquad (2)$$

B is the intensity of the magnetic field, α is a magnet geometry parameter (2α is the separation of the "equivalent wires"), m is the atomic mass, and μ the effective magnetic moment of an atom having hyperfine quantum numbers F,M, given by the Breit-Rabi formula [8]. For hydrogen,

$$\mu = \pm \left(\frac{x + M}{(1 + 2Mx + x^2)^{1/2}}\right) \mu_0 \qquad (3)$$

where μ_0 is the Bohr magneton; the (+) sign corresponds to $F = 0$, $M = 0$, and $F = 1$, $M = -1$; the (-) sign, to $F = 1$, $M = 0$, and $F = 1, M = 1$. x is proportional to the ratio of magnetic-to-hyperfine energies, $x = B/(507\ \text{G})$.

To determine the atomic speed distribution, $f_0(X)$ is measured first, and then $f(X)$ is calculated using Eq. (1) and model velocity distributions. The results of the calculations are compared with the measured $f(X)$, and the model distributions are adjusted for best fit to the experimental deflection data. Fig. 3 shows the measured undeflected beam profile $f_0(X)$, as well as the measured deflected beam profile $f(X)$ at a field of 995 G. Our technique to determine velocity distributions has been previously validated, and the

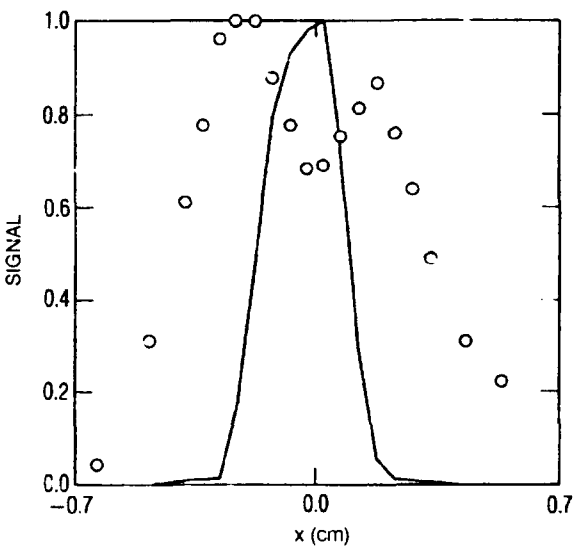


Fig. 3. Detected atomic hydrogen flux vs dissociator slit position. Full line: undeflected beam (zero field). Circles: beam deflected by a 995-G field. Both sets of data have been normalized to unit height. Dissociator pressure: 0.058 Torr.

apparatus has been calibrated, using an effusive rubidium beam [5]. The validation experiment showed that the left side peak in Fig. 3 should be the one used to fit the calculated

deflection profile f(X), since the atoms deflected into the right side peak travel very close to the convex polepiece of the magnet, causing the constant-force approximation used in the derivation of Eq. (2) to break down.

The analysis of preliminary atomic hydrogen deflection data showed [5] that the atomic velocity distributions are significantly narrower than Maxwellians; since then, we have obtained and analyzed additional, higher quality deflection data, and the results support that conclusion. The first model speed distribution used to generate calculated beam deflection profiles was a beam-Maxwellian:

$$g(V) = (2/V_0)(V/V_0)^3 \exp[-(V/V_0)^2] \qquad (4)$$

where the most probable velocity V_0 was treated as a free parameter, since the temperature of the gas within the dissociator might be higher than the wall temperature. An example of "best fit" to the deflection data achievable using a Maxwellian distribution is shown in Fig. 4. Recalling that very fast atoms are not deflected much, and very slow atoms undergo large deflections, Fig. 4 shows that the low and high speed wings of the beam-Maxwellian distribution are both too high. We then attempted to reproduce the deflection data using a Gaussian as a model speed distribution,

$$g(V) = 1/(\sqrt{2\pi}\ \sigma) \exp[-(V-V_0)^2/2\sigma^2] \qquad (5)$$

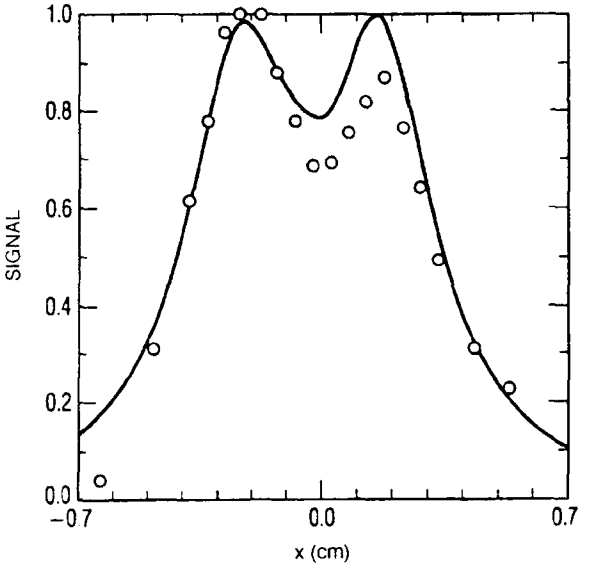


Fig. 4. Atomic hydrogen flux at detector vs dissociator slit position. Circles: as in Fig. 3. Full line: calculated for a beam-Maxwellian velocity distribution, at $T = 350$ K. Both sets of data have been normalized to unit height.

where the most probable velocity V_0 and the width σ are both treated as free parameters. This model reproduced acceptably the preliminary data [5], and fitted the new data better than the Maxwellians, but still failed to provide an acceptable match, due to the smaller error bars in the present measurements. To obtain good fits to the data, we have to allow the speed distribution to be asymmetrical. This was accomplished by using as a model distribution two half-Gaussians joined smoothly at their peaks:

$$g(V) = 1/(\sqrt{2\pi}\ \bar{\sigma}) \exp[-(V-V_0)^2/2\sigma_1^2] \text{ for } V < V_0 \qquad (6a)$$

$$g(V) = 1/(\sqrt{2\pi}\ \bar{\sigma}) \exp[-(V-V_0)^2/2\sigma_2^2] \text{ for } V > V_0 \qquad (6b)$$

where $\bar{\sigma} = (\sigma_1 + \sigma_2)/2$. The most probable velocity V_0 and the two partial widths σ_1 and σ_2 are treated as free parameters. With this model distribution, acceptable matches to the deflection data were obtained over the whole range of dissociator pressures we have explored. Fig. 5 illustrates the match obtained between calculated and measured deflection profiles for the same data shown in Fig. 4, and Fig. 6 shows the speed distributions used to calculate the deflection profiles in Figs. 4 and 5. It is apparent that at low dissociator pressures the atomic speed distributions are nonthermal and significantly narrower than Maxwellians.

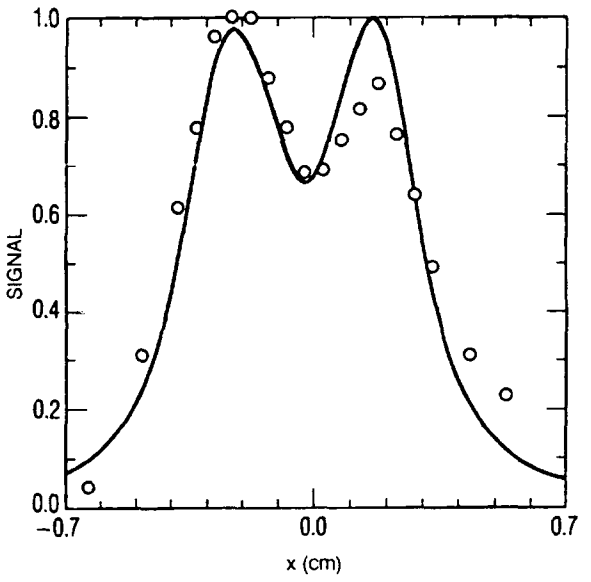


Fig. 5. Atomic hydrogen flux at detector vs dissociator slit position. Circles: as in Fig. 3. Full line: calculated for the asymmetric velocity distribution defined in the text. Both sets of data have been normalized to unit height.

In the range of dissociator pressures we have explored (25 to 350 mTorr), neither the peak velocity nor the partial width of the distribution on the high-speed side change significantly. The dependence of the low-speed side partial width on pressure is shown in Fig. 7. At the high end of our pressure range, the overall width of the distribution is quite close to the Maxwellian width. The peak speed of the distributions corresponds to a kinetic energy of about 0.075 eV, indicating that the hydrogen atoms do lose most of their excess energy before leaving the dissociator. Since, on the other hand, the atoms do not thermalize fully, that energy loss must take place in just a few collisions, requiring a relatively high average energy loss per collision. Impact vibrational excitation of hydrogen molecules within the dissociator bulb, with an ener-

gy loss of 0.546 eV/collision, is a likely mechanism of atomic slowing-down.

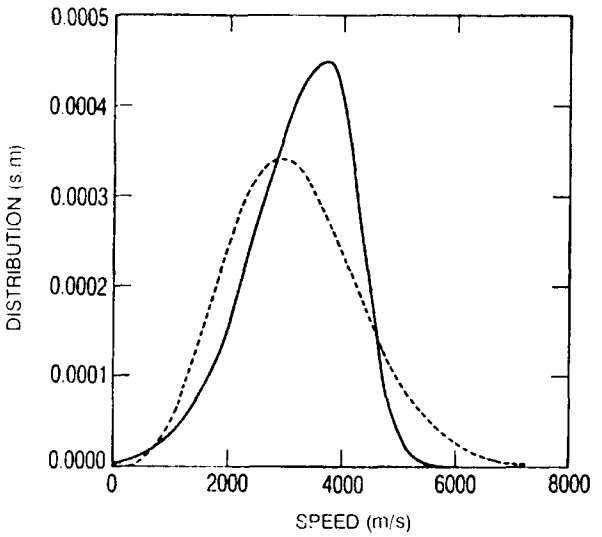


Fig. 6. Velocity distributions used to calculate the deflected beam profiles shown in Figs. 4 and 5. Full line: asymmetric distribution, as defined in text. Dashed line: beam-Maxwellian. Both curves have been normalized to unit area.

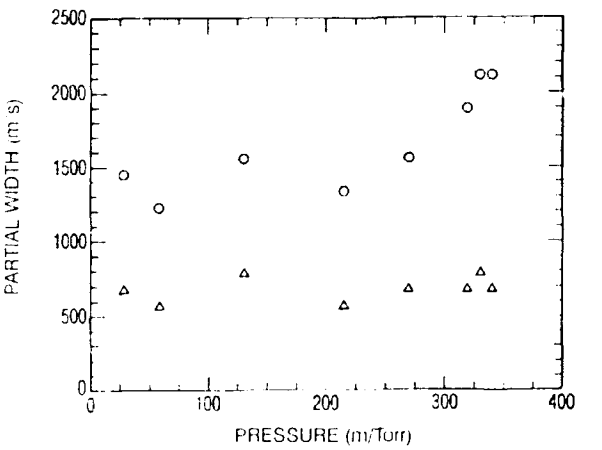


Fig. 7. Partial widths of the best-fitting asymmetric velocity distributions vs total dissociator pressure. Circles: low speed side. Triangles: high speed side.

Hexapole Magnets as Velocity Filters

A hexapole magnet can be used as a state selector in a hydrogen maser because the inhomogeneous hexapolar field will exert a radial force, pointed towards its axis on atoms having a negative effective magnetic moment, and outwards on atoms having a positive effective magnetic moment. In this way, for a well chosen combination of geometry and magnetic field strength, atoms having $F = 1$ and $M = 0$ can be focused at the entrance orifice of the maser bulb, while those atoms having $F = 0$, $M = 0$ will be defocused. The focusing conditions are velocity-dependent, and so the state selector will also act as a velocity filter.

The force acting on an atom immersed in an inhomogeneous magnetic field $\boldsymbol{B}$ is given by $\boldsymbol{F} = \mathrm{grad}(B)$ [9]. The effective magnetic moment, given for hydrogen by Eq. (3), will depend on the atomic quantum numbers F, M and on the magnetic field strength B. The magnetic field strength within a hexapole magnet of bore radius r_0 and poletip field B_0 is given by $B = B_0(r/r_0)^2$ [10]. The radial equation of motion for an atom within the magnet bore is then

$$r = (\mu/m) \times 2(B_0/r_0^2)r \qquad (7)$$

where, for the $M = 0$ states of hydrogen, $\mu = \pm x/(1+x^2)^{1/2}$, and $x = (B_0/507\ \mathrm{G})(r/r_0)^2$. This equation can be rewritten in a form more suitable for numerical solution as

$$\varrho \pm \frac{\beta_0\, \varrho^2}{(1 + \beta_0^2\, \varrho^4)^{1/2}}\, \omega^2\, \varrho = 0 \qquad (8)$$

where $\varrho = r/r_0$, $\beta_0 = B_0/(507\ \mathrm{G})$, and $\omega = (2\mu_0 B_0/m r_0^2)^{1/2}$. The solutions to this equation oscillate about the magnet axis when $\mu < 0$, and are outward bound when $\mu > 0$.

Fig. 8 shows the maser state selector geometry: a hexapole magnet of length L_2 has its entrance plane at a distance L_1 from the source exit plane; the entrance plane to the maser bulb is at a distance L_3 from the exit plane of the magnet. The axial velocity of the atom is constant. The radial velocity is constant in regions a and c, and the radial acceleration in region b is given by Eq. (7). Atomic trajectories leading from the source into the maser bulb are calculated by solving the equations of motion in regions a, b, and c, and matching radial positions and speeds at the boundary planes. The velocity-dependent transmission of the state selector can be calculated by integrating the atomic trajectories leading from source to bulb over starting coordinates and angles. Fig. 9 shows the transmission curves obtained for a given choice of parameters for atoms in both $M = 0$ states. The transmission data have been normalized to the solid angle subtended by the bulb entrance orifice, weighted by the effusive beam angular distribution (i.e., it is measured in units of the total atomic flux per unit speed that would enter the maser bulb in the absence of the magnet). For high speeds, both curves converge asymptotically to 0.25, which is the statistical weight of each of those states. The broad transmission peak for $F = 1$ atoms at 3800 m/s contains those atoms which undergo one radial oscillation in their traverse of the magnetic field; narrower peaks at lower speeds contain atoms which undergo more radial oscillations.

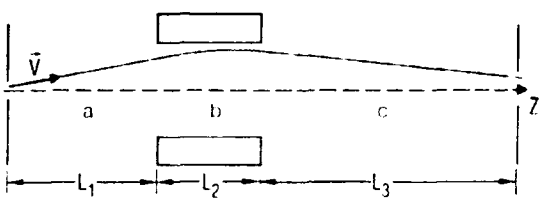


Fig. 8. Schematic view of the state selector geometry (see description in text).

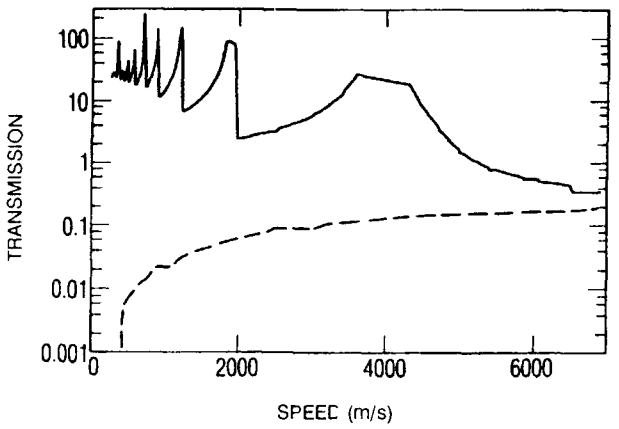


Fig. 9. Transmission of a hexapole magnet (in units of beam flux at the maser bulb entrance orifice in the absence of the magnet) vs atomic speed. Full line: $F=1$, $M=0$ hydrogen atoms (selected state). Dashed line: $F=0$, $M=0$ hydrogen atoms (rejected state).

To obtain the velocity distribution of the transmitted atoms, the velocity-dependent transmission of the magnet has to be folded with the velocity distribution of the atoms entering it. Let us assume that the atoms leaving the dissociator do so with the velocity distribution $g_0(V)$, illustrated by the full line in Fig. 6 (i.e., the distribution yielding a good fit to our beam deflection measurements). The full line in Fig. 10 shows the transmitted distribution when the magnet has been optimized for $g_0(V)$. If, on the other hand, the magnet had been optimized for a beam-Maxwellian distribution at wall temperature, the transmitted distribution would be the one shown by the dashed line in Fig. 10. Integration of the transmitted distributions over atomic speed yields the atomic fluxes into the dissociator. This procedure shows that assuming the atomic speed distribution to be thermal introduces a mismatch between state selector design and atomic speed distribution which causes the waste of 65% of the $F=1, M=0$ atoms leaving the dissociator.

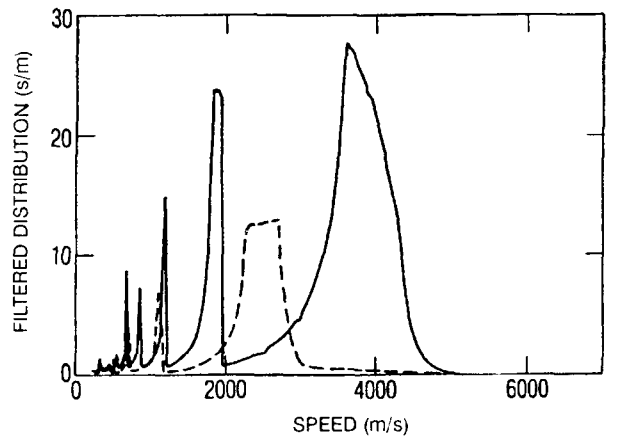


Fig. 10. Velocity distribution of the hydrogen atoms focused at the maser bulb entrance orifice. Full line: state selector optimized for the actual incident beam velocity distribution. Dashed line: state selector optimized for a Maxwellian beam at the dissociator wall temperature.

Conclusions

Our studies of velocity distributions of hydrogen atoms effusing out of rf discharge have shown that while the atoms lose most of their excess kinetic energy rapidly, they may not thermalize fully before exiting. At low dissociator pressures, the atomic hydrogen beam velocity distribution is significantly narrower than the beam-Maxwellian distribution characteristic of thermal equilibrium.

Our analysis of the focusing properties of hexapole magnet state selectors shows that the efficiency of hydrogen use by the maser can be increased significantly by optimizing the state selector design for the actual atomic hydrogen velocity distribution. This finding is of particular importance for the design of space-qualified hydrogen masers, where the maser reliability is enhanced if hydrogen consumption is reduced.

Acknowledgements: This work was supported by the U.S. Air Force Space Systems Division under Contract No. F04701-88-C-0089. The authors would like to thank Dr. R. P. Frueholz for many stimulating discussions.

References

[1] D. Kleppner, H. M. Goldenberg, and N. F. Ramsey, "Theory of the Hydrogen Maser," *Phys. Rev.*, vol. 126, pp. 603–615, 1962.

[2] S. J. B. Corrigan and A. Von Engel, "Excitation and Dissociation of Hydrogen by an Electron Swarm," *Proc. Roy. Soc. A*, vol. 245, pp. 335–351, 1958.

[3] J. T. M. Walraven and I. F. Silvera, "Helium Temperature Beam Source of Atomic Hydrogen," *Rev. Sci. Instrum.*, vol. 53, pp. 1167–1181, 1982.

[4] A. Herscovitch, A. Kponou, and T. O. Niinikoski, "Cold High-Intensity Atomic Hydrogen Beam Source," *Rev. Sci. Instrum.*, vol. 58, pp. 547–556, 1987.

[5] B. Jaduszliwer and Y. C. Chan, "Atomic Velocity Distributions out of Hydrogen Maser Dissociators," *Proc. 21st. PTTI Appl. and Planning Meeting*, Redondo Beach, California, 1989, pp. 223–232.

[6] J. Viennet, P. Petit, and C. Audoin, "Régulateur de Débit d'Hydrogène à Réponse Rapide," *J. Phys. E*, vol. 6, pp. 257–261, 1973.

[7] I. I. Rabi, J. M. B. Kellogg, and J. R. Zacharias, "The Magnetic Moment of the Proton," *Phys. Rev.*, vol 46, pp. 157–165, 1934.

[8] G. Breit and I. I. Rabi, "Measurement of Nuclear Spin," *Phys. Rev.*, vol. 38, pp. 2082–2083, 1931.

[9] N. F. Ramsey, *Molecular Beams*. New York: Oxford University Press, 1956, p. 89.

[10] Ibid., pp. 404–407.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

OPTICALLY PUMPED PRIMARY FREQUENCY STANDARDS

R. E. Drullinger
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway, Boulder CO 80303

Abstract

The use of optical state preparation and detection in atomic beam frequency standards offers tremendous potential for improved short term stability, evaluation and control of accuracy-limiting systematic errors. This paper reviews optical pumping as it pertains to primary frequency standards. The potential benefits and limitations are discussed as is present work on the technology.

Introduction

The development of efficient, inexpensive, tunable, and long-lived diode lasers has made primary clocks with optical state preparation and detection a realistic possibility. Optical state preparation has advantages over conventional magnetic state selection. Magnetic state selection discards most of the atoms in the beam, whereas with optical state preparation, nearly all of the atoms in the beam can be used. This improves the short-term stability of the clock. Optical pumping also produces a homogeneous atomic beam without the spatial velocity dispersion caused by magnetic state selection. The spatial velocity dispersion leads to one of the major accuracy-limiting systematic errors in conventional primary standards. Finally, the replacement of the opaque hot wire detector with a transparent optical detection region makes it possible to run simultaneous counter-propagating atomic beams. This feature, when combined with a frequency control servo system capable of interrogating various parts of the cesium spectrum, makes possible automatic evaluation during continuous clock operation.

The potential for improved accuracy and stability available with this technology has been widely recognized for over a decade, and nearly every national metrology laboratory has at least some effort to investigate it. However, only four labs have efforts involving large, potentially high performance machines. The Communication Research Laboratory in Tokyo is building a machine with about a 1 m interaction length[1]. The National Research Laboratory of Metrology near Tokyo has an operational unit of about the same size which has been evalutated at about 10^{-13} [2]. The Laboratoire Primaire du Temps et des Frequences in Paris is also building a device which is designed to operate at an accuracy of 10^{-13} [3]. The National Institute of Standards and Technology in Boulder is building a device which is designed to become an operational standard with an accuracy of 1 part in 10^{14}. The development of this standard will be used in subsequent sections of this paper to illustrate the technology.

Optical state preparation and detection will be discussed in the first section and used to explain the potential benefits and limitations of this new technology. The potential of these standards for accuracy, combined with their different operational characteristics, has required a careful re-analysis of all errors found in such a standard. This work is outlined in the section on systematic errors. Atomic beam tube design has followed from the error analysis and is briefly described in the next section. Limitations on achievable performance resulting from laser FM noise problems and some solutions are outlined in a section on lasers. Finally, there is a section which discusses some of the requirements on supporting electronics which control the clock frequency and other parameters.

Optical Pumping

Optical pumping to replace state-selecting magnets was first suggested by Kastler[4] in 1950. But it was not practical until tunable lasers were developed. There are a number of ways optical state preparation and detection can be applied in cesium beam tubes and they have been previously described[5-13]. Reference 13 presents an analysis of the various D_2 transitions and gives pumping rates and efficiencies. To illustrate the process, three specific cases will be briefly outlined here. Figure 1 schematically shows two of the electronic energy levels of cesium, the ground state which contains the hyperfine/clock transition and a low lying excited state. An example of the simplest kind of optical pumping would be if the frequency of a laser were tuned to excite the transition $F=4 \rightarrow F'=3$. The $F'=3$ atoms decay back to the ground state in a few nanoseconds. Most go into the $F=3$ state but some return to the $F=4$ state and are excited again. In this way the

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

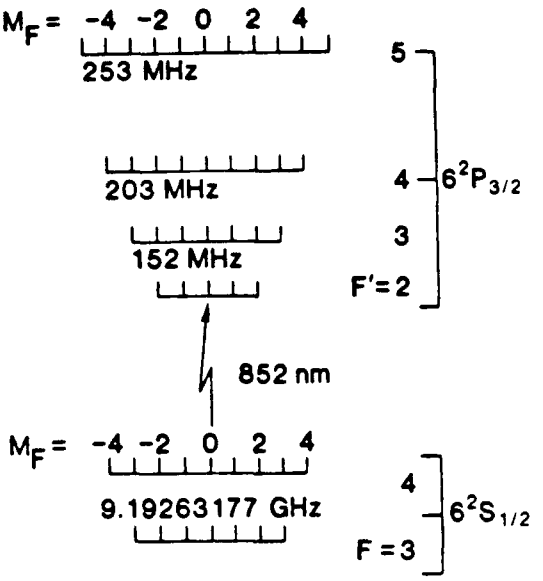


Figure 1. Energy levels of the cesium D_2 transition.

F=4 level is quickly emptied of its population. A slightly more complicated scheme makes use of two lasers[14]. If in addition to the first laser a second laser is used to simultaneously excite the F=3 → F'=3 transition with π polarization, then the atoms are pumped back and forth between the F=3 and F=4 states. However, due to the atomic selection rules applicable to this case, atoms in the F=3, M_F=0 substate cannot interact with the laser radiation and are trapped. As atoms in other substates are forced to jump back and forth between the F=3 and F=4 states they sometimes fall into the trapped state. In this way, the population from all 16 substates can be manipulated into the one substate which is the source state for the clock transition. The last example is that of a cycling transition which can be used to advantage for detection. Because of the selection rule ΔF = 0, ±1, tuning the frequency of a laser to the transition F=4 → F'=5 will produce excited atoms which can decay only to F=4, the state from which they came. In this way the atoms can be excited many times thus insuring their detection by the decay fluorescence.

The advantages and limitations of this technology can now be understood with the aid of the hypothetical clock schematic shown in Figure 2. In this case, 1 or 2 lasers are used to prepare the atomic beam for "clock interrogation." The state-prepared atoms then pass through a conventional Ramsey interrogation zone. Finally, those atoms that make the clock transition can be detected with essentially unit probability by detecting the fluorescence generated in a laser-driven cycling transition which is specific only to atoms in the clock's terminal state.

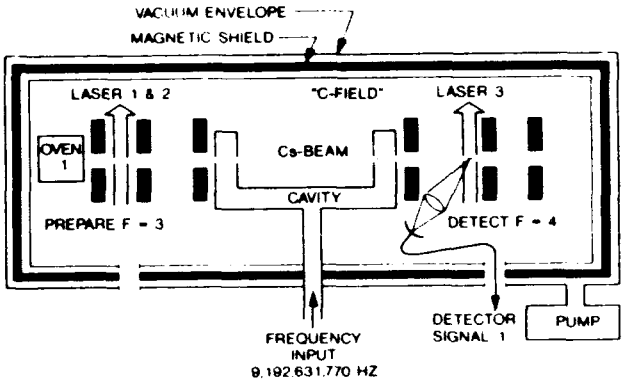


Figure 2. Schematic diagram of an optically pumped, atomic beam frequency standard.

This form of state preparation as opposed to magnetic state selection more efficiently uses the atomic beam flux. For the same atomic flux from the oven, optical pumping can produce or prepare more than 100 times more atoms in the initial clock state than many common magnetic selection schemes. This improved beam flux can directly result in improved short term clock stability.

Other advantages of optical state preparation have to do with accuracy and long term stability of the clock. Optical state preparation does not produce the spatial velocity dispersion in the atomic beam that magnetic state selection does. This comes into effect in the evaluation of end-to-end phase shift, one of the major systematic errors in primary standards. End-to-end phase shift is the result of physical imperfections in the microwave cavity which give rise to a difference in the microwave phase at the two ends of the cavity. This phase difference produces a frequency bias in the standard which is traditionally evaluated by reversing the atomic beam. In magnetically state-selected standards the precision of the beam reversal is limited by the spatial velocity dispersion in the atomic beam. With the homogenous atomic beam produced by optical state selection, the evaluation of cavity phase shift can be performed with greater precision. In fact it can be done in real time without interrupting clock operation.

Figure 2 shows that a second oven can be added to the opposite end of the machine and a counter-propagating atomic beam generated. The two atomic beams can run simultaneously because the flux in each beam is so tenuous that the atoms do not collide. This fact can be better realized if we consider just the velocity spread in one beam. There is a great dispersion between the fastest atoms and the slowest atoms; the faster atoms are constantly overtaking the slower. If the flux were great enough to allow collisions, we would see the results in one beam alone. The atoms in the second beam can pass

through the detection zone for the first beam with no interaction (and vice versa) if they are first optically pumped into the F=3 state. In this way the beam reversal can be done without clock interruption.

Another systematic error can be caused by a spectral line overlap problem known as Rabi tail pulling[15]. The magnetic substate structure of the cesium atom results in a multi-line spectrum on the hyperfine transition. While these lines can be completely resolved spectroscopically, the precision with which the microwave frequency is servoed to the center of the central feature in this spectrum produces a sensitivity to the small residual overlap of the adjacent line wings. In a conventionally state-selected device the population in the magnetic substates is asymmetrically distributed in a way that causes an imbalance in these overlap shifts. Single-laser state preparation, on the other hand, leads to a symmetric spectrum with a cancelation of overlap shifts, and two-laser state preparation eliminates the population in the other substates altogether. This not only improves the accuracy of the clock but gives it added stability against environmentally caused changes in microwave power and magnetic field.

Finally, the elimination of the state-selecting magnets allows the uniform C-field region to be expanded to include the optical state preparation and detection regions. This eliminates Majorana transitions (magnetic field gradient-induced transitions) and frequency errors that may come from them.

All of these potential benefits do not come without limitations. The very aspect of the optical state preparation that gives no velocity selection or dispersion results in a very broad velocity spread in the atomic beam and a comparatively high mean velocity. The slow atoms remain in the microwave field longer than the faster atoms. While the fastest atoms may not be in the field long enough to make the clock transition, the slowest atoms may make the transition, then make it again, and finally end in the starting state. The velocities at which these effects occur are a function of the microwave power and modulation parameters. The result is a slight velocity selectivity that couples microwave power and modulation parameters to the clock frequency through the second-order Doppler shift and cavity phase shift. This effect will show up in the long term stability of the standard. One way to reduce this potential problem is to use modern electronics to manipulate the frequency of the microwave source in ways that produce information about these parameters directly from the cesium spectrum. This will make it possible to operate these parameters under closed loop control.

A new source of error in an optically pumped standard is caused by the fluorescence which comes from the state preparation and detection zones. Some of this light travels directly along the atomic beam path and is present during the interrogation of the clock transition. The result is an AC Stark shift of the hyperfine energy levels. Fortunately, the effect is small and is easily controlled to a level below 1 part in 10^{15} through the choice of geometry and operational conditions[16].

Systematic Effects

The systematic effects which have been analyzed or re-analyzed include: fluorescent light shift[16]; velocity-dependent effects such as second-order Doppler shift and end-to-end cavity phase shift including its dependence on RF power and modulation parameters; Rabi-pulling; cavity pulling; Majorana effects; distributed-cavity phase shift[17]; RF spectral purity and magnetic field uniformity.

Many of the shifts can be expressed as the ratio of two integrals over the velocity distribution containing factors dependent on the microwave power, the modulation parameters, and the particular shift mechanism. For very narrow velocity distributions, the velocity average can be ignored, and the power-and modulation-dependent factors cancel. The shifts then have little or no dependence on microwave power or modulation parameters. An optically pumped standard, however, will use almost all of the broad thermal distribution of velocities emerging from the oven. The shifts then acquire significant dependence on microwave power and modulation parameters. For example, the second-order Doppler shift and end-to-end cavity phase shift can change by 5 to 10% with microwave power changes of only 1 dB.

Second-order Doppler shifts are calculable if the effective velocity profile of the atoms contributing to the signal is known to adequate accuracy. The broad velocity spread in an optically pumped standard not only results in a sensitivity to microwave power but produces a Ramsey resonance with less structure (information content) than conventionally state-selected devices. This has rendered some traditional velocity measurement techniques inadequate. However, a numerical method for extracting both the velocity distribution and the effective microwave power level from Ramsey lineshapes has been developed[18].

Rabi-pulling and Majorana effects should be extremely small in optically pumped standards, but the theoretical studies give new insight into how these effects enter a standard. These studies are briefly outlined in [19], and more detailed publications are in preparation.

Atomic Beam Tube

Optical state preparation and detection offer potential for improved control of the clock resonance line-shape physics primarily through the elimination of magnetic field gradients and spatial velocity dispersion in the atomic beam. The beam tube design should take full advantage of this potential.

Any magnetic field gradients can be nearly eliminated by extending the C-field region to include the entire clock. The non-velocity dispersed atomic beam offers the potential for better beam retrace precision during beam reversal evaluation of end-to-end cavity phase shift errors. Retrace precision is necessary to avoid a sensitivity to distributed cavity phase shift which is not directly evaluated. To realize the full potential of this attribute of optical pumping, a new Ramsey cavity has been developed to minimize the distributed cavity phase shift[17]. Finally, the geometry must be chosen to allow complete evaluation and not produce an unacceptable "fluorescence light shift."

The NIST beam tube is shown schematically in Figure 3. The beam tube is totally symmetric about the central microwave feed point, so only half the tube is shown. The design logic and major sub-systems have been described previously[20]. An axial C-field has been chosen to minimize Rabi-pulling[15], provide a more uniform field and to facilitate the new Ramsey cavity. The fluorescence collection optics are large-radius, spherical mirrors which collect 50% of the fluorescent light and inject it into a light guide for detection outside the vacuum envelope. The imaging nature of this system provides high selectivity against scattered laser light. All laser optics are external to the beam tube.

Laser Systems

Simple, off-the-shelf laser diodes with their inherent FM noise and linewidths of many megahertz are incapable of supporting optically pumped clock operation at full atomic shot-noise-limited performance[21]. To solve this problem a laser line-narrowing technique based on optical feedback from a high-Q cavity has been developed[22]. With this line-narrowing technique, essentially atomic shot-noise-limited performance in an optically pumped standard has been demonstrated[23].

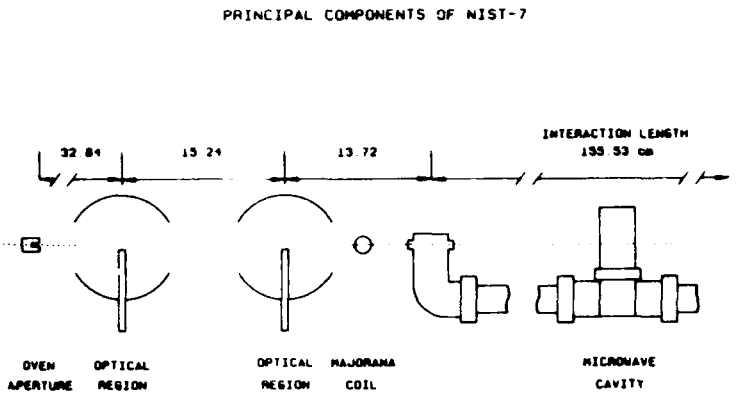


Figure 3. Schematic diagram of NIST-7.

Control Electronics

As pointed out in the discussion of optical pumping, the mean atomic velocity and the velocity spread are greater in optically pumped standards than in most magnetically state-selected standards. This results in a somewhat limited line Q since arbitrarily long beam tubes present other engineering problems, for example, structural, thermal, magnetic and gravitational. The limited line Q combined with greater accuracy goals places a severe burden on the accuracy of the frequency control servo electronics. Additionally, because of the higher sensitivity to microwave power and the higher accuracy we would like to servo such things as microwave power and C-field, which have always been run open-loop in conventional standards. These requirements seem most easily met by a computer-controlled servo that could interrogate the cesium spectrum and develop information about the magnetic field and the microwave power as well as frequency offsets. Correspondingly, for ease of implementation, such a servo would most logically use a slow square-wave modulation scheme. When optical pumping was first considered for use in high performance primary standards, however, such a modulation scheme was not a possibility because the phase noise in available crystal oscillators was too high[24-25]. Fortunately, crystals of adequate performance have recently become available[26].

Conclusions

The technology of optically pumped primary frequency standards should place standards of frequency and time solidly at an accuracy of 10^{-14}. This performance is achievable as a result of the improved short term stability and added control over accuracy-limiting systematic errors. However, the high mean atomic velocity and the broad

velocity spread that result from optical state preparation leads to a limited line Q and comparatively high sensitivity to microwave power. Consequently, to achieve a 10^{-14} accuracy places an extreme burden on the servo electronics, and further improvement in accuracy with thermal atomic beams may not be readily forthcoming. Fortunately, photon pressure cooling and cooled, non-thermal beam technology are rapidly advancing, and new standards should soon be available which overcome these limitations[27].

References

1. T. Maeno, H. Saito, J. Umezu, Y. Ohta. M. Kajita, R Hayashi and E. Morikawa, "Characteristics of a Laser Diode System for an Optically Pumped Cesium Atomic Clock", to appear in Proc. Conference on Precision Electromagnetic Measurements, Ottawa, Canada, June 1990.

2. Shin-Ichi Ohshima, Yasuhiro Nakadan, Takeshi Ikegami Yasuki Koga, R. Drullinger and L. Hollberg, "Characteristics of an Optically Pumped Cs Frequency Standard at the NRLM," IEEE Trans Instrum Meas, vol. IM-38, pp. 533-536 1989.

3. E. de Clercq, A. Clairon, B. Dahmani, A. Gerard, and P. Aynie, "Design of an Optically Pumped Cs Laboratory Frequency Standard," Frequency Standards and Metrology, pp. 120-125, ed. A.DeMarchi, Springer-Verlag, 1988.

4. A. Kastler, "Quelques Suggestions concernant la Production Optique et la Detection Optique d'une Inegalite de Population des Niveaux de Quantification Spatiale des Atomes. Application a l'Experience de Stern et Gerlach et a la Resonance Magnetique", J. Phys. Rad., vol. 11, pp. 255-265 1950.

5. M. Arditi, I. Hirano and P. Tougne, "Optical pumping of a ceasium beam and detection of the 0-0 'clock' transition", J. Phys. D, vol., 11 pp. 2465-2475, 1978.

6. L. L. Lewis, M. Feldman and J. C. Bergquist, "Impact of Lasers on Primary Frequency Standards and Precision Spectroscopy" J. Physique, tome 42, pp. C8-271-281, 1981.

7. L. L. Lewis and M. Feldman, "Optical Pumping be Lasers in Atomic Frequency Standards", Proc. 35th Ann. Freq. Control Symposium, 1981 pp. 612-624.

8. L. L. Lewis, F. L. Walls and D. A. Howe, "Prospects for Cesium Primary Standards at the National Bureau of Standards", Proc. International Conference on Precision Measurement and Fundamental Constants, Gaithersburg, MD 1981 pp. 25-28.

9. M. Arditi, "A Cesium Beam Atomic Clock with Laser Optical Pumping as a Potential Frequency Standard", Metrologia vol. 18, pp. 59-66 1982.

10. Wu Xin-xin, Xu Yue-ting, Tian Kun-yu, Yao Shu-tong, Xia Lin-zhen and Wang Yi-qiu, "Main Aspects in the Design of Optically Pumped Cesium Beam Tube", Acta Metrologica Sinica, vol.6, pp. 15-23 1985.

11. E. de Clercq, M. de Labachelerie, G. Avila, P. Cerez and M. Tetu, "Laser diode optically pumped caesium beam", J. Physique, tome 45, pp. 239-247 1984.

12. G. Avila, E. de Clercq, M. de Labachelerie and P. Cerez, "Microwave Ramsey Resonances from a Laser Diode Optically Pumped Cesium Beam Resonator", IEEE Trans Instrum Meas, vol. IM-34, pp. 139-143 1985.

13. G. Avila, V. Giordano, V. Candelier, E. de Clercq, G. Theobald and P. Cerez, "State selection in a cesium beam by laser-diode optical pumping", Phys. Rev. A, vol. 36, pp. 3719-3728 1987.

14. L. S. Cutler, "Atomic Beam Device Using Optical Pumping," United States Patent 4,425,653.

15. A. DeMarchi, G. D. Rovera, and A.Premoli, "Pulling by Neighbouring Transitions and its Effects on the Performance of Ceasium-Beam Frequency Standards", Metrologia, vol. 20, pp. 37-47 1984.

16. J. Shirley, "Fluorescent Light Shift in Optically Pumped Cesium Standards" in Proc. 39th Annual Symposium on Frequency Control, pp. 22-23, 1985.

17. Andrea DeMarchi, Jon Shirley, David J. Glaze and Robert Drullinger, "A New Cavity Configuration for Cesium Beam Primary Frequency Standards" IEEE Trans Instrum Meas, vol. IM-37, pp. 185-190, 1988.

18. Jon H. Shirley, "Velocity Distributions from the Fourier Transforms of Ramsey Line Shapes", in Proc. 43rd Annual Symposium on Frequency Control, pp. 162-167, 1989.

19. R. E. Drullinger, J. H. Shirley, D. J. Glaze, and L. Hollberg, "An Optically Pumped Primary Frequency Standard", Frequency Standards and Metrology, pp. 116-118, ed A.DeMarchi, Springer-Verlag 1988.

20. R. E. Drullinger, Jon Shirley, D. J. Glaze, L. W. Hollberg, and A. DeMarchi, "Progress Toward an Optically Pumped Cesium Beam Frequency Standard" in Proc. 40th Annual Symposium on Frequency Control, pp. 428-431, 1986.

21. A. Derbyshire, R. E. Drullinger, M. Feldman, D. J. Glaze, D. Hillard, D. A. Howe, L. L. Lewis, J. H. Shirley, I. Pascaru, and D. Stanciulescu, "Optically Pumped Small Cesium Beam Standards: A Status Report", in Proc. 39th Annual Symposium on Frequency Control, pp. 18-21, 1985.

22. B. Dahmani, L. Hollberg, and R. Drullinger, "Frequency Stabilization of Semiconductor Lasers by Resonant Optical Feedback", Optics Letters, vol. 12, pp. 876-878, 1987.

23. S. Ohshima, Y. Koga, Y. Nakadan, L. Hollberg and R. Drullinger, "The Effect of Laser Line Narrowing on the Performance of Optically Pumped Cesium Atomic Beam Frequency Standards", in Proc. 2nd European Frequency and Time Forum, pp. 531-536, 1988.

24. F. L. Walls, "Stability of Frequency Locked Loops," Frequency Standards and Metrology, ed. by A. DeMarchi, Springer-Verlag 1988, pp. 145-149.

25. C. Audoin, V. Candelier and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards", to appear in the Proc. Conference on Precision Electromagnetic Measurements, Ottawa, Canada June 1990.

26. F. L. Walls, private communication, NIST, 325 Broadway, Boulder, CO 80303.

27. See for example articles in "Frequency Standards and Metrology, ed. by A. DeMarchi, Springer-Verlag, 1988.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

LINEAR ION TRAP BASED ATOMIC FREQUENCY STANDARD*

J. D. Prestage, G. J. Dick, L. Maleki

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

Abstract

Our goal at the Time and Frequency Systems Research group at JPL is the development of a trapped ion based fieldable frequency standard with stability $1 \cdot 10^{-13}/\sqrt{\tau}$ for averaging times $\tau > 1 \cdot 10^4$ seconds. To achieve this goal we have developed a hybrid rf/dc linear ion trap which permits storage of large numbers of ions with reduced susceptibility to the second-order Doppler effect caused by the rf confining fields. We have confined $^{199}Hg^+$ ions in this trap and have measured very high Q transitions with good signal to noise. In preliminary measurements we have obtained stabilities of $1.6 \cdot 10^{-13}/\sqrt{\tau}$ ($50 < \tau < 800$ seconds) with a 160 mHz wide atomic resonance linewidth and a signal-to-noise ratio of 40 for each measurement cycle. Atomic resonance lines as narrow as 30 mHz on the 40.5 GHz clock transition have been measured with no appreciable reduction in the ion signal. A stability of $7 \cdot 10^{-14}/\sqrt{\tau}$ is made possible by the signal to noise and line Q of this measured transition. Analysis of fundamental sources of frequency instability indicates that long term stability of $2 \cdot 10^{-16}$ is feasible for this device with existing technology for $\tau \geq 10^6$ 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 millisecond pulsar, PSR 1937+27, shows stability in its rotational period that exceeds that of all man-made clocks for averaging times longer than 6 months. Comparison of this pulsar period with an earth based clock of stability $1 \cdot 10^{-15}$ over averaging periods of one year is expected to show the effects of very low frequency gravitational waves[1,2]. Spacecraft ranging measurements across the solar system would be improved with earth based clocks whose stabilities exceeded $1 \cdot 10^{-15}$ for averaging times of 10^4 to 10^5 seconds. This clock performance would also improve gravity wave searches in spacecraft ranging data. Another use for long term stable clocks in NASA's Deep Space Network would be in maintaining syntonization with UTC.

We are developing a fieldable frequency standard based on $^{199}Hg^+$ ions confined in a linear ion trap which should show unprecedented long term frequency stability. Typically the largest source of frequency offset stems from the motion of the ions caused by the trapping fields via the second-order Doppler or relativistic time dilation effect. Moreover, instability in certain trapping parameters, e.g., trap field strength, temperature, and the actual number of trapped particles will influence the frequency shift and lead to frequency instabilities. Since this offset grows with the number of ions, a trade-off situation results, where fewer ions are trapped in order to reduce the (relatively) large offset which would otherwise result.

A conventional hyperbolic or Paul trap is shown in Fig. 1. With the electrodes biased as shown ions are trapped around the point node of the rf electric field at the center. The strength of the electric field and the resulting micromotion of the trapped particles grows linearly with distance from this node point. As ions are added the size of ion cloud grows until the second order Doppler shift arising from the micromotion in the trapping field dominates the second order Doppler shift from the ion's thermal motion

*This work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

CH2818-3/90/0000-082 $1.00 © 1990 IEEE

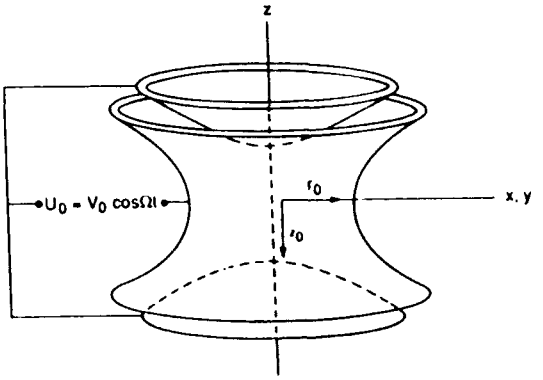


Figure 1: A conventional hyperbolic RF ion trap. A node of the RF and DC fields is produced at the origin of the coordinate system shown.

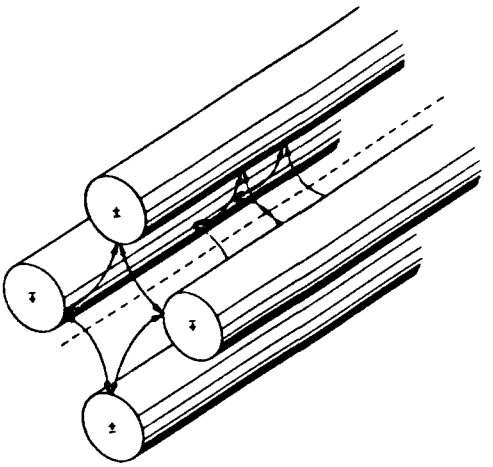


Figure 2: The rf electrodes for a linear ion trap. Ions are trapped along the line of nodes of the rf field with reduced susceptibility to second-order Doppler frequency shift.

at room temperature. For typical operating conditions[3,9] a spherical cloud containing $2 \cdot 10^6$ mercury ions shows a 2nd order Doppler shift of $2 \cdot 10^{-12}$, a value some ten times larger than that for mercury ions undergoing room temperature thermal motion.

We have designed and are currently testing a hybrid rf/dc linear ion trap which allows an increase in the number of stored ions with no corresponding increase in second-order Doppler shift from the micromotion generated by the trapping fields. This trap confines ions along a line of nodes of the rf field (see Fig.2). The trapping force transverse to the line of nodes is generated by the pondermotive force as in conventional Paul traps while the axial trapping force is provided by dc electric fields [3-7].

We can compare the second-order Doppler shift, $\Delta f/f$ generated by the trapping fields for a cloud of ions in a linear trap and a conventional Paul trap [3,4] assuming that both traps are operated so that the ions have the same secular frequency ω. When the same number of ions, N, are held in both traps the average distance from an ion to the node line of the trapping field is greatly reduced in the linear trap. Since the perpendicular distance from the line of nodes determines the magnitude of the rf trapping field the 2nd order Doppler shift of an ion's transition frequency due to motion in the trapping field is reduced from that of a conventional point node trap. If R_{sph} is the ion cloud radius in the Paul trap and L is the ion cloud length in the linear trap the Doppler shift in the two traps are related by[3]

$$\left(\frac{\Delta f}{f}\right)_{lin} = \frac{5}{3}\frac{R_{sph}}{L}\left(\frac{\Delta f}{f}\right)_{sph}. \qquad (1)$$

As more ions are added to the linear trap this shift will increase. It will equal that of the spherical ion cloud in a hyperbolic trap when

$$N_{lin} = \frac{3}{5}\frac{L}{R_{sph}}N_{sph}. \qquad (2)$$

These expressions are valid when the ion cloud radii, R_{lin} and R_{sph}, are much larger than the Debye length. This is the characteristic plasma density fall off length at the ion cloud edge and is about 0.1 mm for typical Hg^+ ion plasmas used in frequency standard work [3,9].

In addition to its larger ion storage capacity the dependence of the 2nd order Doppler shift on trapping parameters in a linear trap is very different from that in a conventional Paul trap. For many ions in a Paul trap this shift is given by[3,12]

$$\left(\frac{\Delta f}{f}\right)_{sph} = -\frac{3}{10c^2}\left(\frac{N\omega q^2}{4\pi\epsilon_o m}\right)^{2/3} \qquad (3)$$

where ω is the secular frequency for a spherical ion cloud containing N ions each with charge to mass ratio q/m. c is the speed of light and ϵ_o is the permittivity of free space. Ions in a linear trap show a 2nd order Doppler shift from the motion generated by the rf confining field given by[3]

$$\left(\frac{\Delta f}{f}\right)_{lin} = -\left(\frac{q^2}{8\pi\epsilon_o mc^2}\right)\frac{N}{L} \qquad (4)$$

where L is the length of the ion cloud.

In contrast to the spherical case as described in Eq.(3), this expression contains no dependence on trapping field strength, as characterized by ω, and depends only on the linear ion density N/L. That is, for an infinitely long linear trap where end effects are negligible, if the rf confining voltage increases and consequently the micromotion at a given point in space increases, the ion cloud radius will decrease (because the radial trapping force increases with rf level) so that the 2nd order Doppler from the trap fields remains constant.

The finite length linear trap depends on variations in radial trapping strength (characterized by ω) as[4]

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = -2\frac{R_t}{L}\frac{\delta\omega}{\omega}. \qquad (5)$$

Similarly, for variations in endcap voltage we find

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{lin}}{\left(\frac{\Delta f}{f}\right)_{lin}} = 2\cdot\frac{R_t}{L}\frac{\delta V_e}{V_e} \qquad (6)$$

where R_t is the trap radius. The Paul trap shows a corresponding sensitivity to trap field strength variations

$$\frac{\delta\left(\frac{\Delta f}{f}\right)_{sph}}{\left(\frac{\Delta f}{f}\right)_{sph}} = -\frac{2}{3}\frac{\delta\omega}{\omega}. \qquad (7)$$

A comparison of Eqs. (5) and (7) shows the linear trap based frequency standard to be less sensitive to variations in trapping field strength than the Paul trap by a factor of 3Rt/L. For the trap described in the next section this factor is about 1/3.

Linear Trap Description

Our linear trap is shown in Fig. 3. The full size of the total system is about 150cm tall by 60cm square as shown in Fig. 4. The operation of the trap as a frequency standard is similar to previous work [8,9]. The 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

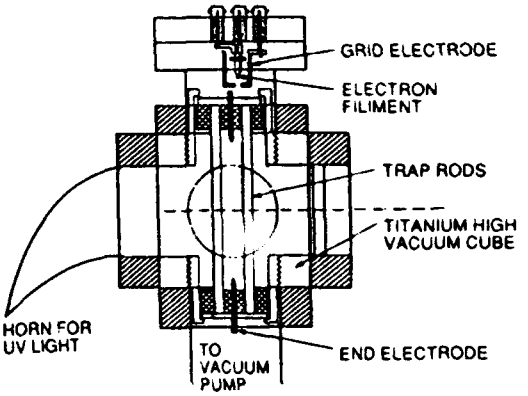


Figure 3: Linear Ion Trap Assembly View. The trap is housed in a 3.375" vacuum cube. State selection light from the ^{202}Hg discharge lamp enters from the right, is focused onto the central 1/3 of the trap and is collected in the horn. Fluorescence from the trapped ions is collected in a direction normal to the page.

discharge lamp optically pumps the ions into the F=0 hyperfine level of the ground state. 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 as shown in Fig. 4. The placement of the LaB_6 electron filament is also chosen to prevent light from the white hot filament from entering the collection system. Its placement and relatively cool operating temperature together with good filtering of the state selection/interrogation UV light in the input optical system have allowed frequency standard operation without the use of a 194 nm optical bandpass filter in the collection arm. This triples data collection rates since such filters typically have about 30% transmission for 194 nm light.

Microwave radiation (40.5 GHz) propagates through the trap perpendicular to the trap axis thereby 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. This radiation enters the trap region through the pyrex horn (see Fig. 4) and propagates in the opposite direction to the UV

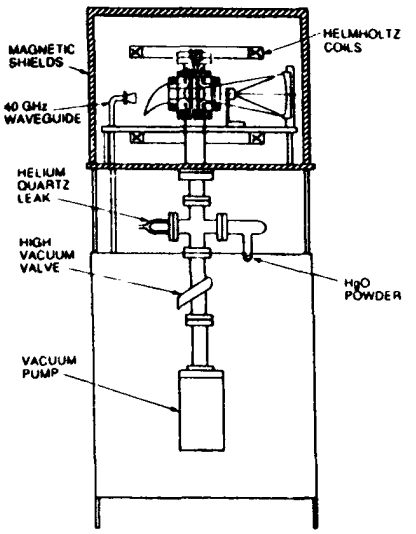


Figure 4: The full trap vacuum system and support stand. The full height is about 5 feet and is about 2 feet on each side. In the measurements described here the magnetic shields were absent.

state selection/interrogation light. This allows fluorescence collection in both directions perpendicular to the plane of the page in Fig.4. For the resonance and stability data shown below fluorescence was collected in only one of these two directions.

Frequency Standard Operation

We have used Ramsey's technique of successive oscillatory fields to probe the approximately 40.5 Ghz clock transition in $^{199}Hg^{+}$ ions confined to the linear trap described above.In these measurements the 40.5 Ghz signal is derived from an active Hydrogen maser frequency source. One of the first resonance lines measured is shown in Fig. 5. The sequence of operations leading to each of the 400 measurements (10 mHz per frequency step) begins with a .5 second electron pulse to maintain a steady state population of trapped Hg ions. The ^{202}Hg discharge lamp is on during this time preparing the ions in the F=0 hyperfine state of the ground level. The rf power driving the lamp discharge is then reduced so that only a dim discharge exists. At the same time a .75 second microwave pulse is started followed by a 1.7 second free precession period. After a final .75 second microwave pulse the lamp is switched to the bright discharge mode. The signal shown in Figs.

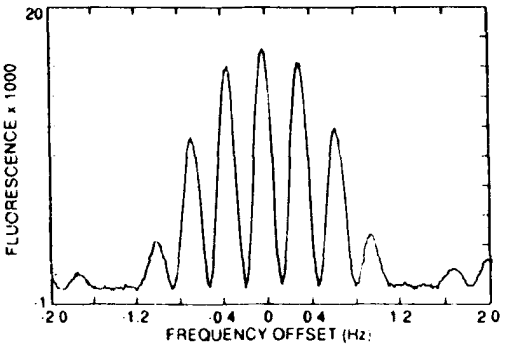


Figure 5: $^{199}Hg^{+}$ clock transition as measured with Ramsey's method of successive oscillating fields. This line shape results from two 0.75 second microwave pulses separated by a 1.7 second free precession period. The central line is about 160 mHz wide.

5, 6, and 7 is the atomic fluorescence counted during a 1 to 1.5 second period following the lamp turn on. The measurement shown in Fig. 5 is the average of 20 scans. The line shape shown in Fig. 6 was obtained with two pulses of .275 second duration separated by a 4 second free precession period. It is the average of 7 scans. Our highest Q line is shown in Fig. 7 and is derived from two pulses .275 seconds long separated by a 16 second free precession. The linewidth $\Delta f = 30$ mHz represents a line $Q = f/\Delta f$ of $1.3 \cdot 10^{12}$ on the 40.5 Ghz transition. The data shown is an average of 4 full scans.

We have locked the output frequency of a 40.5 Ghz source to the frequency of the central peak of the resonance shown in Fig. 6 in a sequence of 512 measurements [8]. The time required for each measurement is about 6.3 seconds. By averaging the frequencies of 2^N adjacent measurements (N=1,2,..,7) we form the modified Allan variance giving frequency stability as shown in Fig.8. Also shown in that figure are other frequency standards including the active hydrogen masers used in JPL's Deep Space Network. The upper dashed curve represents expected performance based on the improved line Q and SNR demonstrated in the 30 mHz resonance of Fig. 7. The lower dashed line shows the expected improvement by the addition of a second set of collection optics, as allowed by our trap geometry. This latter performance is projected to be approximately $5 \cdot 10^{-14}/\sqrt{\tau}$ for $\tau > 150$ seconds. All measurements reported in this work were made with no magnetic shielding of the ion trap re-

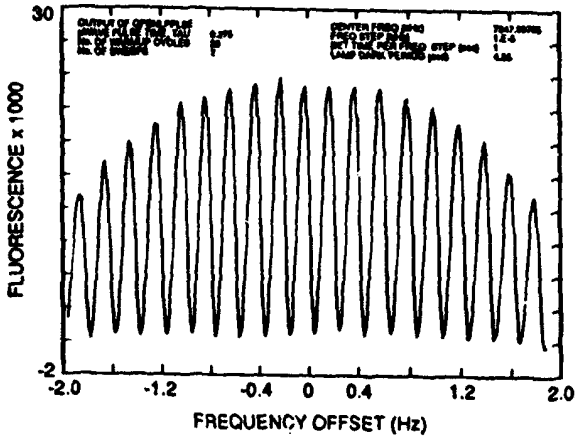


Figure 6: Resonance data for two 0.275 second pulses separated by a 4 second free precession period. The curve shown is an average of 7 scans with a background light level of about 150000 subtracted. The central peak is about 110 mHz wide.

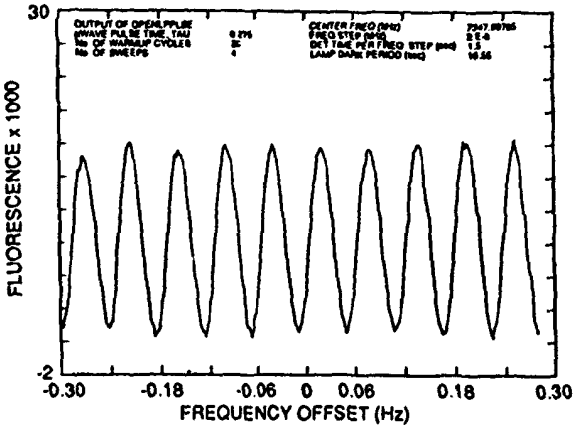


Figure 7: The highest line Q data. In this measurement the two microwave pulses of 0.275 seconds are separated by a 16 sec free precession period. The fringe width is 30 mHz and the data shown is an average of 4 scans.

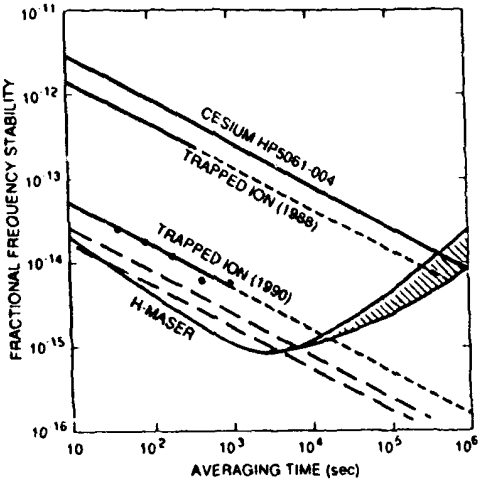


Figure 8: Measured stability for the 160 mHz resonance line of Figure 5. Fractional frequency stability is $1.6 \cdot 10^{-13}/\sqrt{\tau}$ for $50 > \tau > 800$ sec. The dashed lines below the measured data represent projected stability for the 30 mHz data shown in Fig.7. The lowest dashed line is the expected performance when the fluorescence collection system is fully implemented ($5 \cdot 10^{-14}/\sqrt{\tau}$).

gion. Variations in the ambient laboratory fields in the absence of shields degraded long term stability ($\tau > 800$ seconds).

Local Oscillator Requirements

As the performance of passive atomic and ionic frequency standards improves, a new limitation is encountered due to fluctuations in the local oscillator (L.O.). This limitation continues to the longest times, having the same $1/\sqrt{\tau}$ dependence on measuring time τ as the inherent performance of the standard itself. The cause of this effect is time variation of the sensitivity to L.O. fluctuations due to the interrogation process. This limitation was evaluated in a recent calculation for sequentially interrogated passive standards[10]. Since our trapped ion standard is of this type, the analysis should be directly applicable. In a new calculation the limitation has also been evaluated for the sine wave FM interrogation used in cesium and rubidium standards[11].

Roughly speaking, the analysis shows that the local oscillator must have frequency stability at least

as good as that of the standard itself, for a measuring time equal to the cycle time t_c. Thus our trapped mercury ion frequency source with performance of $\sigma_y(\tau)|_{STD} = 1.6 \cdot 10^{-13}/\sqrt{\tau}$ and a cycle time of $t_c = 6$ seconds requires a L.O. with performance of $\sigma_y(6)|_{L.O.} = 1.6 \cdot 10^{-13}/\sqrt{6} \approx 7 \cdot 10^{-14}$. This value is 4 to 12 times lower than that available from presently available quartz oscillators. Operation of the trapped ion standard with a longer cycle time and higher performance places an even more stringent burden on the local oscillator. This can be ameliorated somewhat if the dead time can be kept very small[10].

While stand-alone operation at the highest performance levels may place unattainable requirements on available crystal quartz local oscillators, application as a stabilizer for a hydrogen maser, or other ultra-high stability source such as a high-Q cryogenic oscillator, is straightforward. However, in such an application, the hydrogen maser's frequency would not be steered to that of an independently operating trapped ion source. The maser's output signal would instead be used itself to interrogate the ionic transition. Information thus gathered would be used to compensate for long term variation in the maser frequency.

Sources of Frequency Instability

Figure 9 shows the leading sources of perturbations to the ion hyperfine clock transition in the operating conditions we foresee. By running the clock with about $5 \cdot 10^6$ trapped ions we will have reduced the 2nd order Doppler shift from the trap field to near its minimum value for room temperature operation. Controlling ion number to 0.1% will stabilize this source of frequency jitter to $2 - 3 \cdot 10^{-16}$. This level of ion number control has been demonstrated by Cutler et al[12].

Only modest temperature regulation (0.1 K) is required to reduce variation in the 2nd order Doppler shift due to thermal motion to a value smaller than $1 \cdot 10^{-16}$. Variations in helium buffer gas pressure also influence the clock transition[12] and will require pressure stabilization to about 0.4%. This requirement is less stringent than regulation of hydrogen gas pressure in present day commercial active H-masers.

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. Operating at 0.2μT (2 mG) field would require stabilization of the current in the Helmholtz coils to 0.5% (Fig. 4) to reach the 10^{-16} level. At this field the $^{199}Hg^+$ sensitivity is $1 \cdot 10^{-14}$ per 10 μT (mG) of field change at the position of the ion. To prevent variations in ambient field from influencing the clock's output frequency we will install a triple layer set of shields with shielding factor 10,000.

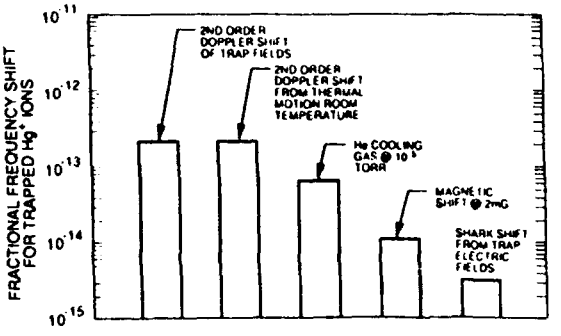


Figure 9: Perturbations to the $^{199}Hg^+$ clock transition for approximately $5 \cdot 10^6$ ions held in the linear ion trap described in the text.

The Stark shift of the atomic transition from the rf electric field[13] is smaller than the 2nd Doppler induced by the same electric field, thus, controlling 2nd Doppler will more than adequately control the Stark shift.

Conclusions

We have demonstrated the increased signal-to-noise and short term stability inherent in a linear ion trap based frequency standard. Clock operation with line $Q = 2.5 \cdot 10^{11}$ has achieved performance of $1.6 \cdot 10^{-13}/\sqrt{\tau}$ for $50 < \tau < 300$ seconds. Measured line Q's of as high as $1.3 \cdot 10^{12}$ have been measured, indicating consequent performance for this trap as high as $5 \cdot 10^{-14}/\sqrt{\tau}$ for $\tau > 150$ seconds. The requirement for local oscillator stability required to achieve this performance is quite stringent. However, an application to stabilize the frequency of a hydrogen maser or cryogenic oscillator for very long times seem straightforward. Analysis of fundamental sources of atomic perturbations and their control shows that stability of 10^{-16} is feasible for $\tau > 10^6$ seconds.

References

[1] D. W. Allan, "A Study of Long Term Stability of Atomic Clocks", *Errata Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 375–380, 1987.

[2] M. M. Davis, J. H. Taylor, J. M. Weisberg,and D. C. Backer, "High-precision timing observations of the millisecond pulsar PSR1937+21", *Nature* **315**, 547-550, 1985.

[3] J. D. Prestage, G. J. Dick, L. Maleki, "New Ion Trap for Frequency Standard Applications", *J. Appl. Phys.* **66**, No. 3, 1013–1017, August 1989.

[4] J. D. Prestage, G. R. Janik, G. J. Dick, T. K. Tucker, and L. Maleki, "Linear Ion Trap for Second-Order Doppler Shift Reduction in Frequency Standard Applications",in *Proceedings of the 43rd Ann. Symposium on Frequency Control*, 135–142, 1989.

[5] G. R. Janik, J. D. Prestage, and L. Maleki, "Simple Analytic Potentials for Linear Ion Traps", *J. Appl. Phys.* **67**, No. 10, 6050–6055, May 1990.

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

[7] D. J. Wineland, J. C. Bergquist, J. J. Bollinger, W. M. Itano, D. J. Heinzen, S. L. Gilbert, C. H. Manney, and C. S. Weimer, "Progress at NIST Toward Absolute Frequency Standards Using Stored Ions", in *Proceedings of the 43rd Ann. Symposium on Frequency Control*, 143–150, 1989.

[8] J. D. Prestage, G. J. Dick, and L. Maleki, "The JPL Trapped Ion Frequency Standard Development, *Proc. 19th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting*, 285–297, 1987.

[9] 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.

[10] 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).

[11] C. Audoin, V. Candelier, and N. Dimarcq, "A Limit to the Frequency Stability of Passive Frequency Standards," this conference.

[12] L. S. Cutler, R. P. Giffard, and M. D. McGuire, "Thermalization of ^{199}Hg Ion Macromotion by a Light Background Gas in an RF Quadrupole Trap", *Appl. Phys. B* **36**, 137–142, 1985.

[13] M. Jardino, M. Desaintfuscien, and F. Plumelle, "Prospects for a Mercury Ion Frequency Standard", *Jour. De Physique*, **C8**, 327–338, 1981.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

IMPROVED PERFORMANCE OF THE SUPERCONDUCTING CAVITY MASER AT SHORT MEASURING TIMES*

R. T. Wang and G. J. Dick

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

Abstract

Recent measurements on the superconducting cavity maser (SCM) oscillator show frequency stability of parts in 10^{15} for times from 1 second to 1000 seconds. Phase noise of approximately $-80dB/f^3$ was also measured. We believe this short- and mid-term performance to be better than that of any known microwave oscillator. In particular, measured stability at 1 second interval is 10 times better than that of a hydrogen maser, and phase noise at 8 GHz is more than 20 dB below that of the best multiplied quartz crystal oscillators.

Introduction

The superconducting cavity maser (SCM) is a helium cooled, all-cryogenic oscillator with superior stability at short measuring times[1,2,3,4]. It differs from other superconducting cavity stabilized oscillator (SCSO) designs[5,6,7] in its use of a very rigid ($Q \approx 10^9$) sapphire-filled stabilizing cavity, and in its all-cryogenic design; excitation being provided by an ultra-low noise cryogenic ruby maser.

A comparison of ultra-stable atomic frequency sources shows active hydrogen masers to be superior to passive atomic standards in short term stability (1 second $< \tau <$ 100 seconds). Performance of the SCM at short measuring times is superior even to the active hydrogen maser. Like the hydrogen maser, the SCM is also an active oscillator. The advantage of the SCM is its larger output signal power ($\approx 10^{-9}$Watt vs $\approx 10^{-12}$Watt for the hydrogen maser). Long term performance is limited by variation of the operating parameters, such as temperature, drive power, output VSWR etc., depending on the sensitivity of the SCM to these various parameters.

Figure 1 shows a block diagram of the improved oscillator. The three cavity oscillator, consisting of a ruby maser, coupling cavity, and a high Q lead on sapphire cavity, have been discussed previously[2]. Oscillation at a frequency of 2.69 GHz results from ruby maser operation with a 13.1 GHz pump frequency to create a population inversion. Energy level splittings in the Ruby are matched to that of the high-Q cavity by means of a bias field provided by a superconducting solenoid. Frequencies of the three modes of the coupled cavity system are spaced relatively close to each other (5% spacing) in order to couple effectively, but are spaced far enough from each other to allow mode selection by adjustment of the bias field[2].

Experimental Aspects

Substantial technical improvements have been made to eliminate frequency instability due to operational parameters. They are temperature, pump frequency, pump power, pump frequency polarization, temperature gradient, coupling strength and output VSWR. We have either stabilized the parameter or minimized the coefficient which couples the parameter to the operating frequency.

* This work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

CH2818-3/90/0000-089 $1.00 © 1990 IEEE

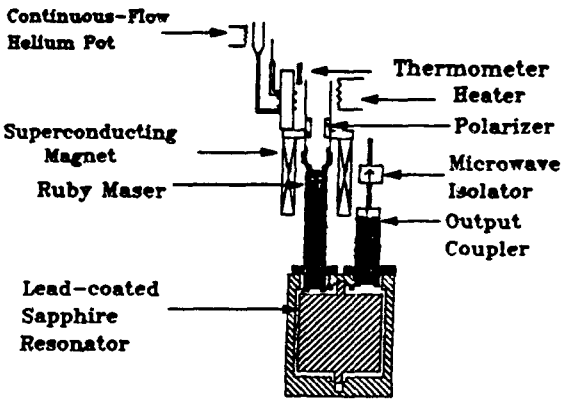


Figure 1: Schematic diagrams of the Superconducting Cavity Maser (SCM) oscillator with improved temperature control system. Recent modifications include consolidation of heating and cooling elements to prevent thermal regulation power from flowing through the oscillator assembly. A direct output coupler and microwave isolator were installed to reduce noise and increase stability. The cryogenic polarizer was added to provide effective and reproducible ruby pumping.

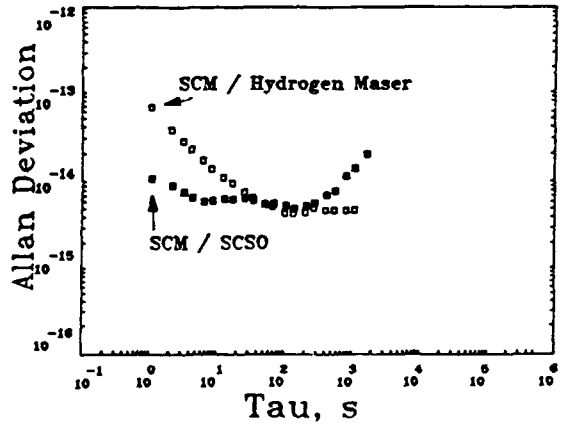


Figure 2: Two sample Allan Deviation of the SCM tested with SCSO and hydrogen maser references. Filled squares show data of SCM/SCSO; open squares are SCM/Hydrogen Maser data.

The temperature dependence of the output frequency for the SCM shows an extremum in the range between 1 and 2 Kelvins[4]. From a functional point of view, the presence of the frequency maximum at about 1.57 Kelvin is an extremely desirable feature since it allows operation of the oscillator in a region of vanishingly small temperature coefficient. The quadratic coefficient in $\delta f/f$ at the maximum is $3.3 \cdot 10^{-9}/\text{Kelvin}^2$. Thus a temperature accuracy of one milliKelvin together with a stability of 30 microKelvins allows a frequency stability of $\frac{\delta f}{f} = 2 \cdot 10^{-16}$.

Frequency dependence on microwave pump frequency and amplitude has also been studied[4]. Since the pump power is very much more difficult to stabilize than its frequency, a major feature of the results to date is a "valley" where the sensitivity to pump power is greatly reduced. In this region the slope is $\leq 2 \cdot 10^{-13}/\text{db}$, a value 100 times smaller than was typically found.

Several recent improvements have made increased stability possible. They are listed as following:

1. Extension of operational period: The 1.57 Kelvin cooling system was changed from closed bath refrigeration to continuous flow. The operational period was extended from 3 days to 7 days, limited only by the storage time of the larger 4.2 Kelvin helium bath from which the small flow is drawn. We expect to extend this to 30 days with a better dewar. This modification also provides continuous operation of the SCM during helium transfer.

2. Improved temperature stability: Temperature fluctuation has been minimized to 40 microKelvins, a factor of 1000 improvement. Previously we measured a parabolic curve of oscillation frequency versus temperature with a frequency maximum near 1.57 Kelvin. With present temperature control capacity, even a temperature offset of 100 milliKelvins would only degrade the frequency stability to $4 \cdot 10^{-14}$. We are able to operate in the region of nominally zero temperature coefficient with a temperature accuracy of one milliKelvin.

3. Reduction of temperature gradients across oscillator: Substantial reduction in thermal gradients was made by modification of the cryogenic temperature control system. Gradients associated with the regulation configuration were eliminated by consolidating the heating and cooling elements to allow a single thermal contact point to the oscillator assembly.

4. More effective ruby pumping: A fixed rectangular waveguide was installed with pump signal B-field

perpendicular to the ruby c-axis. This cryogenic polarizer should eliminate the primary remaining system uncertainty, and allow reliable operation from run to run.

5. Improved pump signal propagation: Elimination of a coaxial signal transmission line within the pump waveguide now allows a more direct pump signal path. A waveguide adaptor was installed and a teflon window was used for vacuum seal and allowing low loss microwave propagation. It is expected that less pump power will be required to obtain oscillation since the un-matched impedance caused by the right angle feed will be eliminated.

6. Reduction of in-oscillator noise due to back coupling from the room-temperature amplifier: We have installed a cryogenic isolator[8], to prevent room-temperature radiation from coupling into the oscillator and also to reduce sensitivity of the operational frequency to output VSWR.

Measurements

The improvements discussed above have made possible excellent stability at both relatively long (10000 seconds) and very short (1 second) measuring times. In order to characterize performance of the SCM at shorter times, we obtained the use of another cryogenic oscillator (SCSO) for use as a frequency reference [5,6]. Substantial improvement in other instrumentation was also necessary. New procedures included bypassing the receiver of the SCSO in order to make direct measurements between microwave frequency signals. Long term measurements primarily made use of a Hydrogen maser as frequency reference.

Figure 2 shows raw data for two tests of the SCM against SCSO and Hydrogen Maser references. The filled squares represents data of the SCM/SCSO test and open squares the SCM/Hydrogen-Maser data. Performance of the SCM is clearly superior to hydrogen maser for measuring times shorter than about 30 seconds, and superior to that of the SCSO for times longer than about 200 seconds. SCM performance can be well characterized for times longer than 30 seconds due to overlap of the two data sets and the well-characterized hydrogen maser stability as shown in Figure 3. However, only the test with the SCSO reference provides detailed information about SCM performance at times shorter than 30 seconds and so the contributions of the two sources cannot be absolutely distinguished for short times. In previous tests of several SCSO sources, a slope at short measuring times was reported, reaching a value of $1 \cdot 10^{-14}$ at 1 second[6]. This SCSO variability is sufficient to explain the slope in our data for times less than about 5 seconds as shown in Figure 2.

Figure 3 shows SCM stability inferred from the two sets of data shown in Fig. 2. Stability for a single hydrogen maser is also shown. A conservative estimate was made for short times of equal contributions by the two cryogenic sources. If the slope at the shortest times is due to the SCSO, as discussed above, SCM performance would be $4 - 5 \cdot 10^{-15}$ for all times from 1 second to 1000 seconds.

In a three corner hat measuring scheme, using the SCSO and a hydrogen maser as references, performance of SCM was shown to be better than hydrogen maser for all times shorter than 70 seconds with long term performance better than the SCSO.

Figure 4 shows the results of phase noise measurement between two signals at 8.1 GHz derived from the cryogenic oscillators. A value of -80dB/f^3 was measured, which is 25dB better than the newly upgraded hydrogen maser and is 20dB better than the best quartz oscillator reported.

In order to combine the short term stability of the SCM with the long term stability of the hydrogen maser we have modified the SCM to allow its frequency to be tuned. A coil has been installed on the ruby housing to allow the bias field to by slightly modified, and so to tune the frequency of oscillation within the passband of the high-Q resonator. This coil, with 60 turns, gives a sensitivity of $7 \cdot 10^{-12}$ per mA with a range of approximately 10^{-10}. This range is sufficient to accommodate the typical SCM drift of $4 \cdot 10^{-13}$/day in long term operation.

Conclusions

We have demonstrated a frequency stability of parts in 10^{15} for all times from 1 to 1000 seconds for the SCM. The measured stability of $8 \cdot 10^{-15}$ at 1 second is 10 times better than the hydrogen maser at the same measuring time, and improvement over the hydrogen maser is shown for all times from 1 to 30 seconds. We believe these results to be better than any known RF or microwave frequency source.

Ultra-stable frequency sources like the SCM will make possible new experiments with high sensitivity. An experiment on gravitational wave search is now planned in 1992 with two SCM units to perform a "three-way" experiment with Galileo spacecraft. Furthermore, the SCM can be used to unam-

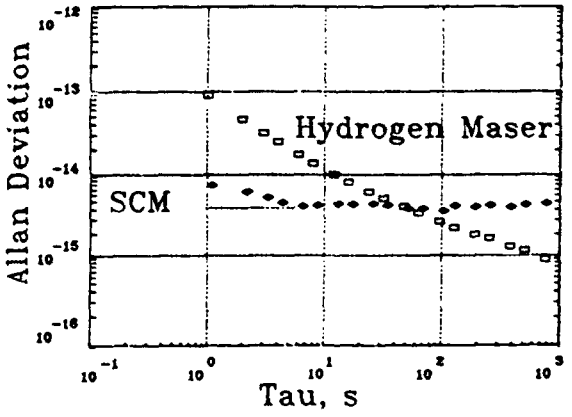


Figure 3: Plot of Allan Deviation of SCM stability after modification, also shown is stability of a single hydrogen maser reference. Improvement over the hydrogen maser is apparent for times from 1 to 30 seconds.

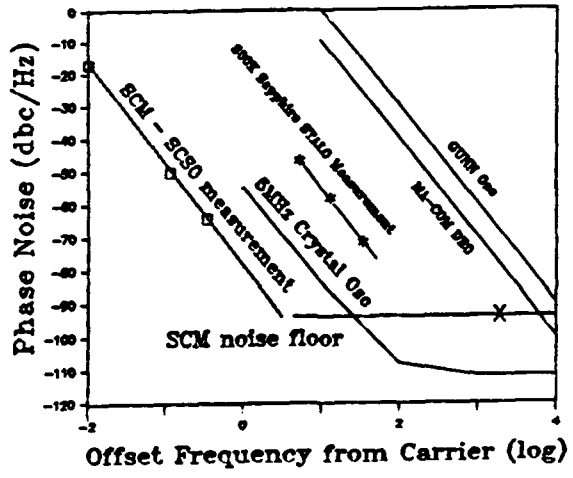


Figure 4: Phase noise measurements of SCM at 8.1 GHz show improvement of 20dB over the best multiplied 5 MHz crystal oscillator performance available at X-band. Noise plot for various conventional X-band frequency sources are also shown.

biguously characterize the performance of a hydrogen maser for short measuring times.

A frequency pulling coil has been implemented and tested to enable the SCM to be slaved to a hydrogen maser. This combination would make possible the excellent long term performance of the hydrogen maser combined with the newly available short term performance of the SCM.

Acknowledgements

Special thanks to Bill Diener for his assistance in ultrastable measurements, Roland Taylor for easy access to the capabilities of the Test Facility, Chuck Greenhall for software enhancements meeting our special requirements and Tom Tucker, for assistance and support with hydrogen maser measurements. The authors would like to thank Dr. L. Maleki for many helpful suggestions. During the time this work was done one of the authors (R. T. Wang) held a National Research Council-NASA/JPL Research Associatship.

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] G. J. Dick and D. M. Strayer, "Development of the Superconducting Cavity Maser as a Stable Frequency Source," in *Proceedings of the 38th Annual Frequency Control Symposium*, 435-446 (1984).

[3] D. M. Strayer and G. J. Dick, "The Superconducting Cavity Maser Oscillator - Toward Higher Stability," in *Proceedings of the 18th Annual Precise Time and Time interval (PTTI) Planning and Applications Meeting*, 601-609 (1986).

[4] R. T. Wang, G. J. Dick and D. M. Strayer, "Operational Parameters for the Superconducting Cavity Maser" in *Proceedings of the 42th Annual Frequency Control Symposium*, 345-354 (1988).

[5] S. R. Stein and J. P. Turneaure, in *Proceedings of the 27th Annual Frequency Control Symposium*, 414-420 (1975).

[6] S. R. Stein, "Space Application of Superconductivity: Resonators for High Stability Oscillator and Other Applications," *Cryogenic*, **20**, 363-371 (1980).

[7] 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," *Proceedings of the 43rd Annual Frequency Control Symposium*, 89-93 (1989).

[8] Cryogenic microwave isolator made by Passive Microwave Technology. Center frequency is 2.7 GHz with 30dB isolation.

FINE-STRUCTURE ARTIFACT OF THE VELOCITY DISTRIBUTION OF Cs BEAM TUBES AS MEASURED BY THE PULSED MICROWAVE POWER TECHNIQUE

W. A. Johnson, S. K. Karuza, and F. J. Voit

Electronics Research Laboratory
The Aerospace Corporation
P. O. Box 92957, Los Angeles, CA 90009

Abstract: An artifact of the pulsed microwave power technique used to measure the velocity distributions of Cs beam tubes is a fine structure on the distribution. In previously reported measurements the signal-to-noise ratio was not high enough to show this structure clearly. This paper presents results that show this structure in detail. To predict these results, a simple theory has been used that considers the Ramsey response function (i.e., output tube current vs. microwave frequency) as being a simple electrical-filter transfer function for the input of the pulse-modulated microwave power. Calculations that assume the principle of superposition give predictions that are in close agreement with the measured results. This theory also predicts how to set the pulse width so that the fine-structure artifact is minimized.

Introduction

The knowledge of the velocity distribution of a cesium (Cs) beam tube provides considerable insight into the performance and alignment of the beam optics, which in turn affects the final performance of the frequency standard. To measure these velocity distributions, early investigators [1,2,3] developed a method that used the pulsed excitation of atomic beam devices that had Ramsey-type interaction regions; this allowed for the observation of signals that were due to very narrow velocity groups. A completely automatic system that employed this pulsed microwave technique was set up in our laboratories. Figure 1 is a block diagram of the equipment used; the set-up is described in detail in [4].

Figure 2 shows the operation of the Ramsey cavity with pulsed microwave power. For a pulse period of T the velocity v selected is L/T, as shown. The pulse width τ is chosen to be less than the Cs ions' time of flight through the interaction regions, i.e., less than l/v.

In the process of making some of our early velocity-distribution measurements, we noticed what appeared to be a periodic structure on the distribution. At first we thought this structure to be noise in the measurement system; however, as we improved our system further, we could see that the structure was clearly there. The amplitude of this structure was found to be a function of the pulse width. A typical example of the structure is shown in Figure 3, which is a plot of the Cs velocity distribution $\rho(v)$.

CH2818-3/90/0000-094 $1.00 © 1990 IEEE

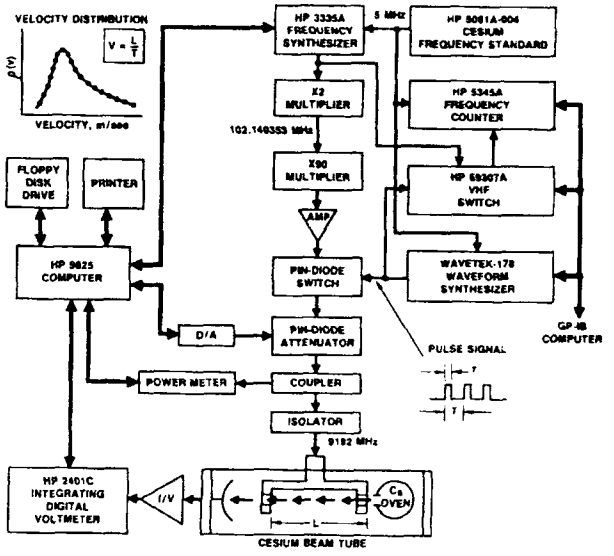


Figure 1. Block Diagram of the System for Measuring the Velocity Distribution of the Cs Beam Tube under Pulsed Microwave Conditions

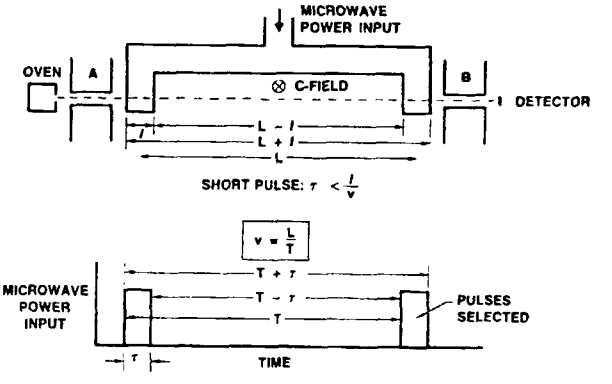


Figure 2. Operation of the Ramsey Cavity with Pulsed Microwave Power. The design of the Ramsey dual-interaction region allows one to select the velocity of the atomic beam by setting the period T of the microwave pulses.

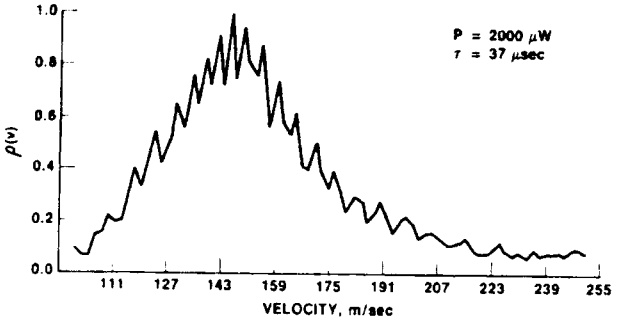


Figure 3. Measured Velocity Distribution for a Cs Beam Tube Using a Pulse Width of 37 μsec and a Peak Microwave Power of 2000 μW

Discussion

The structure shown in Figure 3 can be understood by looking at the pulse-power spectrum and viewing the Ramsey response function as a filter for this spectrum. The Ramsey response for the tube measured in Figure 3 is shown in Figure 4. For a pulse modulation function as shown in Figure 5, the Fourier components are given by

$$F(t) = \sum_{n=0,1,2,3,} A_n \cos 2\pi n \frac{t}{T} \qquad (1)$$

where

$$A_0 = \frac{\tau}{T} \qquad (2)$$

and

$$A_n = 2\frac{\tau}{T}\frac{\sin \pi n \frac{\tau}{T}}{\pi n \frac{\tau}{T}}, \quad n = 1, 2, 3, \ldots \qquad (3)$$

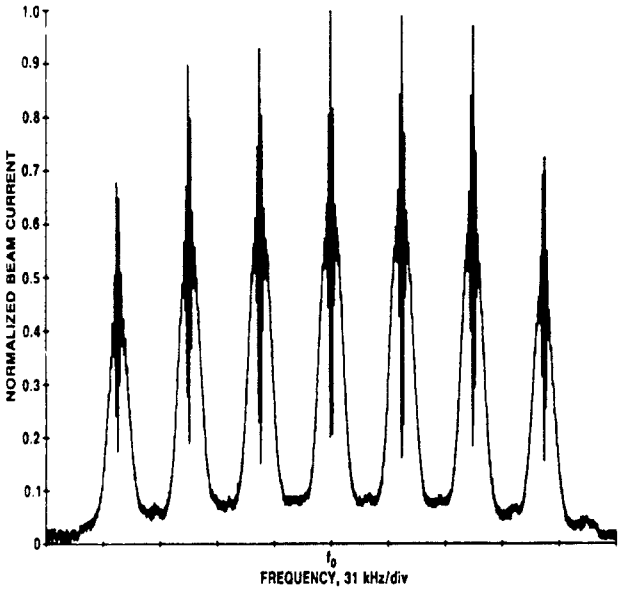


Figure 4. The CW Ramsey Patterns of All Seven Transitions in a Cs Beam Tube

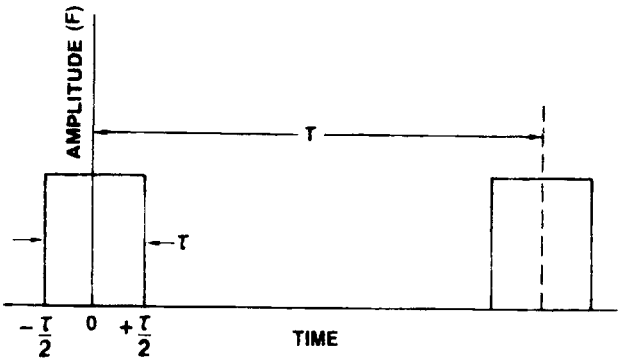


Figure 5. Pulse Modulation Waveform

Each of these components modulates the rf carrier to give a pair of equal components at $f_0 \pm n/T$, where f_0 is the rf carrier. Figure 6 shows this amplitude distribution (which we will refer to as the $\sin x/x$ distribution) in relation to the Ramsey response function. For the particular C-field used for the tube, the separation in the Ramsey responses (i.e., the Zeeman frequency f_Z) was 38.86 kHz. The tube length L is ≈12.5 cm; therefore, for a velocity of 150 m/sec, the modulation frequency would be ≈1200 Hz. There are therefore 38.86/1.2 ≡33 components between the component at f_0 and the component at f_0 + 38.86 kHz. Figure 6 is drawn for a τ of about 10 μsec, which places the first null in the $\sin x/x$ distribution at 100 kHz. For ease of illustration, the separations of the pulse components are depicted as much wider than is actually the case.

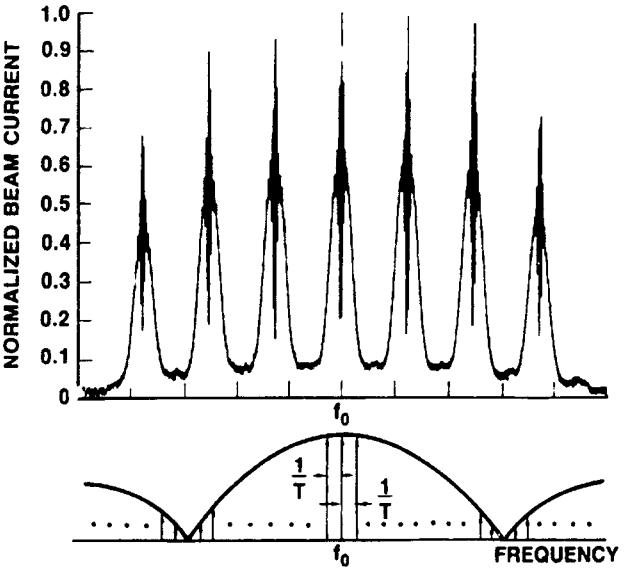


Figure 6. Ramsey Response in Relation to the Pulsed Microwave Spectrum for a Narrow Pulse Width

Figure 7 shows the location of the components for the central Ramsey transition as well as for the first upper transition. Each of the components gives a maximum current and

therefore, to the extent that the principle of superposition applies, the output current will be a maximum. The rf frequency is then swept and the component in the output current at the modulation frequency f_m ($= 1/T$) is then the measure of the density of the atoms in the beam at the velocity v_m, where

$$v_m = l f_m \tag{4}$$

The carrier is then reset to f_0 and f_m is changed slightly to $f_m + \Delta f$, to enable one to look at a slightly different velocity, $v_m + \Delta v$. Then

$$\Delta v = l \Delta f. \tag{5}$$

Note that the components close to f_0 change slowly; i.e., the first sideband pair moves by Δf, the second by $2\Delta f$, and so on. However, the components at the first upper transition move very rapidly; i.e., n is very high. If Δf is chosen such that each of the components at the first upper transition moves by f_z, we will again have a maximum-current situation. Then

$$f_z = n \Delta f, \tag{6}$$

but

$$n = \frac{f_z}{f_m}, \tag{7}$$

Hence

$$\Delta v = l \frac{f_m}{f_z}. \tag{8}$$

In other words, from these very heuristic arguments a fine-structure velocity period of $l f_m / f_z$ has been predicted.

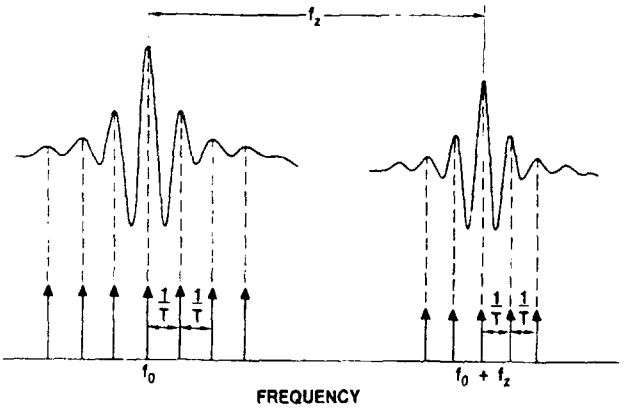


Figure 7. Main and First Upper Ramsey Response in Relation to the Pulsed Microwave Spectrum

We should note that although Δv is not a function of the pulse period τ, the amplitude of the structure certainly will be. For example, a pulse that is much narrower than $1/f_z$ will result in a very broad power spectrum (such as that shown in Figure 6) and will consequently yield a large amplitude in the fine structure. Conversely, if the pulse width is chosen so as to put the nulls in the sin x/x function at the upper Ramsey responses, the effect will be minimized, because a smaller share of the sideband energy is in the upper and lower Ramsey responses. This case is illustrated in Figure 8.

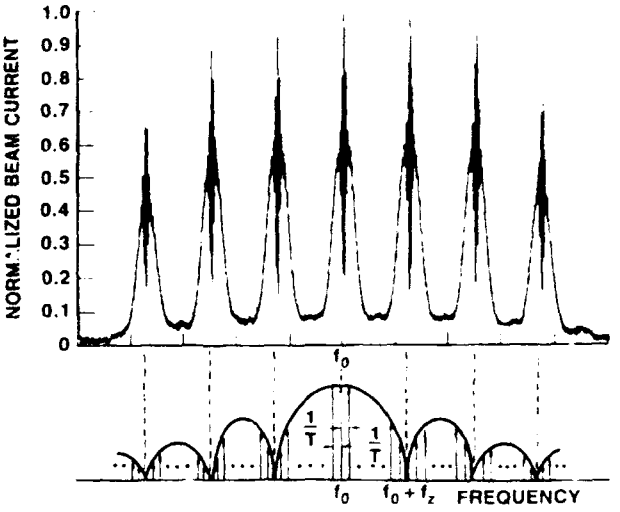


Figure 8. Ramsey Response in Relation to the Pulsed Microwave Spectrum for a Pulse Width that Places the Spectral Nulls at the Centers of the Upper and Lower Ramsey Responses

Measurements

The velocity measurement shown in Figure 3 was for a pulse width of 37 μsec. The Zeeman frequency for this tube was 38.86 kHz and thus the second maximum in the sin x/x distribution is very close to the first upper and lower Ramsey responses. This pulse width also places the third and fourth maxima in the sin x/x distribution close to the second and third upper and lower Ramsey responses. Therefore, there should be a clear fine structure, and indeed, this structure is easily seen in Figure 3.

The same tube with the same C-field was measured for a very narrow (10 μsec) pulse. This should also result in a clear fine structure. Figure 9(a) shows the results of this measurement. It is a little surprising that the minima of the distribution are so close to zero. This might be because the polarity of the components in the first side lobe in the sin x/x distribution is reversed from that in the main lobe. Choosing the pulse width such that the first null in the sin x/x distribution is at the first upper and lower Ramsey responses should result in a minimum structure, and in fact Figure 9(b) shows this result. Similarly, placing the second null at the upper and lower Ramsey responses should also minimize this structure. Figure 9(c) confirms this.

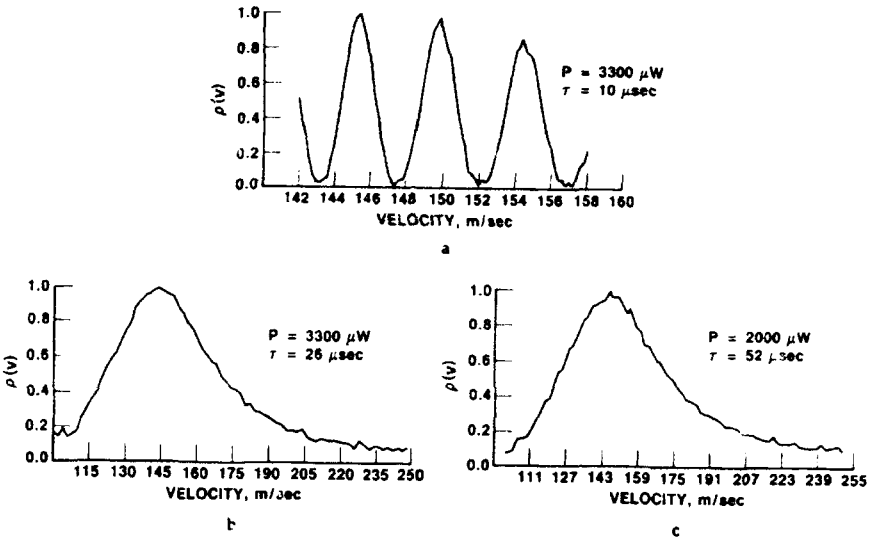


Figure 9. Measured Cs Velocity Distributions for Three Pulse Widths and Peak Microwave Powers, as Indicated. (a) 10 μsec and 3300 μW, (b) 26 μsec and 3300 μW, and (c) 52 μsec and 2000 μW.

Table 1: Fine-Structure Period $\Delta\nu$ for Three Cs Beam Tubes

Tube	τ, μsec	f_Z, kHz	Measured, m/sec	Calculated, m/sec
A	37	38.86	4.42	4.29
A	10	38.86	4.59	4.48
B	10	33.43	2.79	2.81
C	20	30.55	3.67	3.79

Table 1 presents a comparison of the measured and calculated [from Eq. (8)] fine-structure periods for the cases shown in Figures 3 and 9(a), as well as for measurements made on two other tubes from two other manufacturers. Each of the tubes is between 12 and 13 cm long. Note that the two calculated $\Delta\nu$ for the two different τ's for tube A were for two slightly different velocity ranges, and hence are slightly different. In general, the agreement between the calculated and measured Δ's is excellent.

Conclusions

A fine-structure feature on the velocity distribution of Cs beam tubes, as measured by the pulsed microwave power technique, has been identified and quantified. This feature was shown to result from the existence of large spectral components at the frequencies of the upper and lower Ramsey responses. It has been shown that this feature can be minimized by choosing the pulse width such that the nulls in the microwave power spectrum are at these upper and lower Ramsey responses. Specifically, this means that the pulse width should be chosen to be an integer multiple of the reciprocal of the Zeeman frequency. A simple theory was developed to enable one to predict the fine-structure period as a function of the beam velocity, the tube length, and the Zeeman frequency. This theory employed the concept of the Ramsey response pattern as an electrical filter. This concept may be useful in other tube studies, e.g. those of modulation effects. Also, measurements were presented that demonstrated close agreement between the measured fine-structure period and the calculated one. Finally, it was shown that the fine structure virtually disappears if the pulse period is chosen to be an integer multiple of the reciprocal of the Zeeman frequency.

Acknowledgment: The authors thank Mr. Michael Meyer of The Aerospace Corporation for editing and preparing this paper.

References

[1] H. Hellwig, S. Jarvis, D. Halford, and H. E. Bell, "Evaluation and operation of atomic beam tube frequency standards using time domain velocity selection modulation," *Metrologia*, vol. 9, 1973, pp. 107–112.

[2] H. Hellwig, S. Jarvis, D. J. Glaze, D. Halford, and H. E. Bell, "Time domain velocity selection modulation as a tool to evaluate cesium beam tubes," in *Proceedings of the 27th Annual Symposium on Frequency Control* (Ft. Monmouth, N.J.), June 1973, pp. 356–362.

[3] D. A. Howe, H. E. Bell, H. Hellwig, and A. DeMarchi, "Preliminary research and development of cesium beam tube accuracy evaluation system," in *28th Annual Symposium on Frequency Control*, 1974, pp. 362–372.

[4] S. K. Karuza, W. A. Johnson, J. P. Hurrell, M. F. Bottjer, and F. J. Voit, "An Automated System for Measuring Velocity Distribution in Cesium Beam Tubes," in *Proceedings of the 4th European Frequency and Time Forum*, Neuchâtel, Switzerland, March 13–15, 1990.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

TOTAL PROCESS CONTROL AND ITS IMPLEMENTATION IN THE MANUFACTURING OF CESIUM BEAM TUBES

Emil R. Straka, Richard W. Sarrica and Ladell Hansen
Hewlett Packard Co., 5301 Stevens Creek Boulevard
Sata Clara, California 95052-8059

ABSTRACT

The Total Quality Control (TQC) operating philosophy as practiced at Hewlett Packard 1) focuses on customer need and expectation, 2) makes a total commitment to quality, 3) emphasizes universal participation and teamwork and 4) results in continuous process improvement methodology. The resultant Total Process Control (TPC) provides the key to achieving improved quality and higher productivity.

This philosophy is practiced throughout HP and in particular in the manufacturing of atomic frequency standards. TPC training and its implementation involves all personnel: production operators, managers, supervisors, production engineers, manufacturing and component engineers, buyers, schedulers and quality engineers.

In this paper, the specific application of several analytical TQC tools will be discussed. One example deals with the restructuring of the method of handling the cesium beam tube (CBT) work orders. The result was a reduction of manufacturing cycle time as well as a reduction of the total work-in-process inventory. We observed a corresponding improvement of the CBT performance. This paper also discusses TPC influence on existing and potential suppliers of CBT parts. Clear statements of expectations and measures of success have led to benefits for all parties concerned.

IMPLEMENTATION OF TQC

Implementation begins by establishing a sound foundation through extensive training in TQC fundamentals. At HP the class instruction extends over twelve weeks and includes application of each step which is shown in the Process Improvement Method, Figure 1. In order to accomplish this, class members choose projects to which they will apply the steps of the process. This learn-as-you-go technique has proven very effective. Training involvement is universal. All employees in the factory undergo the training. Training is not limited to production personnel; support people also become accomplished which consequently leads to widespread understanding of the TQC vocabulary and analytical tools.

Universal participation leads to several advantages:

1) a common vocabulary is developed for describing processes and performance.

2) technical capabilities are expanded throughout the work force. This provides leverage in interpreting data and in identifying changes in processes and performance.

3) there is quicker response to deviations in process or performance.

4) the work force plays a greater role in monitoring and recording performance data. Production line operator motivation improves as the results of their work become visible.

5) a variety of analytical tools (eg: control charts, Pareto diagrams) are made available. Their usefulness is better appreciated and understood.

6) timely, factual information is available for support in making decisions.

CH2818-3/90/0000-98 $1.00 © 1990 IEEE

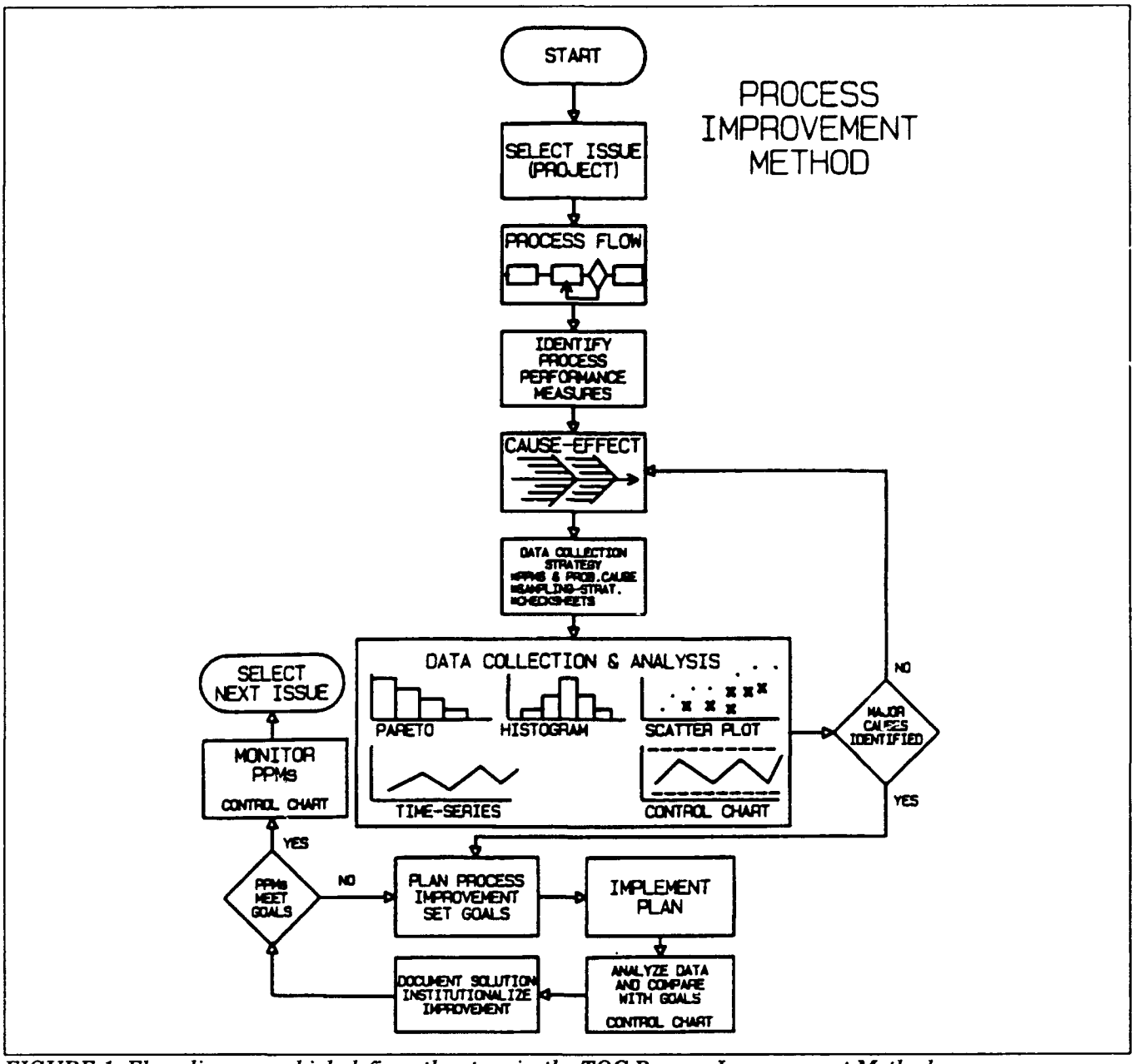


FIGURE 1. Flow diagram which defines the steps in the TQC Process Improvement Method.

Details of the steps in the Process Improvement Method have been discussed in a previous paper[1].

APPLICATION OF TPC TO PRODUCTION OF CESIUM BEAM TUBES

Over the years, HP has promoted efforts to improve manufacturing efficiency. One way to do this is to reduce manufacturing cycle time. Cycle time is defined as the time interval which begins when work commences on a group of products. In our case, the group consists of five CBTs. The cycle interval ends when the first of the five CBTs is complete. One advantage to such a reduction is that the in-process inventory is correspondingly reduced. Longer cycle time implies that more material is lingering in the plant for longer periods. The cost of such inventory becomes a source of increasing concern.

In order to further address the issue of cycle time reduction, a task force was formed. The task force consisted of production supervisors and their manager, production engineers, a production scheduler, an accountant and a consultant engineer

from a different production line. This group worked with the production line operators to employ TQC tools to help define the existing state of affairs. The tools included process flow diagrams, cause and effect charts, time-series charts and Pareto charts.

The task force had been previously exposed to the **Demand Pull** production philosophy which has been practiced at other HP divisions. The existing philosophy employed in CBT production is called the **Push** philosophy. Figure 2 compares the differences in inventory and product flow for some imaginary manufacturing line which has four process stages.

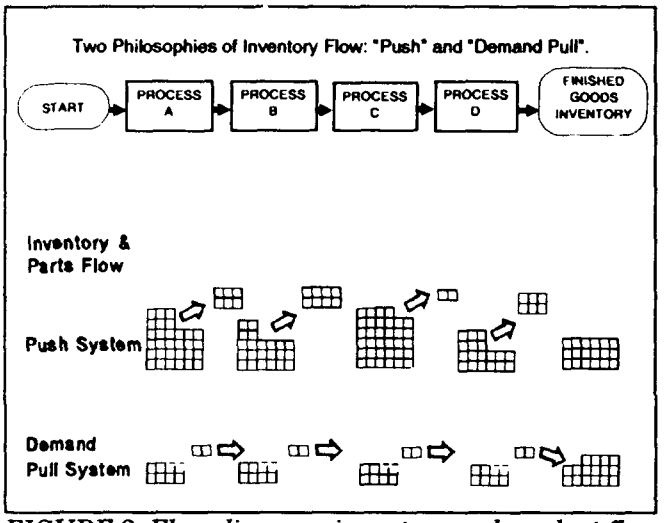


*FIGURE 2. Flow diagram, inventory and product flow for an imaginary production line operated in two manners: **Demand Pull** and **Push** systems. Each square represents a production unit.*

The **Push** philosophy operates according to the following rules:

1) Run each process at 100% speed.

2) Push the product forward, whether the next process needs it or not.

3) Keep people busy at all times by building inventory. (Part of the psychology is that workers feel guilty if they are idle. Supervisors probably feel worse!)

The **Demand Pull** philosophy uses these rules:

1) The signal to build comes from the next process and originates at the output end of the line.

2) An upper limit is set on the amount of inventory which is permitted to exist at any given point in the line.

Let's examine how the **Demand Pull** system works. When an opening occurs for two units in Finished Goods Inventory, a signal is sent for two units to move up from the inventory at Process D. (Finished Goods Inventory **pulls** in two units.) Process D **pulls** material from the inventory at Process C and replenishes the inventory at D. This **Pull** process trickles back through the line. The actual throughput of the line is determined by the process stage with the least capacity. If throughput is to be increased, the capacity of the "bottle neck" process must be increased.

Two advantages to the **Demand Pull** system are obvious:

1) The inventory is well controlled and the cost of the inventory should be lowest in the **Demand Pull** system.

2) If a defective part is found, the **Demand Pull** system has fewer parts to inspect, reject or repair. Costs due to this type of problem are minimized.

After having reviewed the status of the CBT production system, the task force decided to implement the **Demand Pull** system. In order to establish a better common understanding of the concepts of **Demand Pull**, all members of the task force were required to read **The Goal**[2].

The strategy would be to bring in **Demand Pull** in phases. The goal was to reduce cycle time by 35% and accomplish this within 1 year of beginning the project.

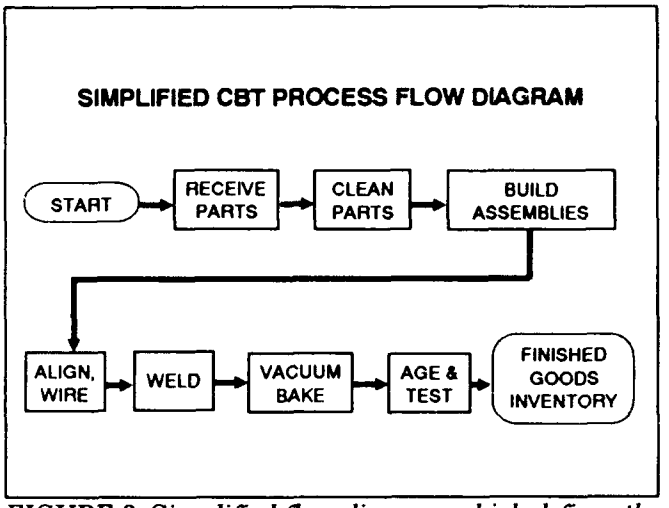


FIGURE 3. Simplified flow diagram which defines the CBT Production process.

Figure 3 is useful in describing how the project proceeded. First, an upper limit was established for the stock level in Finished Goods Inventory. The **Demand Pull** process was applied to the Vacuum Bake and Age & Test processes. It was here that we were able to work on methods for signaling backwards in the line when materials were to be moved forward. After the system was operating smoothly at these two stations, we expanded **Demand Pull** to include Weld and Align,Wire. This expansion went quite smoothly.

When we considered applying the **Demand Pull** method to the Build Assemblies stage, the situation was complicated by the nature of that stage: Build Assemblies takes in parts and subassemblies from several process areas. Build Assemblies is not the result of a simple linear flow of parts; it is better described as a "merging" of parts. It was at this point that we decided to diverge from a pure **Demand Pull** process. We instituted a **Push-Pull** process. We had expected at the beginning of the project that we would have to develop a "custom fitted" system to match our manufacturing process. Each production line is unique and should require some form of specialization. Currently, the hybrid **Push-Pull** method extends from Build Assemblies back through Cleaning. The impact of the changes have been significant.

RESULTS OF THE DEMAND PULL METHOD

We used cycle time, amount of work in process (WIP) and aging time as measures of the success of the project. Each of these improved markedly as the following graphs show.

Cycle Time

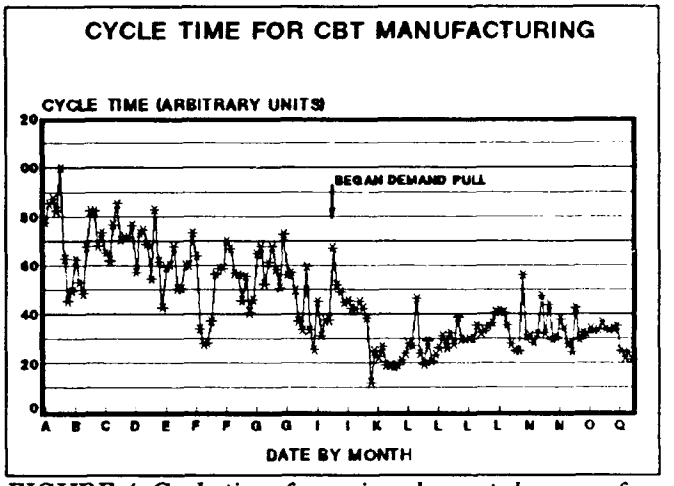


FIGURE 4. Cycle time for cesium beam tube manufacturing. The results of attempts to reduce cycle time are shown before and after the Demand Pull process was instituted.

Figure 4. is a time-series chart which demonstrates that early attempts to reduce cycle time had a positive effect. Even though the trend is downward, there was an erratic nature to the cycle time. When **Demand Pull** was instituted, we observed a significant drop and even more important, the consistency from group to group was much better. This improved consistency is one of the benefits of Total Process Control. The cycle time is reduced approximately 40% due to the **Demand Pull** method and nearly 60% as a result of overall efforts.

Work In Process (WIP)

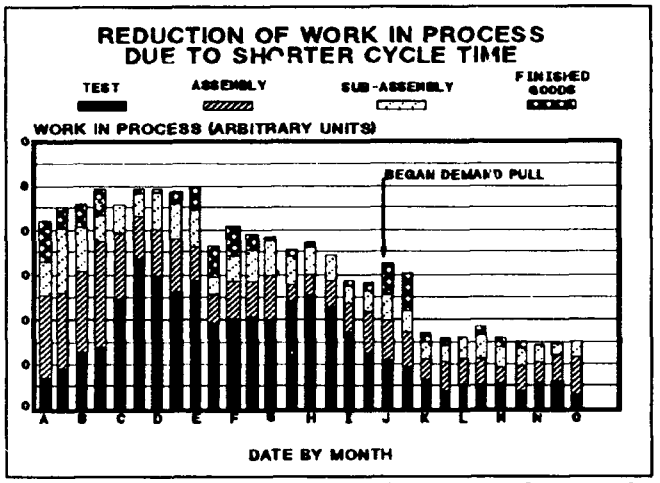


FIGURE 5. Work In Process for cesium beam tube manufacturing. Inventory levels at three process stages and Finished Goods Inventory. WIP declines as cycle time shortens and drops dramatically when Demand Pull is instituted.

Aging Time

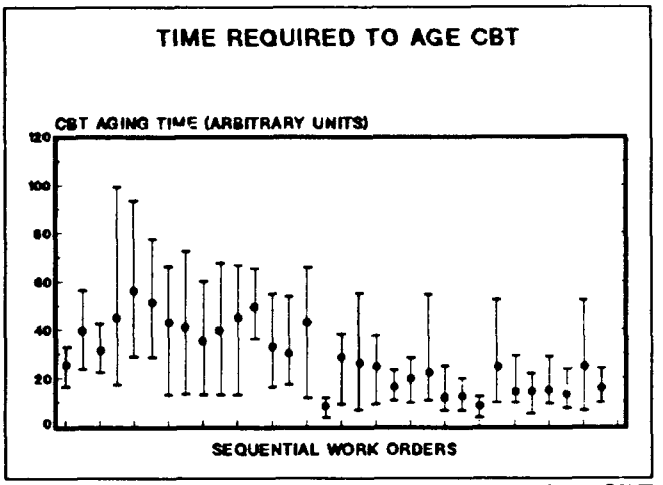


FIGURE 6. CBT aging time. The time required for CBT output to stabilize declined as the cycle time was reduced. Each dot indicates the average aging time for each group of 5 CBTs. The error bars indicate the longest and shortest aging times for the group.

It is clear from Figure 5 that the greatest level of WIP occurred in the Age and Test stage. Efforts to reduce cycle time had most impact on the Age and Test stage although all stages were affected. The WIP was reduced about 70% due to overall cycle time reduction. WIP was reduced about 50% relative to the level indicated when **Demand Pull** began.

Figure 5 is particularly meaningful in terms of the financial impact on the department. Simply stated, 70% fewer dollars were required to be invested in inventory. Previously, a larger investment was required for the department to operate.

It also was made clear to production line management that the Age and Test stage was no longer the bottleneck that it had once been. The volume processed through this stage was fairly constant over the time period displayed in Figure 5. Figure 6 helps us to understand how the bottleneck disappeared.

This time-series chart shows the improvement in aging cycle as the cycle time was reduced. If we consider the average aging times, overall time is reduced more than 70%. If we consider the behavior after the beginning of **Demand Pull**, the improvement is about 65%. Of most significance is the tightening up of the aging ranges once **Demand Pull** was instituted. Early on, the longest aging time in a group was 5 or 6 times the shortest aging time. Statistical variations of this ratio are evident. After the **Demand Pull** began, the variation was reduced to a factor of 2 or three, with less frequent occurrence of wide spreads.

The significance of the tighter spread in aging times is that we can now more easily predict the throughput in the Aging and Test stage. That stage is no longer a bottleneck.

We postulate that the improvement in aging is related to the fact that the parts now spend a much shorter time in process following cleaning. Consequently, less adsorbed gas is introduced to the CBT. Therefore, the vacuum bake process is not so hard pressed to eliminate gases.

Having achieved a shorter aging time (a benefit not foreseen) we decided to use aging time as a measure of process consistency. This is best monitored in the form of an X-Bar, R control chart which is shown in Figure 7.

The time-series chart (Figure 6) is excellent for determining the extent to which an improvement is made. Once the improvement goals have been achieved, it is necessary to establish a measure of the statistical variation of the new process. The control chart, as in Figure 7, does this and aids us in maintaining control of the new process.

SUMMARY, PRODUCTION LINE ACCOMPLISHMENTS

The goals were achieved. Cycle time was reduced 40% compared to a goal of 35%. This was accomplished in 9 months compared to a goal of 12 months. Other benefits include a reduction of WIP by 50%, reduction of aging time by 65% and a tremendous boost in motivation and sense of accomplishment with production line operators because of these results and the recognition they received.

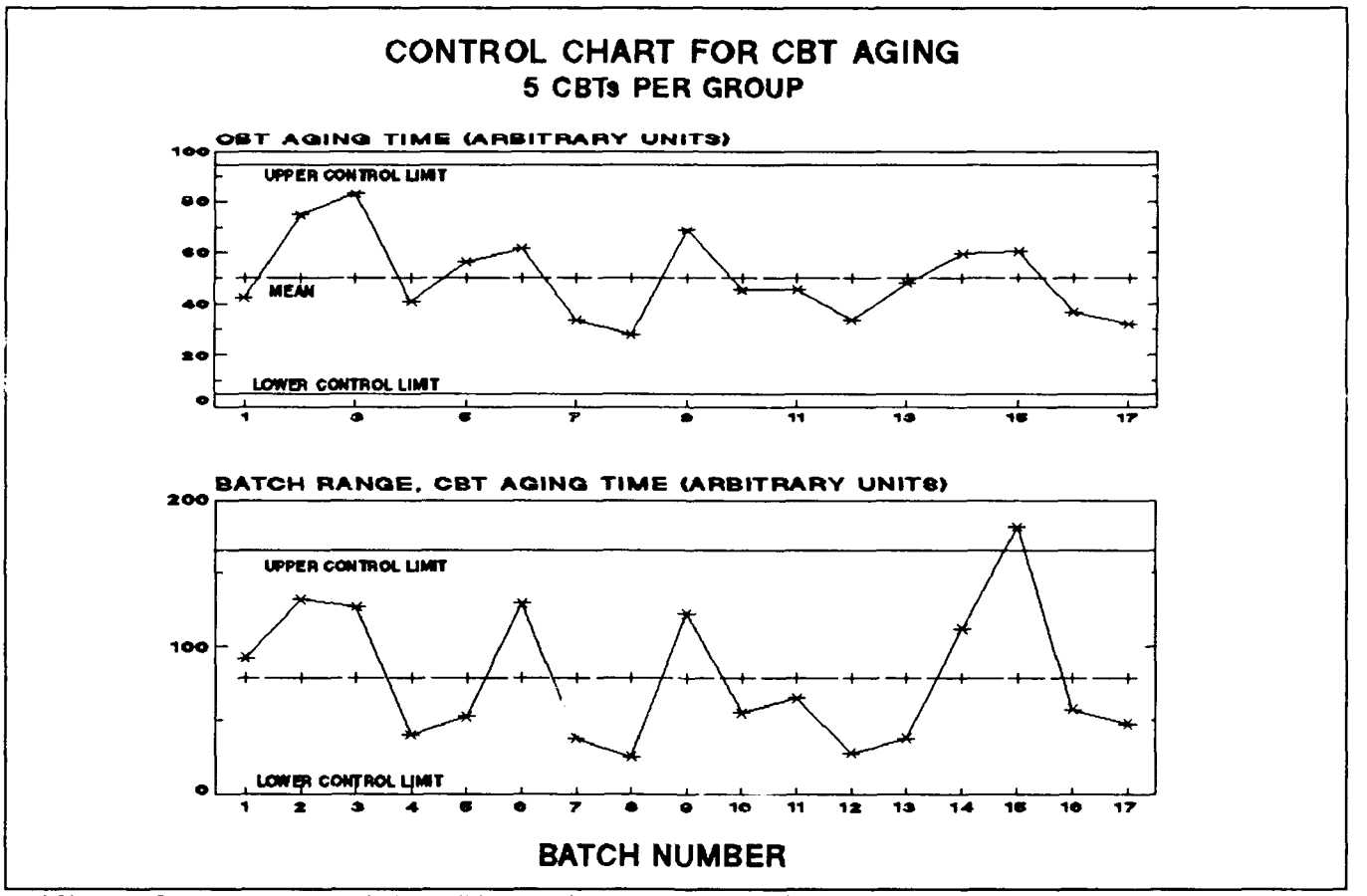


FIGURE 7. CBT aging control chart. The top chart is an X-Bar chart where each point is the average aging time for each group of 5 CBTs. The lower chart is a range chart (R chart) which shows the range of aging for each group of 5 CBTs.

APPLICATION OF TPC TO AN EXTERNAL SITUATION

By including this application with this discussion it is possible to give an example by what we mean by Total Process Control. Here we emphasize that controls are applied to the total process, including what takes place with our suppliers.

At issue was the performance of a supplier which we shall identify as Supplier A. Our quality records indicated that this supplier did not meet the basic incoming quality standards namely, the rejection rate was greater than 1.0%.

The first action on the part of HP was to establish a commodity team to deal with Supplier A. The team consisted of a manufacturing engineer, a quality engineer, a cesium beam tube engineer and the buyer who deals with Supplier A.

The first action of the commodity team was to specifically identify the concerns. This set of concerns was then transmitted to Supplier A.

After Supplier A had time to digest this information, the commodity team scheduled a meeting which was to take place at Supplier A's plant. A questionnaire was sent to Supplier A several weeks beforehand so that they would have an idea of the nature of the questions and concerns. The objective of the visit was four-fold: 1) to establish rapport, trust and common understanding of the concerns, 2) to discuss the issues and identify the problems, 3) to assist Supplier A in setting up TQC methods and tools as needed and 4) to initiate corrective action.

We used TQC tools to help define and prioritize the defects. Specifically, Pareto charts were helpful in this respect. In presenting the data to Supplier A, Figures 8 and 9 were used. Of all the HP parts that Supplier A makes, Figure 8 shows which pieces had the most defects. The next stage in the analysis involved creating a Pareto chart for each specific part. In Figure 9, the characteristic failure modes for one part number are presented.

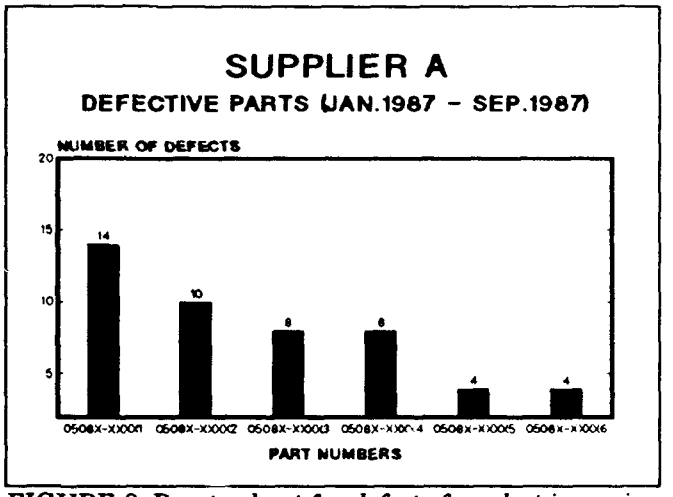


FIGURE 8. Pareto chart for defects found at incoming inspection. Each column refers to one particular part type.

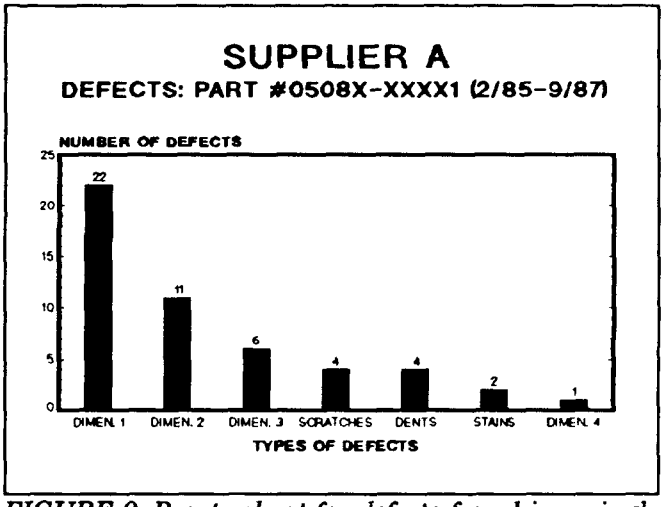


FIGURE 9. Pareto chart for defects found in a single HP part type. Each column refers to a specific dimension or other attribute which was an inspection criterion.

In this example, it is clear that a majority of defects are dimensional. The few remaining defects are cosmetic. Prior to making the visit, the commodity team suspected Supplier A's manufacturing methods or lack of process control as potential sources of trouble. As a result of the cooperative visit, the source of the problem was determined to be differences in inspection methods at the two facilities.

Corrective action was initiated before the visit was concluded. Both parties agreed to specific actions and a decision was made as to who would take responsibility for the various tasks. The summary of the corrective action items was as follows: 1) a common inspection tooling and methodology was established. 2) HP and Supplier A were to use identical written procedures to perform inspections. An important feature of the agreement was to emphasize the use of readily available, generic inspection tools. This avoided any complications which might arise from the use of specialized tools which would require an unusual level of expertise in order to be used correctly.

In addition, the commodity team confirmed the expectation that the incoming inspection defect rate must be less than 1.0%. A plan was agreed upon wherein Supplier A would strive to meet modified defect rate requirements during the ensuing 12 months. HP would deliver a bimonthly progress report to Supplier A.

Figure 10 shows the performance goal and the actual performance of Supplier A. After 4 months, the performance of Supplier A was excellent. Figure 11 shows the 24 month performance of Supplier A. These two graphs represent the type of progress report which was provided to Supplier A.

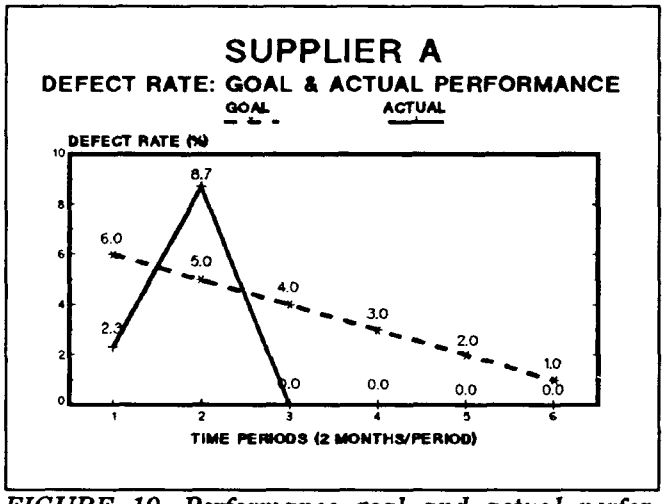


FIGURE 10. Performance goal and actual performance of Supplier A.

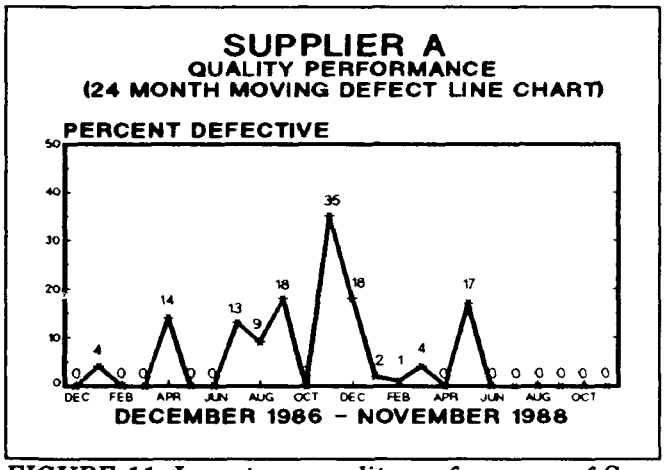


FIGURE 11. Long term quality performance of Supplier A.

SUMMARY, EXTERNAL SITUATION

By working with Supplier A, the commodity team was able to focus the efforts of the two companies. This constructive effort resulted in developing an understanding so that the supplier is now able to meet our quality requirements on a sustained basis. The two parties took a potentially antagonistic environment and developed it into one of cooperation, trust and success.

CONCLUSION

Application of TQC principles and tools, both inside the factory and out, has led to improved quality, lower manufacturing cost and an improved motivation level for all involved. The Total Process Control which has resulted from the application of these tools is evidence of the value of such methods.

ACKNOWLEDGMENTS

Successful execution of the details of these projects involved many persons including the entire Cesium Beam Tube Production Line and the task force which consisted of Abe Bromberg, Carole Kable, Helene Patton, Helga Perez, Steve Rockhold, Al Wooldridge and the authors of this paper.

The commodity team consisted of Bob Cerruti, Bob Corley, Lou Mueller and Joe O'Banion. Cerruti and O'Banion provided historical and quality information in support of this paper.

REFERENCES

[1] J.A. Kusters and C.A. Adams, "Application of Total Process Control Techniques in the Production of High Precision Quartz Resonators." Proceedings of the 39th Annual Symposium on Frequency Control, 1985, pp. 475–480.

[2] E.M. Goldratt and J. Cox, The Goal. Croton-on-Hudson, New York, North River Press, 1986.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NEW INEXPENSIVE FREQUENCY CALIBRATION SERVICE FROM NIST

by

D. W. Allan, D. D. Davis, J. Levine, M. A. Weiss,
N. Hironaka, and D. Okayama

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

Abstract

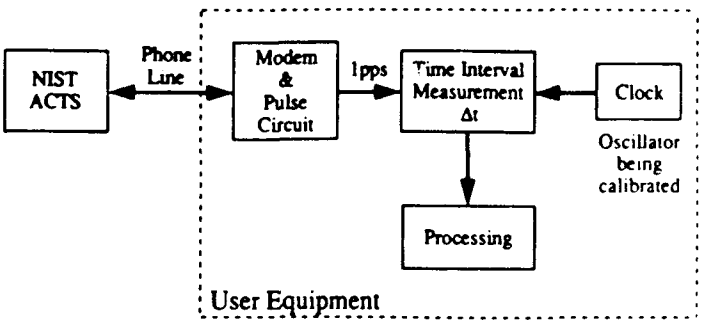
A new inexpensive frequency calibration service from NIST is now available. This service takes advantage of the operation of the NIST Automated Computer Time System (ACTS), which was begun in 1988. Software to access the service from several types of computers was released at the same time. Time and frequency dissemination by this modest-accuracy service depends on the reciprocity of the telephone system. The round-trip delay is measured by the NIST equipment. The advance of an on-time marker is adjusted so as to arrive at the user's site on time. A frequency calibration method taking advantage of this service has been designed and preliminary tests conducted. A computer is not required to access this service. All that is required is a telephone modem, a simple peripheral circuit to generate an on time marker and standard time and frequency measurement and data processing equipment.

Introduction, Review and Relevant Statistical Measures

The intent of this paper is to demonstrate an extremely cost-effective, modest-accuracy method of remotely obtaining time and frequency via a telephone modem. The time and frequency so obtained are traceable to the National Institute of Standards and Technology's (NIST) time scale, UTC(NIST). We will demonstrate the accuracy and stability of both time and frequency available with this technique. [1]

The Automated Computer Time Service (ACTS) system has been described previously [1-3]. The accuracy and stability of ACTS is based on the assumption that the telephone path for the sending and receiving signal is reciprocal. Under this assumption, the round-trip time is measured by an automatic system at NIST, and a time marker in the time-code is then advanced in order to arrive at the user's site on time. The basic equipment needed is a telephone line that can call (303) 494-4774, a modem and a simple circuit outlined in reference 3 to generate an on-time one pulse-per-second (1 pps) or one pulse-per-two-seconds. It was evident from the preliminary measurements that the modems contributed substantially to the uncertainties of the measurements.

A block diagram of this method is as follows:



Proper processing is essential for the best calibration accuracy. The accuracy is a function of the calling rate, the modem used, and the quality of the oscillator employed. The time accuracy limit for NIST ACTS is about a millisecond. A good choice of modems gives a time stability of NIST ACTS of better than 0.1 ms.

In order to test the reciprocity of the telephone transmission system, we used the same brand of modems at

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

each end. We tested at both 300 bits/s and 1200 bits/s on a local telephone switching network. Finding the 300 bits/s to be better, we then used it over a long distance (from WWVH, Kauai, Hawaii to NIST, Boulder, Colorado) where both land-links and a satellite-link were involved.

We use a basic statistical theorem to evaluate the time and frequency stability and accuracy. If a time series has a white spectrum (the sequential values of the time series are random and uncorrelated), then the optimum estimate of the mean is the simple mean.

The confidence on the estimate of the mean is the standard deviation of the mean, which is the standard deviation divided by the square root of the number of values in the series. If the series does not have a white spectrum, then the standard deviation of the mean will not be an unbiased estimate. In some cases, such as a 1/f spectral density for the random deviations, the standard deviation and the standard deviation of the mean are not convergent, and hence are not useful measures.

We define the following measure, which is equal to the standard deviation for τ equal to the data spacing and to the standard deviation of the mean for τ equal to the data length [4,5]

$$\sigma_x(\tau) = \tau * \mathrm{mod}\sigma_y(\tau)/\sqrt{3} \qquad (1)$$

The $\mathrm{mod}\sigma_y(\tau)$ denotes the square root of the "modified Allan variance." [4,5] This quantity in equation 1 has the following properties:

1) it is equal to the classical standard deviation of the time residuals for $\tau = \tau_0$, the data spacing interval, if the process has a white spectrum;

2) it equals the standard deviation of the mean of the time residuals for $\tau = N\tau_0$, the data length (N is the number of data points), if the process has a white spectrum;

3) it is convergent and well behaved for most of the random processes encountered in time and frequency metrology;

4) the τ dependance indicates the power-law spectral density model appropriate for the data;

5) and the amplitude of $\sigma_x(\tau)$ at a particular value of τ provides an estimate of the spectral density coefficient for any one of the common kinds of power-law spectra.

We have the following relationships

$$\begin{aligned} S_x(f) &\sim f^{\beta}, \\ \sigma_x^2(\tau) &\sim \tau^{\eta}; \\ \beta &= -\eta - 1, \\ -2 &< \eta < 4, \end{aligned} \qquad (2)$$

where $S_x(f)$ is the spectral density of the time residuals, x, and β denotes the kind of power-law spectrum. Given a log $\sigma_x(\tau)$ versus log τ plot, the slope of the values will nominally follow a slope equal to $\eta/2$ for power-law spectra in the range indicated. The spectral levels also can be calculated using information supplied in NIST Technical Note 1337 [4].

If a set of time or phase residuals has a white noise spectrum ($\beta=0$, $\eta=-1$), a linear regression to the residuals gives the optimum estimate for the mean time (the mid-point to the fit) and to the frequency (the slope of the fit). If the linear regression equation for the nth value is given by

$$\hat{x}(n) = a_o + a_1 n, \qquad (3)$$

for a set of N measurements with standard deviation σ, standard deviation of the mean σ_m and a data spacing τ_o, then the mean value is

$$\bar{x} = a_0 + \frac{a_1 N}{2}. \qquad (4)$$

The confidence of its estimate is σ_m,

$$\begin{aligned} \sigma_m &= \frac{\sigma}{\sqrt{N}} \\ &= \sigma_x(N\tau_o) \end{aligned} \qquad (5)$$

The confidence, s_1, on the frequency estimate, a_1, is given by :

$$\begin{aligned} s_1 &= \sqrt{12}\ \sigma\ N^{-\frac{3}{2}}/\tau_0, \\ &= 2\ \mathrm{mod}\sigma_y(N\tau_0). \end{aligned} \qquad (6)$$

If the residuals do not have a white spectrum, then $\sigma_x(N\tau_0)$ and $2\text{mod}\sigma_y(N\tau_s)$ will take on different values than the first equation listed in (5) and (6), respectively. These different values will be in a direction to accomodate the improvement or the degradation in the confidence intervals appropriate to the particular power-law spectra noise process.

Time and Frequency Stability and Accuracy of ACTS

Measurements Using a Local Switching Network

For the modems we used, the 1200 bits/s rate had an instability which was about two to three times that obtained using 300 bits/s modems. Figure 1 is a plot of the frequency stability using a pair of 300 bits/s modems of the same model. As stated in the previous section, the frequency accuracy obtainable is given by $2\text{mod}\sigma_y(\tau)$. We repeated the experiment for sets of 7 000 points while still connected to the same line to see how far we could push the long term stability. Each sequential set gave a frequency stability plot nearly identical to that shown in Figure 1.

Notice that the data exhibit a slightly steeper slope than $\tau^{-3/2}$. The $\tau^{-3/2}$ model corresponds to white phase or time modulation (PM). That the slope is steeper indicates the presence of more high frequency noise than white noise should have. This is probably due to a combination of processes such as the digitization granularity and the asynchronous nature of modems, which are built for communications rather than for timing. This suggests that less short term noise could be obtained if a modem were designed and built for accurate and stable timing. In any case, we see that the data are averaged at least as quickly as white noise, and that accepting a white noise model is conservative.

We found that the approximate $\tau^{-3/2}$ behavior continued to beyond 10^4 s. We measured $\text{mod}\sigma_y(\tau$ = 7 3/4 hours) = 2.1×10^{-11}. At this integration time and longer the spectrum was no longer white. This measure implies a frequency calibration accuracy of better than one part in 10^{10} with an integration time of $\tau \geq 10^4$ s.

The time stability, $\sigma_x(\tau$ = 2 s), as shown in Figure 2, was about 55 μs for the local network, which would yield $\sigma_x(\tau = 2 \times 10^4$ s) = 550 ns. This turned out to be statistically significant as long as we maintained the same connection. We measured $\sigma_x(\tau$ = 14 000 s) = 230 ns--a smaller number than predicted with the assumption of white noise PM. As before we assume this smaller number is due partially to the digitization and asynchronous noise in the modems.

One application of time stability is the maintenance of synchronization. In principle it seems that one could maintain synchronization to less than 1 μs once a path (particular connection) was calibrated. If connection is lost and a good enough clock exists to fly-wheel time to 1 μs over the next 10^4 s while recalibrating, then the 1 μs synchronization accuracy could be maintained. This method may be convenient in a local calling area.

If connection is lost and no fly-wheel method for recalibrating exists, then the time stability degrades markedly from the white PM model. We will show the amount in the long-link analysis in the next section.

Measurements via long-link

We chose the path between Boulder, Colorado (NIST) and Kauai, Hawaii (WWVH) because of known reference clocks on each end and because the link involved both land and satellite paths. We made measurements more than once a day on some days as well as nominally once per day over a few weeks. As is the case with white noise the variance appeared to be interval independent. The variance was much larger from call to call than during a call.

Figures 3 and 4 for this long-link measurement using ACTS correspond to Figures 1 and 2 for the local switching network. In Figure 3 we measured $\text{mod}\sigma_y(\tau = 512\text{s}) = 1 \times 10^{-8}$ with the $\tau^{-3/2}$ behavior (white PM) being a reasonable model. Some departure from this power-law spectra is observed, and again we assume this departure to be driven in part by data quantization and the asynchronous nature of communication modems. If the trend in Figure 3 continued, we should achieve a stability of 10^{-10} at $\tau = 10^4$ s.

Figure 4 is representative of several such curves we plotted. We obtain the following results for time stability and frequency accuracy calibrations. The values of $\sigma_x(\tau = 2\text{s})$ ranged from 65 μs to about 90 μs, and reached values less than 10 μs in almost all cases at τ = 100 s. As with the local network, the standard deviation of the mean was only statistically significant as long as connection was maintained.

Figure 5 is a plot of one set of the time difference measurements. Each diamond symbol represents one measurement of the NIST ACTS time transfer, taken every

2.5 seconds. The symbols are simply plotted at the relative time of the measurement. The digitization and asynchronization effects are obvious in the apparent curves that the eye resolves, though these curves do not represent consecutive measurements.

Figure 6 is a plot of $\sigma_x(\tau)$ using as input the mean values from several sets of data, the connection being broken after each measurement. We measured a white noise level of $\sigma_x(\tau = 1.9d)$ = 270 μs, a degradation of about 3 or 4 from the 2 s measurements. The mean value of 15 measurements was 219 μs with a standard deviation of the mean of 70 μs, indicating that there are some systematic errors in the ACTS system. Thus time transfer accuracy, or synchronization, at the 1 ms level is quite reasonable with ACTS, but systematics appear to limit the time accuracy at the few hundred μs level.

Figure 7 shows a mod$\sigma_y(\tau)$ diagram for the average values. It is apparent that, with about a week's worth of measurements taken a few seconds per day to assure statistical veracity, we can obtain NIST traceability of frequency accuracy at better than 1 part in 10^9. This could be done automatically using low rates of calling, that is, a rate such that the telephone charge would be less that $2 for the whole time.

The confidence on the frequency calibration can also be written as follows:

$$s_1 = \frac{\sqrt{12}\ \sigma_x(\tau_0)}{\sqrt{N\ T}}, \qquad (7)$$

where $T = N\tau_0$, is the data length. Equation (7) holds if the spectrum of the time deviation is white, which appears to be the case from Figure 7. The confidence improves as the data length and as the square root of the number of measurements. Various options can therefore be among calling cost, desired accuracy and the amount of time to obtain the calibration. The fastest way to obtain a frequency calibration is to stay connected to take advantage of the better short-term stability. If a user wishes to maintain connection, please call (303) 497-3294 for arrangements for a special line; otherwise, an automatic disconnect occurs after 55 s. It appears that an accuracy of better than 1 part in 10^9 could be reached in less than 10^4 s for the long link. Time accuracy, on the other hand, cannot be improved much by averaging because of the apparent biases present.

Conclusions

A comparison of the time accuracy, the time stability, and the frequency calibration accuracy is plotted in Figure 8 for several current time and frequency dissemination systems. We see that the accuracies and stabilities for ACTS are quite competitive with some of the other traditional ways of obtaining traceability to NIST. In addition, Figure 9 indicates the cost effectiveness, which makes the ACTS calibration approach extremely attractive for the range of accuracies and stabilities it provides.

Figure 10 is a plot of the fractional frequency stability of the ACTS calibration approach compared with a wide variety of other techniques. The figure shows that a calibration accuracy of better than 1 part in 10^9 is readily and inexpensively obtainable either from a single long call or from a sequence of short calls averaged over a few days. With 300 bits/s modems of the same model, time accuracies better than 1 ms are available either by land and/or satellite links. Time stabilities pulse-to-pulse are better than 0.1 ms, but from call to call degrade to about 0.3 ms.

The ACTS telephone number is (303) 494-4774. Example user software can be obtained for $35 by calling (301) 975-6776 (refer to RM 8101, software for Automated Computer Time Service) [2,3]. The broadcast time code includes the Modified Julian Date, the year, month and day, the hour, minute and second, advanced alerts for daylight savings time and for leap seconds, the time difference between UTC and UT1 (earth time), and the amount by which the time code is advanced in order to arrive at the user's site on time.

The system will work at 1200 or 300 bits/s. Using a variety of modems, we observed much larger inaccuracies and instabilities than reported above - amounting to a few milliseconds. We found that 300 bits/s modems from a single supplier were the best combination.

Acknowledgements

The authors wish to thank Tom Weissert and Trudi Peppler for significant assistance in proofing the paper as well as in data processing and reduction.

References

[1] D. W. Allan, Keynote Address, "Synchronization of Clocks," Proc. of the 21st Annual Precise Time and Time Interval (PTTI) Applications and

Planning Meeting, Redondo Beach, CA, Nov. 28-30, 1989, pp. 1-10.

[2] J. Levine, D. W. Allan, and D. B. Sullivan, "The NIST Digital Time Service," Proc. of the 21st Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Redondo Beach, CA, Nov. 28-30, 1989, pp. 181-190.

[3] J. Levine, M. Weiss, D. D. Davis, D. W. Allan, and D. B. Sullivan, "The NIST Automated Computer Time Service," Journal of Research 94, 311-321 (1989).

[4] "Characterization of Clocks and Oscillators," NIST Technical Note 1337, Ed. by D. B. Sullivan, D. W. Allan, D. A. Howe, and F. L. Walls.

[5] D. W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, TN-121-TN-128 (1987).

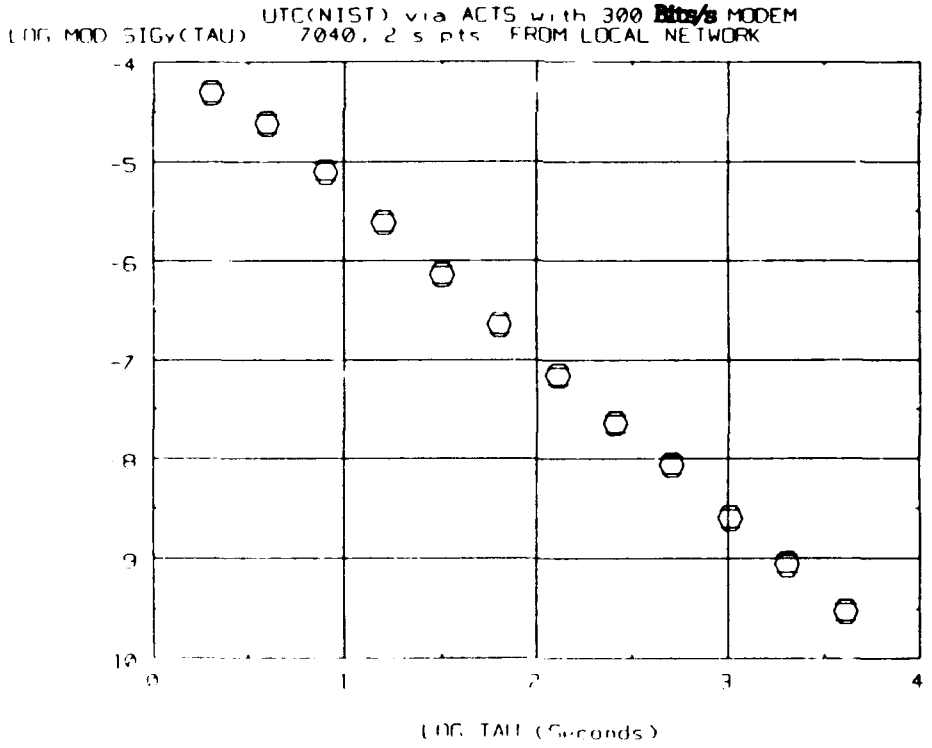


Figure 1. A plot of the fractional frequency stability, $\mathrm{mod}\sigma_y(\tau)$, for UTC(NIST) as accessed over a local telephone switching network via the NIST ACTS system. The confidence on a frequency calibration is given by $?\mathrm{mod}\sigma_y(\tau)$ The stability was analyzed for τ values longer than those shown, and the same nominal white PM ($\tau^{-3/2}$) model continued down to a level of about 3×10^{-11}.

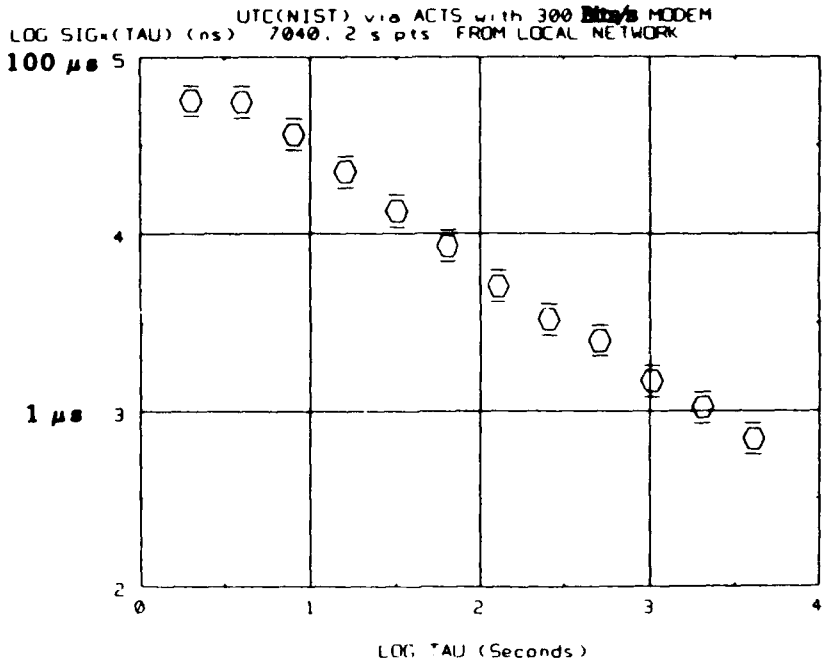


Figure 2. A plot of the time stability, $\sigma_x(\tau)$, of UTC(NIST) as accessed through the NIST ACTS system over a local telephone switching network. The stability improves to better than 1 μs for long enough integration times. This stability is lost when connection is broken. If, for example, a quartz oscillator could fly-wheel and maintain a microsecond accuracy while the system was being recalibrated via ACTS, then a system accurate to less than 1 μs could be maintained in the long term. This idea might be useful in a local calling network where there would be little expense for a dedicated line.

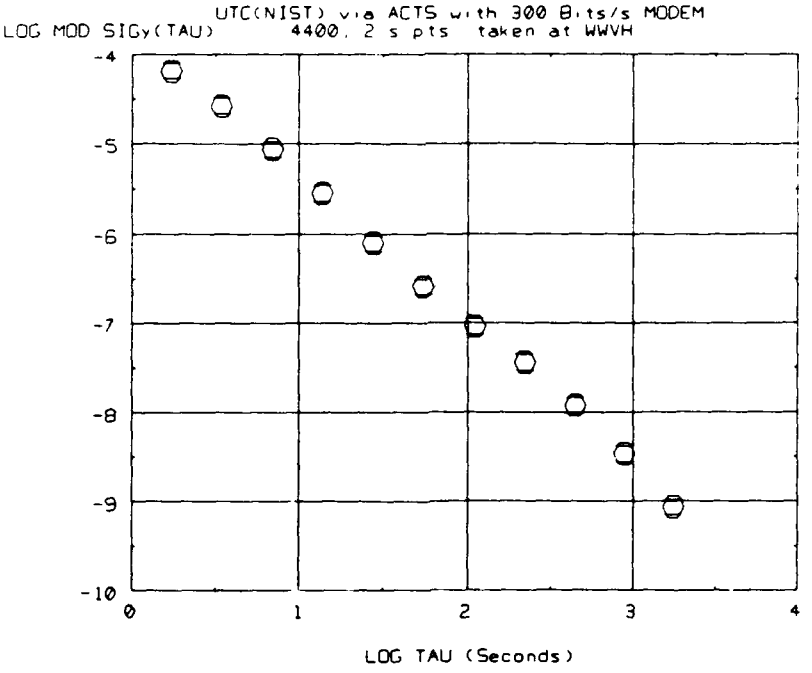


Figure 3. A plot of the fractional frequency stability, mod$\sigma_y(\tau)$, for UTC(NIST) as accessed over a long-link telephone network using the NIST ACTS system. The long-link involved both land-links and a satellite link between NIST Boulder, CO and WWVH Kauai, HI. The confidence for a frequency calibration is given by 2mod$\sigma_y(\tau)$. The data in this figure would indicate that a frequency calibration accuracy of less than 1 part in 10^9 is available from a continuous long-link connection with an averaging time greater than about an hour.

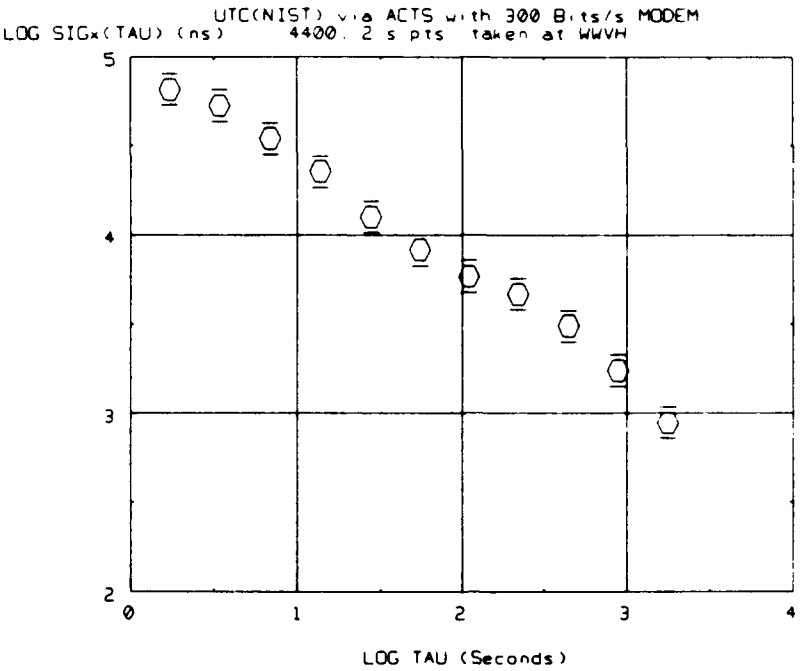


Figure 4. A plot of the time stability, $\sigma_x(\tau)$, of UTC(NIST) as accessed using the NIST ACTS system over a long-link telephone network using the NIST ACTS system. The long-link involved both land-links and a satellite link between NIST Boulder, CO and WWVH Kauai, HI. The stability improves to less than 1 μs for long enough integration times. This stability is lost when connection is broken. A single measurement has less instability than the noise introduced in repeated calls.

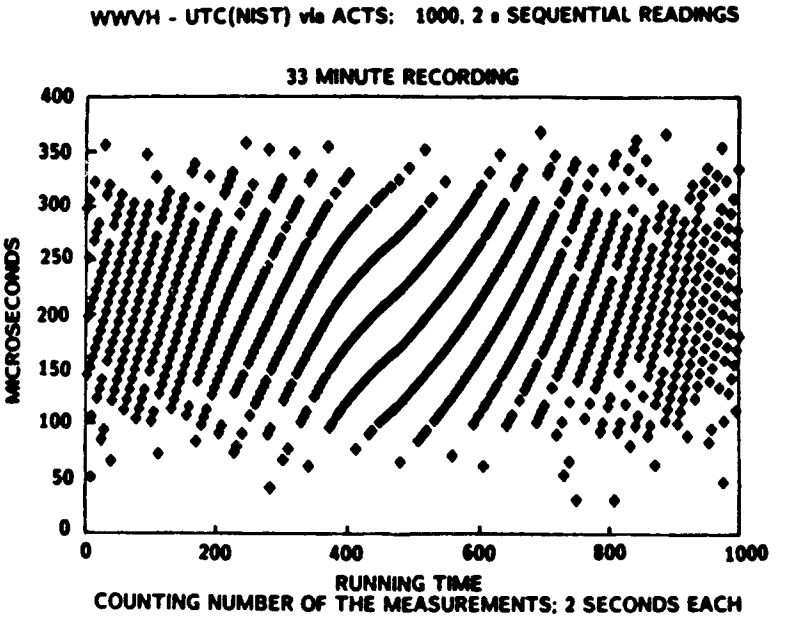


Figure 5. A plot of the individual ACTS measurements, each taken with two seconds spacing and with 300 bits/s modems of the same model at NIST Boulder and at WWVH Kauai. Each diamond symbol represents one measurement of the NIST ACTS time transfer. The symbols are simply plotted at the relative time of the measurement. The digitization and asynchronization effects are obvious in the apparent curves that the eye resolves, though these curves do not represent consecutive measurements. The mean value is 209 μs and $\sigma_x(N\tau_0) < 10$ μs. Hence, biases in ACTS limit its accuracy. The accuracy appears to be well under a millisecond.

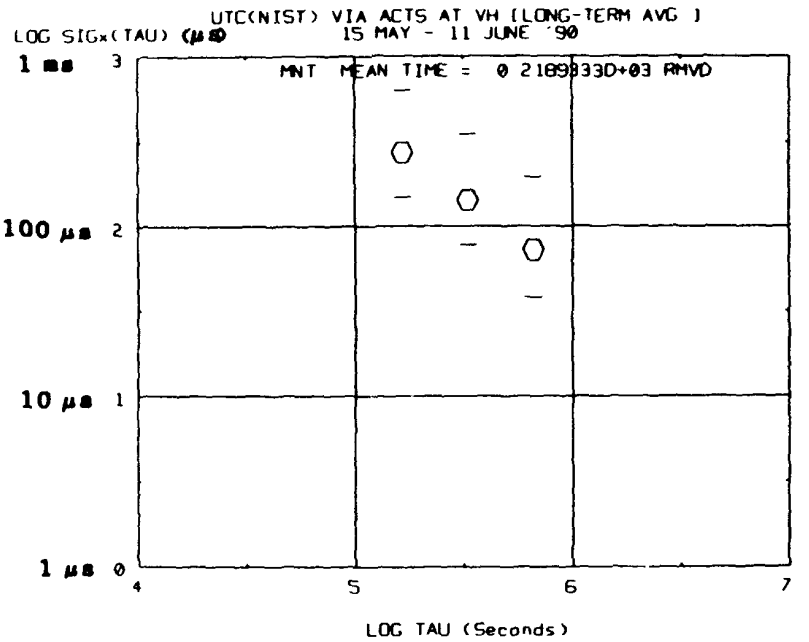


Figure 6. A plot of the time stability of the average values, $\sigma_x(\tau)$. Each call accesses UTC(NIST) via ACTS through a 300 bits/s modem at WWVH Kauai. The benefit of averaging the individual call's average value is illustrated by this plot since the values have a white spectrum ($\tau^{-1/2}$). This plot shows that a repeated set of short calls, reasonably spaced, is more economical for a frequency calibration; see equation (7).

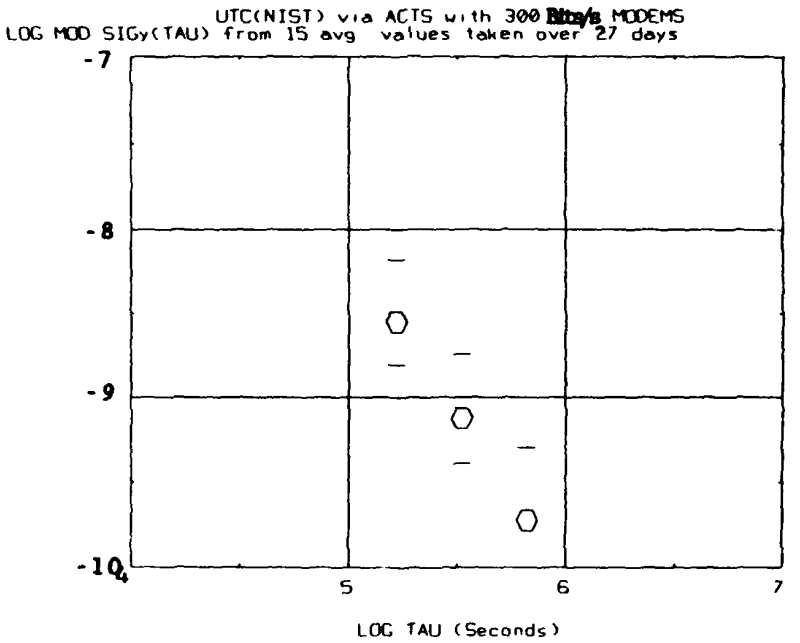


Figure 7. A plot of the fractional frequency stability, mod$\sigma_y(\tau)$, obtained from the average values of a series of calls with average spacing 1.9 days. Since the spectrum appears to be nominally white ($\tau^{-3/2}$) and the calibration accuracy is 2mod$\sigma_y(\tau)$, we see that calling once each night (when calls are less expensive) for about a week yields a frequency calibration of better than 1×10^{-9}.

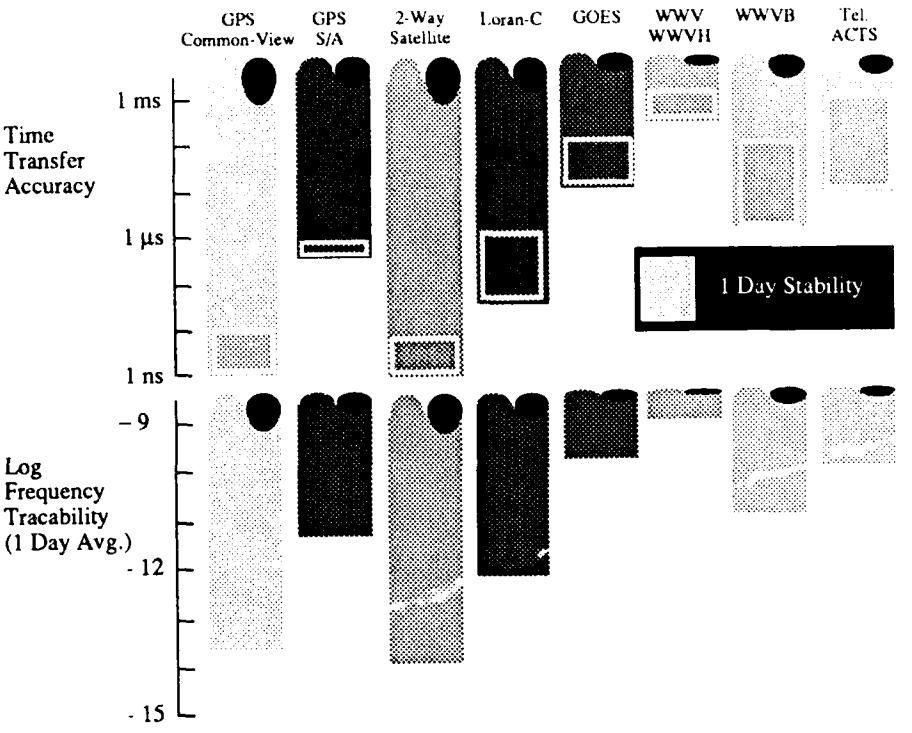


Figure 8. This bar chart compares the time accuracy, the time stability from day to day and the frequency calibration accuracy of several of our current dissemination systems.

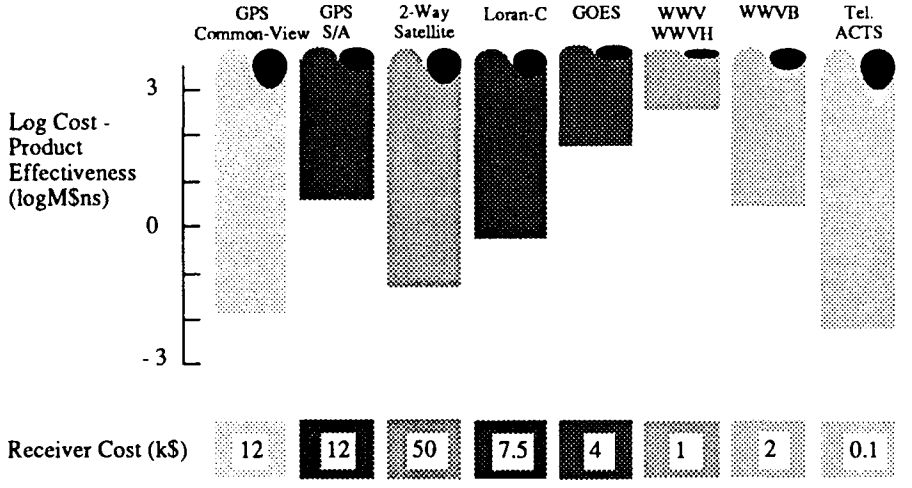


Figure 9. This chart illustrates the cost times the day-to-day time stability as a measure of the cost effectiveness of various time and frequency dissemination systems that are currently available. The units are mega-dollar nanoseconds. The further the bar goes down - the more cost effective the system.

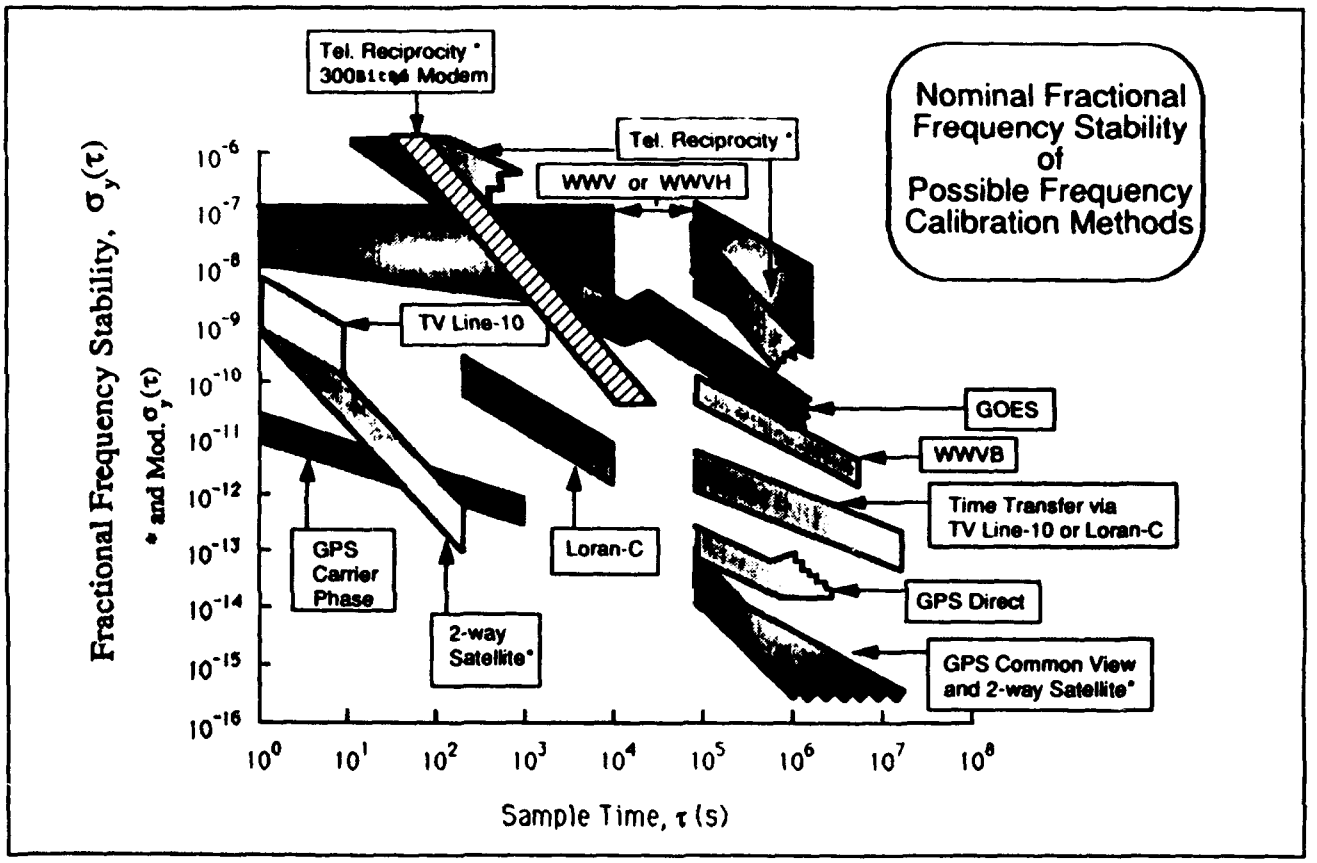


Figure 10. This fractional frequency stability plot compares most of the currently available time and frequency dissemination systems. The "*" indicates where mod$\sigma_y(\tau)$ has been used in order to distinguish if the deviations have a white PM spectrum. Otherwise, the conventional $\sigma_y(\tau)$ is used as the stability measure. The stability of the reciprocity of the telephone lines clearly makes this service a competitive one for the low accuracy user.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A REDUNDANT TIMING SOURCE FOR DIGITAL TELECOMMUNICATION NETWORK SYNCHRONIZATION

STEVEN M. BASS, PETER VLITAS, R. MICHAEL GARVEY

Frequency and Time Systems, Inc. 34 Tozer Road Beverly, MA 01915

Abstract

The synchronization of digital telecommunication networks requires solutions different from those traditionally used in the time and frequency industry. New approaches are needed to solve the special problems inherent in the distribution of timing between system nodes across communications lines.

This paper describes the problems which occur in synchronization of digital telecommunication networks. In addition the architecture of the FTS 3800 Timing Signal Generator is described. The 3800 has been designed to respond to some of the problems with existing synchronization equipment. Details of the design and performance will be presented.

Introduction

The telecommunication network in the US has largely changed from an analog system to a digital system over the past 10-20 years. The transition to a digital network has imposed a requirement that all equipment in the system be synchronized in frequency. There are tens of thousands of pieces of equipment in the telephone system which must all run at the same frequency (to within 1e-11). The requirement comes from the practical necessity of having finite sized data buffers. Disparate data rates result in buffer slips, which in turn disrupt data transmission.

It would be prohibitively expensive to provide each piece of equipment with its own primary frequency standard. The telecommunications industry instead relies on a limited number of primary frequency standards to which all the other equipment is synchronized. A standard approach and system architecture has evolved over the past decade, which makes it possible to inexpensively distribute timing throughout the network.

CH2818-3/90/0000-117 $1.00 © 1990 IEEE

This paper will describe the basic approaches used in providing timing in such a diverse environment. Emphasis will be placed on aspects which may be new to individuals familiar with precision time and frequency generation.

The current system is not perfect, and new applications and requirements require that the timing in the network be held to tighter and tighter tolerances. This has generated a need for new equipment, which can provide improved performance to meet the needs of the system as it evolves.

The remainder of the paper will discuss the architecture, features, and performance of a new instrument, the FTS 3800. The 3800 is a clock which has been designed to synchronize to a master timing source and distribute timing to other equipment. The 3800 was developed to meet the current needs and support the requirements demanded by some of the new features being introduced into the telephone system.

Telecommunication Synchronization

Synchronization Hierarchy

There are two basic approaches used by the industry to distribute timing in a network. The two approaches are not exclusive, and some mixture is usually used.

The most common model is a hierarchic system. A single master clock sits at the top of the network, and produces the reference frequency. The master clock is called a Stratum One clock. Timing is passed down from the master clock to the other clocks. One step down from the Stratum One clock are Stratum Two clocks. These are slave clocks which are phase

(occasionally frequency) locked to the Stratum One master clock. The Stratum Two clock typically contains a high quality ovenized quartz oscillator, or rubidium frequency standard. There are a limited number of nodes which contain Stratum Two clocks, from which timing is distributed to Stratum Three nodes. A Stratum Three clock contains a lower quality OCXO or a TCXO. Stratum Four clocks are normally used only at the end of a synchronization chain and do not pass timing to other clocks. The basic rule in this hierarchic system is that no clock can provide timing to a clock of a higher stratum.

The other common approach is plesiochronous timing, in which each node contains its own master clock and no synchronization is used between nodes.

Today, a mixture of the two techniques is commonly used. A network will often contain several plesiochronous islands with their own master clocks. Within each island timing is distributed hierarchically.

Inside a single network node there are many pieces of equipment which need to be synchronized. The preferred architecture is based on a Building Integrated Timing Supply (BITS) clock. The BITS clock is synchronized to an incoming data stream to generate the master timing. It distributes timing signals to synchronize all the other equipment at the node. This approach simplifies the synchronization engineering and trouble shooting.

A key element of the BITS architecture is a device called a Timing Signal Generator (TSG). A TSG is a Stratum Two or Three slave clock. In addition to its synchronizing function, a TSG also produces multiple timing outputs to synchronize the other equipment at the node. The TSG can either be the BITS clock, or is synchronized to the BITS clock to provide timing for all the other equipment.

Master Timing Source (Stratum One)

The master timing for a telecomm network is derived from a primary reference standard. A primary reference standard must have a long term accuracy of 1e-11 or better when compared to UTC [1]. In practice, the primary reference source can be a cesium frequency standard, or a disciplined oscillator controlled by a LORAN-C or GPS timing receiver.

Stratum Two and Three Clocks

The Stratum Two and Three clocks are synchronized to an external source which is traceable to a primary reference standard (Stratum One clock). Since the BITS clock will be either a Stratum Two or Three clock and will provide timing for a whole node, high performance and reliability are essential. Redundancy must be built in at all stages to ensure that valid timing signals are generated. The clock must accept multiple synchronization inputs (usually two), contain multiple oscillators, provide for redundancy at the timing distribution ports, and allow for maintenance and repair without disrupting operation.

The essential differences between Stratum Two and Three clocks are related to the quality of the oscillators in them. Table 1 lists some of the key performance parameters for these clocks. Many aspects of the clock performance are not specified (e.g. stability, phase noise, long term accuracy).

One key parameter is the clocks' ability to maintain an accurate frequency when the synchronizing reference is not available. Without a reference, the clock enters a mode referred to as holdover, in which it tries to maintain its oscillator at the last correct frequency. The clock performance in holdover is a function of three factors: the accuracy with which the clock estimates the frequency, the aging rate of the internal oscillator, and the thermal sensitivity of the clock.

The capture range guarantees that the clock can synchronize to another clock of its own stratum which is free running. There is no requirement on phase noise, stability, wander, loop bandwidth, etc. (Some of these parameters will be specified in updates to current specifications. There is a general consensus that existing specifications allow for clocks which will not adequately perform in the evolving network.)

Table 1. Slave Clock Performance Specifications

Parameter	Stratum Two	Stratum Three
Holdover [1]	< 1e-10	< 4e-7
Accuracy [2]	1.8 e-8	4.8 e-6
Capture range	3.6 e-8	9.6 e-6
Phase steps [3]	< 1us	< 1us

Notes:

1 Holdover is the frequency error after one day with a temperature change of ± 15 ° C at 3° C/hr

2 Accuracy is defined as the nominal frequency of the clock without synchronization

3 Clocks are allowed to have "rare" events in which the phase changes by 1 us, due to internal rearrangement or diagnostic operations

Timing Distribution

Timing is distributed on the data streams which transit the network. Most of the time a data carrying signal is used to pass the timing; commonly these data carrying signals run at 1.544 Mbits/sec and are designated DS1 in the industry signal hierarchy. The DS1 signals are framed (partitioned into discrete data segments) to allow the receiving equipment to align itself [2].

DS1 signals can be transmitted directly between network nodes. Frequently, multiple DS1 signals are multiplexed into a higher rate signal, transmitted to the next node, and de-multiplexed to recover the original DS1.

During the transmission of the DS1 signal between offices, the signal and timing can be degraded. In a DS1 transmission system, repeaters are needed every few miles to regenerate the signal to compensate for transmission loss. The repeaters can add substantial amounts of phase jitter. Network specifications allow up to 3.25 us of jitter at a modulation rate of 10 Hz on a DS1 signal (Note that this is 10π radians of phase excursion). The signal can also be completely lost, due to a cable break, component failure, or human error. Systems which multiplex the DS1's into a higher rate and de-multiplex them at the receiving end add their own sources of noise. The multiplex, de-multiplex process will usually add jitter, due to bit (or byte) stuffing needed to align the DS1 signals in the higher rate stream.

Problems with the Existing Network

The preceding discussion on telecomm synchronization describes the way things are designed currently. Concepts like BITS, and detailed examination of the clocks used in the system, has allowed telecommunications networks to normally maintain a frequency accuracy of better than 1e-10.

While synchronization of all network elements is critical to making digital communications systems work, it receives only the minimum attention possible from the people in the industry. One factor is the novelty and unfamiliarity of precision timing to telephone engineers. The concepts of Allan variance, phase noise, and other standard methods for evaluating clocks are foreign to them. This makes it hard to sell the value of improved timing to many telephone engineers. They know they need synchronized timing systems, but they don't understand the difference between implementations. One symptom of this is the crudity of some of the clocks used in digital switches. The switch may cost $10 million, but the manufacturer will scrimp on the oscillators to save $100, thereby degrading the timing output of the switch.

The long service life of telephone equipment creates problems for timing systems. Much of the equipment in the network is relatively ancient, and was designed to meet timing standards that are now inadequate. The need to work with all existing equipment in the network sets a limit on the quality of timing that can be achieved in the current network. It will be hard to improve this situation. Since the specifications must support the existing base of crude clocks, there is little pressure on equipment manufacturers to improve their clocks.

Maintenance of the existing timing system is difficult. Since much of the network is comprised of equipment and lines designed before digital communications was common, the network can suffer from many faults which produce symptoms similar to synchronization problems. Differentiating between clock faults and other network failures is difficult, and requires senior personnel to visit the site and evaluate the situation.

The piggybacking of the timing distribution upon the data streams creates its own problems. Most telephone personnel are primarily concerned that the data traffic doesn't fail. Lines which carry timing can be rearranged by craftsmen to solve traffic problems. This disrupts the synch network, and makes tracking the timing paths

difficult. In fact, it is easy to generate timing loops, in which whole sections of the network are isolated from the master timing source and rapidly go off frequency. Tracking down these sorts of problems is very difficult and time consuming.

The Need for Improved Clocks

Current clocks in the network range from quite good to abysmal. New clocks must be able to work with any of the current clocks, yet still provide good timing. This requires more intelligence in the clock, so it can monitor the input timing references more carefully, and reject them when their timing is unacceptable. It is easy to tell when a backhoe has cut the line, as the signal goes away. Very few clocks can tell when the upstream clock has been left in holdover by a craftsman, and is off frequency by 1 e-8. The ability to detect off frequency or noisy timing inputs will improve the quality of the timing system.

The goal of most network providers is to allow for centralized monitoring and control of their equipment. A centralized Operational Support System (OSS) allows status information to be gathered at a single location, simplifying the operations and maintenance of the system. Most current timing equipment lacks any support for centralized monitor facilities.

The telecommunications industry is continually looking for ways to increase the reliability and lower costs for communications. Each advance to higher data rates usually requires advances in the quality of the timing. The newest system being discussed is SONET, the Synchronous Optical Network. SONET is probably the first system that has placed a requirement on the stability of the timing source for its operational elements [3]. This requirement, while not severe, exceeds the performance of many clocks deployed in the network.

One way to improve performance and reliability of the network is to build more intelligence into each piece of network equipment, and to have the equipment communicate status over the data streams. The newest standard for framing DS1 signals contains a data channel embedded in the framing structure [4]. This can be used as a network management tool to signal failures, provide performance information, and possibly even remotely control other equipment. A portion of the data channel has been allocated for synchronization messaging. While the exact message format has not been defined, at some point in the near future a standard will be established. No current equipment can handle this messaging for synchronization. SONET has also allocated overhead bandwidth for synchronization, but again no messages have been defined yet. Direct transmission of synchronization status by the source clock will enhance the reliability of the timing network.

FTS 3800 Timing Signal Generator

Architecture

The FTS 3800 is an advanced microprocessor controlled TSG. As mentioned previously, the basic function of a TSG is to lock an oscillator to a digital data stream in the presence of large and noisy phase excursions as well as occasional losses of the data stream. The TSG uses a modular design for flexibility and maintenance. While the 3800 meets all existing requirements for Stratum Two and Three clocks, it has been designed to support the new applications that are being developed.

Figure 1 shows a block diagram of the 3800 clock. Modules in the 3800 fall into five categories: input, processor, oscillator, frame generator, and distribution. Table 2 lists the modules currently available for the 3800.

Table 2. 3800 Module Types

Input	DS1 CC Sine (1, 5, 10 MHz) E1 (2048 kHz CEPT protocol)
Processor	
Oscillator	Stratum 2 Stratum 3 Stratum 2.1
Frame Generator	DS1
Distribution	DS1 CC 1544 kHz RS-422 56 kHz V.35

All modules except the distribution modules are connected via a computer bus to the processor module. This allows the processor to extract data and control the operation of the modules. In addition, there may be

as many as 40 distribution modules, in the main chassis and two expansion chassis.

The 3800 modules have been designed to be removed and installed while the system is operating. This allows the unit to be upgraded or repaired without having to take it out of service.

Input Modules

The 3800 accepts up to four input modules. All input modules contain the same basic functional blocks. An input signal enters the module and the timing is extracted from it. The input is monitored for amplitude and proper data format. All signals are checked for loss of signal (LOS), while different types of input signals are monitored for faults specific to their format (e.g. DS1 signals are monitored for Alarm Indication Signal (AIS) and Out of Frame (OOF) conditions, and Bipolar Violations (BPV)). The phase of the input is measured against both of the oscillator modules simultaneously and all the information is transmitted to the processor module.

The input signal phase is measured on each input module so all inputs (not just the selected reference) can be continuously monitored. Each oscillator generates a 1 pps signal. The input signal clock is also divided to 1 pps. The oscillator and input 1 pps signals generate a phase reading once a second with a resolution of 54 ns.

Processor Module and Operating Software

The processor module contains an 8088 microprocessor. The processor module software runs the 3800, monitors for faults, and controls two RS-232 ports for interaction with the operator, or a remote monitoring system.

Low level routines collect the status of each module over the processor bus. This is used to determine if any modules have failed. The phase of the derived input signal timing and status of the input signals is also read. The phase data is used to control the oscillators and to monitor the frequency of each input signal. Signal faults (LOS, AIS, etc.) are used to qualify which inputs are acceptable for use as a reference.

The phase information for the selected reference input is used to control the oscillator modules. A digital phase locked loop, containing both proportional and integral terms, is implemented in the 3800 software. The software implementation allows for the use of variable and extremely long time constants. For the Stratum Two oscillator, time constants in the range of 100 - 10,000 seconds are allowed. This allows the operator to optimize the performance of the unit based on the system configuration. Since all inputs are monitored simultaneously, the 3800 can switch between inputs without introducing any errors into the phase lock loop.

The phase information is also used to calculate the frequency over the past 100, 1000, and 10000 second intervals. This data is stored for retrieval by the operator. The software compares the measured frequency against operator set limits and generates an alarm when any input exceeds the limit. If the faulted input is currently the reference, the 3800 can be configured to automatically switch to a different input.

Since the 3800 is already collecting a large quantity of information about the input signals, its own operation, and the actions of the operator, an event log was added to assist in monitoring the state of the timing system. The event log time stamps and maintains a record of the last 500 significant events which have occurred. This allows the operator to look for correlations when problems are detected.

The 3800 is highly configurable. The operator can configure the limits at which signal faults and frequency errors will cause the input to be considered unacceptable, and how the 3800 will respond to the failure. The PLL time constant can be set (independently for each oscillator) anywhere in the range allowed for the oscillator. The oscillator operating mode can be set (holdover, acquire, etc.) for times when it is necessary to override the default operating procedure.

The phase data for each input is stored for 24 hours, along with the oscillator control voltages. This information can be extracted over the RS-232 interface and analyzed to provide a detailed record of the input signal behavior. Frequency measurements are also stored for subsequent analysis.

Oscillator Module

Each 3800 can contain up to two oscillator modules. The oscillators are configured redundantly. A failure in one initiates an automatic switchover to the other oscillator thereby providing continuous operational

signals. All the output timing signals are generated from the master oscillator.

All oscillator modules are similar, regardless of the type of oscillator used in the module. Each module contains a 16 bit DAC to control the oscillator frequency, an 18.528 MHz synthesizer, and a 1 pps generator.

Oscillators of varying quality can be used in the 3800. Status data from the oscillator module is used to inform the processor about the type of oscillator used, so it can set up the phase lock loop parameters correctly.

A Stratum Three module uses an AT-cut FTS 2510 OCXO as the oscillator. The Stratum 2.1 module uses a SC-cut FTS 1000B oscillator. Its performance greatly exceeds the Stratum 3 module, but falls short of the Stratum Two requirements for thermal stability. The Stratum Two module uses an FTS 1130 oscillator. The 1130 oscillator utilizes a double oven technique to provide the required thermal stability. The inner oven is hermetically sealed to provide immunity to humidity changes [5]. Table 3 lists the specifications for the different oscillator modules.

Table 3. Oscillator Module Specifications

	Str 3	Str 2.1	Str 2
Tuning Range	10e-6	3e-7	3e-7
Thermal stability(per ° C)	7e-9	1.7e-11	3e-12
Aging (per day)	1e-8	1e-10	5e-11
Time constant range(secs)	20-100	100-3,000	100-10,000

Frame Generator Module

There is one frame generator associated with each oscillator module. The frame generator uses the signals from the oscillator module to generate properly framed signals used by the distribution modules to produce the timing outputs.

The frame generator also acts as an interface between alarm information generated on the distribution bus and the processor bus.

Distribution Modules

The distribution modules generate the multiple output signals which are used to synchronize other equipment at the node. All the distribution modules are similar in design, with differences mostly in the output stage to create the appropriate waveforms and levels.

All distribution modules derive their outputs from signals generated by the frame generator module. Each module can utilize the signals created by either frame generator. Normally the processor module selects which frame generator signals should be used to produce the outputs. Each distribution module monitors its inputs, and if one disappears the module will switch to the other input automatically, thus providing a fail-safe redundancy configuration.

The timing outputs are the principal output of the 3800. Loss of a TSG output will cause the downstream equipment to become unsynchronized and will disrupt communication. Redundancy of the timing outputs is critical. The 3800 utilizes a one-for-one "hitless" approach. Modules are configured in pairs, each module producing a full set of outputs, nominally 20. The modules produce the same complement of output signals which are passively combined to drive one set of outputs. If one of the modules fails, it removes its output signals and signals the other module of its failure. The remaining module boosts its output level to compensate for the failure. The failed module can be removed and replaced without ever disrupting the timing signal.

Performance

Jitter

The telecommunications network is largely susceptible only to phase noise at frequencies above 10 Hz. The industry denotes this high frequency noise region as jitter, and for frequencies below 10 Hz as wander. The ability to tolerate large amounts of jitter, and to attenuate jitter on the TSG timing outputs are important.

A clock must be able to work with signals which are highly corrupted by noise. Current specifications require that a Stratum Two or Three clock accept up to 18 us of phase jitter at 10 Hz. The 3800 performance and system specification are shown in Figure 2.

Because jitter is defined to only include noise above 10 Hz, the jitter attenuation in the 3800 is essentially perfect. The update rate for the oscillator control voltage is 1 Hz, so no input jitter is passed through to the timing signals. (Of course, jitter (high frequency noise) is translated into phase fluctuations at low

frequencies, but the industry doesn't care much about this, yet.)

An alternative way to view the high frequency noise is as phase noise. Figures 3 and 4 show the phase noise for Stratum Two and Stratum Three 3800's locked to a DS1 signal. The DS1 signal is modulated with 1 us of white noise (10-150 Hz band limited) to simulate a real signal. This approach to measuring the performance of slave clocks is being considered by telecomm standards organizations and may become standard.

Short Term Stability

There is currently no requirement on the stability of the clock at frequencies below 10 Hz. New applications, especially SONET, are sensitive to the phase stability in the range of 1 to 1000 seconds. Figure 5 shows the stability for the 3800 when locked to a DS1 signal with 1 us of white phase noise, along with the current value for the stability specification. The stability specification is being reviewed, and is likely to change.

Stress Test

DS1 signals are never perfect, and frequently terrible. In addition to substantial amounts of jitter and wander, there are usually several episodes a day in which the input is lost entirely for short periods. This could be due to maintenance, automatic re-routing of communications lines, transient faults, or other causes. These brief episodes make it impossible to maintain a continuous phase history of the input. Each time the signal is lost and reappears, the clock must establish a new baseline for the phase locked loop. Noise makes this process imperfect, and the nominal phase to lock to will appear to undergo a random walk. This can lead to a net frequency offset. The maximum frequency offset in the presence of stressed signals is currently not specified. The performance for the 3800 was measured, using a DS1 input with 1 us of white noise and a 100 ms interruption every 7 minutes. (This level of interruptions is four times larger than the level recommended for stress testing and is much higher than what is seen in the network.) The results are shown in Figure 6. The frequency offset due to the random walk is effectively zero for the Stratum Two clock, and 4e-12 for the Stratum Three clock. This type of performance is adequate to maintain the 1e-11 frequency accuracy requirement, and is vastly superior to what many clocks currently in use can do.

Holdover

The ability of the clock to hold an accurate frequency when the input reference is lost is the ultimate differentiation between Stratum Two and Three clocks. The aging and thermal sensitivity are largely set by the type of oscillator used in the clock. The error in the initial estimation of the frequency is determined by the software in the 3800. For a Stratum Two or 2.1 oscillator with a 1000 second time constant, the frequency estimation error is specified as 3e-11. Typical performance, even with jittered input signals is 5e-12. The Stratum Three estimation error is specified at 3e-9. These levels of accuracy are sufficient to allow the 3800 to achieve the one day holdover requirements of 1e-10 for Stratum Two and 4e-7 for Stratum Three clocks.

Conclusion

Digital telecommunication network timing has problems which differ greatly from those seen in more traditional time and frequency environments. New approaches are needed to cope with the special features of this environment. The FTS 3800 Timing Signal Generator has been designed to provide high quality timing in the present network, and to be adaptable to the new requirements that are being developed.

References

[1] ANSI T1.101-1987 American National Standard for Telecommunications - Synchronization Interface Standards for Digital Networks

[2] ANSI T1.102-1987 American National Standard for Telecommunications - Digital Hierarchy - Electrical Interfaces

[3] ANSI T1.105-1988 American National Standard for Telecommunications - Digital Hierarchy - Optical Interface Rates and Formats Specifications

[4] ANSI T1.403-1989 American National Standard for Telecommunications - Carrier to Customer Installation - DS1 Metallic Interface

[5] F.L. Walls, "The Influence of Pressure and Humidity on the Medium and Long-Term Stability of Quartz Oscillators" in Proceedings of the 42nd Annual Frequency Control Symposium, 1988, pp 279-283

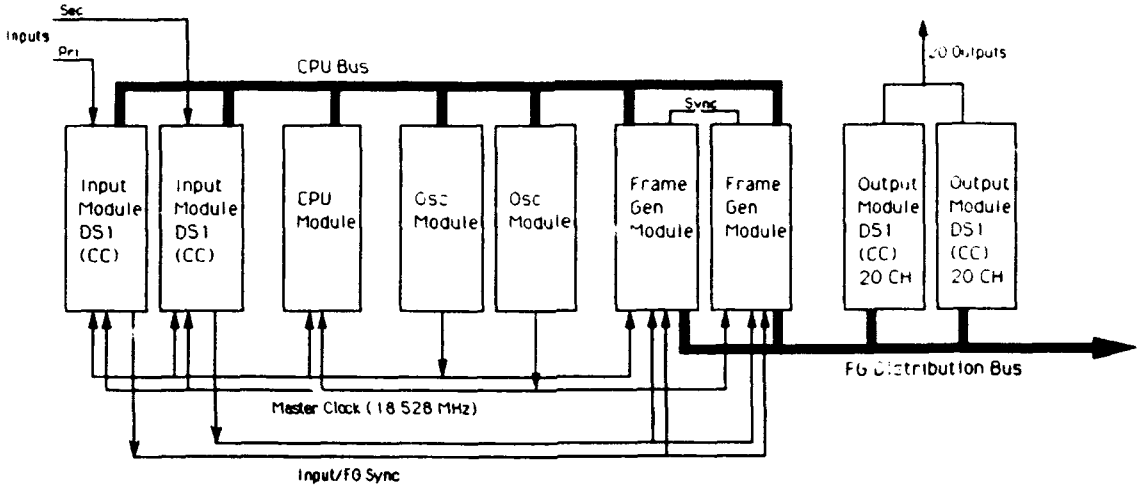


Figure 1. FTS 3800 Timing Signal Generator

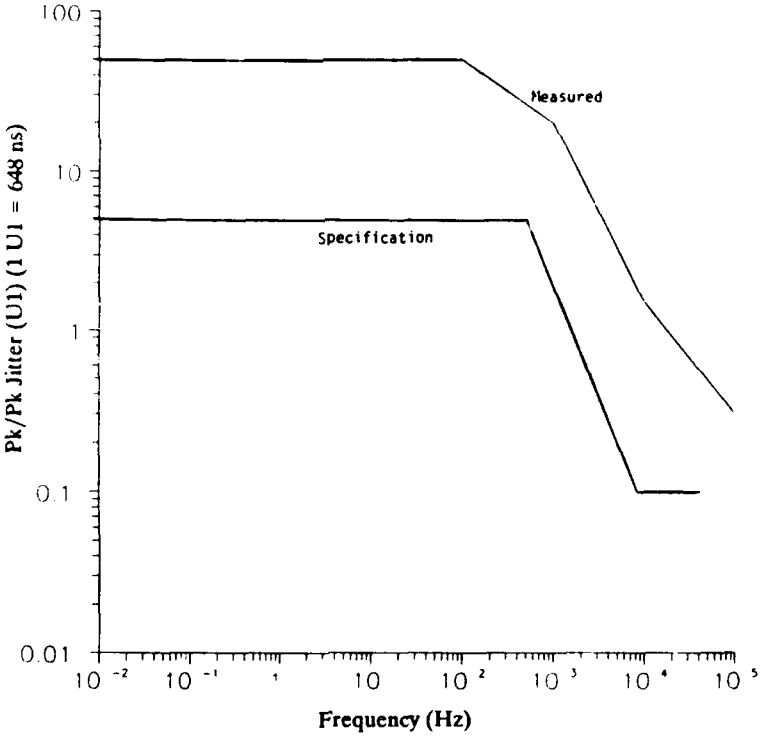


Figure 2. Jitter Tolerance

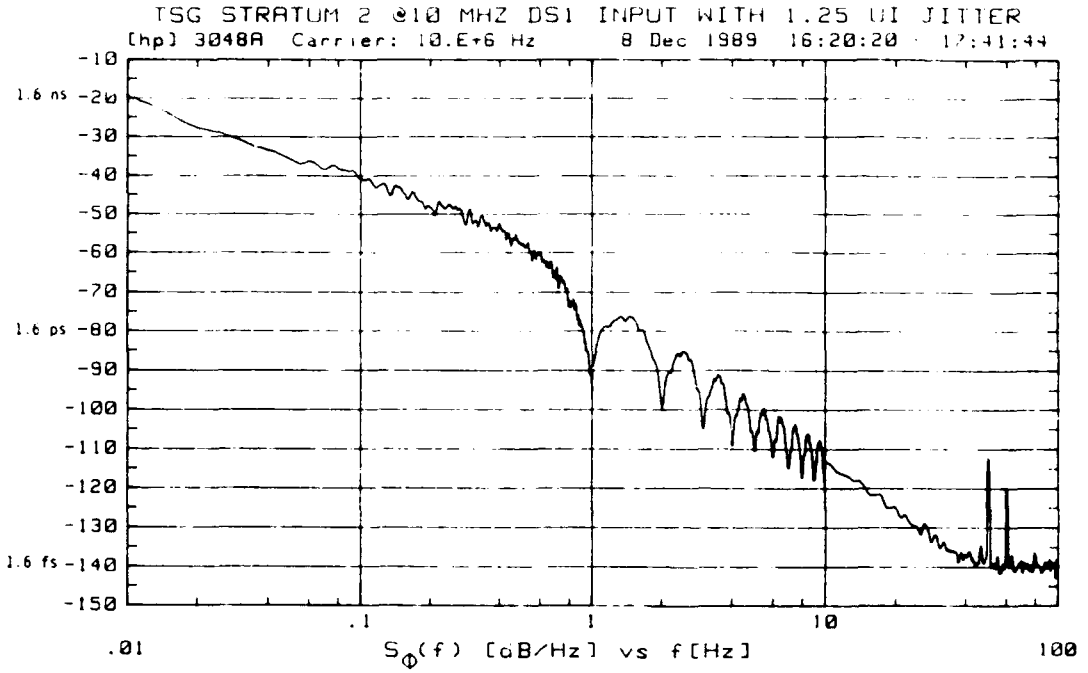


Figure 3. FTS 3800 Phase Noise, Stratum 2, Jitter

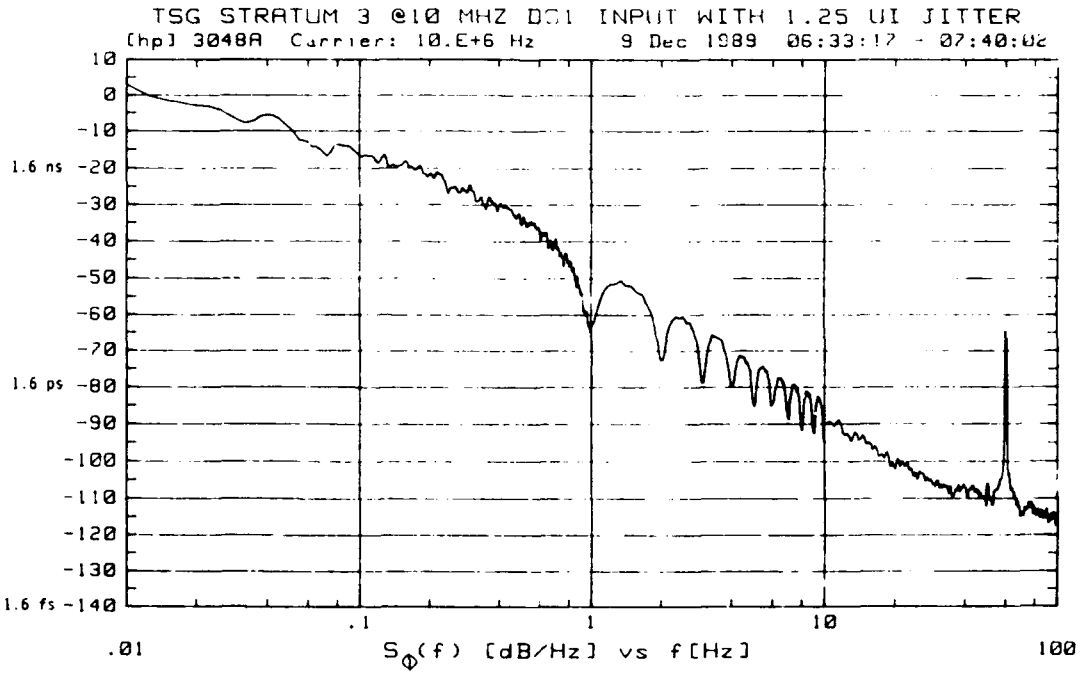


Figure 4. FTS 3800 Phase Noise, Stratum 3, Jitter

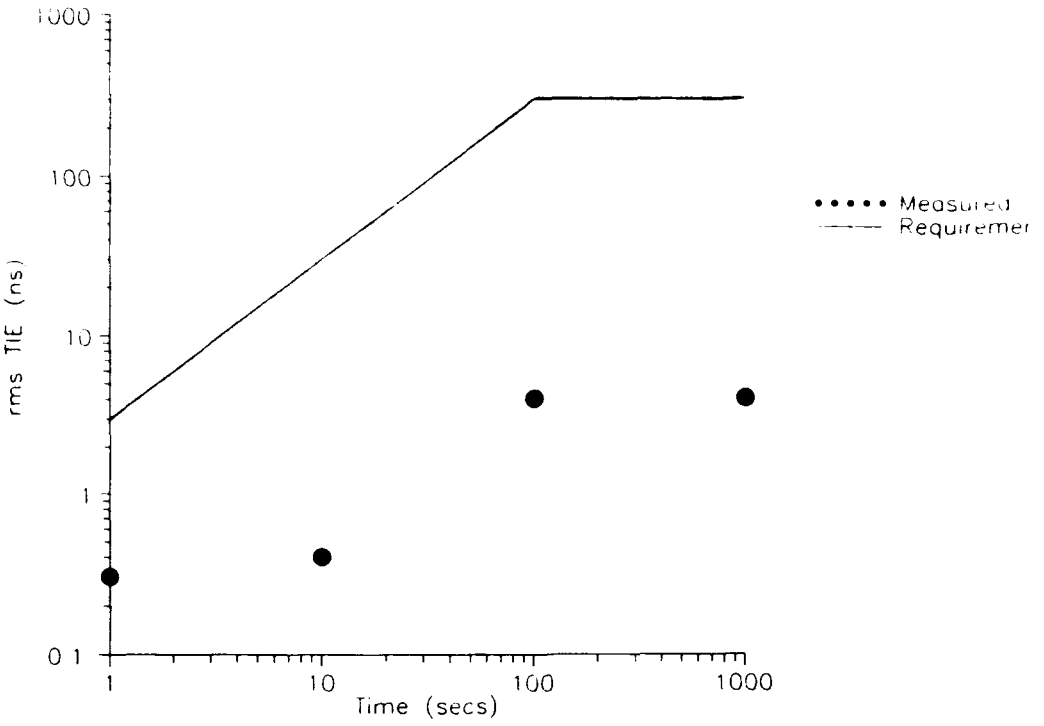


Figure 5. Stratum 3 Short Term Stability

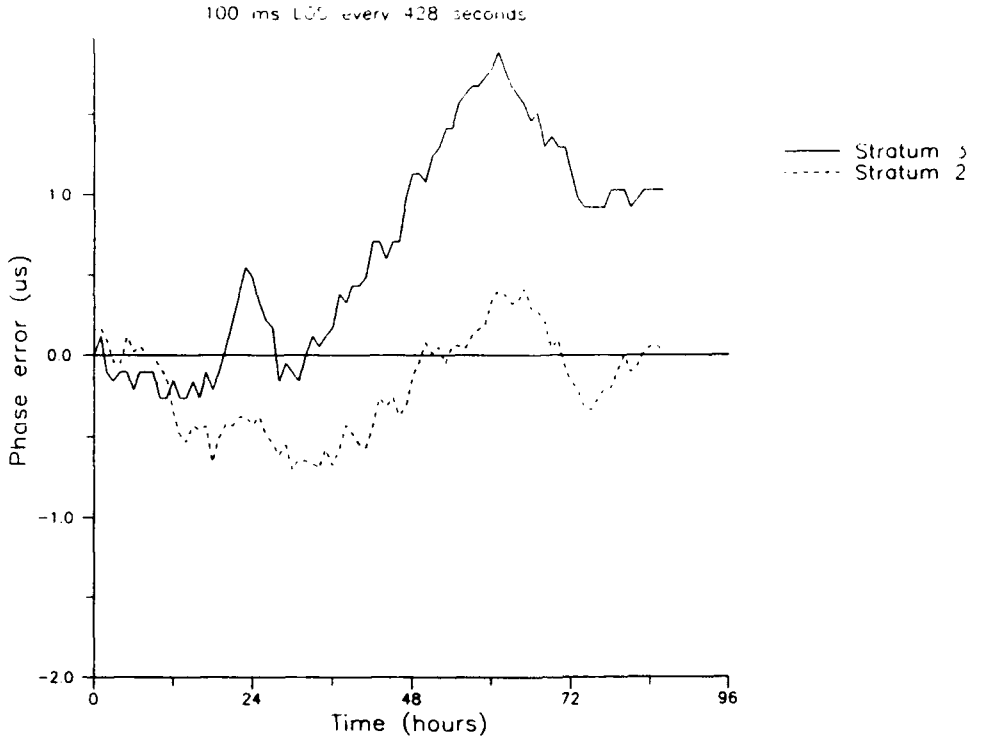


Figure 6. Stress Test

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FREQUENCY AND TIME STABILITY OF GPS AND GLONASS CLOCKS*

by

Peter Daly and Ian D. Kitching
The University of Leeds
Leeds LS2 9JT, United Kingdom

and

David W. Allan and Trudi K. Peppler
National Institute of Standards and Technology
Boulder, Colorado 80303, U.S.A

ABSTRACT

The frequency stability and reliability of the clocks are critical to the success of the GPS and GLONASS programs. We will show some of the similarities and differences between the clocks involved in these two systems. Because both systems plan to be operational in the next few years, the data leading up to this operational stage is of significant interest. On-board clocks and the stability of the master control clocks for these systems are analyzed.

We will discuss the attributes of these two systems as time and frequency references. Their relationship to UTC will also be illustrated. More data over a longer period of time was available for the authors from GPS than from GLONASS. Even so it is obvious that both systems have matured. Though the GLONASS system was developed later, its overall clock performance has improved more rapidly. Some of the more recent GLONASS clock performance is at about the same level as that of the GPS clocks.

The analysis has yielded some very interesting contrasts, comparisons, and changes in these systems that should be of great interest for time and frequency users, as well as for clock vendors and receiver vendors.

INTRODUCTION

The Global Positioning System (GPS) and the GLONASS (USSR satellite navigation system) are the first systems to use atomic clocks in the space environment.

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

The reasons for using atomic clocks to assist in forming an accurate navigation solution for a receiver set are obvious. In the error budget for navigational accuracy, the performance of satellite clocks is an important entry. As a direct source of time and frequency from space, the performance of the satellite clocks, again, is an important consideration, as are the methods of controlling the systems.

The purpose of this paper is a cursory study of the performance of these two systems as time and frequency reference sources. We do not have access to all of the data from these systems for their full time of operation but do have significant samples from which it is possible to assess the general performance. For this paper the GLONASS data base is that taken by the receivers at the University of Leeds. [1] The data base for GPS is that of the National Institute of Standards and Technology (NIST) in Boulder, Colorado.

We will study the time inaccuracy and time instability, and the frequency inaccuracy and frequency instability, of both systems. Contributing to the investigation are the frequency offsets, the frequency drifts, and methods of estimating these states as seen through clock noise, measurement noise, and uncertainties associated with other relevant system parameters.

GPS SATELLITE-CLOCK PERFORMANCE

The timing reference used for the analysis of the GPS were: UTC as generated by the Bureau International des Poids et Mesures (BIPM), UTC(NIST) and UTC(USNO). UTC is the official international reference scale, and we will use it wherever data permit and values are relevant. The frequency stabilities of all three of these reference time scales are better than a few parts in 10^{14}. In the long

term, both UTC(NIST) and UTC(USNO) are kept synchronous within a few microseconds of UTC.

The receivers used for measuring GPS are L1 clear-access timing receivers. The inaccuracies in determining the time of a GPS satellite clock are the sum of the uncertainties associated with the broadcast estimate of the satellite's position and ionospheric correction for the signal path, the tropospheric delay, multipath perturbations, receiver hardware delays, and software perturbations. From experience, the size of these inaccuracies can amount to a few tens of nanoseconds.

A given GPS satellite position with regard to a fixed receiver on the earth remains the same from day to day when measurements are made once per sidereal day. With the data taken in this way the time instabilities from day o day are typically less than 10 ns.

The frequency inaccuracy and the systematic trends in the frequencies of the GPS satellite clocks are determined by an appropriate filtering of the daily time readings. The frequency instabilities are measured in the usual manner using a $\sigma_y(\tau)$ diagram to characterize the time-domain performance.

Figures 1a and 1b are plots of the time accuracy and the frequency accuracy of the GPS-received signal, with respect to UTC. GPS time is estimated with a Kalman-Bucy filter for each operating satellite clock. The filter determines a correction to be applied for each satellite clock and broadcasts this information. Biases and random variations of the order of a few nanoseconds have been observed between the GPS time as given by the different GPS satellites. [2] The time stability from day to day of a given satellite is typically better than if the measurement is made on different satellites. Figure 2 is a plot of the frequency stability of GPS time over the entire interval shown in Figure 1.

When a weighted set of the GPS satellites are used for time transfer in a common-view mode, day-to-day stabilities less than 1 ns have been achieved. [3] The common-view mode is most commonly used for transferring clock time to the BIPM (pertinent to the UTC generation). Internationally the range of time transfer stabilities from day to day is from 0.8 to about 6 ns when using a weighted set of the GPS satellites.

Under a Department of Defense directive, US military time is to be synchronous with UTC(USNO). Figure 3 is a plot of how well the GPS has achieved that goal. Because of the low Fourier frequency processes present in the data, the standard deviation of the data can be deceptive in meaning. Figure 4a is a plot of $\sigma_x(\tau) = \tau * \mathrm{mod}\sigma_y(\tau)/\sqrt{3}$ for these data, which provides a way of dealing with these low frequency processes in a statistically valid way. The GPS data word provides a correction to GPS time; this correction provides an estimate of UTC(USNO). Plotted in Figure 4b is the time stability of the estimate of UTC(USNO) as obtained from GPS satellite NAVSTAR 10.

Figure 5 is a plot of the frequencies of several of the GPS satellite clocks corresponding to rubidium-gas cell clocks and to cesium-beam clocks. The plots are with respect to the rate of UTC(USNO) which is typically within a few parts in 10^{14} of the rate of UTC.

Figure 6 shows plots of the frequency stabilities of samples of the GPS satellite clocks. Each kind of clock seems to have a characteristic kind of "fingerprint." In addition, the frequency drifts--as can be observed in Figure 5--are very different between the different kinds of clocks.

Figure 7a is a plot of the time of GPS satellite No. 14's clock with the corrections applied to give an estimate of UTC(USNO). It is obvious where selective availability (SA) is turned on and off. SA is the purposeful degradation of the broadcast signal in order to deny full GPS accuracy to a non-cleared receiver. This degradation is accomplished by modulating the effective output of the satellite clock and/or by degrading the broadcast ephemeris (satellite position information). Figure 7b is a frequency stability plot with and without selective availability (SA) present. This plot employs the statistical measure $\mathrm{mod}\sigma_y(\tau)$ to show that the long term variations, though at a high level, can be characterized as a white phase or time modulation process. [4] With SA on it takes a couple of weeks of averaging before the instabilities of a cesium-beam clock become measurable.

March 25, 1990, GPS went officially to the SA mode of operation on all of the block 2 satellites (those satellites launched since 1988). The effects of SA are readily apparent. We have been given to understand that as long as the block 1 satellites last, there will be no SA on them. We performed an experimental test of the effects of SA on GPS common-view time transfer via satellite 14. We measured the once per sidereal day time instabilities, $\sigma_x(\tau = 1\ \mathrm{d})$ over a three month period as compared to the block 1 satellites. The results are tabulated in Table 1. We conclude from this that little or no ephemeris degradation was present on SVN 14 during this period. In other words, if the SA present on SVN 14 were mostly placed in the modulation of the on board clock's output, this would probably cancel when the differences are computed in the common-view mode and if the

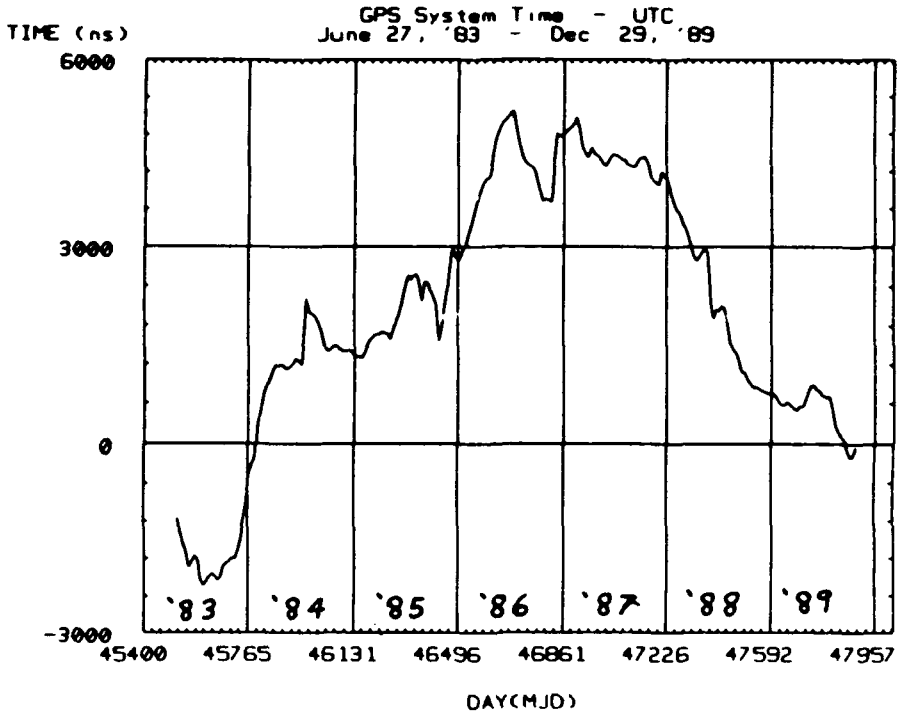


Figure 1a

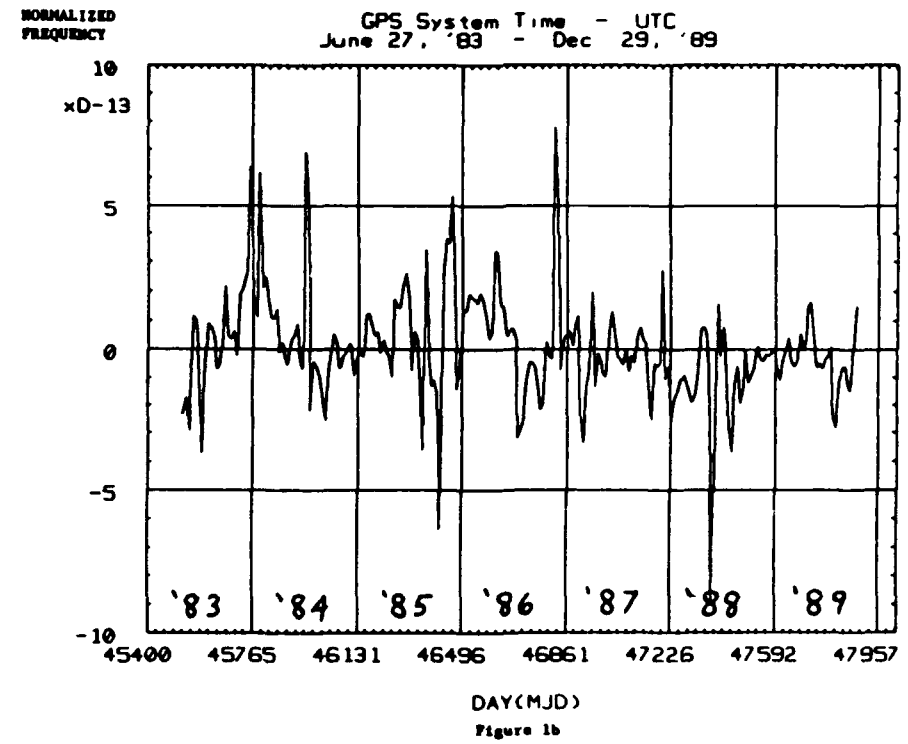


Figure 1b

Figures 1a and 1b. GPS time and normalized frequency from 1983 onward as measured against the international time and frequency reference, UTC.

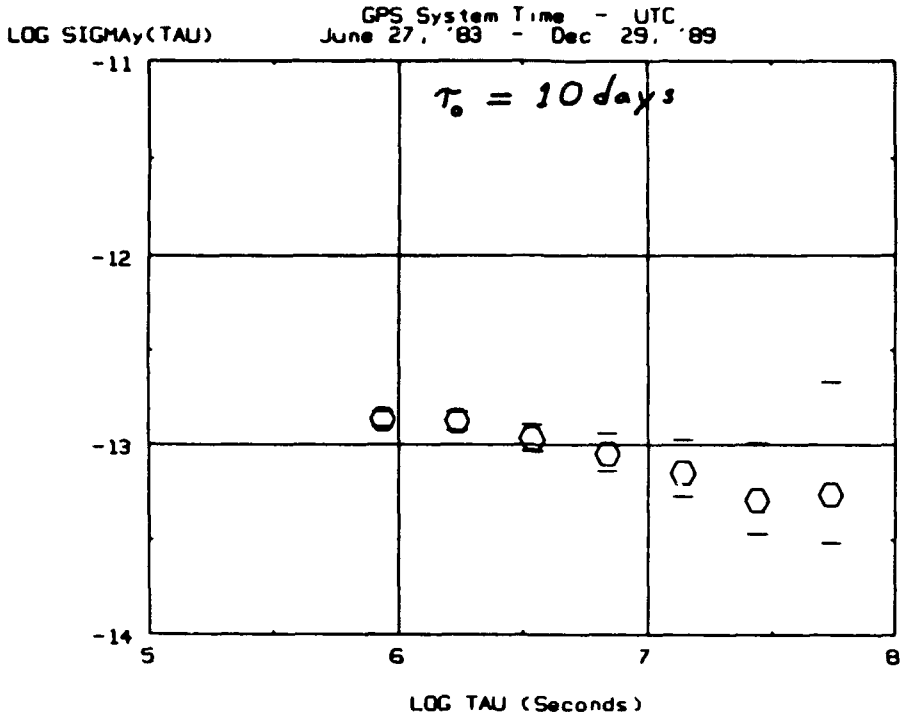


Figure 2. A plot of the fractional frequency stability, $\sigma_y(\tau)$, of the data shown in Figure 1 - the stability of GPS with UTC as the reference.

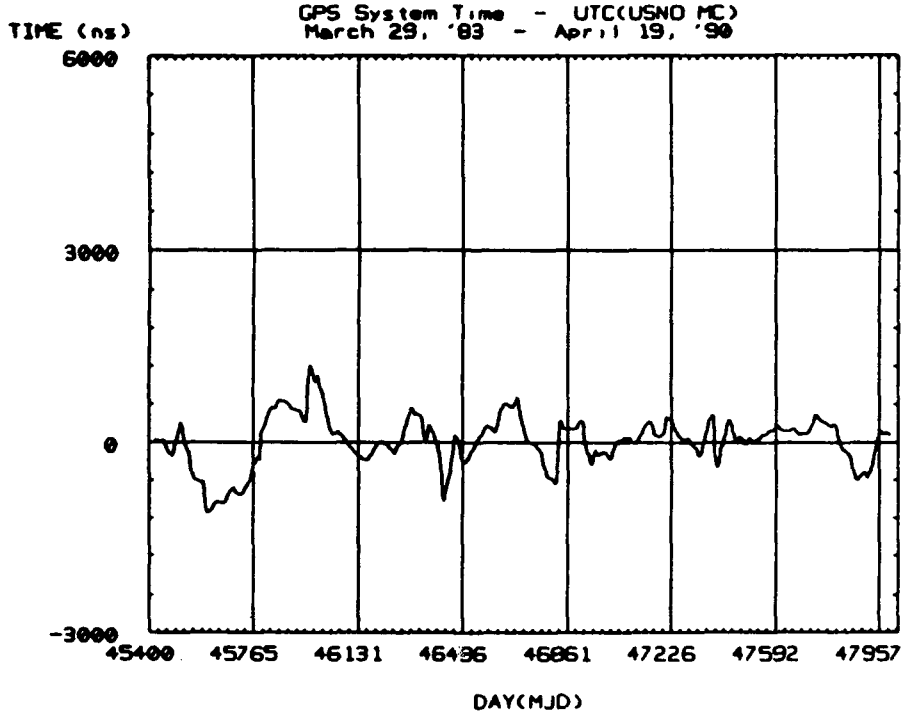


Figure 3. A plot of GPS time against UTC(USNO MC). Over the last few years it has stayed within its ∓ 1 μs goal. UTC(USNO) and UTC(USNO MC) are the same.

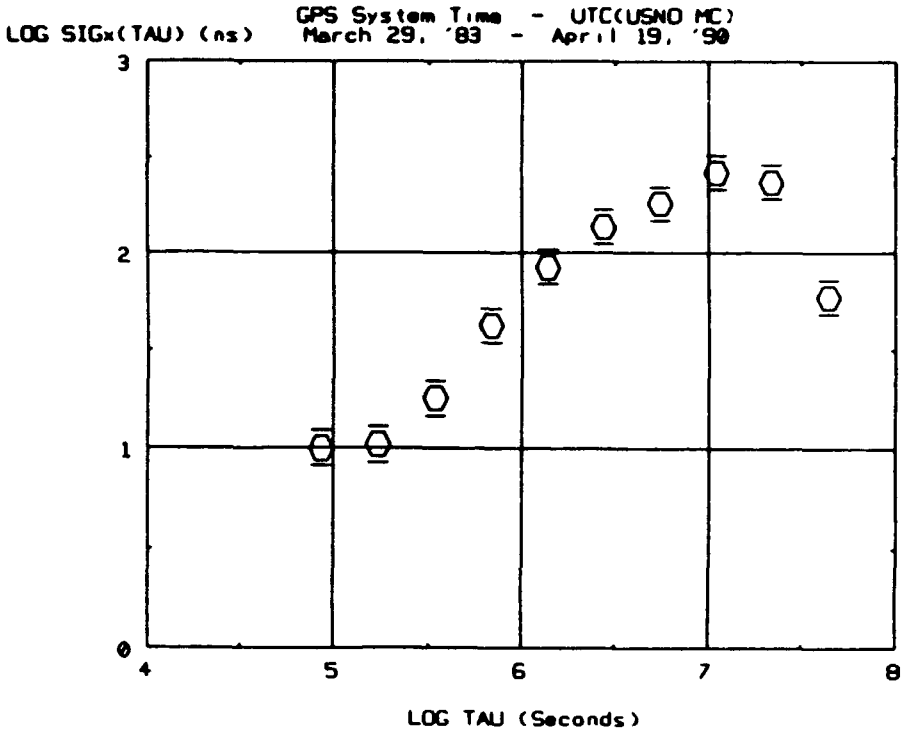


Figure 4a

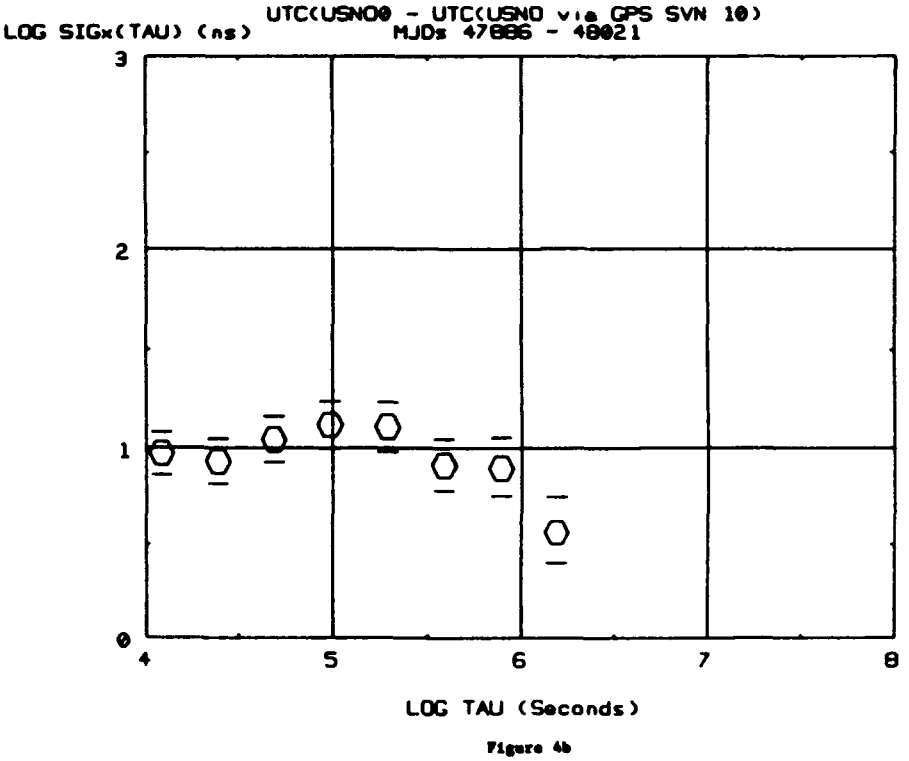


Figure 4b

Figure 4. A plot of the time stability ($\sigma_x(\tau)=\tau*\mathrm{mod}\sigma_y(\tau)/\sqrt{3}$) of GPS time and of GPS time with the UTC(USNO MC) corrections with respect to UTC(USNO MC).

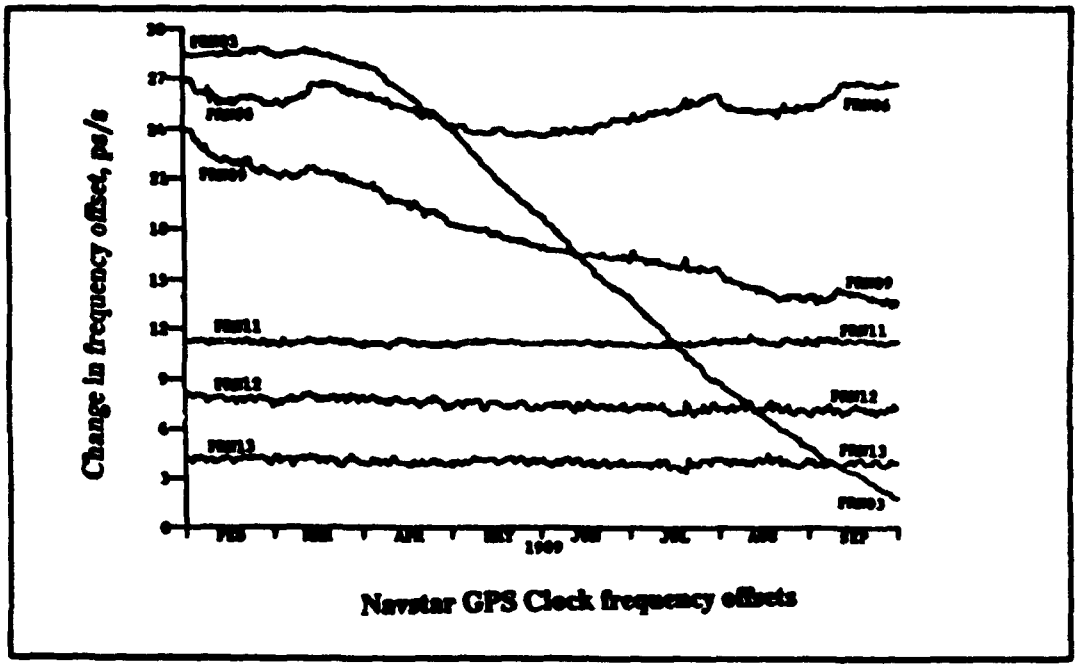


Figure 5. A plot of the normalized frequency offset in parts in 10^{12} with an arbitrary offset subtracted from each GPS satellite clock's frequency for plotting convenience.

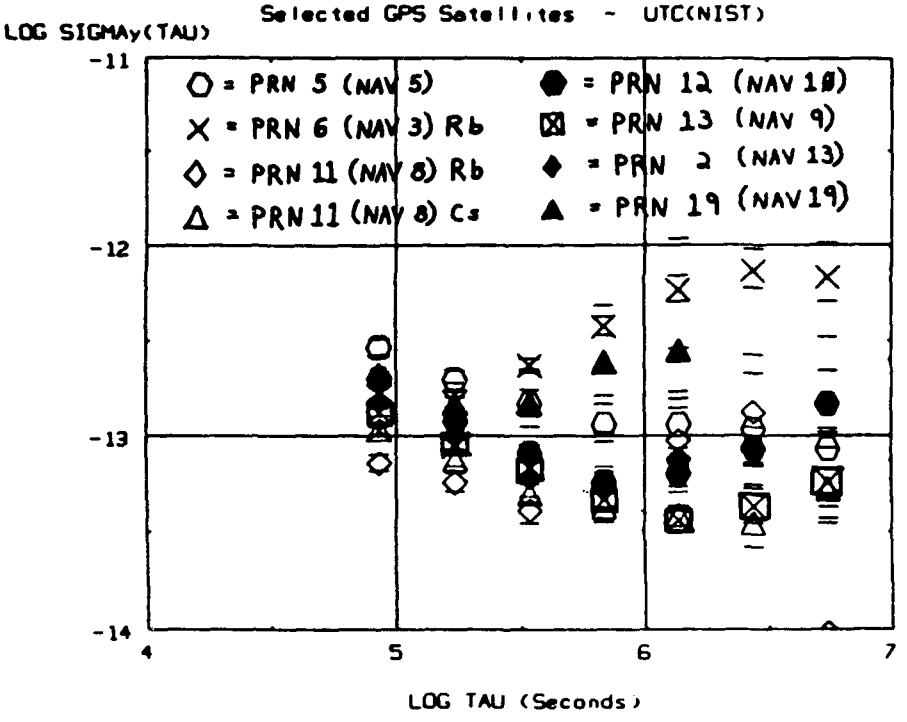


Figure 6. A plot of the fractional frequency stability, $\sigma_y(\tau)$, of most of the GPS satellite clocks with respect to UTC(NIST).

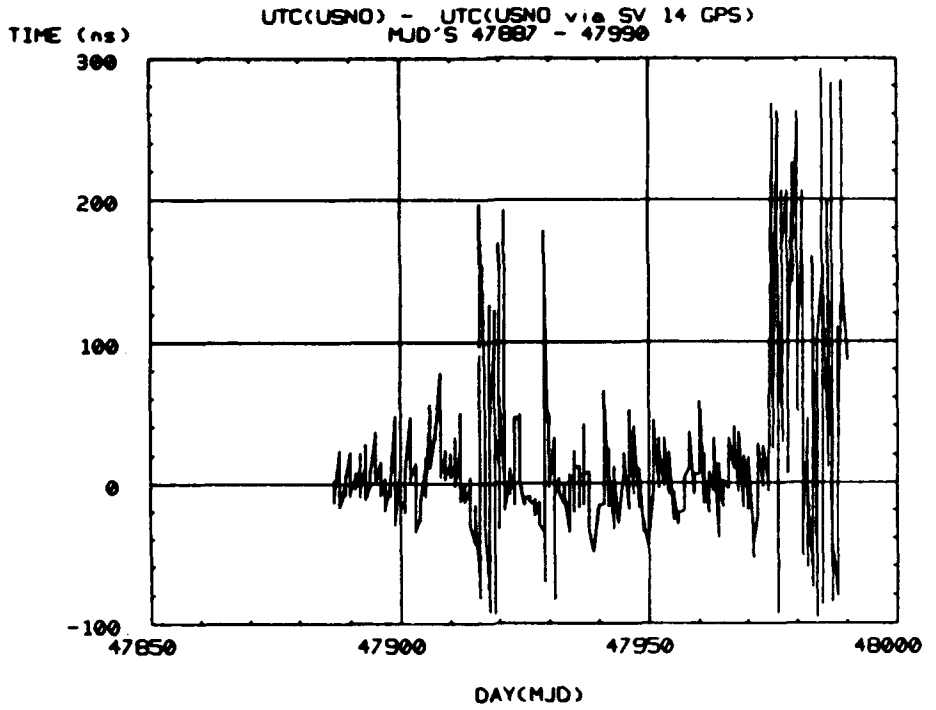


Figure 7a

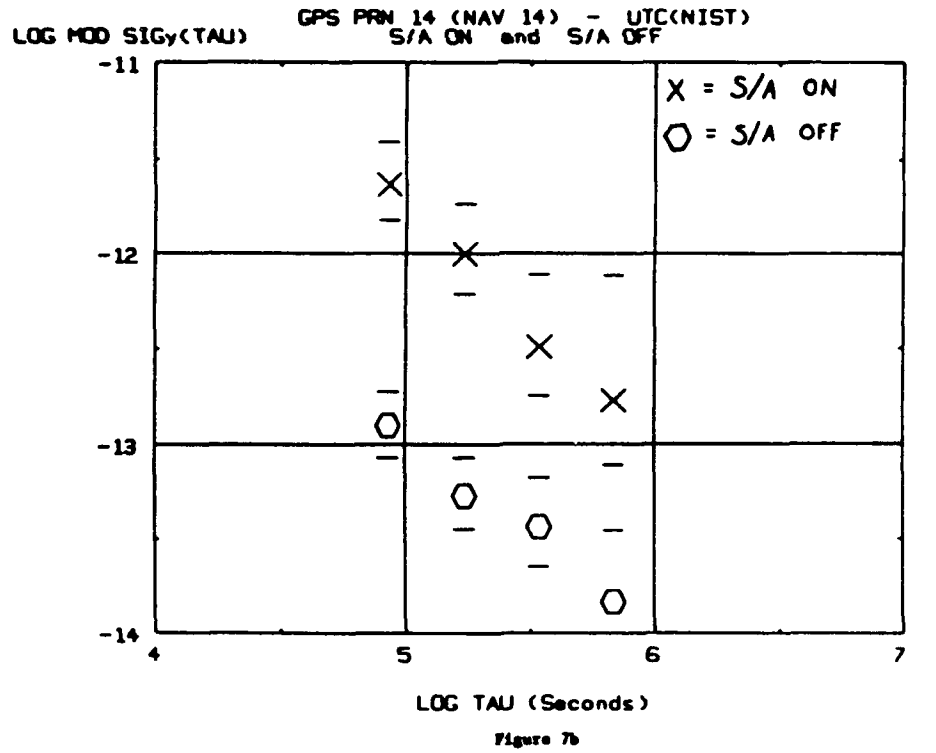


Figure 7b

Figure 7. A plot of the time (a) and the fractional frequency stability, $\sigma_y(\tau)$, (b) of GPS vehicle No. 14 showing the effects of selective availability (SA) as it is turned off during certain periods and on during others.

measurements were symultaneous at the two sites involved. The very small time instabilities observed would confirm this hypothesis. The common-view sites chosen were Boulder, Colorado to Washington D.C. and Boulder, Colorado to Paris, France.

TABLE I

COMPARISON OF MEASUREMENT NOISES

January 1, 1990 - March 31, 1990

UTC(OP) - UTC(NIST)

SVN	MEASUREMENT NOISE (ns)	% VALID
12	7.9	99
8	6.7	95
10	12.5	99
9	2.1	99
14	10.6	43

UTC(USNO MC) - UTC(NIST)

SVN	MEASUREMENT NOISE (ns)	% VALID
12	7.3	96
3	4.7	96
6	3.5	93
8	2.1	94
10	8.1	96
9	3.0	96
14	2.1	97

GLONASS SATELLITE CLOCK PERFORMANCE

The receivers used for the measurement of GLONASS satellite clocks were developed and built at the University of Leeds, and this work is described elsewhere [1]. The frequency band used by GLONASS is the same, nominally, as that of GPS, and the timing accuracies and stabilities will be limited by the same sets of phenomena as for GPS. As to differences, because the GLONASS orbits are slightly lower the geometry does not repeat from day to day but repeats every eight days.

For many of the measurements reported herein, it has been convenient to use the ao term as broadcast. This term is the time difference GLONASS system time minus GLONASS satellite clock time. Like GPS, GLONASS transmits correction terms which allow the user to obtain Moscow time, and Moscow time in turn is kept synchronous with UTC(SU). SU is official Standard Time for the USSR as kept by VNIIFTRI - their standards laboratory 40 km north of Moscow. VNIIFTRI is the nominal equivalent of NIST in the US. Over the last four years UTC(SU) has moved from a time difference with respect to UTC of about 30 μs to now about 10 μs.

Figure 8 is a plot of GLONASS system time versus UTC(USNO). We chose to analyze the quiet segment during the March-June, 1989, period to see how good the performance (in terms of frequency stability) might be. Figure 9 is a mod$\sigma_y(\tau)$ plot for this period. The reference clock at the USNO is a synthesized output from a hydrogen maser. We understand that the reference for the GLONASS control facility is also a hydrogen maser. The one-day instability is too large for the clocks involved. Also, the slope of the data plotted in Figure 9 can indicate the type of noise, and in this case it is nominally well modelled by flicker noise phase modulation (PM). Neither the amplitude nor the type indicates clock noise. Instead, the observations could be explained by short-term receiver instabilities or clock estimation noise. In the latter case, the source is likely to be GLONASS, because the estimation noise for GPS has been traditionally lower than this for typical time transfer receivers.

Figures 10a and 10b are plots of the frequencies from several of the GLONASS satellite clocks. These data are derived from the GLONASS system estimate of the time of the clock with respect to GLONASS system time. There are periods where there appear to be long term correlated frequency drifts between the satellite clocks. If the system clock were drifting, that would explain some of these segments. Or if the satellite clocks had correlated production dependencies which could effect the frequency drift, such a performance might be observed.

Figure 11 shows $\sigma_y(\tau)$ plots for several of the GLONASS satellite clocks. We observe a significant improvement in performance with time. Whereas the GPS data, outside of measurement noise, are direct measurements of the satellite clocks with respect to UTC(NBS/NIST), these GLONASS satellite clock data are estimates of the clocks' time with respect to GLONASS system time. The improvement observed with time for sample times, τ, of a few days could be due to an improvement in the estimates or an improvement in the clocks. If performance in this region of sample times is

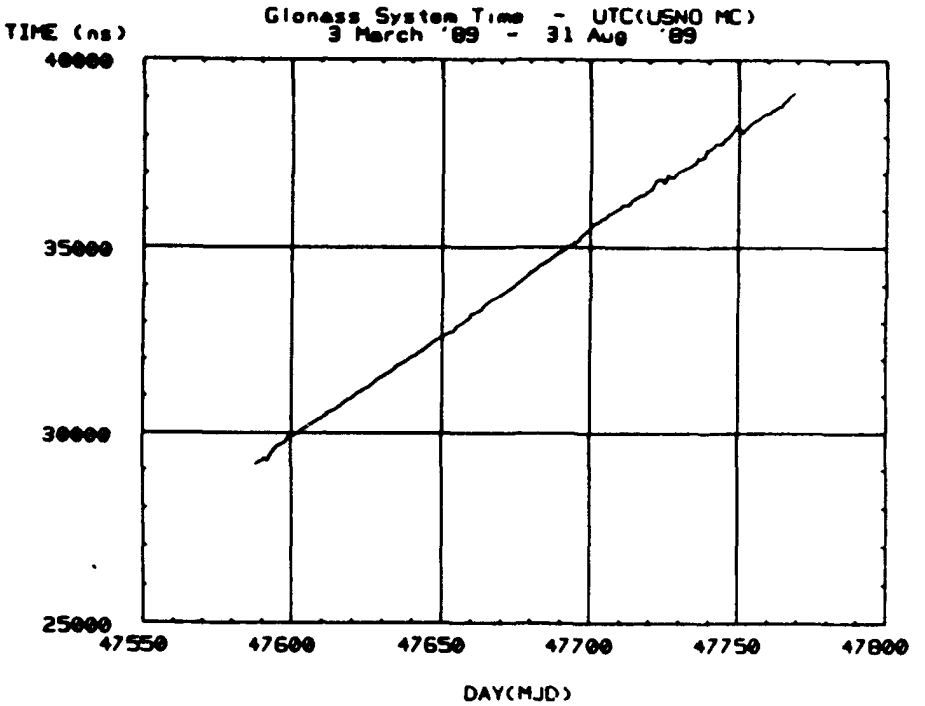


Figure 8. A plot of GLONASS system time as measured against UTC(USNO MC). These data were measured at Leeds deriving an estimate of UTC(USNO MC) with a GPS receiver.

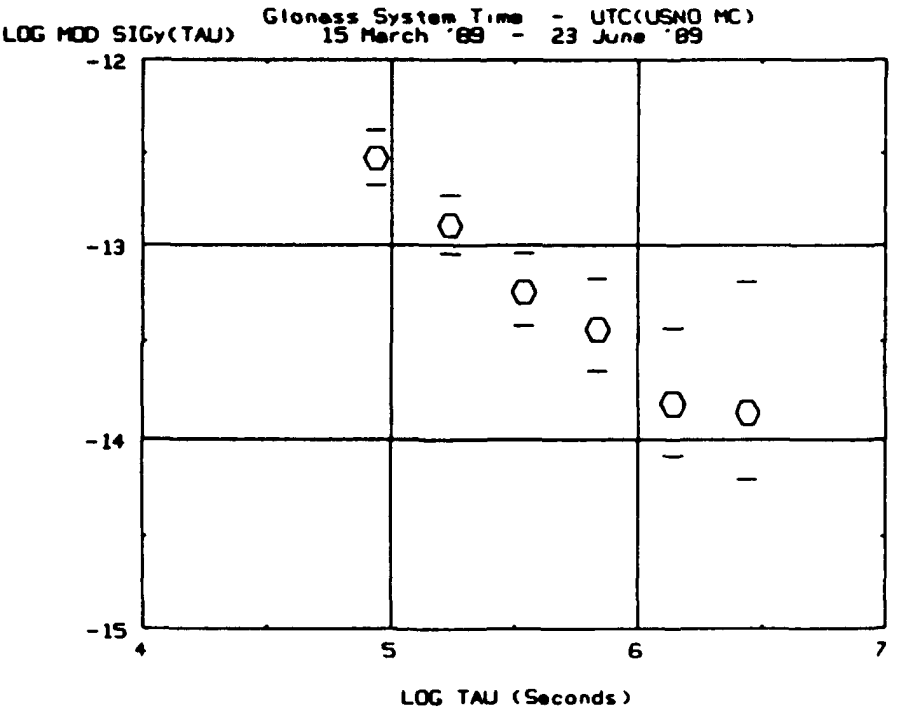


Figure 9. A plot of the fractional frequency stability, mod$\sigma_y(\tau)$, of a the smoothest part of the data plotted in Figure 8. The τ^{-1} behavior is modeled by flicker noise PM. This is not characteristic of the clocks involved.

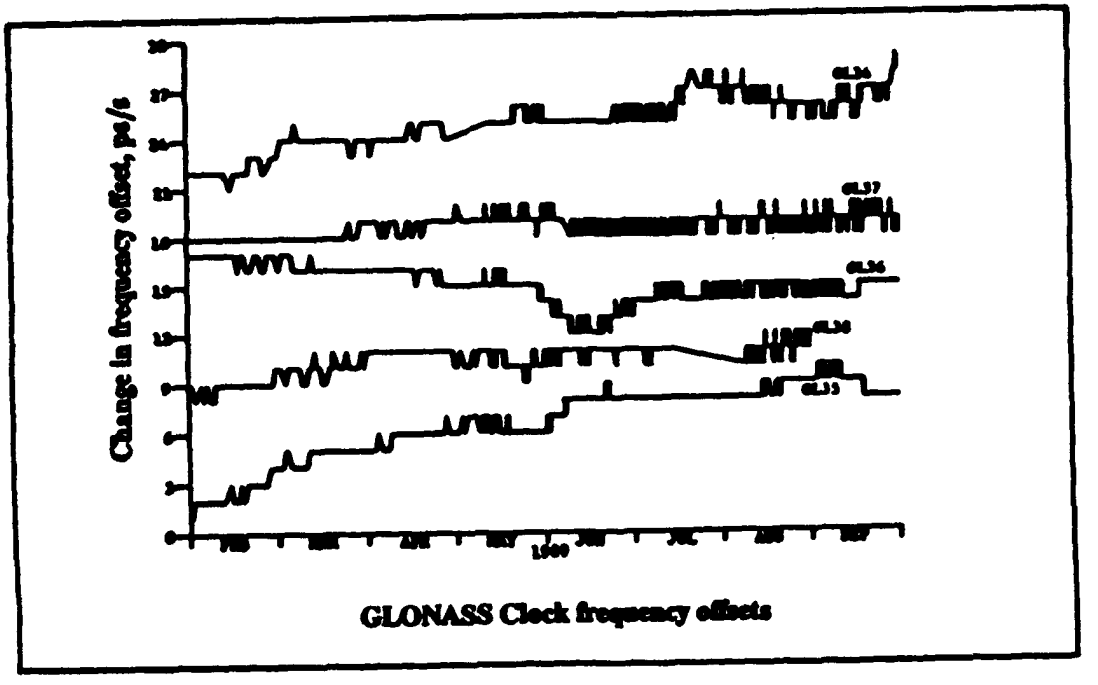


Figure 10a

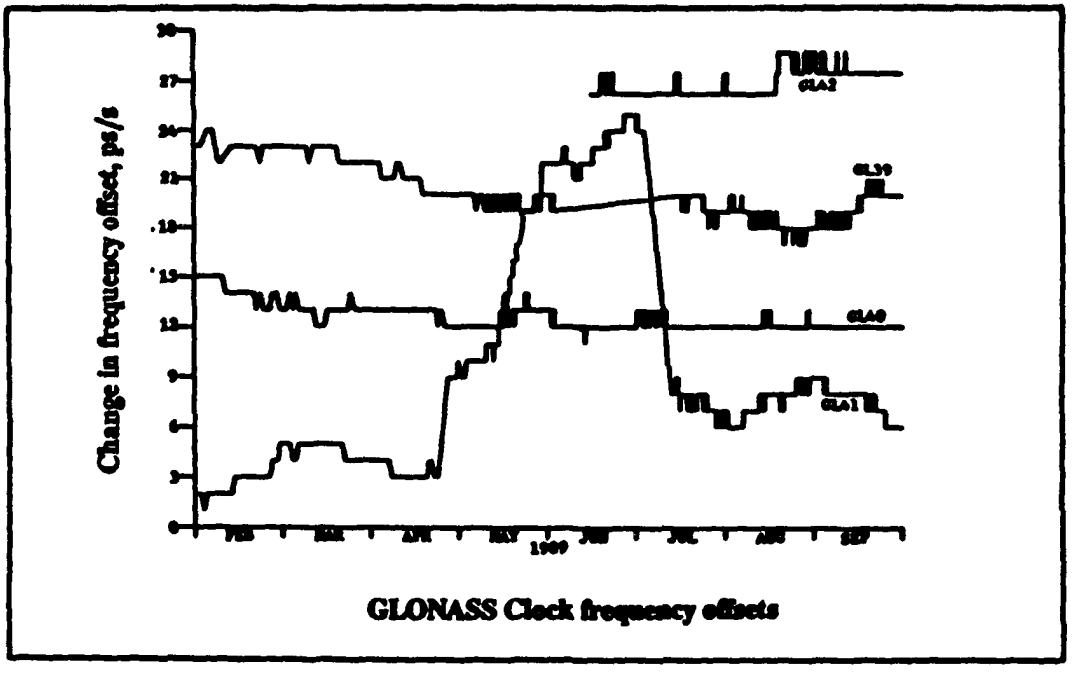


Figure 10b

Figures 10a and 10b. A plot of the normalized frequencies in parts in 10^{12} of most of the GLONASS satellite clocks.

due to the estimate, it is likely that the slope on the $\sigma_y(\tau)$ plot would be steeper than $\tau^{-1/2}$.

The longer-term stabilities shown are undoubtedly due to the clocks involved - the clocks being either the reference (probably a hydrogen maser clock) or the satellite clocks. The long-term frequency drifts measured are fairly characteristic of cesium-beam frequency standards.

COMPARISONS OF GPS AND GLONASS CLOCK PERFORMANCE

The data bases, as they were available to us for this paper, do not allow a direct and fair comparison, since different reference standards and different measurement systems were used. If the reader takes into account the conditions stated in the previous sections, the following comparisons yield some interesting perspectives.

By way of review, the reference for the GPS data was UTC(NBS/NIST). This reference has a one-day stability of about 1.5×10^{-14}, and a flicker floor of less than 1×10^{-14} and a frequency drift of about 3×10^{-16}/d or less. Flicker floor means the best stability achieved in a $\sigma_y(\tau)$ plot. For the GPS and GLONASS satellite clocks the τ values for the flicker floor were a few weeks. UTC(USNO) has comparable performance to UTC(NBS/NIST). Other than the data shown in Figures 8 and 9, we have no other direct comparisons of the GLONASS system time to any other UTC time scale.

The frequency accuracies of UTC(NBS/NIST) and UTC(USNO) are a few parts in 10^{14}. The frequency offset of GLONASS system time is about 6×10^{-13} for the period shown. Figure 12 is a plot of the time accuracy of the various systems. UTC(SU) is official USSR time as determined by their primary standards laboratory VNIIFTRI near Moscow. What is plotted for UTC(SU) is obtained from Loran-C measurements as published by the BIPM.

If the random deviations of the frequency of a clock have a white spectrum, then a linear regression to the frequency is the optimum estimate for the frequency drift. If the random deviations are random walk (if they have an f^{-2} spectrum), then the mean finite second difference is the optimum estimator for the frequency drift. Since in practice we may have combinations of these processes, special algorithms for drift estimation may be necessary. [4]

Table 2 gives a comparison of the one day stabilities, the flicker floors and the frequency drifts for the GPS and GLONASS clocks. Please note again that these are not direct comparisons because of the different reference standards and measurement systems. The flicker floors and frequency drifts should be fairly representative numbers. Notice the dramatic improvement in the performance of the GLONASS clocks from the earlier to the more recent satellites.

TABLE II

GPS Satellite PRN/NAV	1 Day Stability $*10^{14}$	Flicker Floor $*10^{14}$	Drift/Day $*10^{15}$
5/5	29.4	11.6	1.6
6/3 (Rb)	12.3	50.0	-6.0
11/8 (Rb)	6.5	2.8	-168.0
11/8 (Cs)	8.8	3.5	-0.8
12/10	11.1	3.5	-0.8
13/9	13.9	4.0	-0.3
2/13	15.0	5.5	-4.2
14/14	12.5	< 3	-0.5
16/16 (Rb)	26.4	12.5	-330.0
19/19	21.2	15.1	22.0

Glonass Satellite	1 Day Stability $*10^{14}$	Flicker Floor $*10^{14}$	Drift/Day $*10^{15}$
G16	119.0	100.0	-8.1
G23	114.0	89.0	-1.8
G28	91.1	78.0	-4.1
G36	38.9	31.0	-4.7
G37	20.4	18.0	-6.1
G40	28.7	24.0	+14.1
G42	28.0	< 4	-0.2

CONCLUSIONS

Our procedure for determining the performance of both GPS and GLONASS on-board clocks is identical - we examine the time series of daily phase and frequency offsets transmitted by the spacecraft themselves. In the case of ground-based system references, our GLONASS reference is a 1 pps reference itself locked to within 100 ns of

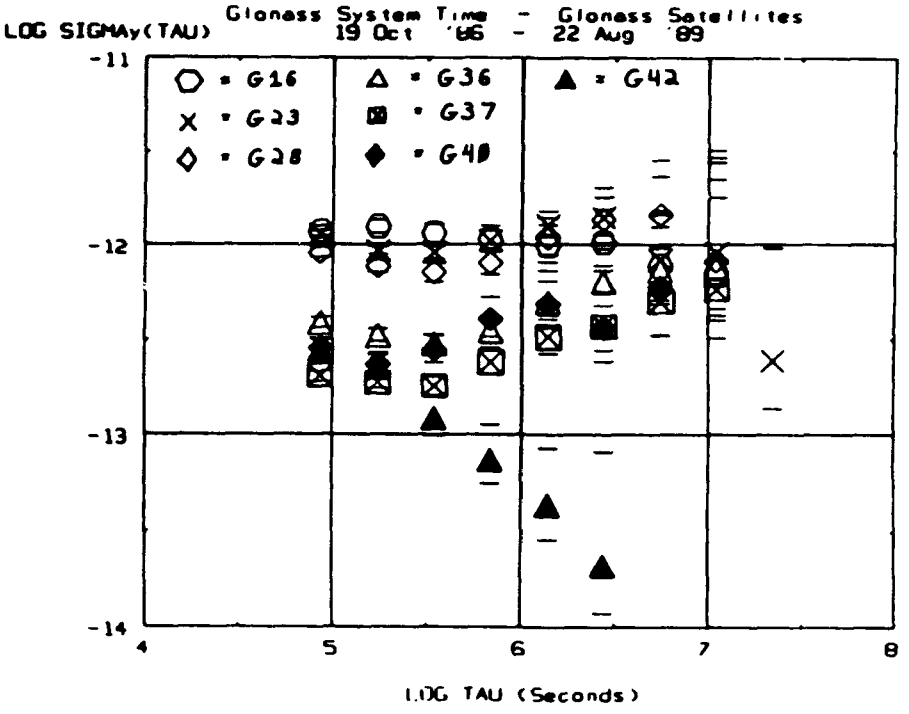


Figure 11. A plot of the fractional frequency stability, $\sigma_y(\tau)$, of several of the GLONASS satellite clocks.

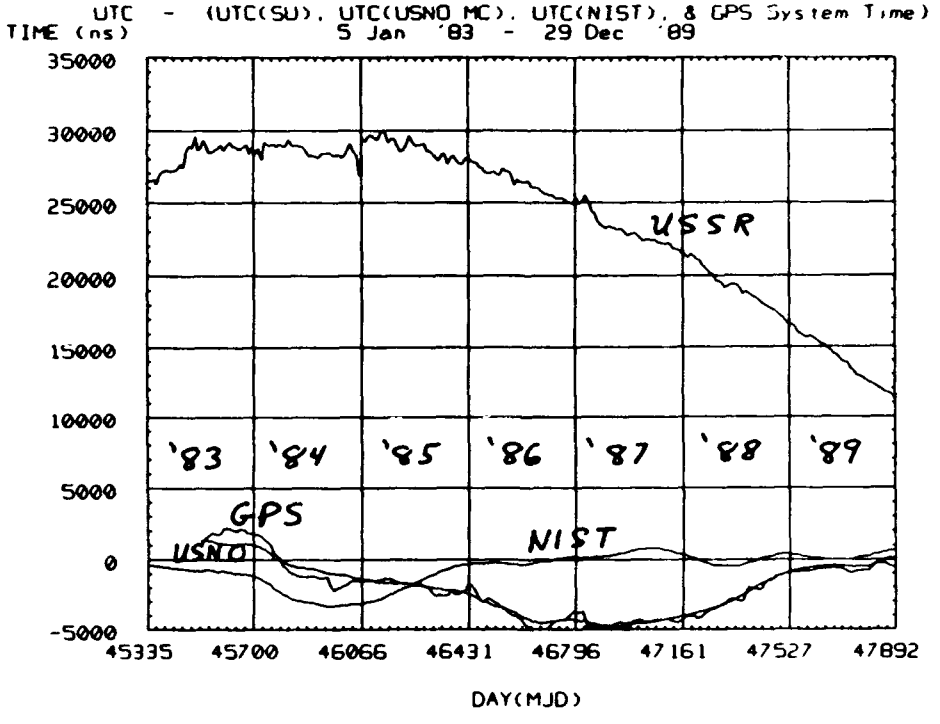


Figure 12. A plot of the principal reference time scales of the USSR and the USA that were available to the authors over the last six years with respect to UTC. All of them appear to be improving in performance with time.

UTC(USNO). As far as on-board clocks are concerned, our analysis can be compared with the known performance of individual GPS clocks. In the case of GLONASS it is not known a priori what types of clocks are carried and in this sense our analysis of GPS can be seen as a kind of precalibration.

The time series analysis which results in $\sigma_y(\tau)$ versus τ plots shows the individual characteristics of each satellite clock. Rubidium and cesium clocks are carried by the different GPS satellites. It also leads us to discuss the performance and types of on-board standards carried by GLONASS. In general terms, the time series analysis shows a consistently higher performance for GPS clocks over GLONASS clocks. This statement is taken to refer to satellites active during a period of several years. However, while the GPS clock performance has been consistently high, the GLONASS clocks started from a mediocre performance and have improved steadily with time so that some of the recently launched GLONASS clocks are comparable to GPS clocks. This conclusion is reinforced by the data presented in Table 2.

GPS satellite clocks and their more recent GLONASS counterpart clocks meet specification for global navigation satellites and are capable of transferring time with day-to-day time stabilities of 10 ns or better and long-term frequency stabilities of a few parts in 10^{14}.

Our analysis of measurements of GLONASS system time versus UTC(USNO) give evidence in the short term (1 day) of estimation or receiver errors not properly understood. In the long term (>8 days) we see clear evidence of the high-quality of both reference clocks. The UTC(USNO) reference is known to be a hydrogen maser and it seems more than likely, based on the evidence presented, that the GLONASS system clock is also a hydrogen maser at least during certain periods of time.

REFERENCES

1) S. A. Dale, I. D. Kitching, and P. Daly: "Position-fixing using the USSR's GLONASS C/A code," IEEE Plans '88, Las Vegas, pp. 13-20.

2) D.W. Allan and Lin Ping Ping, Estimating the GPS Ephemeris Plus Propagation Errors and their Effect on Time and Frequency Transfer, submitted to the 41st Annual Symposium on Frequency Control, 1987.

3) M.A. Weiss and D.W. Allan, An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data, IEEE Trans. Instrum. & Meas., IM-36, 572-578 (1987).

4) D.W. Allan, Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 647-654, 1987.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A NOVEL APPROACH TO THE SIMULTANEOUS MEASUREMENT OF PHASE AND AMPLITUDE NOISE OF OSCILLATORS

KIN-WA WAN, JOHN AUSTIN & ENRIC VILAR

Microwave Telecommunications & Signal Processing Research Group, School of Systems Engineering, Portsmouth Polytechnic, Anglesea Building, Anglesea Road, Portsmouth, PO1 3DJ, England.

Abstract:

The paper is concerned with one aspect of our recent development carried out at Portsmouth Polytechnic (UK) regarding a novel approach for the simultaneous measurement of amplitude and phase nose in oscillators. A description of the measurement system and the associated processing techniques is presented.

The measurement system includes a conventional complex (I & Q) demodulator, a high speed sampler and a digital computer. The techniques have been named time-domain phase unwrapping and real amplitude de-enveloping. They permit the reconstruction of the demodulated I and Q signals as a vector rotating at a convenient beat frequency randomised, in phase and magnitude, by noise. The problems encountered in conventional AM and PM detection can be overcome using the afore-mentioned approaches. As a result, direct and true amplitude and phase noise can be measured and characterised.

Keywords: Phase Noise, Amplitude Noise, Phase Unwrapping, Amplitude De-enveloping, Vector Demodulation

1 Introduction

The use of phase detectors, together with feedback control loops, for measuring phase noise in oscillators is well established. The dynamic range and linearity of the phase detector, the sensitivity of the control loop, as well as other aspects, are the main problems associated with this approach. Even though developments in advanced digital signal processing (DSP) techniques offer alternative ways to carry out phase noise measurements, the importance and capability of DSP techniques in the measurement and analysis of noise in oscillators are still not well recognised. This paper attempts to overcome this and an approach is presented combining advanced instrumentation computer technology with fast digital signal processing techniques. A brief introduction for phase-only measurements was given in [3] with greater details given in [4].

The measurement system includes a conventional in-phase and quadrature (I-Q) demodulator, a dual-channel high speed sampler (12 bits) and a digital computer. The demodulated signals are converted to digital form and then analysed further in the computer. The block diagram of the measurement system is shown in figure 1. The digital signal processing techniques which follow after ***complex demodulation*** have been called ***time-domain phase unwrapping*** (**TDPU**) and ***real amplitude de-enveloping*** (**RADE**). They permit the simultaneous measurement of the phase and amplitude noise of the oscillator under test provided the sampling speed is sufficiently high.

In AM measurements, the clipping problem which occurs in the conventional AM envelop detectors can be eliminated using the present system. We note that the response of a conventional envelop detector can be very complex and that it is non-linear. As a result, the response to signal plus noise is not simply the sum of the separate responses to signal and to noise. For phase noise detection, the dynamic range limitation of $\pm 90^{\circ}$ or $\pm 180^{\circ}$ found in a conventional phase detector can be overcome using the time-domain phase unwrapping technique applied to reconstruct the random phase from the measured I and Q signals.

It is important to emphasise that the LO signal of the I-Q demodulator is not phase or frequency locked to the RF signal under test, and the LO frequency is only adjusted to be close to the RF frequency for convenience. In such a phase or frequency unlocked measurement system the first advantage is that the limited time response of a phase or frequency lock loop system is avoided. Secondly, the complex transfer function of the phase/frequency lock loop which exhibits different characteristics inside and outside the loop bandwidth, need not be taken into account in the spectral analysis of oscillator noise. Sometimes, the decision to use either a tight or a loose phase/frequency lock loop system is not easy. However, the measurement bandwidth is only related to the sampling frequency frequency f_s. The selection of f_s depends only upon how far from or close to the carrier the characterisation of the noise present is taken provided aliasing is not introduced.

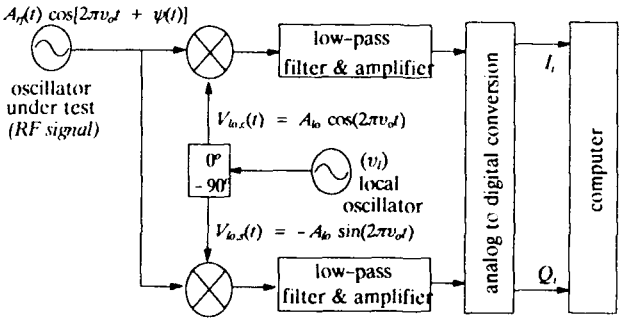


Figure 1: The block diagram of I-Q measurement system.

CH2818-3/90/0000-140 $1.00 © 1990 IEEE

2 Theoretical Background

2.1. Phase Noise in RF Signals

Let us consider the simple model of an RF signal

$$V_{rf}(t) = A_{rf}(t)\cos[2\pi v_o t + \psi(t)] \quad , \qquad (1)$$

where v_o is the nominal (carrier) frequency, and $\psi(t)$ is the phase noise associated with the random frequency fluctuation. $A_{rf}(t)$ is assumed to be constant, that is, $A_{rf}(t) = A_{rf}$. If this signal is demodulated as shown in figure 1 using I and Q amplitude and phase noise-free local oscillating signals $V_{lo,c}(t) = A_{lo}\cos(2\pi v_o t)$ and $V_{lo,s}(t) = -A_{lo}\sin(2\pi v_o t)$ (followed by low-pass filtering) then, the ***sampled*** outputs of the filters, I_i and Q_i, at the time t_i will be

$$I_i = A_{rf}\, a_i \cos(\psi_i) \qquad (2)$$

$$Q_i = A_{rf}\, a_i \sin(\psi_i) \qquad (3)$$

where ψ_i is the sample of the random phase noise $\psi(t)$ at the time t_i ($i \geq 1$). ψ_0 is assumed to be zero with no loss of generality. a_i is the factor associated with the conversion loss in the mixing process, and the attenuation/gain in the filtering/amplifying process. It is noted in here that the mixers must be hard driven, that is, the LO power is at least 10dBm greater than the RF power. In the hard driven mode, the mixer diodes will be fully turned on and off for lowest distortion during the mixing process. Now, let us define the complex voltage output V_i at the time t_i as

$$V_i = I_i + j\, Q_i \quad , \qquad (4)$$

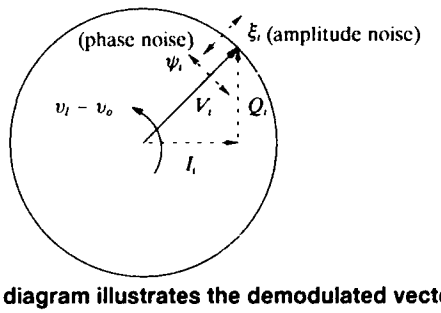


Figure 2: The diagram illustrates the demodulated vector.

as shown in figure 2, where $j^2 = -1$. The demodulated I and Q signals represent a vector $(I_i\ ,\ Q_i)$ ***randomised*** with ψ_i in ***phase*** at the ith time interval. For convenience, let us assume for the moment that the LO frequency, v_l, is equal to the RF frequency, v_o, that is, $v_l = v_o$ in figure 2. The product $V_i V_{i-1}^*$ is

$$\begin{aligned} V_i\, V_{i-1}^* &= (I_i + j\, Q_i)\,(I_{i-1} - j\, Q_{i-1}) \\ &= A_{rf} a_i e^{j\psi_i}\, A_{rf} a_i e^{-j\psi_{i-1}} \\ &= \left(A_{rf}\, a_i\right)^2 e^{j\Delta\psi_i} \quad , \end{aligned} \qquad (5)$$

where $*$ denotes the complex conjugate. Therefore, one can find $\Delta\psi_i$ from the following equations:

$$\log_e(V_i V_{i-1}^*) = 2\log_e(A_{rf} a_i) + j\Delta\psi_i \qquad \text{that is} \qquad (6)$$

$$\Delta\psi_i = Imag\left[\log_e(V_i V_{i-1}^*)\right] \qquad (7)$$

where $Imag$ represents the imaginary part of the expression. ***As long as the sampling frequency is fast enough, and the in-phase I_i and quadrature Q_i outputs are band-limited, there is no phase increment from i-1th to ith time interval greater than π***. Thus, if each incremental phase change $|\ \Delta\psi_i\ |$ is less than π, we can find the angle of the vector $V_i V_{i-1}^*$ by algebraical manipulation of the successive samples I_{i-1}, Q_{i-1}, I_i, Q_i and obtain the sampled random phase noise process ψ_i from the incremental results $\Delta\psi_i$ ***without any 2π ambiguity in the reconstructed random phase noise curve***. The phase reconstruction process is called ***time-domain phase unwrapping*** (**TDPU**). The algorithm is shown in figure 3.

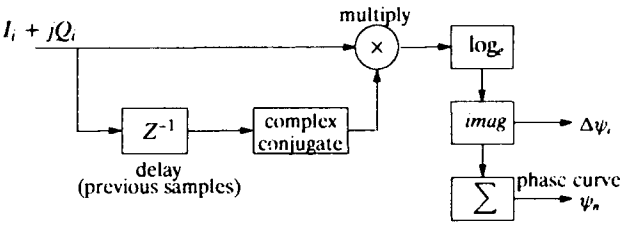


Figure 3: The signal processing diagram of the time-domain phase unwrapping process.

Furthermore, the averaged frequency fluctuations $\Delta\overline{v}_i$ over the sampling period Δt can be found by

$$\Delta\overline{v}_i = \frac{\Delta\psi_i}{2\pi\Delta t} \quad . \qquad (8)$$

Because of (8) and the condition $|\ \Delta\psi_i\ | < \pi$, the absolute value of the fractional frequency deviations $\Delta\overline{y}_i$ (or $\frac{\Delta\overline{v}_i}{v_o}$) must be smaller than $\frac{1}{2\Delta t v_o}$ in order to unwrap the measured random phase with no ambiguity. One can also think in terms of $\Delta\overline{v}_i$ as being equivalent to the frequency deviations measured by a frequency counter with a gate-time $\tau = T$ and dead-time $\tau_d = 0$. Thus, time-domain frequency analyses of the measured signal, like using the Allan variance [1] or the Hadamard variance [2], can be carried out.

So far it has been assumed for convenience that the LO frequency, v_l, is equal to the RF frequency, v_o. However, it is very important to note in here that the local oscillator frequency is only adjusted to be close to the frequency of the oscillator under test; ***it is not necessary to lock*** the LO and RF signals to the same frequency. If the LO and RF frequencies are not equal, the demodulated signals will contain a "beat" signal of frequency $|v_l - v_o|$. Therefore $V_i V_{i-1}^*$, (5), becomes

$$V_i V_{i-1}^* = (I_i + j Q_i)(I_{i-1} - j Q_{i-1}) \quad (9)$$

$$= A_{rf}\mu_i e^{j[2\pi(v_l - v_o)t_i + \psi_i]} A_{rf}\mu_i e^{-j[2\pi(v_l - v_o)t_{i-1} + \psi_{i-1}]}$$

The incremental phase change, $\Delta\Phi_i$, from the $i-1$th to ith time interval can be written as

$$\Delta\Phi_i = 2\pi(v_l - v_o)(t_i - t_{i-1}) + |\psi(t_i) - \psi(t_{i-1})|$$

$$= 2\pi(v_l - v_o)\Delta t + \Delta\psi_i \quad (10)$$

where Δt is the sampling period. Therefore, the unwrapped phase will contain the random phase $\psi(t)$ superimposed to a linear phase trend represented by $2\pi(v_l - v_o)t$. In the other words, ***the demodulated vector*** (I_i, Q_i) ***rotates at the frequency*** $(v_l - v_o)$, ***and is also randomised with*** ψ_i ***in phase at the*** ith ***time interval***. In this case, the unwrapped phase is given by $\Phi(t)$ and expressed as

$$\Phi(t) = 2\pi(v_l - v_o)t + \psi(t) \quad (11)$$

which is illustrated in figure 4. One should notice from (11) and figure 4 that the "beat" frequency $(v_l - v_o)$ can be found by estimating the slope of a straight line fitting best $\Phi(t)$.

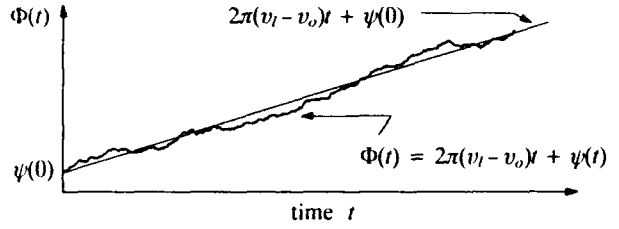


Figure 4: The illustration of the unwrapped random phase $\psi(t)$ superimposed with the linear phase $2\pi(v_l - v_o)t$.

2.2. RF Amplitude Noise

In section 2.1., we have assumed that the AM noise of the RF signal was negligible, which is usually the case for high quality oscillators. However, the measurements of amplitude noise, $\xi(t)$, of the RF signal can also be carried out using the I-Q analysis system shown in figure 1. For this we re-write (1) emphasizing the (random) amplitude component $\xi(t)$ as follows:

$$V_{rf}(t) = A_{rf}(t)\cos[2\pi v_o t + \psi(t)]$$

$$= [A_{rf} + \xi(t)]\cos[2\pi v_o t + \psi(t)] \quad (12)$$

and with the assumption of no amplitude noise in the LO signals, the I_i and Q_i components will be expressed as

$$I_i = [A_{rf} + \xi_i]\, a_i \cos(\psi_i) \quad (13)$$

$$Q_i = [A_{rf} + \xi_i]\, a_i \sin(\psi_i) \quad (14)$$

which complement (2) and (3). ξ_i is the discrete form of the amplitude noise of the RF signal at the ith time interval. In the other words, ***the demodulated vector*** (I_i, Q_i) ***is randomised with*** ξ_i ***in magnitude***, and (5) of section 2.1. becomes

$$V_i V_{i-1}^* = a_i^2 (A_{rf} + \xi_i)(A_{rf} + \xi_{i-1}) e^{j\Delta\psi_i} \quad (15)$$

which shows that no error is introduced by the RF amplitude noise ξ_i in the estimation of $\Delta\psi_i$. Since $A_{rf} \gg \xi(t)$, we can also assume that the conversion loss/gain a_i is fairly constant all the time.

Furthermore, from the results of the I_i and Q_i components, we can calculate ξ_i/A_{rf} by the following equation:

$$\frac{\xi_i}{A_{rf}} = \frac{1}{a_i}\sqrt{\left(\frac{I_i}{A_{rf}}\right)^2 + \left(\frac{Q_i}{A_{rf}}\right)^2} - 1$$

$$= \frac{1}{a_i A_{rf}}\sqrt{(I_i)^2 + (Q_i)^2} - 1 \quad (16)$$

as $\cos^2(\,) + \sin^2(\,) = 1$. The vector (I_i, Q_i) of figure 2 is supposed to exhibit constant magnitude. Any deviation of the magnitude of the vector (I_i, Q_i) from $A_{rf}\, a_i$ can be considered as the noise contributed by $\xi(t)$. This approach applied to the measurement of amplitude noise present in oscillators, as demonstrated in (16), has been named ***real amplitude de-enveloping*** **(RADE)**. The corresponding algorithm is illustrated in figure 5. With the assumption of zero mean of ξ_i, one can obtain

$$< (A_{rf} + \xi_i)\, a_i > = < \sqrt{(I_i^2 + Q_i^2)} >$$

$$= A_{rf}\, a_i \quad (17)$$

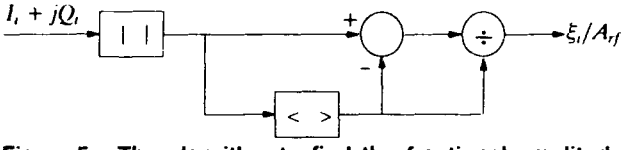


Figure 5: The algorithm to find the fractional amplitude noise.

where $<\ >$ denotes ensemble average. Therefore, by means of (16) and (17), one can calculate the power spectrum of the ***fractional amplitude deviation***. Calling $z(t) = \frac{\xi(t)}{A_{rf}}$, $S_z(f)$ permits the comparison of the amplitude noise of various signals. The additional advantage of studying the amplitude noise of a signal in terms of z is that the accuracy of the noise estimation does not rely on the calibration result of the conversion loss/gain a_i but on estimating $A_{rf}\, a_i$ in (17). An accurate value for $A_{rf}\, a_i$ can be easily obtained provided the measured data set of (I_i, Q_i) is sufficiently large.

Finally, $\frac{1}{2}A_{rf}^2$ represents the signal power, $10\log_{10} S_z(f)$ represents the single side based amplitude noise power ξ^2 relative to the carrier power in, again, dBc/Hz and one can use the

symbolic representation

$$L_c(f)\ (\text{dBc/Hz}) = 10\log_{10}[S_x(f)] - 3 \qquad (18)$$

which complements the conventional

$$L(f)\ (\text{dBc/Hz}) = 10\log_{10}[S_\psi(f)\ (\text{rad}^2/\text{Hz})] - 3 \quad . \qquad (19)$$

3 Experimental Results and Discussion

A photograph of the arrangement for the amplitude/phase noise measurements is shown in figure 6. The test oscillator consisted of a synthesizer AM or FM modulated by Gaussian white noise degrading the spectral purity and introducing amplitude or phase noise in a a controlled manner. In the "AM" and "FM" experiments, the carrier frequency and the amplitude of the RF signal from the signal synthesizer were set to 2.5MH and 4.5dBm respectively. The "rms amplitude open circuit" and the Gaussian noise bandwidth of the noise generator were 0.1volt and 150Hz respectively. A cut off frequency of 300Hz was chosen in the low-pass filter unit.

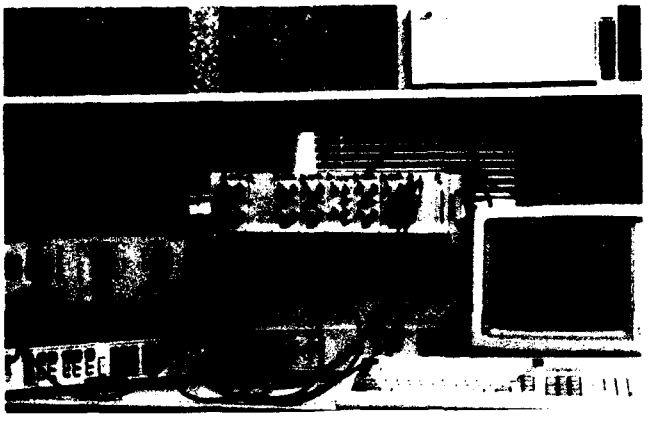


Figure 6: Photograph of the experimental layout for phase and amplitude noise measurements.

For the "AM" experiments, 10%, 30% and 50% AM depths in the signal generator were selected. The sampling frequency f_s was 2kHz. The results of the analyses offered by the **RADE** and then the conventional Fourier transform are shown in figure 7. We note from the spectra of figure 7 that the white AM noise is present in the RF signal under test as expected, and the "beat" frequency is about 7.8125Hz. On the other hands, the "beat" frequency is also found to be -7.3517Hz using the **TDPU**. The negative sign indicate that the carrier frequency of the LO signal is smaller than that of the RF. A difference of less than about 0.4Hz (7.8125Hz - | - 7.3517| Hz) between the frequencies indicated by the **TDPU** and the Fourier transform technique was found.

In the "FM" experiments, the FM deviation set up in the synthesizer was 0.1kHz/volt. The sampling frequency was 80kHz. The result of the analysis offered by the **TDPU** followed by the Fourier transform is shown in figure 8. Since the RF signal is frequency modulated by white noise, the phase noise spectrum of figure 8 has about -20dB/decade roll-off within the bandwidth set in the noise generator. The noise floor is about -123dB/Hz, with f_s = 80kHz .

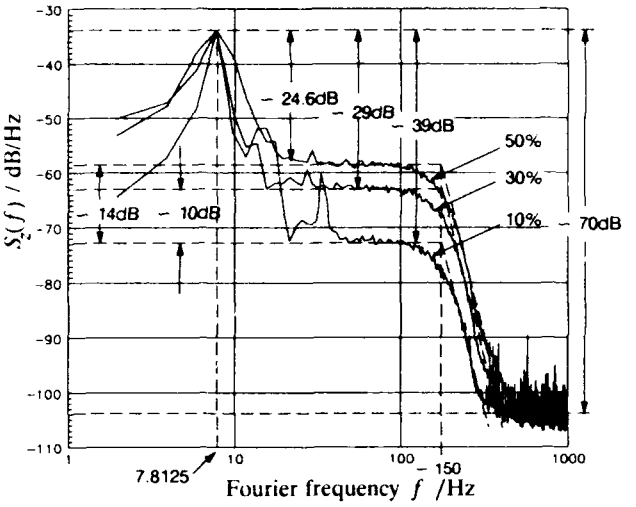


Figure 7: Power spectra of the measured fractional amplitude noise. The AM modulation indexes selected in the signal generator were 10%, 30% and 50%.

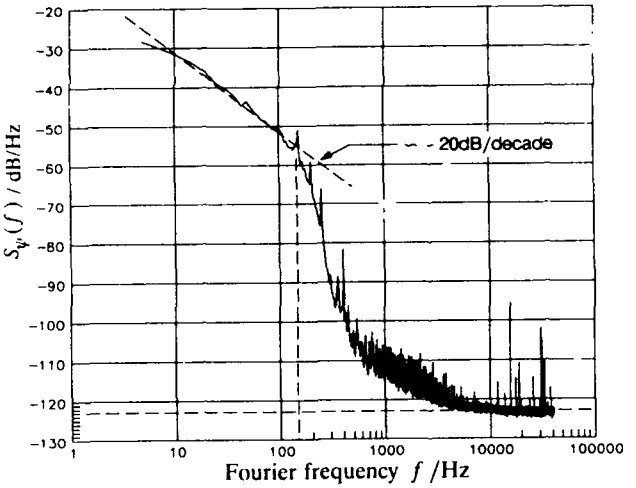


Figure 8: The measured phase noise spectrum under the second test condition. The sampling frequency f_s is 80kHz.

Finally, in order to investigate the dynamic range limitations of the present measurement system associated with the sampling frequency f_s, experiments and computer simulations have been carried out. This provides a straight forward method of determining the sensitivity of the system when $\Delta\psi_i$ is calculated by means of $\log_e(V_i V_{i-1}^*)$. The results of the investigation are shown in figure 9. One concludes from figure 9 that as f_s increases, the spectral density of the "noise floor" or

quantization noise, $S_q(f)$, is reduced by the appropriate factor $f_s/2$. There is less than 5dB difference between simulations and experimental results. The difference is due to the noise introduced by filters and amplifiers in the experimental measurements but not in the computer simulations. The ***relative reduction*** as f_s increases is exactly as predicted by assuming the quantization noise power "evenly" distributed over the bandwidth $f_s/2$.

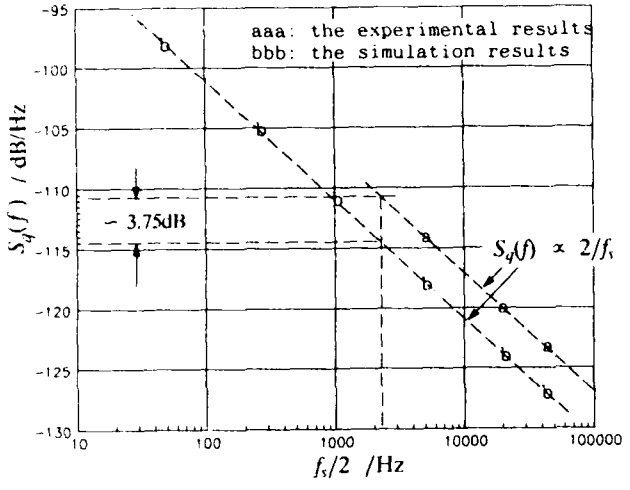


Figure 9: The experimental and simulation results of the noise floor as a function of the sampling bandwidth $f_s/2$.

4 Summary and Conclusion

In this paper, after an overview of the current measurement system, the concepts of the techniques named **TDPU-RADE** for the reconstruction of random phase and amplitude noise processes have been presented. For the measurements, the RF signal under test is demodulated by the in-phase and quadrature LO signals. That is, complex demodulation is involved. It is very important to note that the LO signals are not locked to the RF signal. It is only necessary that the I and Q LO signals themselves maintain an accurate 90° phase relationship over the measurement bandwidth. The demodulated I and Q signals can be considered as a vector rotating at the frequency $(v_t - v_o)$, which is randomised with ψ_t in phase and with ξ_t in magnitude. I and Q components are digitisied and then analysed by means of the **TDPU-RADE** techniques followed by conventional Fourier analysis.

Only the concept of the reconstruction of a linear phase superimposed to a random phase has been demonstrated in this paper, and the frequency drift of the signal under test has not been considered. The authors believed however that time-domain phase unwrapping can handle the drift situation using a second order polynomial fitting to the reconstructed phase signal as long as the drift is not too large or too fast and the sampling frequency is high enough. As a result, the rate of the frequency drift can be determined. Furthermore, a periodic signal present in the phase of the RF signal under test can be identified by fitting a sinusoidal curve to the unwrapped random phase or from the result of the corresponding spectral analysis.

Finally, the experimental and simulation results of figure 9 confirm that the dynamic range of the measurement system is sampling frequency (f_s) dependent, that is, $S_q(f) \propto 2/f_s$, even though vector multiplication and logarithm operations are involved in the calculation of ψ_t.

Acknowledgement

The financial support of the UK National Advisory Body which has sponsored this research programme is gratefully acknowledged.

References

[1] D.W. Allan and F.L. Walls, "Measurements of Frequency Stability," in *Proceedings of the IEEE, Vol. 74, No. 1, PP. 162-168*, Jan. 1986.

[2] R.A. Baugh, "Frequency Modulation Analysis with the Hadamard Variance," in *Proc. 25th Ann. Freq. Control Symp., Atlantic City NJ, PP. 222-225*, April 1971.

[3] K.W. Wan, J. Austin and E. Vilar, "A Novel Approach to the Simultaneous Measurement of Phase and Amplitude Noises in Oscillator," in *the 19th European Microwave Conference Proceedings, PP. 809-813*, 4-7 Sept. 1989.

[4] K.W. Wan, "Advanced Numerical and Digital Techniques in Frequency Stability Analysis," in *Ph.D Thesis, Microwave Telecommunications & Signal Processing Research Group, Portsmouth Polytechnic, England*, 1990.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

TIME & FREQUENCY SYSTEM FOR SATELLITES GROUND STATIONS

E. Peled, M. Zelitzki, A. Nemesh,

and A. Stern

TFL Time & Frequency Ltd. P.O.B 1792 Holon 58117
Israel

Abstract

The paper describes a high redundant Time & Frequency generation and distribution system which has been developed at TFL for Satellites Ground Station (or Stratum 1 for digital telecommunication). The system features extremely high reliability (millions of hours MTBF), where a majority vote occurs among 3 frequency sources which are phase aligned and from which a Master is selected under microprocessor control. No dead time, amplitude or phase jump, occurs during a master change over. The selected Master is used as a source for a distribution amplifier. The system is designed to meet CCITT Rec. G811 and features exceptional phase stability, low noise and high isolation between outputs. All three frequency sources are phase locked to the jittered output of GPS receiver, using a very slow time constant to filter out the jitter. Thus both excellent: short-and-long-term-stabilities ($5 \cdot 10^{-12}$ per month, 1×10^{-11} per life) are achieved.

Introduction

Accurate Time and Frequency source is a must in every satellite control ground station.
Some of the uses for such source (master clock) are:
a. Measuring and recording slow and fast phenomena relating to the satellites various tasks.
b. All transmitters and receivers in the ground station must have one common time & frequency reference. The same goes for synthesizers, counters and other measuring and generating equipment in the station, which are phase-locked to the master frequency.
(Lack of synchronization and time skew is prevented this way).
c. When data stream is passed from one station to another, via satellites, it is essential to use one time base for both stations (or more) plesiochronousley.

Thus, one needs a highly stable (jitterless) reliable and uninterruptible, Time & Frequency generating and distributing system.

For these applications T.F.L have designed a high redundant, multiple output whose frequency & time source is locked to the UTC (Coordinated Universal Time) via a GPS receiver. Thus a long-term-stability of 1×10^{-11} per life is achieved. This system is modeled TF-3100A and is described in figure 1.

Some of the system's features are:

A. UTC traceable via GPS receiver

B. Filtering and eliminating jitter and noises of the GPS receivers outputs and tracking it's main frequency.

C. The use of oven control voltage controlled crystal oscillator (OCVCXO) (or Rubidium standard).

D. Redundancy of 2:1 sources (3 clocks)

E. Comparing and aligning the three sources outputs, combining and distributing them to the various users.

F. An automatic master change-over when one source malfunctions or out of phase (Majority vote).

G. Audible and visual Real time alarm indications.

H. Inner and outer battery backup for the whole system.

CH2818-3/90/0000-145 $1.00 © 1990 IEEE

General Description

The primary purpose of the TF-3100A is to generate and distribute a high precision, low-noise, redundant, standard frequency signals, time display and time code.

The system can provide the following signals:
* up to 50 5MHz isolated sinusoidal outputs.
* up to 50 1MHz isolated sinusoidal outputs.

* IRIG-B time code output.
* 1pps (1 pulse per second) output.

These signals are traceable to the UTC by means of the frequency and time signals transmitted by the NAVSTAR Global positioning System (GPS) satellites. The GPS signals are received by a GPS receiver, (Kinemetrics, Model GPS-DC) capable of automatically tracking up to four satellites. The signals, provided by the GPS receiver, have excellent long-term-stability (traceable to the UTC), but high jitter and poor short-term-stability (STS) which is caused by the transmission from satellites to the receiver. Another problem is that when satellites are not in view, the receiver provides the signals from it's own inner TCXO (precision of ±1ppm). In order to prevent this unstable signal from getting into the synchronous clocks, TF101A (see figure 1) the receiver was specially designed in such a way that it's output signals are automatically cut-off while "unlock" situation occurs.
TF-101A will automatically change over to "Free run" mode, when input signal does not exist and the system output will stay stable (10^{-11}/day).

To provide stable and highly-reliable, uninterruptible signals, the TF-3100A uses 3 redundant disciplined ovenized quartz frequency sources (or Rubidium frequency standard). These are installed in the 3 synchronous clocks (TF101A) which are locked to the GPS signals, by means of software and unique algorithm that filter the jitter and short-term instabilities.

Advanced signal processing techniques are used to evaluate the quality of the signals provided by the frequency sources. The system optimally combines in-phase (chosen by the majority) the output signals of the operational sources with 6db amplitude difference in accordance with predetermined priorities. Figures 2 and 3 are the schematic drawings of the phase aligner & switch module (PSAM) and the combiner amplifier module (CAM).

The PASM tests for phase errors and jumps in the three inputs signals of the same frequency. It corrects phase errors to an accuracy of 1/16 unit interval (CCITT Rec.G811 compatible) and aligns the phase of the other inputs with reference to the Master input signal. An algorithm is used to choose the Master input (in case it is not supplied by the next module CAM).

If an input malfunctions or an external alarm is received or it is not possible to correct the phase, the PASM will automatically cut-off that channel. (Channels may be connected or cut-off manually by push buttons on the front panel of TF-3007A-Distribution Amplifier System).

The three "in-phase" signals are fed into the CAM which consists of two main sections:

The first section combines the three input signals according to a given priority into a single output with continuous amplitude and phase. The second section, amplifies and distributes the output from the combiner, and provides two groups of 4 preamplified outputs (with isolation between outputs of more than 100db). In addition, the module includes logic circuits (microprocessor) that indicates which of the signals present is treated as "Master".

Warning is given if there is a malfunction in the combiner, or an input signal is missing. In such cases the output is taken from one of the other two standby inputs, with no dead time, phase jump or amplitude change at change over(CCITT Rec.G811).

The CAM outputs are fed into the Distribution Amplifier modules (DAM) that devides the input signal into 8 outputs isolated by more than 100db, 12dbm of power each.

This technique ensures high quality, continuously available signals even under fault conditions, with excellent long term (UTC) and short term stability, and a very low jitter.
Figure 4 shows two of the TF-101A (synchronous clock) outputs offset from TFL's software clock (ensemble of Cesium clocks) during 31 days. The average offset obtained is $\pm 1 \cdot 10^{-12}$ while they were tracking the GPS output. Figure 5 shows the time interval error of one TF-101A. The time error is 700 nsec after 30 days ($2.7 \cdot 10^{-13}$).

TF-3100A SYSTEM CONFIGURATION

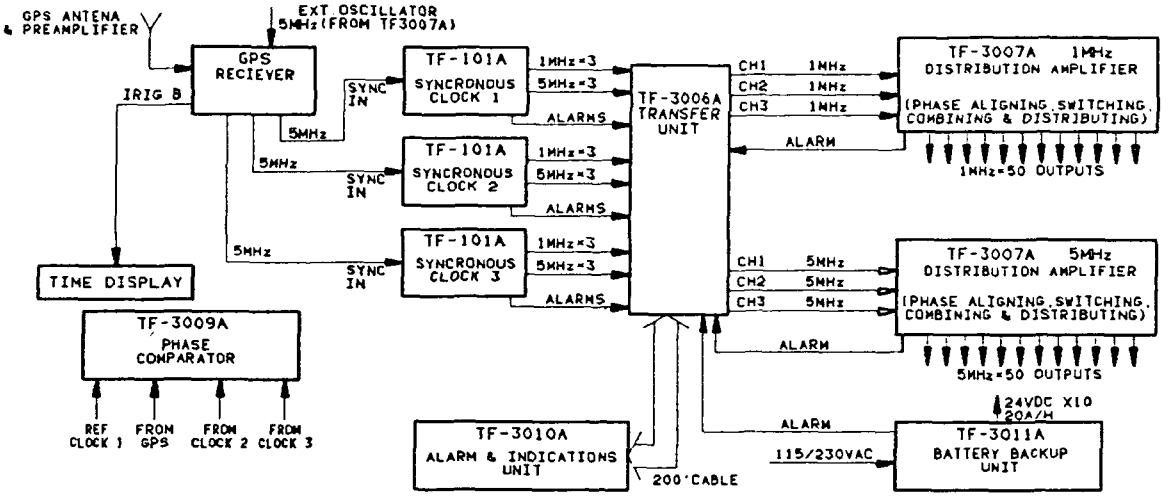
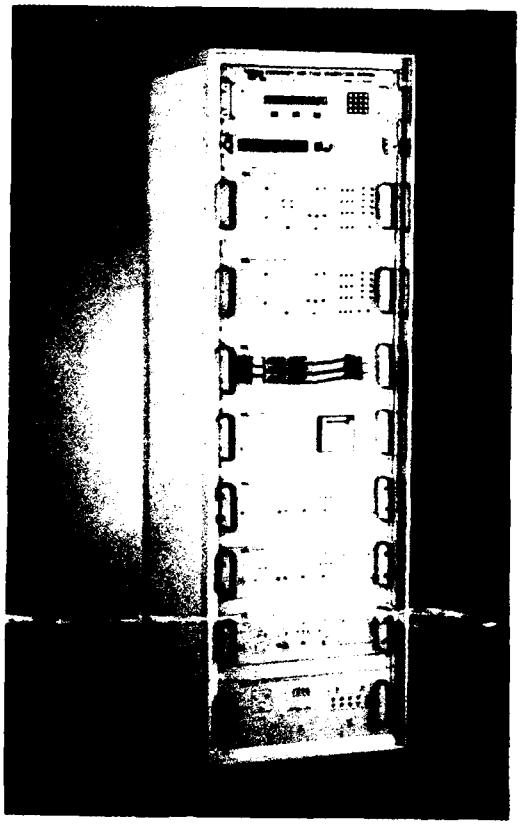


FIG 1:REDUNDANT TIME & FREQUENCY SYSTEM WITH GPS RECIEVER



TF-3100A FREQUENCY & TIME GENERATING SYSTEM

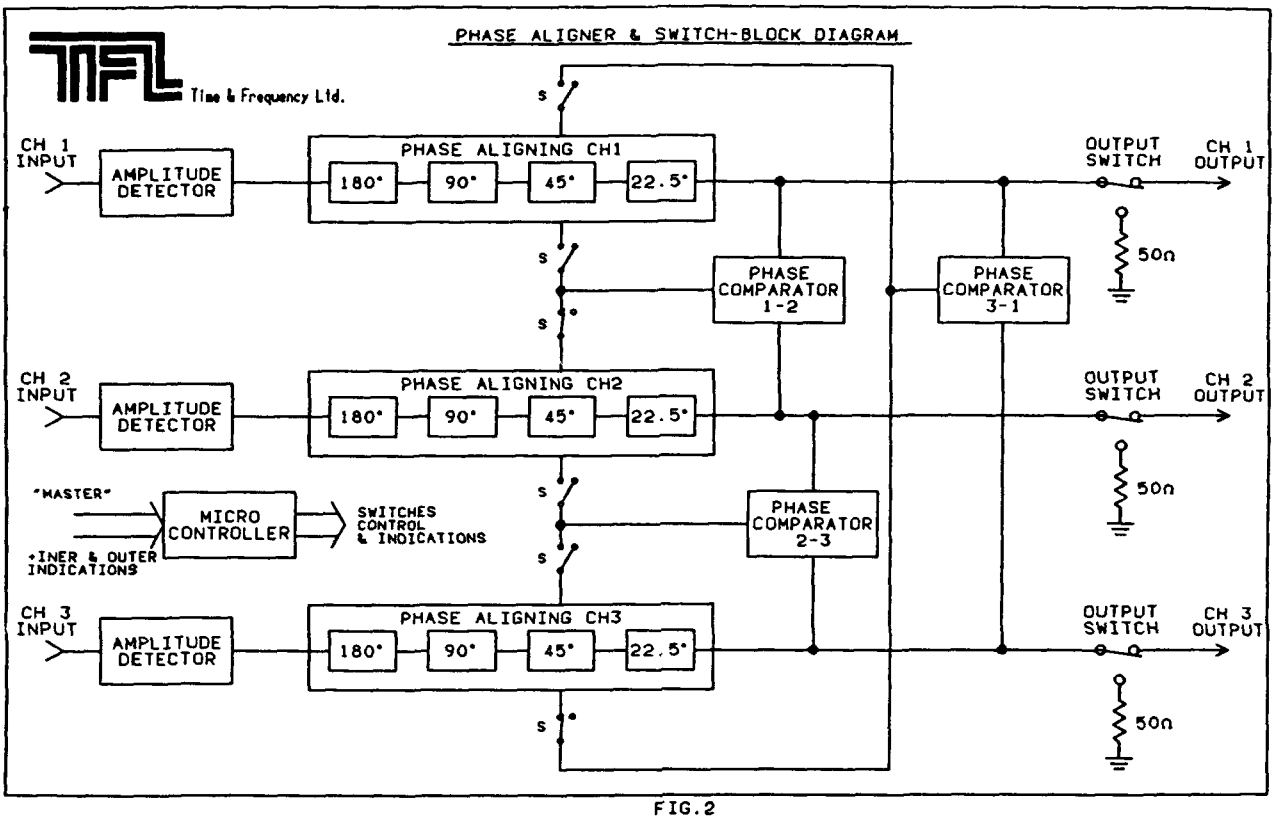

PHASE ALIGNER & SWITCH-BLOCK DIAGRAM
Time & Frequency Ltd.
CH 1 INPUT
AMPLITUDE DETECTOR
PHASE ALIGNING CH1
180°
90°
45°
22.5°
OUTPUT SWITCH
CH 1 OUTPUT
50Ω
PHASE COMPARATOR 1-2
PHASE COMPARATOR 3-1
CH 2 INPUT
AMPLITUDE DETECTOR
PHASE ALIGNING CH2
180°
90°
45°
22.5°
OUTPUT SWITCH
CH 2 OUTPUT
50Ω
"MASTER"
+INER & OUTER INDICATIONS
MICRO CONTROLLER
SWITCHES CONTROL & INDICATIONS
PHASE COMPARATOR 2-3
CH 3 INPUT
AMPLITUDE DETECTOR
PHASE ALIGNING CH3
180°
90°
45°
22.5°
OUTPUT SWITCH
CH 3 OUTPUT
50Ω

FIG.2

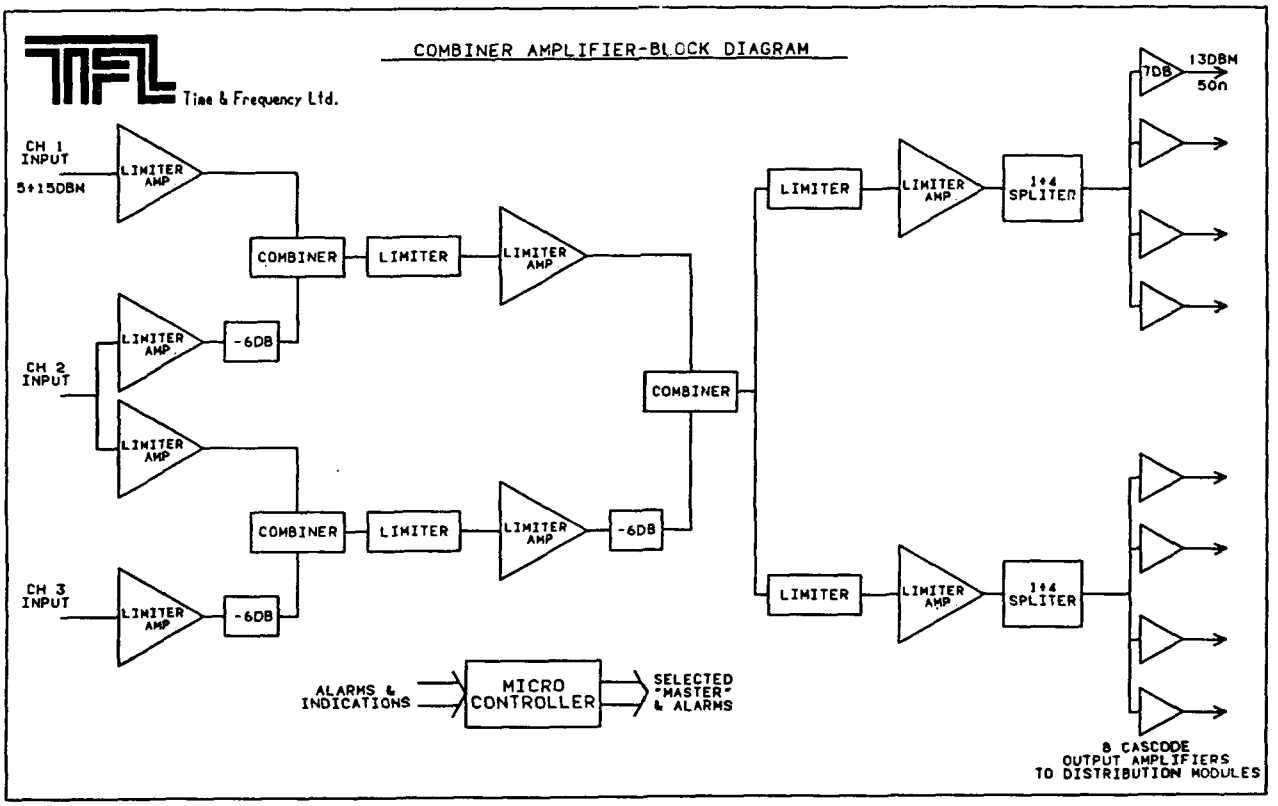

COMBINER AMPLIFIER-BLOCK DIAGRAM
Time & Frequency Ltd.
CH 1 INPUT
5÷15DBM
LIMITER AMP
COMBINER
LIMITER
LIMITER AMP
CH 2 INPUT
LIMITER AMP
-6DB
LIMITER AMP
COMBINER
LIMITER
LIMITER AMP
-6DB
CH 3 INPUT
LIMITER AMP
-6DB
COMBINER
LIMITER
LIMITER AMP
1÷4 SPLITER
7DB
13DBM
50Ω
LIMITER
LIMITER AMP
1÷4 SPLITER
ALARMS & INDICATIONS
MICRO CONTROLLER
SELECTED "MASTER" & ALARMS
8 CASCODE OUTPUT AMPLIFIERS TO DISTRIBUTION MODULES

FIG.3

```
1Combined graph of clocks:
32: 101-4
34: 101-5
0
```

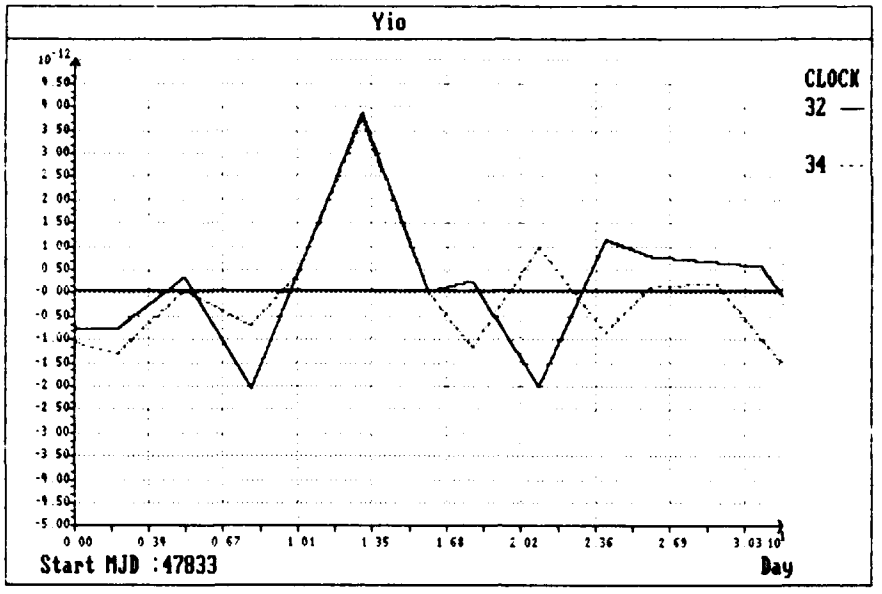


Fig. 4: Two TF-101A Synchronous Clocks frequency offset from TFL's software clock, during 31 days tracking the GPS receiver.

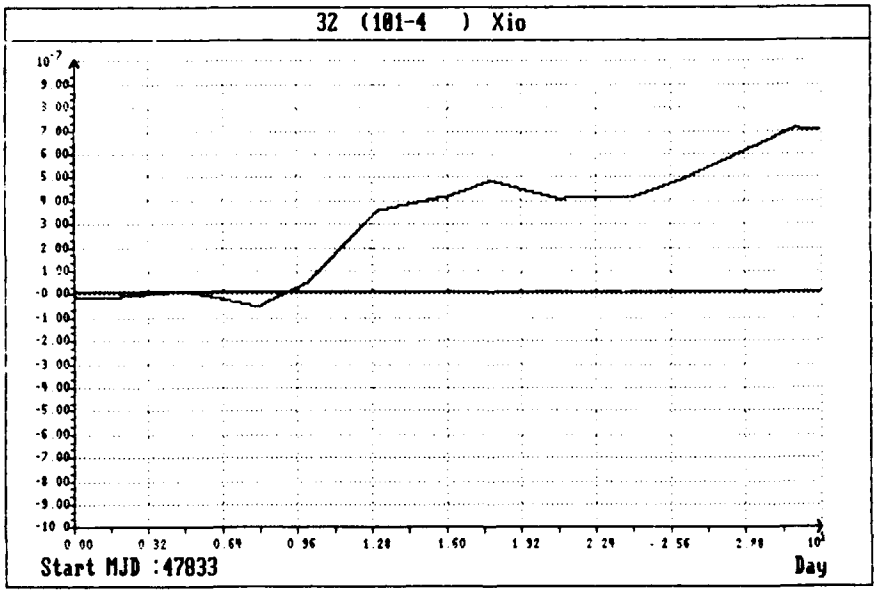
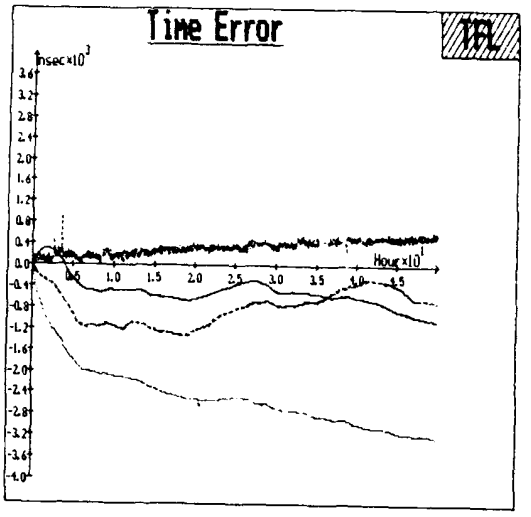


Fig. 5: Time interval error of TF-101A Synchronous Clock, tracking GPS receiver's 5 MHz output, 30 days period measurement. TFL's software clock - as reference.



Ser No. Date Time
101-4 05/03/1990 06:43:44 ————————
Average Offset -5.0*10⁻¹⁰

101-5 05/03/1990 06:43:44 - - - - - - - - - - - - - - - -
Average Offset -1.0*10⁻¹⁰

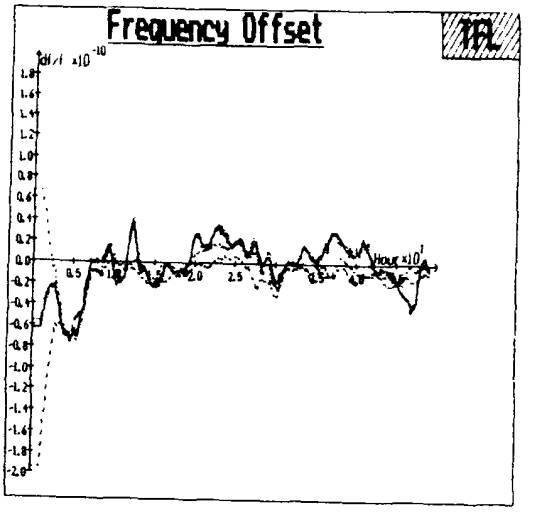
101-3 05/03/1990 06:43:44 .....
Average Offset -3.0*10⁻¹⁰

GPS 05/03/1990 06:43:44 .....
Average Offset 3.2*10⁻¹⁰

3007-1 05/03/1990 06:43:44 ----------
Average Offset -3.0*10⁻¹⁰

Reference: 8005E

Fig. 6: Three TF101A Synchronous Clocks tracking the GPS receiver (50 hours of observation). 3007-1 is the system 1MHz output. It is tracking the TF-101A S/N#3 "one by one" which was the "Master".



Ser No. Date Time
101-3 05/03/1990 06:43:44 ————————
Average Offset -3.7*10⁻¹⁰

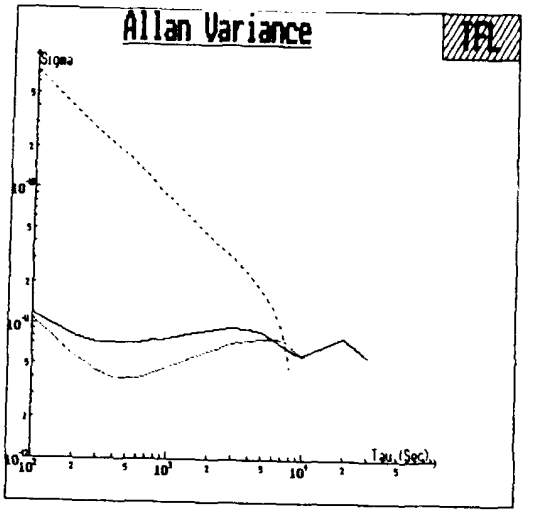
3007-1 05/03/1990 06:43:44 ----------
Average Offset -3.0*10⁻¹⁰

101-5 05/03/1990 06:43:44 .....
Average Offset -1.0*10⁻¹⁰

101-4 05/03/1990 06:43:44 .....
Average Offset -5.0*10⁻¹⁰

Reference: 8005E

Fig. 7: df/f calculation of three TF-101A (offset from TFL's reference) while tracking GPS receiver from "free run" mode (when no input was supplied to the TF-101A) to synchronous mode, 50 hours of measurements.



Tau 10s 1000s 10000s
Ser No. Date Time
101-3 05/03/1990 06:43:44 ————————
Sigm 1.2*10⁻¹¹ 7.2*10⁻¹² 6.0*10⁻¹²

101-4 05/03/1990 06:43:44 ————————
Sigm 1.1*10⁻¹¹ 4.0*10⁻¹² 5.0*10⁻¹²

GPS 05/03/1990 06:43:44 .....
Sigm 7.2*10⁻¹⁰ 8.7*10⁻¹¹

Reference: 8005E; Auxiliary References: 101-3, 101-4

Fig. 8: STS of two TF-101A and GPS receiver at 100 seconds observation period. GPS is noisy 7 X E-10 and TF-101A are 1 X E-11. At 10000 seconds observation period they are united.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A RUBIDIUM FREQUENCY STANDARD AND A GPS RECEIVER: A REMOTELY STEERED CLOCK SYSTEM WITH GOOD SHORT-TERM AND LONG-TERM STABILITY

by

David W. Allan and Judah Levine
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

Abstract

The short-term stability of a rubidium gas-cell frequency standard is usually better than that of commercial cesium-beam frequency standards. In the short-term stability region (from a few seconds to several thousand seconds) and for the specific value $\sigma_y(\tau = 100\ s)$, the range of short-term stabilities for rubidium is from about 4 to 10 x 10^{-13}, and for cesium about 6 to 30 x 10^{-13}. In the short-term, the stability improvement for rubidium and cesium is proportional to $\tau^{-1/2}$. Cesium almost always has better stability in the long-term because cesium has less sensitivity to environmental perturbations. For example, cesium has little or no frequency drift, whereas rubidium usually does. Improving a clock's environment invariably improves the long-term performance, especially in the case of rubidium.

Satellite time transfer shows a day-to-day stability of about a nanosecond. The spectrum of the fluctuations implies that $\sigma_y(\tau)$ should decrease as τ^{-1}. If a rubidium standard, in a good environment with the above performance in the short-term, were married to a satellite time-transfer system, then the combined performance of the system could have better short-term and better long-term stability than a stand-alone, free-running, commercial cesium standard.

We have taken some data to test this idea. The conclusions confirm the hypothesis. We have also replaced the rubidium oscillator with a quartz oscillator and with a high-performance commercial cesium standard. In both cases the system had significantly improved long-term stability over what otherwise would be obtainable from either oscillator by itself.

Introduction

Most commercial rubidium frequency standards have better short-term frequency stability than commercial cesium-beam frequency standards, where short-term implies averaging times (τ) up to a few thousand seconds. Placing a rubidium standard in a good environment usually extends its short-term stability to longer τ values (of the order of a day). For τ longer than a day, the opposite is most often true; that is, cesium is better than rubidium in the long-term.

The measurement noise of transferring time and frequency to a remote location using the Global Positioning System (GPS) common-view (C-V) technique can reach $\sigma_y(\tau = 1\ day)$ values approaching 1 x 10^{-14}, where $\sigma_y(\tau)$ is the square-root of the two-sample or Allan variance.[1] Therefore, an environmentally controlled rubidium standard, together with a GPS time transfer receiver, has the potential for the following advantages (taken as a system):

1) For a lesser cost, both the short and the long-term stability can be better than that of a commercial cesium standard;
2) in the long-term the system can reflect the stabilities of some of the best time and frequency standards in the world, in that it can be both syntonized and synchronized to a primary reference;
3) the system can be made fully automatic;
4) the system can remove the usual frequency accuracy limitation of rubidium including the frequency drift--allowing the system to be as accurate as the reference standard; and
5) the system can remove the usual time accuracy limitation of either cesium or rubidium--allowing the system to be synchronized to within a few nanoseconds of the reference standard.

Some of the disadvantages are:

1) The system's feedback is by a daily telephone call between the system and the reference standard used in the GPS common-view (GPS C-V) mode; and
2) the rubidium standard needs a good

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

environment to achieve the desired intermediate stability goals.

Figure 1 is a sketch of the GPS-common-view, remote-clock servo concept. Figure 2 shows some more detail for the remote clock at site B. We assume that the clock at site A has good long-term stability. One of our goals is to transfer the long-term stability of clock A to the clock at site B with the time transfer capability of GPS C-V. The theory we used for designing a digital servo for this system is based on a thesis of Alain Guetrot [2]. The essential concept of his thesis is that if the measurement noise spectrum, $S_N(f)$, can be represented as

$$S_N(f) \sim f^{\beta}$$

and the signal spectrum, $S_S(f)$, as

$$S_S(f) \sim f^{\beta-2},$$

then an exponential filter is optimum for obtaining the best signal-to-noise ratio. Here β is typically an integer between 3 and -5. What is measured, of course, is signal plus noise. If the measurements are a discret time series then a practicle realization of an exponential filter is the following,

$$z_i = \frac{kz_{i-1} + \hat{z}_i}{k+1}, \qquad (1)$$

where k is proportional to the exponential filter time constant, z_{i-1} is the last best estimate, $\hat{z}_i$ is the current measurement (signal plus noise) and z_i is the current best estimate of the signal, that is, the true difference of clock A compared to B.

We assume flicker noise time or phase modulation (PM) for the measurement noise (GPS C-V) and flicker noise frequency modulation (FM) for the long-term stability of the remote clock. This model, which satisfies the assumption in Guetrot's analysis, was tested and was found to be reasonable for the set of common-view data we analyzed and for the kinds of clocks we may use at A and/or B (quartz, rubidium, cesium and/or hydrogen frequency standards) for selected regions of sample times, τ, as determined from a $\sigma_y(\tau)$ diagram.

We assume clock B has frequency drift, so that the time prediction equation can be written as

$$t) = x(t-\tau) + y(t-\tau) \cdot \tau + \tfrac{1}{2}D(t-\tau) \cdot \tau^2, \qquad (2)$$

where $x(t-\tau)$, $y(t-\tau)$ and $D(t-\tau)$ are the best estimates of

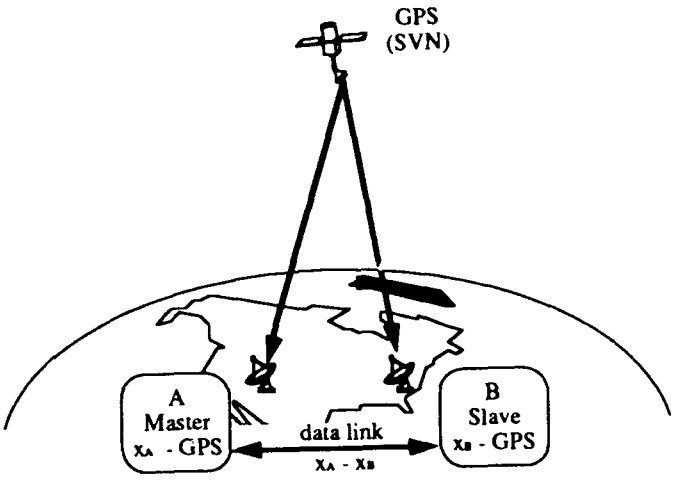


Figure 1. A schematic diagram showing how to provide accurate and stable time and frequency at a remote site. If the clock at site B is a rubidium gas-cell frequency standard in a good environment, and given the levels of GPS common-view measurement noise, the net output can be that time and frequency at the remote site is better in long-term as well as in short-term than that from a commercial cesium standard by itself.

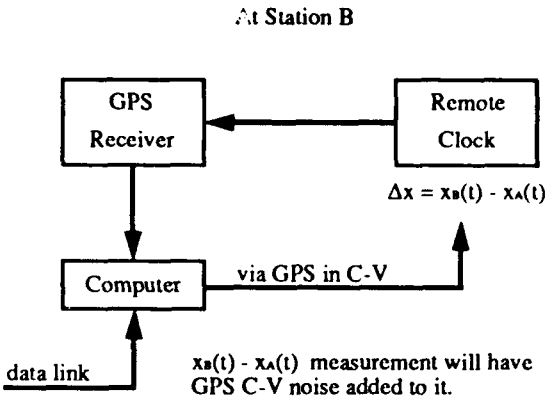


Figure 2. A block diagram of the GPS common-view time transfer receiver combined with a clock at site B. The system achieves better short-term and long-term, time-and-frequency stability than would be available from a commercial cesium standard by itself.

the time offset, the normalized frequency offset and the normalized frequency drift of the remote clock at the last measurement, and τ is the prediction interval. In our case, the prediction interval was usually one sidereal day - using a single satellite in the GPS C-V mode. For a general prediction interval $\tau = n\tau_0$, where τ_0 = 1 sidereal day, we may write the best estimate of the exponential time difference as

$$x(t) = (k_x\hat{x}(t) + n\,\Delta x)/(k_x + n), \quad (3)$$

where k_x is proportional to the exponential filter time constant for the best time difference estimate. n gives more weight to the measurement to compensate for flicker FM in the clock where the rms time error of prediction is proportional to n, and Δx is the measured common-view time difference between clock B and clock A. The exponential filtered drift is

$$D(t) = [k_D D(t-\tau) + \hat{D}]/(k_D + 1), \quad (4)$$

where $\hat{D}$ is the drift estimated over the last two measurement intervals, and is given by

$$\hat{D} = \frac{\dfrac{x(t) - x(t-\tau)}{\tau} - \dfrac{x(t-\tau) - x(t-\tau-\tau_{-1})}{\tau_{-1}}}{\dfrac{\tau + \tau_{-1}}{2}}, \quad (5)$$

where τ_{-1} is the measurement interval before the last one. The normalized frequency offset estimate is given by

$$y(t) = \{k_y[y(t-\tau) + D(t)\cdot\tau] + \hat{y}(t)\}/(k_y + 1), \quad (6)$$

where k_y is proportional to the exponential filter time constant for the best normalized frequency offset estimate, and

$$\hat{y}(t) = \frac{x(t) - x(t-\tau)}{\tau} + \tfrac{1}{2}D(t)\cdot\tau. \quad (7)$$

The k_x, k_y and k_D parameters are functions of τ_0, the measurement noise type and level and the noise type and level of the clocks at A and B. The parameters can be optimized through simulation or with real data. We used the latter. The above recursive equations assume τ is the time since the last measurement. The measurement intervals need not be equal.

Experimental Results

The rubidium frequency standard used for the experiment was an engineering development model for the GPS program. It was placed in vacuum with a temperature controlled environment so that variations were less than 0.1°C. The quartz oscillator frequency standard was a special one designed for short-term and long-term stability. It was also placed in a special environment [3].

A plot of the free-running fractional frequency of the rubidium is shown in Figure 3a and its frequency stability is plotted in Figure 3b. If a frequency drift is subtracted, the spectrum for the long-term stability is reasonably modeled by flicker FM. Figure 4 shows a plot of the frequency stability of the quartz oscillator. A frequency drift has been subtracted from the quartz data before plotting the stability. The nominal τ^0 behavior corresponding to flicker FM is apparent. The frequency stability of the GPS C-V measurement noise is plotted in Figure 5, and the τ^{-1} behavior is indicitive of flicker noise PM. Hence, we see that the assumptions needed for the Guetrot thesis are satisfied.

The two GPS receivers were colocated for convenience in checking the truth of the hypothesis. Colocating the receivers will cause a cancellation of the broadcast ephemeris errors, ionospheric delay errors and tropospheric delay errors as well. Errors due to multipath effects and to the receivers will remain about the same for long-baseline separations between the clocks. Therefore, the experimental results obtained in this paper may tend to give better stabilities than would be obtained where there is a large distance between clocks A and B. However, similar levels of GPS C-V measurement noise are often achieved over long baselines if several satellites are used and the data are properly combined, weighted and filtered [1,4].

Figures 6 and 7 are plots of the predicted values minus the measured values ($\hat{x}(t) - \Delta x$) for rubidium and the quartz, respectively, using the GPS C-V servo illustrated in Figures 1 and 2. The errors between the times of the measurements will usually be smaller than the errors plotted. There was a several day break in the continuity of the data between MJD 47615 and 47621 as can be seen in Figures 6 and 7. Before this break, there were a few days for the servo parameters to initialize. This initialization allowed for a reasonable prediction across this break in

data continuity. We can learn two important things from this fortuitous break in data continuity. First, we can see the ability of the system to predict time over several days; and second, we can see the transient response of the system as it re-acquires the daily signal.

Figures 8 and 9 are plots of the frequency stabilities of the errors plotted in Figures 6 and 7 for rubidium and quartz, respectively. We see that after only a few days of integration, this system has the capacity to track the best primary frequency standards in the world.

Figures 10 and 11 are the corresponding time stability plots where the time stability, $\sigma_x(\tau)$, is defined as $\tau * \mathrm{mod}\sigma_y(\tau)/\sqrt{3}$. [5,6] For rubidium (Figure 10) the system stability for single one-day measurements is within 10 ns (rms) of clock A. And again, for times between the measurement times, the stabilities will usually be less than this number. If we average the time readings for a few days, we can approach a nanosecond of time stability. The average value for the data in Figure 6 is 0.05 ns. There will typically be biases of a few nanoseconds in the GPS C-V measurement technique; hence, the time accuracy cannot be better than these biases.

The corresponding single one-day measurement stability number for the quartz oscillator stability plotted in Figure 11 is 75 ns (rms). We see, in this case, that we have to average for very long periods to approach a time stability level of 1 ns. That one can reach these levels at all with a quartz oscillator in the system is impressive. The mean value for the time errors plotted in Figure 7 was 0.5 ns. Again, please note that biases in the system will limit the time accuracy to a few nanoseconds.

We tested the system with a high-performance, commercial, cesium-beam frequency standard as the clock at site B and observed long-term improvements. The improvements were not nearly so dramatic as for the above data for rubidium and quartz. The main benefits for using cesium in the system is to keep the time at site B synchronized. The cost effectiveness for such a system is obviously better with rubidium and quartz kept in a good environment.

Methods of Implementation

This servo could be implemented in several different ways. The outputs of the model could be used as a "paper" time or frequency output. This is simple to implement but has the disadvantage that electrical output of the oscillator does not directly reflect the improvement in performance produced by the servo. This disadvantage may be overcome by using the model parameters to drive a microstepper or other similar device to produce a continuously corrected output. A small computer would be required to control the microstepper and to cope with missing data.

Another novel alternative is to provide an input to the system from an unknown clock a user may wish to calibrate. The system would measure the time and frequency difference between the clock being calibrated and the clock, which is part of the system at B. The system, at anytime, has the information for the best estimate of the time and frequency difference between clocks A and B. Hence, the system could straight forwardly calculate the optimum estimate of the clock being calibrated against the reference clock at site A. This information could be provided as a real-time read out of the system and would be more precise and accurate than that obtainable from a steared micro-phase stepper.

The designer can vary parameters k_x, k_y and k_D so as to optimize the transient response of the servo or its rms offset error (but not both simultaneously). Likewise, the threshold for rejecting a measurement must be chosen as a compromise between detecting time or frequency steps and providing immunity to glitches.

The values of the parameters chosen for the data presented in this paper were $k_x = 25$, $k_y = 1$ and $k_D = 0.2$ for rubidium and $k_x = 20$, $k_y = 0$ and $k_D = 0.2$ for quartz. These values were chosen to obtain a best transcient response rather than optimizing steady-state stability. From the results obtained, this choice caused little degradation in the steady-state performance.

The time transfer system need not be the GPS C-V method. Any system, which would allow the time difference comparison of clocks remote from each other at the few nanoseconds level, could be made to provide comparable results as those reported in this paper. In addition, the performance of the system when using GPS C-V is unaffected by GPS selective availability (SA) if only clock dither is turned on (no degradation of the satellite broadcast ephemeris). If, however, GPS SA were fully implemented, including the degradation of the broadcast satellite ephemeris, then some of this degradation would not be cancele.' in the GPS C-V approach. The amount of increased measurement noise is a function of the baseline between the clocks and the level of SA.

We are currently studying more complex algorithms in which the k_x, k_y and k_D parameters are adjusted dynamically in response to changing conditions.

Acknowledgements

The authors wish to thank Dr. Donald B. Sullivan, Dr. Marc Weiss and Mr. Tom Weissert for helpful suggestions on the manuscript. We are especially grateful to the sponsors: the Joint Program Office of Air Force Space Division and the Naval Research Laboratory for supplying the rubidium clock.

References

[1] M. A. Weiss and D. W. Allan, "An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data," IEEE Trans. on Instrum. and Meas., IM-36, 572-578 (1987).

[2] A. G. Guétrot, "Optimum Smoothing Techniques on VLF Time Signals," Thesis, University of Colorado, 1969.

[3] F. L. Walls, "Environmental Effects on the Medium and Long Term Frequency Stability of Quartz Oscillators, Proc. of the 2nd European Frequency and Time Forum, Neuchatel, Switzerland, March 16-17, 1989, pp. 719-727.

[4] M. A. Weiss, "The Design of Kalman Smoothers for Global Positioning System Data," IEEE Trans. and Meas., 38, 652-657 (1989).

[5] D. W. Allan, "Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 647-654 (1987).

[6] D. W. Allan, D. D. Davis, J. Levine, M. A. Weiss, N. Hironaka, and D. Okayama, "New Inexpensive Frequency Calibration Service from NIST," these proceedings.

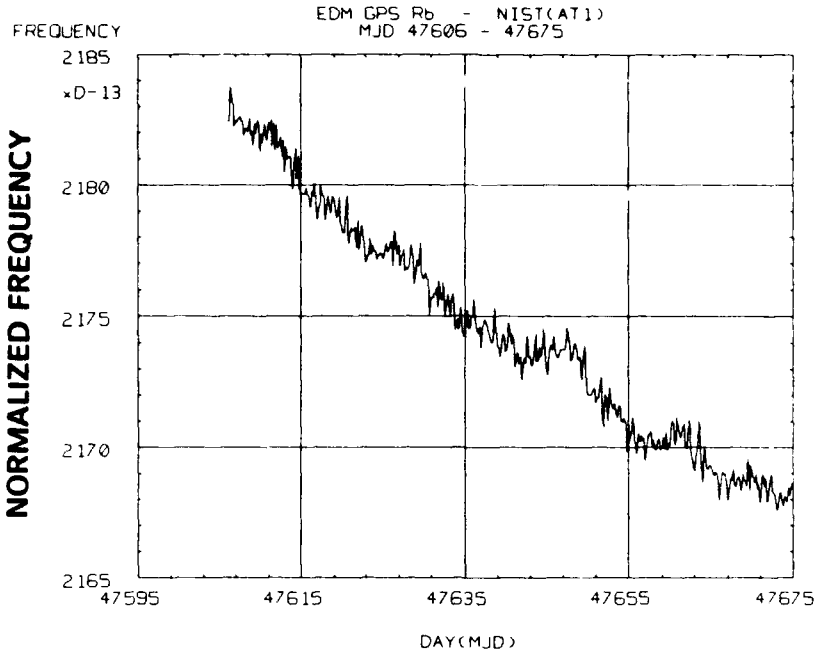
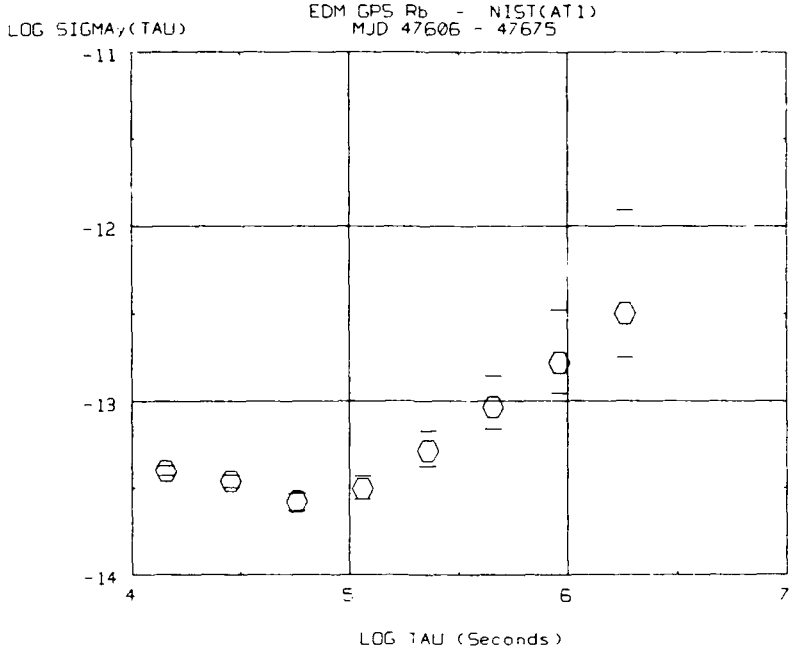


Figure 3a. A plot of the normalized frequency output of an engineering-development-model, rubidium-gas-cell frequency standard. This EDM standard was prepared as part of a GPS program.



3b. A plot of the frequency stability of the data plotted in Figure 3a with no frequency drift removed. The long-term stability is limited by the drift: $\sigma_y(\tau) = \tau D/\sqrt{2}$.

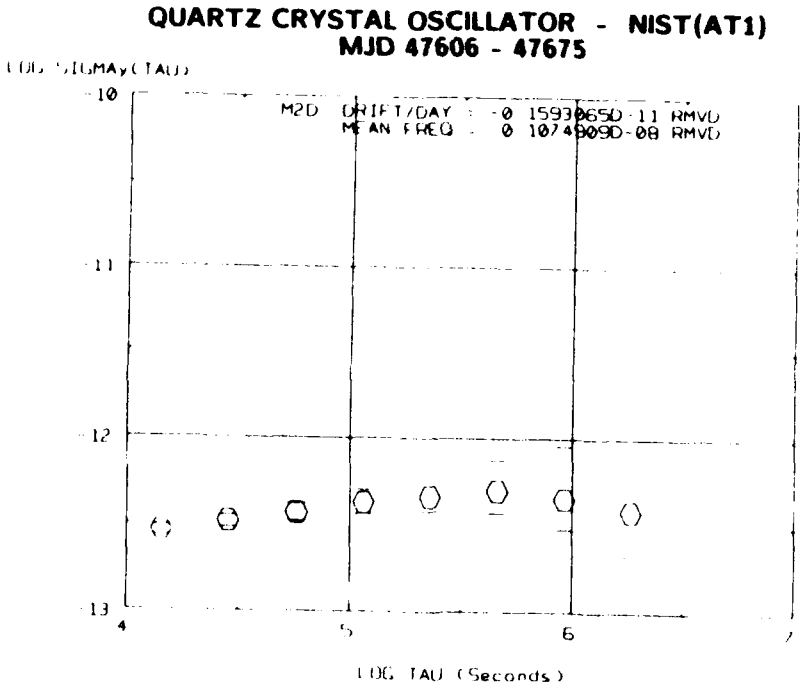


Figure 4. A plot of frequency stability for a special quartz oscillator frequency standard placed in a good environment. A frequency drift was subtracted from the data before calculating the frequency stability as plotted.

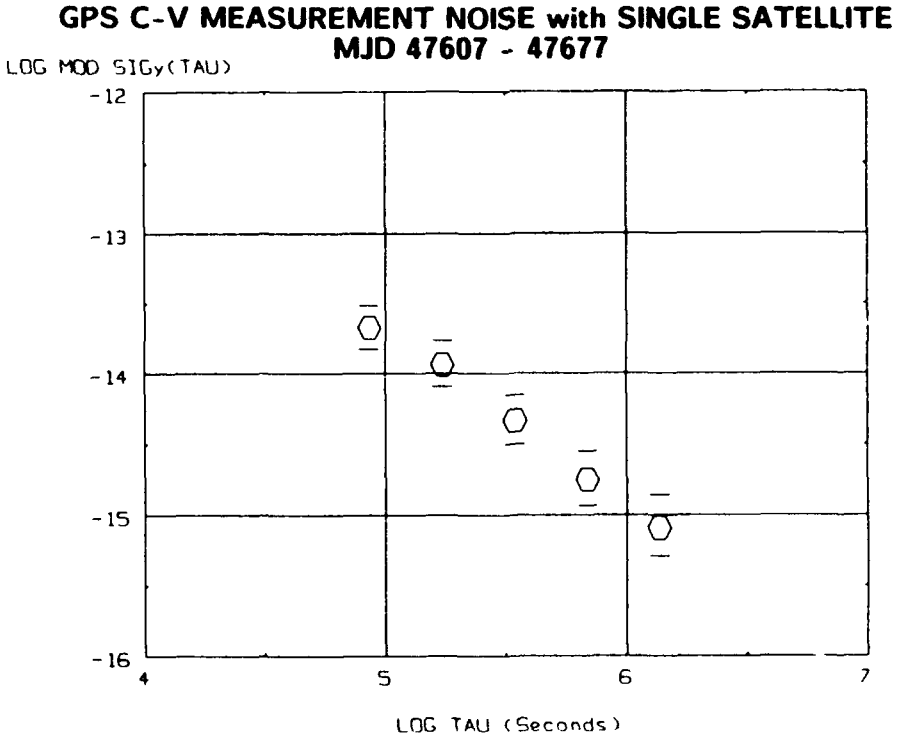


Figure 5. The frequency stability measurement noise for the GPS common-view system used for providing time and frequency at a remote site.

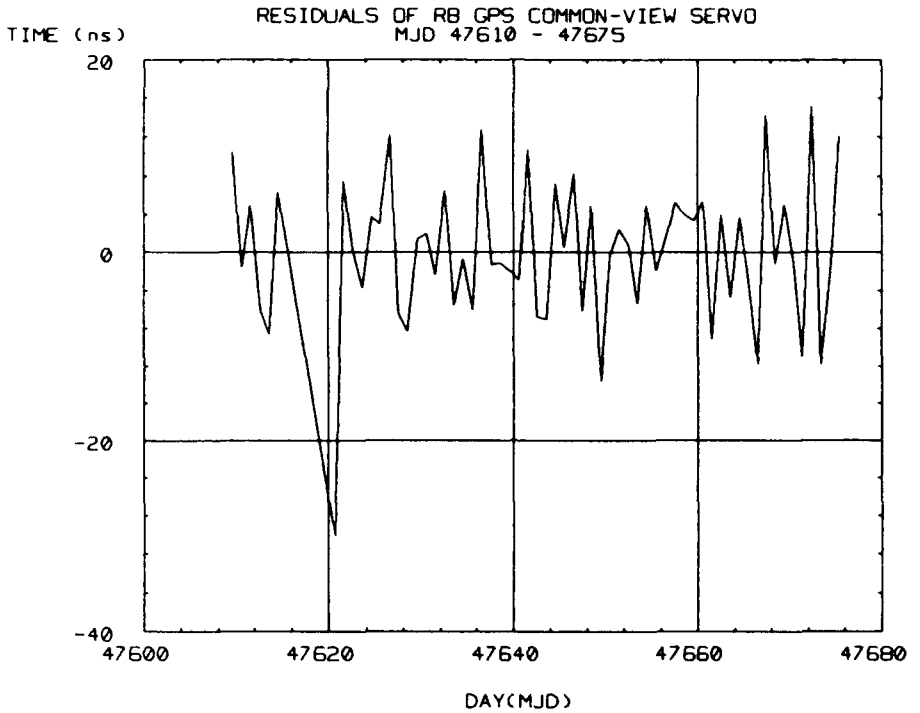


Figure 6. A plot of the predicted time values minus the measured values for a rubidium-gas-cell frequency standard in a good environment at site B.

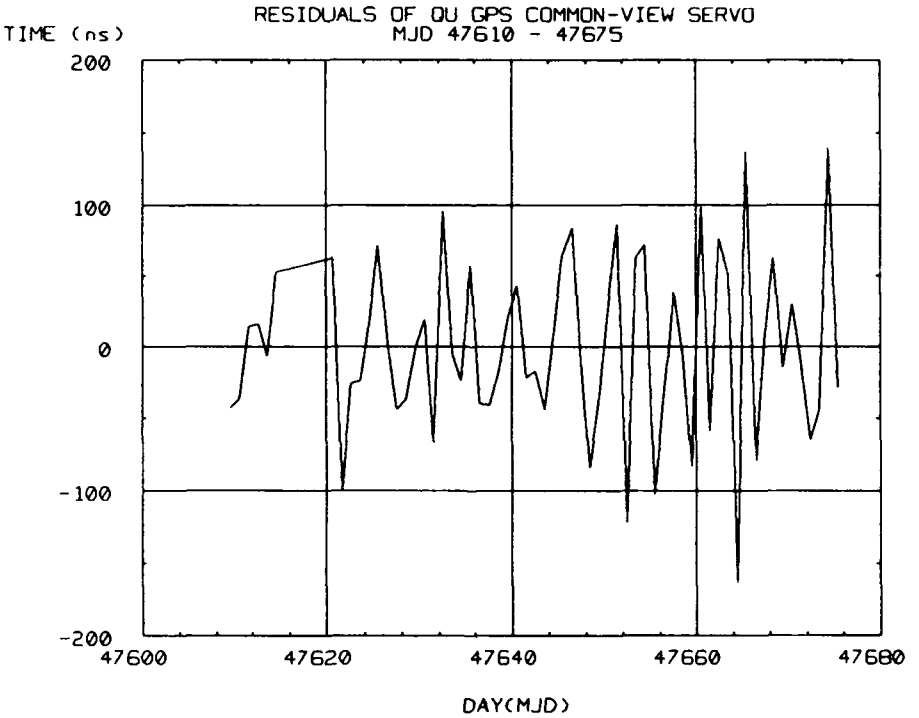


Figure 7. A plot of the predicted time values minus the measured values for a quartz-crystal oscillator in a good environment at site B.

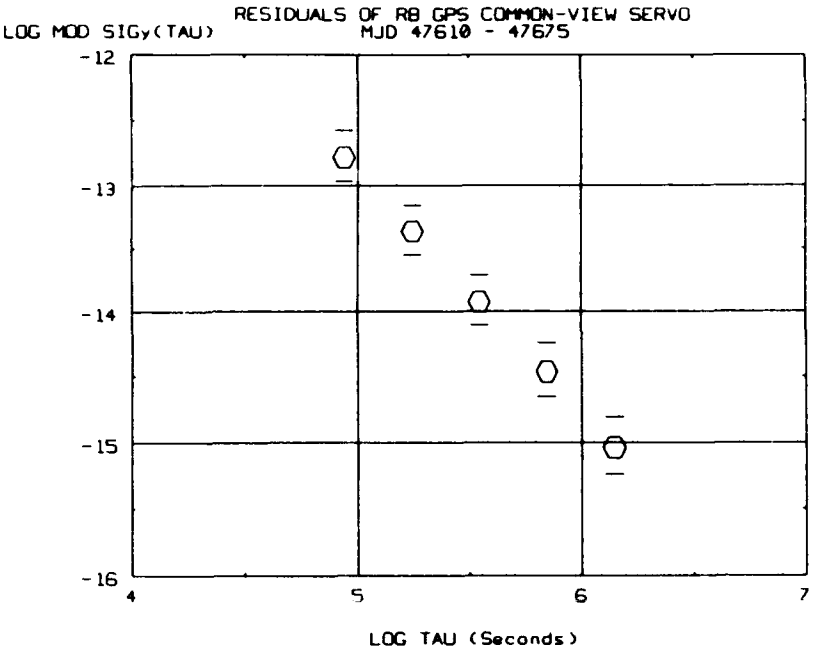


Figure 8. A plot of the frequency stability of the errors plotted in Figure 6.

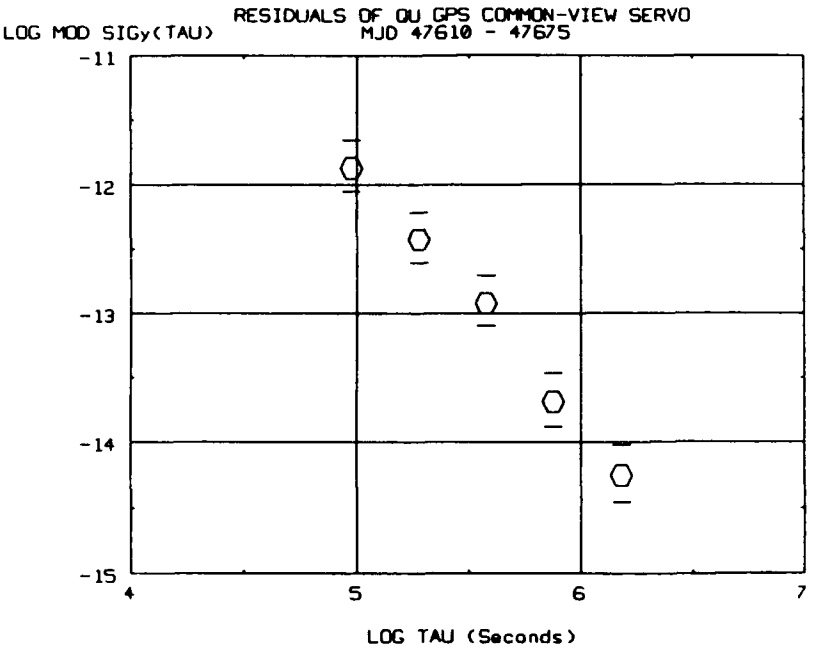


Figure 9. A plot of the frequency stability of the errors plotted in Figure 7.

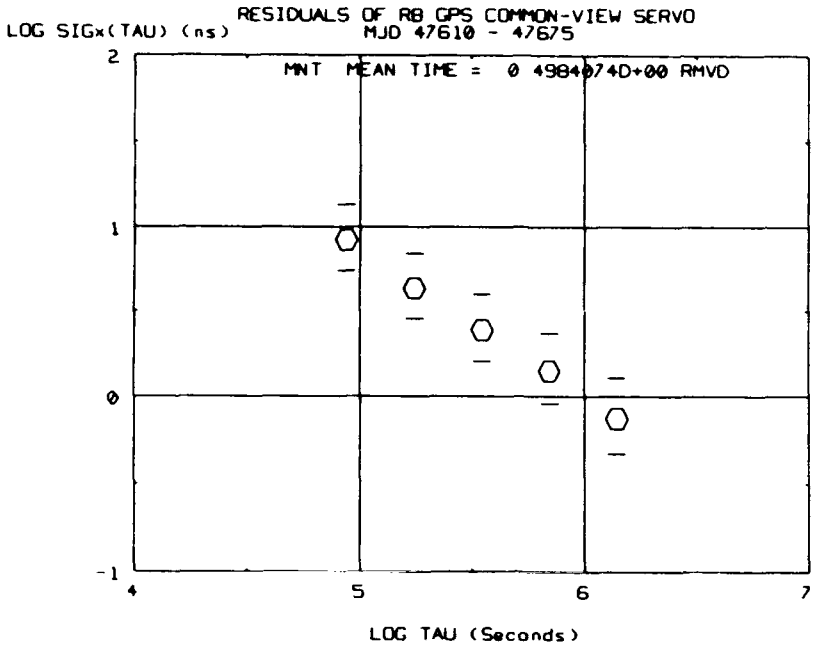


Figure 10. A plot of the time stability of the errors plotted in Figure 6.

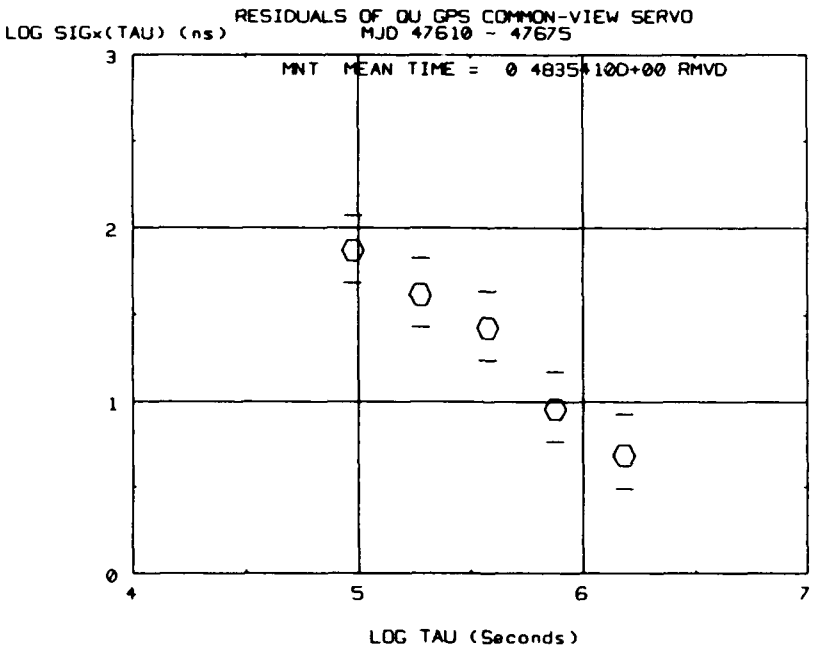


Figure 11. A plot of the time stability of the errors plotted in Figure 7.

RADAR STABILITY SELF-TEST USING BULK ACOUSTIC WAVE DELAY LINES

Perry A. Sorrell

Westinghouse Electric Corporation
P.O. Box 746, MS-709
Baltimore, MD 21203

Abstract:

This paper describes a novel application of sapphire Bulk Acoustic Wave (BAW) microwave delay lines to provide built in test that measures both phase and amplitude noise sidebands for radar stability calculations. Stability measurements using a 15 microsecond sapphire BAW delay line have been made on 3 GHz Moving Target Indicator (MTI) radars that support 65 dB clutter cancellation. Prior to this study one would have thought that a delay larger than 15 μs was required to include the STALO noise contribution to stability. Analysis of the 15 microsecond decorrelation of the associated -135 dBc/Hz STALO noise shows that this technique for measuring radar system stability produces a STALO noise contribution that is only 1.2 dB less than independent CW measurements made on the double sideband STALO noise.

I. Introduction

MTI radars require excellent stability to accomplish high clutter (undesired terrain return) cancellation performance. Measuring this stability is a cumbersome and difficult task; therefore, a technique suitable for built in testing that accurately measures the system instabilities is a highly sought after feature.

The radar stability measurement is difficult because: (1) the high peak transmitter power disrupts receiver performance, and (2) the Stable local Oscillator (STALO) noise is range dependent and thus a microwave frequency delay is required in the test configuration so that the STALO noise does not cancel. A sapphire BAW delay line solves these problems by providing the proper microwave delay. With this microwave delay, all noise of the radar can be considered, including the transmitter and the elusive range dependent STALO noise. The delay must be longer than the transmit time to eliminate the transmitter-to-receiver interference problem. A stability measurement must consider all potential unstable areas in the radar, especially these two since they are the most likely areas to produce instabilities.

The largest available BAW delay line from industry is 15 μs. The large insertion loss and maximum power handling capability of the BAW delay line prevent the use of 15 μs units much above 4 GHz; therefore, smaller delays must be used at higher frequencies to maintain manageable losses. The thin film transducers on the BAW delay lines are power limited and the bulk acoustic material tends to saturate or cause dispersion when driven much above a few watts of peak input power. Thus, one cannot raise the input power to make up for the high insertion loss. However, at L Band and S Band the 15 μs BAW delay line provides an adequate delay with acceptable insertion loss for testing MTI radar stability in the 65 dB range.

II. Stability Measuring Implementation

Stability of a radar is the ratio of the total double sideband integrated spectral energy about the carrier, limited by the radar filters and further modified by the range factor, to the carrier power. Stability is usually expressed in dB.

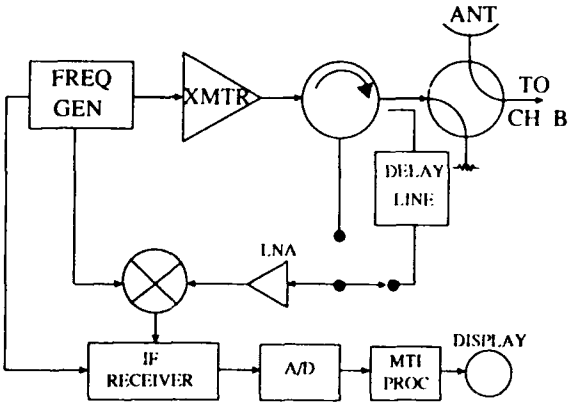


FIG.1 MTI RADAR CONFIGURATION FOR STABILITY TESTING

Figure 1 shows the stability test configuration. A sample of the attenuated transmitter output signal is coupled through the BAW delay line to the low noise amplifier

CH2818-3/90/0000-161 $1.00 © 1990 IEEE

of the radar receiver. Ideally this is done on an off-line channel where the receiver and transmitter are switched to dummy loads. If an off-line channel is not available, the radar is normally taken off-line for this measurement which eliminates the problem of returns from nearby real-life clutter that could contaminate the stability measurement. The radar A/D outputs with and without MTI filtering are then compared for a measure of the system stability. The former yields the noise spectrum power, the latter measures the carrier plus noise power spectrum. The A/D output may also be sent to a buffer, then to a DFT, not only for a measure of stability, but for a display of the signal's spectrum.

III. STALO Delay

The STALO is the only hardware in the radar whose noise or instabilities are inherently range dependent. This occurs because of the coherent detection feature of the radar and takes place at the first down-converter in the receiver. Since the bulk of the STALO noise is frequency and phase modulation, the noise on the STALO signal and the STALO-derived transmit signal are identical. With no delay for either of these two signals, all noise energy on these two signals will cancel at the first down-conversion in the receiver, thus producing a noise-free IF signal. But in real radar performance all live targets are delayed (12.34 μs/nmi); therefore, STALO noise always appears.

The effects of STALO noise have been demonstrated in the laboratory. The IF signal including STALO noise of a laboratory CW model radar is shown in Figure 2 with a 15 μs delay line in the transmit-receive path as shown in Figure 1. The 15 μs delay produces nulls every (1/15) x 10^6Hz (66.6 kHz) due to cyclic 360° phase differences in the two noise sidebands at the receiver first down converter. Therefore, nulls occur in the noise energy sidebands at the IF due to in-phase sideband noise from the two signals, the same as if the delay was zero. The noise frequencies at 33.3 kHz will be 180° out of phase thus adding coherently (6 dB) to the total noise voltage. This repeating of in-phase and out-of-phase noise addition creates the noise sideband lobes in the CW signal shown in Figure 2. The discrete sidebands on the third lobe on both sides of the carrier result from an injected 166.6 kHz modulation signal for reference only. If the microwave delay of one signal is changed from 15 μs to infinity to create total decorrelation, then the noise sideband power in both signals add to give a 3 dB increase in double sideband energy at the radar IF. However if zero delay exists between the two signals the resultant sideband noise energy is cancelled completely.

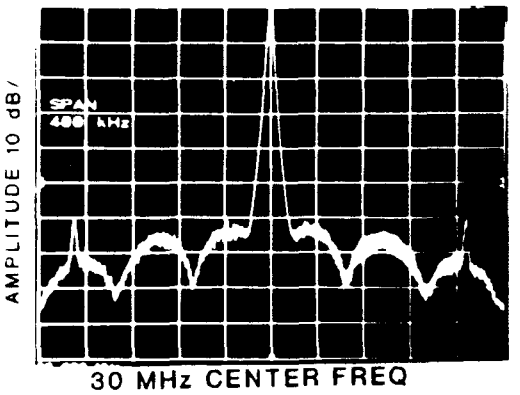


FIG 2 DECORRELATION OF STALO NOISE IN A CW LAB RADAR

IV. STALO Decorrelation

The relationship between decorrelation and delay is given by the spectral characteristic of the range-dependent (decorrelation) filter below [1], which is plotted in Figure 3 for selected values of delay.

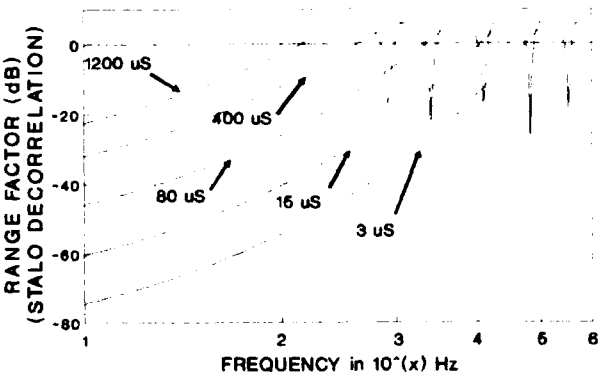


FIG. 3 DECORRELATION vs DELAY

Decorrelation

in dB = 10 log [4 $\sin^2$ (π fm T_d)]

where:

fm = Modulation Frequency (Hz)

T_d = Time Delay = 2 R/c (seconds)

R = Range (meters)

c = Propagation Velocity (3 x 10^8 m/sec)

Notice the small decorrelation at the low frequencies (close to the carrier). As delay increases, the decorrelation of the identical noise sidebands on the two STALO-derived signals at the first down-converter in the receiver increases until at infinite delay the low frequency curve takes on the same cyclic form as the high frequency curve, resulting in maximum decorrelation. This produces an RMS addition of the two sideband energies for a 3 dB increase in the double sideband power at the IF as opposed to total cancellation of sideband noise with zero delay.

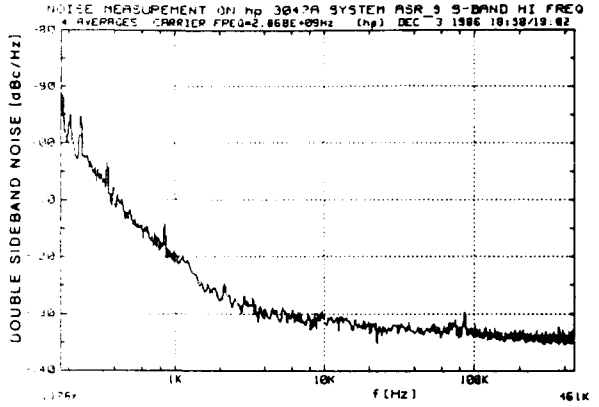


FIG 4 TYPICAL STALO NOISE

A typical STALO double sideband noise curve with noise decreasing to the -135 dBc/Hz level, is shown in Figure 4. Since STALO noise power distribution asymptotically increases at low frequencies close to the carrier, and delay decorrelation decreases as frequency decreases, one may ask if a 15 µs delay is large enough to produce acceptable and accurate results in a stability test. Section V answers this question in the affirmative.

V. 15 us Delay Measurement Accuracy

The sideband noise power density spectrum that is to be measured is affected by all frequency shaping components from the IF filter through the MTI filter; therefore, the calculation of the STALO sideband noise power must include the effects of these filters to evaluate the accuracy of the radar stability measurement using a 15 µs delay.

Figure 5 shows that the IF filter provides the high-frequency cutoff and the composite velocity response provides the low-frequency cutoff. The noise spectrum of the STALO in Figure 4 is shaped by these filters to produce the top curve in Figure 6. This is the total double sideband STALO noise power seen by the radar and the integrated energy is -75 dBc. This is the radar stability limitation imposed by the STALO,

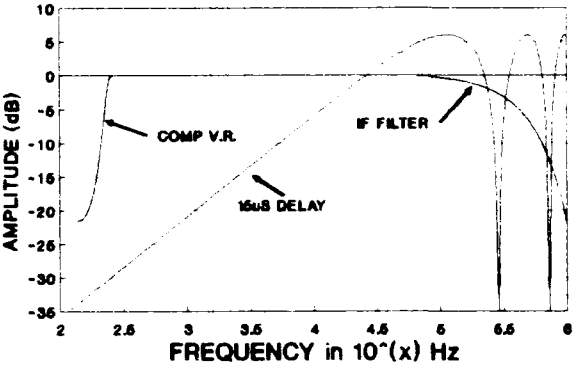


FIG. 5 RADAR FILTERS & 15 uS DELAY

independent of range. However this limitation varies with the Range Factor as discussed in Section III. The standard method of presenting STALO stability limitation does not include this range factor, but it appears in real-life radar operation.

The amount of STALO noise that is visible in a 15 µs delay radar stability test, is determined by modifying this -75 dBc noise curve by the 15 µs range factor curve shown in Figure 5. The resulting spectrum is shown as the lower curve in Figure 6. The integrated energy is now -76.2 dBc and is the total STALO spectral energy seen by the radar with a 15 µs delay. The introduction of the range-dependent filter introduces a total integrated energy difference of only 1.2 dB.

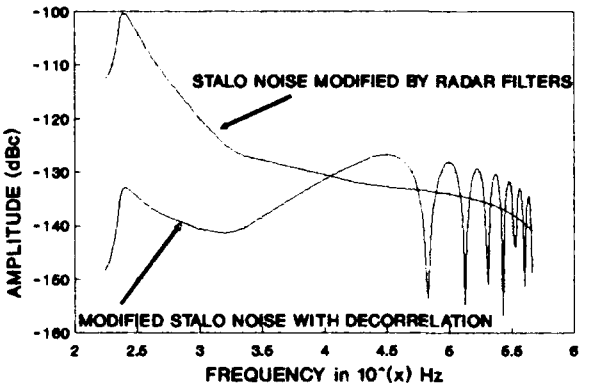


FIG. 6 STALO NOISE MODIFIED BY RADAR

Thus an accurate system stability test can be made with a 15 µs delay because the STALO noise contribution is within 1.2 dB of the measurement made in the laboratory on the STALO noise alone using conventional laboratory noise test equipment. The laboratory measurements are, of course, made

early in the design phase to predict the stability value.

VI. Conclusion

The use of the sapphire Bulk Acoustic Delay line as a practical tool for testing radar stability is now a proven technology and a 15 μs delay line is being incorporated into the built-in-test equipment (BITE) on a large family of S Band ground based Mil radars. Radar stability measurements in the 65 dB range have been made that include the SIAIO noise contribution to within 1.2 dB of that predicted by independent CW SIAIO noise measurements.

Acknowledgment

The author wishes to express appreciation to Richard M. Morrison for his help, especially in the computer analysis of the spectral power integration and the filter curves, and to Dr. Herman Blinchikoff for his careful review of this paper.

References

[1] Taylor, J.W., Jr., 1990 * Radar Handbook, edited by M.I. Skolnik, McGraw-Hill, New York, Chapter 3.5.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

THERMAL HYSTERESIS IN QUARTZ RESONATORS - A REVIEW

John A. Kusters* and John R. Vig**
*Hewlett-Packard Co.; **U.S. Army Electronics Technology and Devices Laboratory, LABCOM

Abstract

The frequency vs. temperature characteristic of a quartz crystal resonator does not repeat exactly upon temperature cycling, i.e., resonators exhibit "thermal hysteresis." This paper reviews the subject of thermal hysteresis. The subject has been studied only sporadically until recently. A search of the literature has revealed only about a dozen Frequency Control Symposium papers, a couple of contract reports, and a few other papers that deal with the subject. Books dealing with oscillators either do not mention the subject at all, or devote only a few sentences to the phenomenon.

The causes of hysteresis are not well understood. The evidence to date is inconclusive. The mechanisms that cause hysteresis include: contamination redistribution, strain changes, changes in the quartz, oscillator circuitry hysteresis, and apparent hysteresis due to thermal gradients.

Introduction

With the advent of the microcomputer compensated crystal oscillator (MCXO), thermal hysteresis has become the dominant factor limiting the stability achievable with temperature compensated oscillators. Since the next largest limiting factor is orders of magnitude smaller than hysteresis, future improvements in MCXO stability depend primarily on reducing the hysteresis. Similarly, as sensor technology has improved, hysteresis has become a limiting factor in the accuracies achievable with quartz resonator sensors.

The purpose of this paper is to review the subject of thermal hysteresis. The subject has been studied only sporadically until recently. A search of the literature has revealed only about a dozen Frequency Control Symposium papers,[1-6, 9, 11-15] a couple of Army contract reports,[16, 18] and a few other papers[19-23] that deal with the subject. Books dealing with oscillators either do not mention the subject at all, or devote only a few sentences to the phenomenon.

An ideal quartz crystal resonator's frequency vs. temperature characteristic can be described by a single-valued function. Frequency can then be uniquely determined from knowing the temperature. In "real-world" resonators, however, the frequency vs. temperature characteristic does not repeat exactly upon temperature cycling.

The lack of frequency vs. temperature (f vs. T) repeatability has been referred to at various times as "retrace," "hysteresis," "restart," and "thermal memory." No formal definitions existed before the advent of MIL-O-55310B, the military specification for crystal oscillators.[24] MIL-O-55310B defines "retrace" and "hysteresis." It distinguishes between the two terms by defining "retrace" as the nonrepeatability of the f vs. T characteristic, at a fixed temperature, upon on-off cycling an oscillator under specified conditions, and "hysteresis" as the maximum value of the nonrepeatability in the f vs. T characteristics during a temperature cycle, i.e., as the difference between the up-cycle and the down-cycle f vs. T characteristics at the temperature where that difference is maximum. Hysteresis is determined during a complete quasistatic temperature cycle between the specified temperature limits. "Quasistatic" means that the temperature is changed in such a manner as to ensure that the frequency offsets due to thermal gradients are much smaller than the specified f vs. T stability, including hysteresis.

"Retrace" is usually applied to specifying oven controlled crystal oscillators (OCXO), whereas "hysteresis" is usually applied to specifying temperature compensated crystal oscillators (TCXO). Fig. 1 illustrates TCXO hysteresis; Fig. 2 illustrates OCXO retrace.

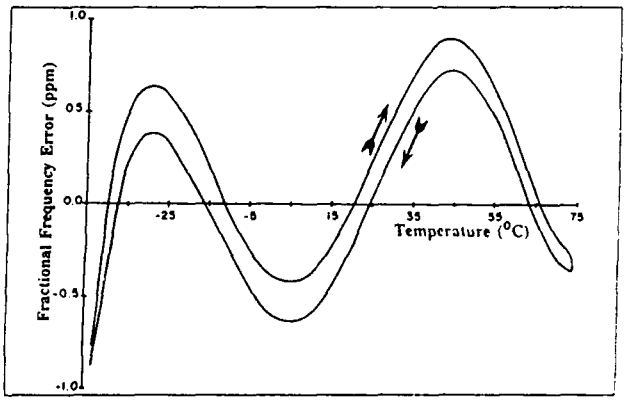


Figure 1. TCXO Thermal Hysteresis

US GOVERNMENT WORK IS NOT PROTECTED
BY US COPYRIGHT

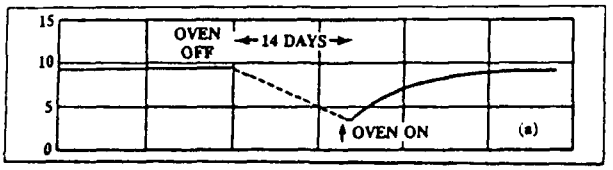


Figure 2. OCXO Retrace

A Review of the Literature

The following summarizes the literature on hysteresis. Included are three papers that deal with frequency vs. pressure hysteresis, as these may possibly have relevance to frequency vs. temperature hysteresis.

Early researchers who studied the stabilization and startup of precision ovenized resonators noted that, when the oven was turned off after stabilization and turned on at a later time, the frequency after restabilization at the original temperature was slightly different from the resonator's frequency just before oven shutdown,[25] as is illustrated in Fig. 2. According to Sykes,[26] et al., "This frequency shift has been attributed to sorption of minute quantities of gas on the crystal surface during an oven or oscillator shutdown." Also recognized as sources of instabilities during stabilization were "relaxation of strain, or defects in the quartz itself."

The "Union Thermoelectric Handbook" mentioned hysteresis as follows:[16] "Temperature changes can result in mechanical changes within the unit. For example, the mounting supports and bonding material may be altered more or less permanently by a change in temperature, resulting in a difference in the stress applied to the quartz plate. Some of the apparent hysteresis phenomenon [sic] have this origin."

Hammond et al. reported the first (and, until recently, the only) in-depth studies of hysteresis.[1, 19, 20] Their research was aimed at improving the performance of an LC-cut resonator that had been developed for a quartz thermometer. Although the resonators were cycled over various temperature ranges between +250°C and -200°C, the "hysteresis" was reported as the frequency shift at 0°C only (due to the high temperature coefficient of the LC-cut and the difficulty of obtaining accurate enough temperature measurements at other temperatures.) Fig. 3 shows the hysteresis of an LC-cut.

Hammond et al. observed that: 1) the fundamental-mode and third overtone c-modes exhibited nearly identical hysteresis, while the fundamental b- and a-modes exhibited a smaller hysteresis than the c-modes; 2) identically fabricated AT-cuts showed a five times smaller hysteresis than the LC-cuts and the hysteresis of the AT-cut was of opposite sign; 3) there was no systematic variation of hysteresis with electrode material or electrode thickness - the electrodes tried included both higher- and lower-than-quartz thermal expansion coefficient materials, and ductile as well as stiff materials; 4) there was no systematic variation of hysteresis with mounting type, stiffness, and rotation of the mounting orientation by 90°, or with the amplitude-of-vibration distribution; beveled plano-plano fundamental-mode resonators, contoured fundamental-mode resonators and third overtone resonators showed the same hysteresis; 5) there was no systematic variation of hysteresis with quartz type or concentration of defects - natural quartz, cultured quartz, optical grade cultured quartz, and swept quartz were tried; 6) there was no correlation with X-ray induced frequency changes; 7) there was a systematic variation of hysteresis with angle of cut in rotated Y-cuts - a change of sign of the hysteresis effect was found at $\theta = 32°$; and 8) although contamination in the resonator enclosure could aggravate the hysteresis, when proper contamination control measures were applied during processing, the remaining hysteresis was not related to contamination.

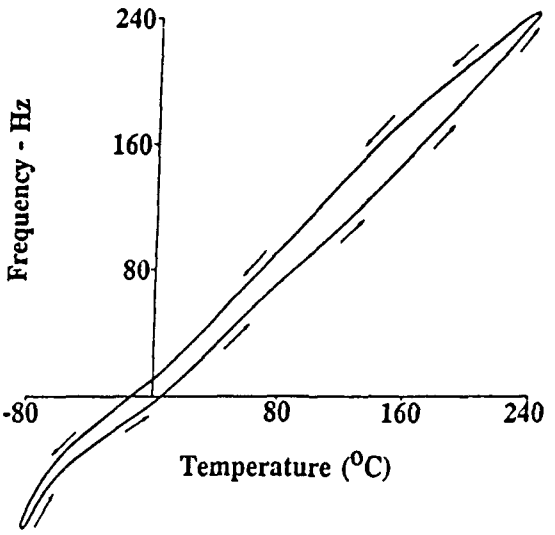


Figure 3. Typical Hysteresis Curve

Hammond et al.[1] concluded that "all the experiments to date indicate that the hysteresis effect in the quartz thermometer must be a property of crystalline quartz. However, it is not clear whether it is an intrinsic property of quartz or attributable to a defect structure. The variability from resonator to resonator and from observation to observation would infer a defect structure."

In a slightly later paper,[19] however, the authors state that hysteresis is "related to the differential expansion between the quartz plate and the thin film, as well as micro contamination [sic] within the crystal holder."

In 1970, Dick and Silver[2] showed the result for a single fundamental-mode 5 MHz crystal, the f vs. T characteristic of which repeated to 3 X 10^{-8} for one cycle between -40°C and +80°C.

Buroker and Frerking[3] developed a digitally compensated TCXO the f vs. T stability of which was ± 5 X 10^{-8} from -40°C to +80°C. In the conclusion of their report, however, they stated that, the "technique is capable of even greater stabilities, but even ± 5 X 10^{-8} cannot be maintained in practical environments due to thermal hysteresis in crystals and other components."

Mroch and Hykes[4, 17] evaluated a variety of purchased 4.5 MHz fundamental-mode AT-cut crystal units in connection with the development of a high stability TCXO. They found that the retrace at the lower turnover point varied from 3 ppm to less than 1 X 10^{-9}, "with no correlation from lot to lot or vendor," and that "...few of the crystals received met the ±3 X 10^{-8} retrace [sic] requirement" of the research program. They also found that the hysteresis was "dependent on the highest temperature to which the crystal had been recently exposed," it was maximum at the low temperature limit of the temperature cycle, as is illustrated in Fig. 4, and, after 30 minute exposures to 100°C, hysteresis was a function of storage time at room temperature. The ceramic capacitors used as bypass capacitors were identified as potential sources of TCXO hysteresis.

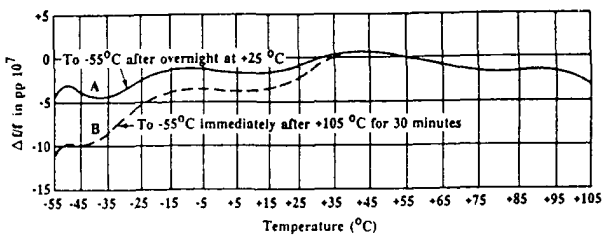


Figure 4. Hysteresis Effects in the Frequency-Temperature Charateristic of 4-MHz TCXO

Förster[5] measured the hysteresis of AT- and SC-cut 10 MHz 3rd overtone crystal units in HC-27/U glass enclosures, "from three German manufacturers." He found that "With nearly all the test specimens it is striking that quite large hysteresis values are obtained in the first cycle. The reproducibilities of the second and third cycles is in general excellent...namely ± 1 X 10^{-8}. In the first up run a 'calming phase' evidently sets in." He concluded that the hysteresis of "...SC-cut crystals measured were not vastly superior to those of AT-cut specimens," and that "The frequency stability attainable with digitally temperature-compensated quartz crystal oscillators is limited to around ..±1 X 10^{-7} solely by the thermal hysteresis of the crystal resonators."

Vig et al.[6] measured nine ceramic flatpack enclosed four- point-mounted SC-cut crystal units and two similarly fabricated AT-cut control units, then compared the hysteresis at the lower turnover temperatures. The worst-case hysteresis of the SC-cut units was 8.5 X 10^{-8}. The AT-cuts were about 10 times worse. The hysteresis was measured during two complete cycles between -45°C and +75°C.

Kaitz[7] and Kusters and Kaitz[8] studied the frequency vs. pressure hysteresis of BT-cut pressure transducer resonators. Kaitz found that "Hysteresis...increases con lerably as temperature and pressure are increased," and extensively cycling" the transducers, which "relaxes the residual stresses incurred during processing of the unit," improved hysteresis "in almost all instances...up to 50%." Kusters and Kaitz found a correlation between pressure hysteresis and the quartz material, and between the temperature of worst pressure hysteresis and the quartz material. Unswept natural quartz showed the lowest hysteresis. Sweeping natural quartz increased the hysteresis. Pure Z-growth material showed a hysteresis similar to that of swept natural quartz. The highest hysteresis was found when X-growth material was used. Anomalously high hysteresis in some "X-growth" transducers was traced to the fact that the transducers "were found to contain regions of Z-growth material and clearly defined X-Z growth boundaries." No correlation was found between inclusion density and hysteresis.

Ueda et al.[9] reported that when a tuning fork thermometer resonator was made relatively stress free by making a "narrow neck between the resonator support and vibrating beams to further reduce the stress transmitted from the support to the vibrating beams...," the hysteresis, when cycled between 4.2°K and 0°C, was reduced by a factor of two at 4.2°K and a factor of three at 0°C, when compared to the conventional tuning fork, and, at 0°C, it was "one tenth that of thickness shear-mode resonator." No details are provided with respect to the number of units tested or the type of thickness shear resonator that was used for the comparison.

Tuladhar, et al.[21] studied the hysteresis of 5 and 10 MHz AT-cut resonators with gold electrodes and with silver electrodes. Their conclusions were that: "Hysteresis at the lower turn-over- temperature is higher than that at the upper turn-over- temperature...Hysteresis for gold

electroded crystals is considerably worse than that for silver electroded crystals...Crystals having a double adjusting layer show greater hysteresis than those with a single layer."

Ward and EerNisse[10] found a correlation between pressure hysteresis and quartz material, as did Kusters and Kaitz. Transducers made of pure Z-growth cultured quartz exhibited higher hysteresis than those made from natural quartz.

Tartakovskii[22] analyzed the results of Hammond et al. and of U.S.S.R. researchers Varfolomeeva et al. (reference 2 of Tartakovskii[22]). He states that "...by taking into account the two different manifestations of thermal inertia in a piezoelectric element it is possible to explain all experimental results" of Hammond et al. One manifestation of "thermal inertia" is "the lag of the average temperature T_p in the piezoelectric element behind that of the medium T_m by ΔT_1," and the other is the "difference ΔT_2 in temperatures at the edge and the center of a piezoelectric element creates in the central region of the plate compressive stresses..." which produce a frequency shift that is a function of the plate's angles of cut and mounting orientation. He also explains the AT-cut hysteresis results of Varfolomeeva et al. by showing that the results are consistent with frequency changes to be expected from "a change of the mechanical stresses in the ring of silver paste" that was used to bond the mounting clips to the quartz plate. His "calculated results showed why a simple turn in the direction of the ring towards the axis Z' (i.e., a shift from $\Psi = 45°$ to $\Psi = 75°$) caused a sharp reduction of frequency hysteresis following a temperature treatment..."

Beaussier[23] drew a similarity between mechanical hysteresis and thermal hysteresis and concluded that "the phenomenon of non-linear elasticity in the presence of crystalline faults, and in particular moving dislocations appear to be the cause." He further reported that continued thermal cycling, after proper selection of the quartz itself, reduces thermal hysteresis. Beaussier reported the development of a model which explained observed experimental data where: $\Delta f/f \approx K_{II} (\Delta T)^2$ where K_{II} equals 0.5 X $10^{-10}/°C^2$ for SC-cuts, and 1.5 X $10^{-10}/°C^2$ for AT-cuts.

Filler[11] analyzed TCXO thermal hysteresis and showed that thermal lag between the resonator and the TCXO's thermometer is a major contributor to the observed TCXO hysteresis. He developed a model "that accounts for both the normally-encountered and anomalous thermal hysteresis. This model can separate apparent hysteresis from 'true hysteresis'." He showed that "the inherent hysteresis of the resonator is much lower than that of the TCXO and has a different temperature dependence," and that "thermal hysteresis is independent of rate of change of temperature, when thermal lag is accounted for..."

Symonds and Wacker[18] encountered excessive f vs. T hysteresis during the initial stages of a TCXO development program. They attempted to vary the design and fabrication of the (10 MHz fundamental-mode AT-cut) resonators in order to reduce the hysteresis. The results to date indicate that the resonator's mounting plays an important role in the observed hysteresis. When the stiffness of the mounting clips and the bonding areas of four-point mounted resonators were reduced significantly, the average hysteresis was reduced from 0.64 ±0.35 ppm to 0.20 ±0.07 ppm.

Filler and Vig[12] developed dual-mode SC-cut resonators for microcomputer compensated crystal oscillators (MCXO). They, together with Schodowski,[13] showed that when the dual-mode self- temperature sensing method[13] is used as the "thermometer" during f vs. T measurements, the effects of thermal gradients can be made negligible even when the temperature is changed rapidly. In the self-temperature sensing method, the fundamental-mode and third overtone c-mode frequencies are excited simultaneously. The frequencies of these c-modes are measured vs. a thermometric beat frequency that is derived from the two c-modes. The resonator acts as its own thermometer; no external thermometer is needed.

Filler and Vig applied the dual-mode self-temperature sensing method to measuring the hysteresis of 10 MHz/3.3 MHz dual-mode SC-cut resonators in a -55°C and +85°C temperature range. "The measured hysteresis ranged from parts in 10^9 for the best units, to about 2 X 10^{-8} for the typical 'good' unit, to several parts in 10^8 for the 'bad' units." Filler[14] continued these hysteresis studies and showed that for 10 MHz 3rd overtone SC-cut resonators obtained from four manufacturers: 1) hysteresis varies with temperature excursion, but for a given temperature cycle, the hysteresis repeats, 2) hysteresis is not always worse at low temperatures, 3) resonator manufacturing lots exhibit "signatures," i.e., resonators within a lot showed similar hysteresis characteristics, but the characteristics varied significantly from lot to lot, and 4) a factor of two change in drive current did not affect the hysteresis.

Filler, Messina and Rosati[15] studied the performance, including hysteresis, of microcomputer compensated crystal oscillators (MCXO) that used dual-mode SC-cut resonators similar to those developed by Filler and Vig.[12]

Hysteresis of the seven MCXO's evaluated ranged from $\pm 5 \times 10^{-9}$ to $\pm 4 \times 10^{-8}$.

Benjaminson[27] investigated circuit contributions to oscillator hysteresis. He analyzed the effects of hysteresis in the most critical components of a dual-mode crystal oscillator (for use in a MCXO). One oscillator of the pair was a 10 MHz 3rd overtone bridge oscillator, the other a 3.4 MHz fundamental-mode impedance-inverting Colpitts oscillator. He found that a 1% change in the tuned circuit inductance of the bridge oscillator changed the frequency of oscillation by 1×10^{-8}, while a similar variation in the PI-network inductance of the Colpitts oscillator caused a change of 7×10^{-8}. Since both oscillators are series resonant circuits, additional tests were performed in order to isolate the effects of circuit component hysteresis from crystal hysteresis, by replacing the crystal with fixed resistors in each oscillator, enabling operation as L-C oscillators at the nominal operating frequencies.

Hysteresis measured during cycling between -55°C and +85°C was typically less than 50 ppm, which translated to a contribution less than 1×10^{-10} f vs. T hysteresis in the 10 MHz oscillator (when operating as a crystal oscillator) and less than 6×10^{-10} hysteresis in the 3.4 MHz crystal oscillator.

The low hysteresis effects demonstrated by the L-C components were achieved by careful analysis to select the most stable inductors and capacitors available and equally careful circuit analysis to minimize frequency pulling by the reactances. To accomplish this, the bridge oscillator was designed with as wide an L-C bandwidth as b-mode rejection permitted; b-mode traps per se, were avoided because their pulling effects are much worse.

The impedance-inverting Colpitts oscillator was designed with the lowest possible reactance values. This was limited by restrictions on transistor current consumption. Higher current operation can provide higher transconductance, permitting lower reactance with a consequent reduction in frequency pulling. An added benefit, particularly in the low resistance fundamental-mode oscillator, was produced by the lower equivalent series resistance of small, high Q inductors that reduce loaded Q degradation and further minimize pulling due to reactance changes.

Hysteresis Mechanisms

So what causes hysteresis? Since the evidence reported to date is inconclusive, we will discuss the various mechanisms that can possibly cause the phenomenon.

Contamination Redistribution - Adsorption-desorption phenomena can cause hysteresis if, during temperature cycling, contamination inside the resonator enclosure is redistributed so as to change the mass loading on the active area of the resonator. Sykes, et al.[28] stated that "A study of the stabilization characteristics of precision oscillators following interruptions of oscillation lead to the conclusion that residual contamination within the crystal unit enclosure is the most likely cause of frequency change during the first several days of operation after an interruption." Armstrong, et al.[29] showed that the retrace of "clean" 5 MHz thermocompression-bonded, high-temperature processed resonators was superior to that of similar but solder-bonded resonators. Hammond, et al.[1] stated that "with inadequate vacuum baking or inadequate cleaning of the crystal mounts, header, or can, the hysteresis effect can be aggravated. However, ...the remaining hysteresis effects are not related to contamination."

When one examines precision resonators' hysteresis curves, it is difficult to see evidence that contamination transfer is a significant factor. Desorption rates generally have an exponential dependence on temperature. Therefore, if adsorption-desorption phenomena played a major role, then hysteresis would show a strong temperature dependence, which has not been reported. When surfaces are heated to produce a uniformly rising temperature, desorption (of a single adsorbent) occurs in a narrow temperature interval, with a pronounced peak.[30] The temperature of maximum desorption rate is a function of desorption energy, and is often used to investigate the interaction of gases with metal surfaces. The characteristic "moisture dip" observed when resonator enclosures contain water vapor is a manifestation of adsorption and desorption occurring in a narrow temperature range. In order for contamination transfer to explain the observed hysteresis, a variety of contaminant molecules, with an appropriate range of adsorption energies, would need to be present in the resonator enclosure.

Ideally, in order to eliminate adsorption-desorption as a hysteresis mechanism, the electrodes should either be highly active or inert. In the first case, all the contamination would be permanently adsorbed (i.e., the contaminant's lifetime on the surface would be infinite). In the second case, none of it would be adsorbed (i.e., the contaminants' lifetime would be zero). In real resonators, however, the contaminants' lifetimes on surfaces are finite. The lifetimes depend on the surfaces, contaminant molecules and temperatures.

Hysteresis in isothermal adsorption-desorption (as a

function of gas pressure) has been reported, e.g., during the adsorption and desorption of water on the gold electrodes of quartz crystal resonators.[31] Such hysteresis, being a small perturbation of a small effect, is probably a negligible second-order effect in f vs. T hysteresis.

Strain Changes - It is well known that changes in the stresses on a resonator plate can produce frequency shifts. The stresses experienced by resonators include mounting stresses (via the force-frequency effect[32, 33] and bending effects[34]), bonding stresses,[35] and electrode stresses.[36] It is clear that temperature cycling can produce changes in these stresses, and can, thereby, result in hysteresis.

If the mounting clips were perfectly elastic, or perfectly soft, then they would not contribute to hysteresis. If, however, the clips undergo stress relief during temperature cycling, then hysteresis can result. The magnitude of the hysteresis produced by a given amount of stress relief is a function of the orientation of the mounting clips with respect to the crystallographic axes of the quartz plate,[32] and the types of stresses.

For in-plane diametric forces, the force-frequency coefficient K_f vs. azimuth angle Ψ have been found to have zeroes for all the commonly used cuts, such as the AT- and SC-cuts.[32] Therefore, one might conclude that hysteresis due to stress relief in the mounting clips can be eliminated by mounting the crystals where $K_f = 0$. Unfortunately, it is not possible to completely eliminate the effects of mounting stresses in conventional resonators for the following reasons. First, the azimuthal angles where $K_f = 0$ are functions of temperature,[33] so that there is no Ψ where $K_f = 0$ over the whole temperature range of a TCXO or MCXO. Second, the Ψ where the effects of bonding stresses are zero is different from the Ψ where $K_f = 0$, at least for the AT-cut, the only cut for which bonding stress effects have been reported.[35] Third, the forces due to the mounting clips are generally not purely in-plane diametric forces. This is especially true for three and four-point mounted resonators because, since the thermal expansion coefficient of quartz is anisotropic whereas that of the typical package base is isotropic, the forces due to temperature cycling will have tangential components. Similarly, for two-point mounted resonators, the base's change of dimensions during temperature cycling will apply shear-type forces in addition to the in-plane diametric forces.

Theoretically, properly mounting the quartz resonator plate on an identically oriented quartz plate, as is attempted in BVA resonators,[37] ought to greatly reduce the hysteresis due to stress relief in the mounting structure. Our literature search did not reveal evidence that BVA resonators exhibit hysteresis or retrace that is superior to those found in high quality conventional resonators.

For an example of the frequency changes that can be caused by stress relief, consider a 5 MHz 3rd overtone, 14 mm diameter resonator. If one were to intentionally mount this resonator at the Ψ where K_f is maximum, then the frequency shifts due to changes in the in-plane diametric forces would be 2.9×10^{-8} per gram for an AT-cut resonator, and 1.7×10^{-8} per gram for an SC-cut resonator[38] (where "per gram" refers to the force due to a one gram weight, on earth).

The effects of electrode stress relief can be minimized by using the SC-cut,[36] or by not having electrodes in contact with the active area of the resonator, as can be done with BVA-type resonators,[37] and, to a lesser extent, with lateral field resonators.[39] Everything else being equal, using a single metal (e.g., Au) as opposed to two or more metals (e.g., Cr-Au, Ti-Pd-Au) is also likely to produce lesser stress relief effects because using two or more metals introduces additional interfaces where stress relief and diffusion can occur.

Changes in the Quartz - Changes in the quartz due to the stresses induced by temperature cycling are among the conceivable causes of hysteresis, although no direct evidence of such changes could be uncovered in the literature. Perfect quartz would not be expected to be affected by temperature cycling. The imperfections that are subject to change include surface defects, dislocations, impurities, inclusions, and twins.

Surface defects, such as the microcracks produced by lapping, can change upon temperature cycling,[40] however, by properly etching the surfaces[41] subsequent to mechanical treatment, the possibility of changes can be greatly reduced or eliminated.

That dislocation motion due to temperature cycling is a factor in hysteresis is unlikely at the typical TCXO temperatures. Even in sweeping experiments[42] which are usually conducted far above the normal operating temperatures of oscillators, no evidence of dislocation motion has been reported. The energy needed to anneal quartz damage due to neutron irradiation may be a clue to the energies needed to move dislocations. When quartz is irradiated with fast neutrons, displacement damage occurs. At high doses, the quartz gradually becomes disordered into an amorphous form. Annealing studies on neutron damaged quartz indicate that the annealing temperature of quartz is above the inversion temperature.[43-45] The activation energy for structure annealing is 0.75 eV.[43]

It takes a great deal less energy, however, to move impurities in quartz to new lattice sites. Impurity motion due to temperature cycling is, therefore, a more probable hysteresis mechanism than dislocation motion. It is possible to induce impurity dependent effects, at "normal" temperatures. For example, some radiation- induced effects can anneal,[46] low DC voltages produce changes in the aging rates of resonators,[47] and sweeping can be produced at room temperature (≈23°C) when 25 V/mm is applied to a quartz bar.[48] Kusters,[49] observed that when a 500V/mm DC field was applied to a doubly rotated quartz plate at 80°C, the frequency shift decayed with a time constant of 7 seconds. Moreover, there is evidence in the literature that strain gradients due to ultrasonic vibrations can produce changes in quartz properties, probably due to the motion of impurities down the gradients.[50]

Recent experimental evidence developed at Hewlett-Packard seems to support material defects being a significant mechanism. In two experimental groups of 38 LC-cuts each (see Hammond[1]), significant differences were observed which correlate with the quartz material. All fabrication processes other than the material were identical. Optical grade quartz was irradiated to show the presence of defect centers. LC-cuts made from darkened areas of the quartz formed the first group. LC-cuts from totally clear areas formed the second group. As shown in Figure 5, Group 1 had a total yield of 34% with a high percentage of dead units and hysteresis rejects. Group 2 had a total yield of 84% with no hysteresis rejects. These crystals are operated on the end of a 1/2 wave cable. Dead units are those for which the effective motional resistance becomes so high at a particular temperature that the oscillator stops.

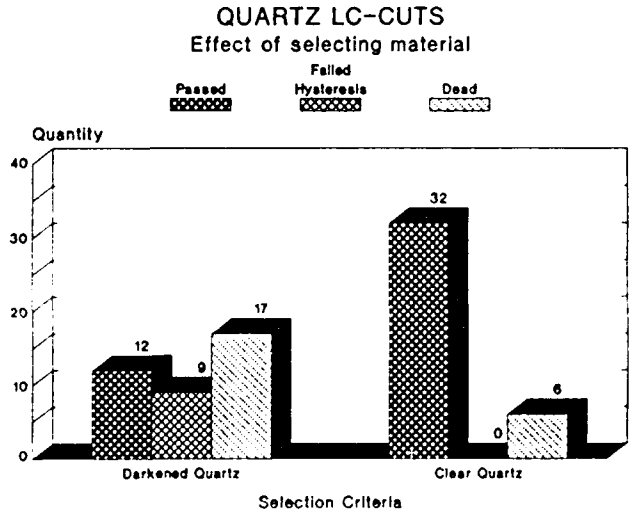


Figure 5. Quartz Thermometer Crystals, 28.2 MHz, 3rd Overtone. Group 1: Darkened Quartz, Yield 34%; Group 2: Clear Quartz, Yield 84%

In a continuation of the work reported in Kusters and Kaitz,[8] normal sweeping was done on one group of natural quartz specimens, and extended sweeping done on a second group. All other processing parameters were identical. Sweeping was continued on the second group until the indicated ion current became stable and remained constant for at least 48 hours. Resistance data of finished pressure transducer resonators are shown in Figures 6 and 7. There is a sharp difference of the measured motional resistance. There is also a sharp difference in the overall yield. Group 1 (Figure 6) has at present a maximum yield of 67%. Group 2 (Figure 7) has at present a 100% yield. Final testing is not yet complete but the trend is significant.

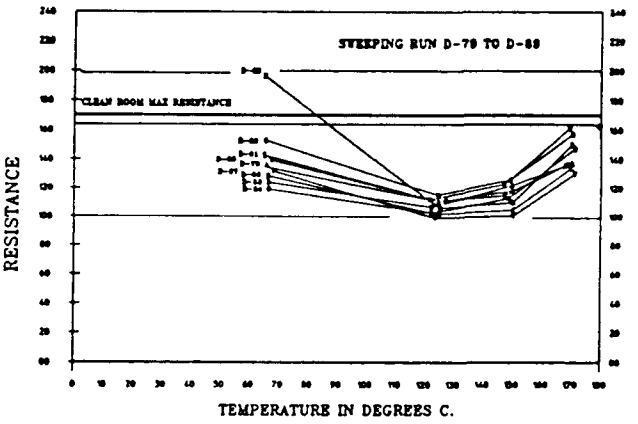


Figure 6. Quartz Pressure Transducer Resonator Resistance Readings, Standard Electrodiffusion Sweeping (5 MHz, 3rd Overtone, BT-Cuts)

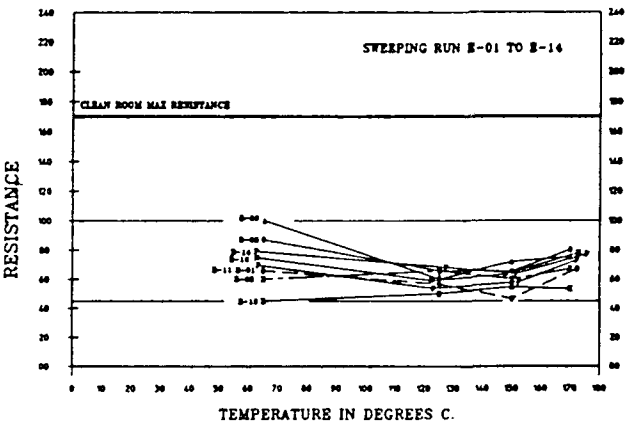


Figure 7. Quartz Pressure Transducer Resonator Resistance Readings, Enhanced Electrodiffusion Sweeping (5 MHz, 3rd Overtone, BT-Cuts)

Experimentally, in quartz LC-cuts and in the quartz pressure transducer resonators, the hysteresis decreases as a result of continued thermal cycling thus confirming the observations of Beaussier[23] that improved hysteresis performance can result from thermal cycling.

Twins in a quartz plate can change under the influence of (high) stresses. For example, when a sufficiently high stress is applied (0.5 GPa), it is even possible to switch a single-domain crystal to a single domain of opposite polarity. The removal of the stress leads to a complete restoration of the original orientation. The stress-strain relations in a "ferrobielastic" material such as quartz can show hysteresis due to ferrobielastic switching resulting from Dauphiné twin formation.[51] Twinning phenomena, however, are unlikely to be a significant hysteresis mechanism because resonators that contain (macroscopic) twins would be rejected during the usual inspection and testing procedures, because changes in twins tend to occur suddenly, and because the stresses needed to move twin boundaries are probably higher than the stresses due to temperature cycling. Changes in microtwins can possibly occur at high pressures, in pressure transducers.

Other interesting thermal hysteresis phenomena in quartz take place near 573°C, during the α-β transition,[52, 53] in the incommensurate phase of quartz. Various physical properties of quartz, such as the heat capacity and refractive index, exhibit thermal hysteresis in this phase.

Apparent Hysteresis - Filler[11] showed conclusively that when the thermometer in a compensated oscillator is separated from the resonator (as it is in TCXO), the thermal gradients between the thermometer and the resonator can produce "apparent" hysteresis. The apparent hysteresis can be eliminated by using dual-mode self-temperature sensing.[12, 13]

Angles-of-cut Dependence - Early data on SC-cut crystals seemed to indicate that the SC-cut has lower hysteresis than AT and BT-cuts. A recent experiment on 30 SC-cut, 10 MHz, 3rd overtone units and on 15 BT-cut, 5 MHz, 3rd overtone units in identical enclosures produced the data shown in Table 1. The material in all was unswept cultured quartz with an IR "Q" of nominally 2.5 million. The units had been cycled from 30°C to 150°C to 30°C. Hysteresis of either sign was seen in both batches. The maximum seen was about ±5 X 10^{-8}. This is essentially the same as reported in other work.[12-15] The experimental data is shown in Fig. 8.

Oscillator Circuit Hysteresis - Hysteresis in circuit components can cause oscillator hysteresis. For example, if a 20 pF load capacitor C_L is in series with a resonator and the resonator's C_1 = 14 fF and C_0 = 5 pF, then a 5 X 10^{-4} hysteresis in C_L produces a 1 X 10^{-7} f vs. T hysteresis.[38]

Table 1. Hysteresis and Orientation

	SC-Cut	BT-Cut
Absolute Mean	1.1 X 10^{-8}	1.1 X 10^{-8}
Standard Deviation	1.3 X 10^{-8}	1.2 X 10^{-8}

Inductors are notorious for their instabilities; e.g., the windings of inductors can stretch and move due to the stresses experienced during temperature cycling. It is possible to minimize the circuit contributions to hysteresis by appropriate resonator and circuit design, and circuit component selection. For example, the need for inductors in the b-mode trap of SC-cut oscillators can be eliminated with lateral field resonator designs that suppress the b-mode.[54]

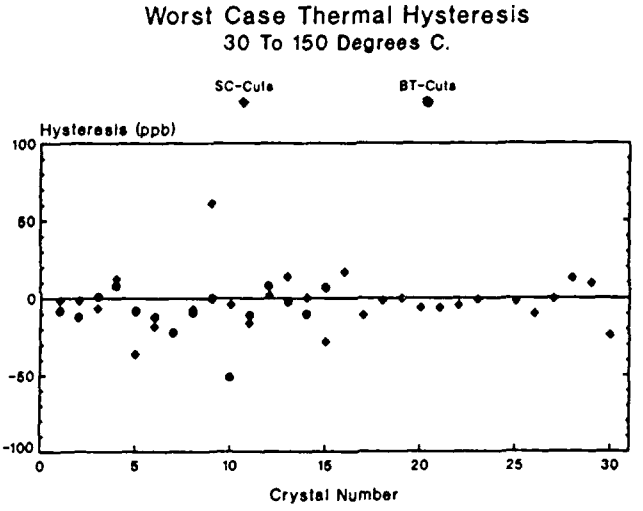


Figure 8. Experimental Hysteresis Data: SC-Cut Resonators are 10 MHz, 3rd Overtone; BT-Cut Resonators are 5 MHz, 3rd Overtone

Summary and Conclusions

The results to date seem to indicate that lattice defects are somehow related to thermal hysteresis. Stress relief in the mounting structure can also produce significant hysteresis. As crystal processing techniques have improved, contamination has become less of a problem. This is shown in Fig. 9. The points represent a rough mean of the published data and demonstrate a two-order-of-magnitude reduction in the observed levels of hysteresis during the past 25 years.

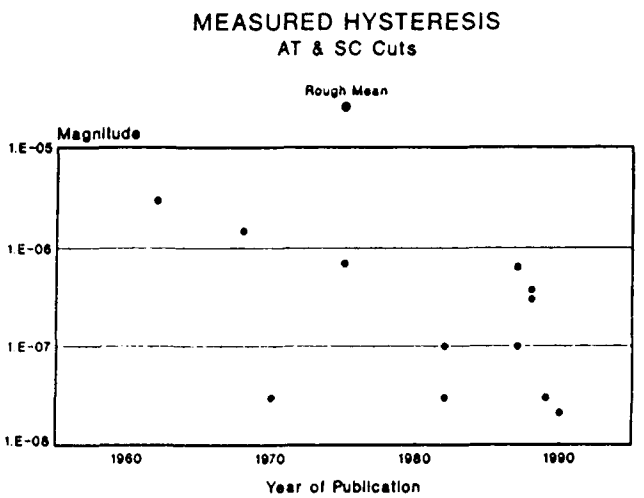


Figure 9. Hysteresis Data Quoted in Publications (Rough Averages Computed when not Available in Original Source)

That parts in 10^9 hysteresis has been observed in some resonators is encouraging. Once the effects of material defects and strain relief due to the mounting structure are better understood, it may be reasonable to expect parts in 10^9 hysteresis in a reproducible manner. MCXO's of parts in 10^9 f vs. T stability will then be obtainable.

Acknowledgements

The authors thank Charles Adams, Jim Collin, and Leo Steindorf of Hewlett-Packard for sharing the results of the hysteresis reduction experiments in the LC-cut and pressure sensor, and for making the measurements on the SC-cut and BT-cut resonators, and Al Benjaminson of General Technical Services, Inc. for editing the section that reviews his work.

References

[1] D. L. Hammond, C. A. Adams, and A. Benjaminson, "Hysteresis Effects in Quartz Resonators," Proc. 22nd ASFC, pp. 55-66, 1968.

[2] L. A. Dick, and J. F. Silver, "Low Aging Crystal Units for Use in Temperature Compensated Oscillators," Proc. 24th ASFC, pp. 141-146, 1970.

[3] G. E. Buroker, and M. E. Frerking, "A Digitally Compensated TCXO," Proc. 27th ASFC, pp. 191-198, 1973.

[4] A. B. Mroch, and G. R. Hykes, "A Miniature High Stability TCXO Using Digital Compensation," Proc. 30th ASFC, pp. 292-300, 1976.

[5] H. J. Förster, "Thermal Hysteresis of AT + SC-Cut Quartz Crystal Resonators: Automated Measurement Method and Results," Proc. 36th ASFC, pp. 140-147, 1982.

[6] J. R. Vig, R. L. Filler, and J. A. Kosinski, "SC-Cut Resonators for Temperature Compensated Oscillators," Proc. 36th ASFC, pp. 181-186, 1982.

[7] G. Kaitz, "Extended Pressure and Temperature Operation on BT-cut Pressure Transducers," Proc. 38th ASFC, pp. 245-250, 1984.

[8] J. A. Kusters, and G. S. Kaitz, "Characteristics of Natural, Swept Natural, and Cultured X- and Z-Growth Quartz Material in High Temperature, High Stress Applications," Proc. 39th ASFC, pp. 223-229, 1985.

[9] T. Ueda, F. Kohsaka, T. Iino, and D. Yamazaki, "Temperature Sensor Using Quartz Tuning Fork Resonator," Proc. 40th ASFC, pp. 224-229, 1986.

[10] R. W. Ward, and E. P. EerNisse, "A Reduced Hysteresis, Extended Range Quartz Pressure Transducer," Proc. 41st ASFC, pp. 344-349, 1987.

[11] R. L. Filler, "Measurement and Analysis of Thermal Hysteresis in Resonators and TCXO's," Proc. 42nd ASFC, pp. 380-388, 1988.

[12] R. L. Filler, and J. R. Vig, "Resonators for the Microcomputer Compensated Crystal Oscillator," Proc. 43rd ASFC, pp. 8-15, 1989.

[13] S. S. Schodowski, "Resonator Self-Temperature-Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," Proc. 43rd ASFC, pp. 2-7, 1989.

[14] R. L. Filler, "Thermal Hysteresis in Quartz Crystal Resonators and Oscillators," elsewhere in these Proceedings.

[15] R. L. Filler, J. A. Messina, and V. J. Rosati, "Frequency-Temperature and Aging Performance of Microcomputer Compensated Crystal Oscillators," Proc. 43rd ASFC, pp. 27-33, 1989.

[16] R. E. Bennett, ed., "Quartz Resonator Handbook - Manufacturing Guide for "AT" Type Units," prepared for the Dept. of the Army by Union Thermoelectric Div., Contract DA36-039-SC-71061, p. 195, 1960, AD 251289.

[17] A. Mroch, and G. Hykes, "High Stability Temperature Compensated Crystal Oscillator Study," Research and Development Technical Report ECOM-73-0137-F, Final Report, pp. 33, 36-42, 63-70, February 1977, AD A036908.

[18] D. Symonds, and M. Wacker, "MACI Adaptation of Moderate Precision TCXO," Research and Development Technical Report SLCET-TR-86-LO94-1, First Interim Report, pp. 1-10, 13-14, 23, 37-38, 63-70, July 1988, AD B126579. The final report is due to be published in late-1990.

[19] D. L. Hammond, C. A. Adams, and A. Benjaminson, "Hysteresis Effects in Quartz Resonators," Frequency Technology, Vol. 7, No. 1, pp. 19-22, January 1969.

[20] D. L. Hammond, and A. Benjaminson, "The Crystal Resonator - A Digital Transducer," IEEE Spectrum, Vol. 6, No. 4, pp. 53-58, April 1969.

[21] K. K. Tuladhar, B. J. Adams, and E. D. Fletcher, "Hysteresis Effects in Quartz Crystal Resonators," Proc. IERE (British) Int'l Conf. on Frequency Control & Synthesis; U. Surrey, Guildford, pp. 19-27, April 1987.

[22] I. I. Tartakovskii, "Temperature Hysteresis of Frequency in Quartz Resonators," Meas. Tech. (USA) Vol. 30, No. 6, pp. 563-566, June 1987, translated from Izmeritel'naya Tekhnika (USSR) Vol. 30, No. 6, pp. 39-41, June 1987.

[23] J. Beaussier, "Hysteresis Thermique dans les Resonateurs et les Oscillateurs a Quartz," Proc. 2nd European Frequency and Time Forum, Neuchatel, pp. 803-815, 16-18 March 1988.

[24] MIL-O-55310B, Military Specification, Oscillators, Crystal, General Specification for, 10 May 1988; this specification is currently undergoing a significant revision. Copies of the latest published revision of MIL-O-55310 are available from Naval Publications and Forms Center, 5801 Tabor Ave., Philadelphia, PA 19120.

[25] T. C. Anderson, et al., "Ground Station Frequency Standard," Bell Telephone Lab. Final Report, RADC-TR-58-83, 9 January 1958, AD 148769.

[26] R. A. Sykes, W. L. Smith, and W. J. Spencer, "Studies on High Precision Resonators," Proc. 17th ASFC, pp. 4-27, 1963.

[27] A. Benjaminson, "Advanced Crystal Oscillator Design," Final Report for Army Contract DAAL01-88-C-0804, to be published in late-1990.

[28] R. A. Sykes, W. L. Smith, and W. J. Spencer, "Performance of Precision Quartz-Crystal Controlled Frequency Generators," IRE Transactions on Instrumentations, Vol. I-11, pp. 245-247, 1962.

[29] J. H. Armstrong, P. k. Blomster, and J. L. Hokanson, "Aging Characteristics of Quartz Crystal Resonators," Proc. 20th ASFC, pp. 192-207, 1966.

[30] R. Glang, et al., "High-vacuum Technology," in Handbook of Thin Film Technology, edited by L. I. Maissel, and R. Glang, pp. 2-39 to 2-49, 1970.

[31] J. H. Thomas, III, and S. P. Sarma, "Adsorption and Desorption of Water on Au by the Quartz-Crystal-Oscillator Method," J. Vac. Sci. Technol., Vol. 13, pp. 549-551, 1976.

[32] A. Ballato, E. P. EerNisse, and T. Lukaszek, "The Force-Frequency Effect in Doubly Rotated Quartz Resonators," Proc. 31st ASFC, pp. 8-16, 1977.

[33] E. P. EerNisse, "Temperature Dependence of the Force Frequency Effect for the AT-, FC-, SC-, and Rotated X-cuts," Proc. 34th ASFC, pp. 426-430, 1980.

[34] E. D. Fletcher and A. J. Douglas, "A Comparison of the Effects of Bending Moments on the Vibrations of AT and SC (or TTC) Cuts of Quartz," Proc. 33rd ASFC, pp. 346-350, 1979.

[35] R. L. Filler, and J. R. Vig, "The Effect of Bonding on the Frequency vs. Temperature Characteristics of AT-cut Resonators," Proc. 30th ASFC, pp. 264-268, 1976.

[36] E. P. EerNisse, "Quartz Resonator Frequency Shifts Arising from Electrode Stress," Proc. 29th ASFC, pp. 1-4, 1975.

[37] R. J. Besson, and U. R. Peier, "Further Advances on B.V.A. Resonators," Proc. 34th ASFC, pp. 175-182, 1980.

[38] J. R. Vig, "Quartz Crystal Resonators and Oscillators - For Frequency Control and Timing Applications - A Tutorial," Technical Report SLCET-TR-88-1 (Rev. 3.0), January 1990, AD A218090. A later edition of this document is in print, copies are

available from the author.

[39] A. Ballato, E. R. Hatch, M. Mizan, T. Lukaszek, and R. Tilton, "Simple Thickness Modes Driven By Lateral Fields," Proc. 39th ASFC, pp. 462-472, 1985.

[40] B. M. Darinskii, N. V. Izmailov, et al., "Anelastic Relaxation in Solids Caused by Surface Damage," Sov. Phys. Solid State, Vol. 29, pp. 2023-2025, 1987.

[41] J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemically Polished Quartz," Proc. 31st ASFC, pp. 131-143, 1977.

[42] D. W. Hart, J. Frank, D. Smith, J. J. Martin, and J. G. Gualtieri, "A Study of the Electrodiffusion Process in Quartz," elsewhere in this Proceedings volume.

[43] F. B. Johnson, and R. S. Pease, "The Pile Irradiation of Quartz Crystal Oscillators," Phil. Mag., Vol. 45, pp. 651-654, 1954.

[44] W. Primak, "Fast-Neutron-Induced Changes in Quartz and Vitreous Silica," Phys. Rev., Vol. 110, pp. 1240-1254, 1958.

[45] J. C. King, and D. B. Fraser, "Effects of Reactor Irradiation on Thickness Shear Crystal Resonators," Proc. 16th ASFC, pp. 8-31, 1962.

[46] J. J. Martin, "Aluminum-Related Acoustic Loss in AT-cut Quartz Crystals," Proc. 38th ASFC, pp. 16-21, 1984.

[47] R. L. Filler, J. A. Kosinski, V. J. Rosati, and J. R. Vig, "Aging Studies on Quartz Crystal Resonators and Oscillators," Proc. 38th ASFC, pp. 225-231, 1984.

[48] J. G. Gualtieri, private communication, April 1990.

[49] J. A. Kusters, "The Effect of Static Electric Fields on the Elastic Constants of α-Quartz," Proc. 24th ASFC, pp. 46-54, 1970.

[50] R. R. Sharp, and E. L. Pace, "Direct Observation of Impurity Motion in Quartz Using the Raman Effect," J. Phys. Chem. Solids, Vol. 31, pp. 2275-2279, 1970.

[51] E. Bertagnolli, E. Kittinger, and J. Tichy, "Ferrobielastic Hysteresis in α-Quartz," J. Appl. Phys., Vol. 50, pp. 6267-6271, 1979.

[52] P. Bastie, and G. Dolino, "γ-ray-diffraction Study of an Incommensurate Phase: Application to Quartz," Phys. Rev. B, Vol. 31, pp. 2857-2861, 1985.

[53] M. Matsuura, H. Yao, K. Gouhara, I. Hatta, and N. Kato, "Heat Capacity in α - β Phase Transition of Quartz," J. Phys. Soc. Japan, Vol. 54, pp. 625-629, 1985.

[54] M. Bloch. M. Meirs, and J. Ho, "The Microcomputer Compensated Crystal Oscillator (MCXO)," Proc. 43rd ASFC, pp. 16-19, 1989.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

THERMAL HYSTERESIS in QUARTZ CRYSTAL RESONATORS and OSCILLATORS

Raymond L. Filler
US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703

Abstract

With the introduction of the microcomputer compensated crystal oscillator (MCXO), the technology of compensation for frequency vs. temperature (f-T) variations has advanced to the state where thermal hysteresis is the largest contributor to frequency error. In this paper data are presented on the hysteresis of resonators and oscillators. The dual-c-mode self-temperature sensing technique was used for thermometry. Temperature measurement inaccuracies are eliminated with this technique because the resonator and the thermometer are one and the same. Measurements were made using both a passive "pi" network/vector-voltmeter system, and dual-mode oscillators of two designs.

The experimental variables investigated were 1) the end point temperatures, 2) the rate of change of temperature, 3) the sequence of temperatures (i.e., low-high-low, high-low-high, or mid-low-high-mid etc.), and 4) the drive level. Resonators were also measured over several different cycles in both a "pi" network and a dual-mode oscillator.

The results show that the thermal hysteresis of the crystal resonators studied is independent of temperature sequence and drive level but is dependent on the temperature extremes. The contribution to the thermal hysteresis from a poorly designed oscillator may be larger than the contribution from the resonator. Resonators from several different manufacturers and different lots from the same manufacturer exhibit thermal hysteresis signatures.

Introduction

The limiting factor in temperature-compensated oscillators is the non-repeatability of the frequency vs. temperature characteristic. All compensation schemes rely on a one-to-one correspondence between temperature and frequency. Figure 1 shows the fractional frequency offset (from the nominal frequency) vs. temperature for a well compensated crystal oscillator. The arrows represent the direction of temperature change. The difference between the f-T curve measured while the temperature is increasing and the f-T curve measured while the temperature is decreasing is the temperature dependent hysteresis. The maximum difference over the entire operating range of the oscillator is defined as the thermal hysteresis for that device.[1] Kusters and Vig, elsewhere in these proceedings, give a comprehensive review of the literature on thermal hysteresis in quartz crystal devices.[2]

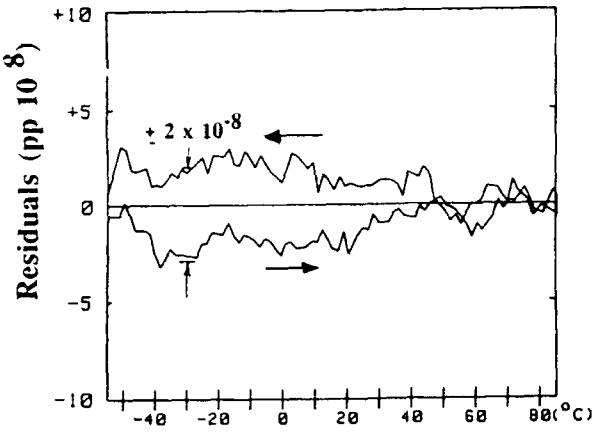


Figure 1 - Δf/f of a well compensated oscillator.

Formerly, data on thermal hysteresis were obtained by analyzing f-T curves of conventional temperature compensated crystal oscillators (TCXO). In a TCXO the crystal resonator is physically separated from the temperature sensing device, usually a thermistor, resulting in different thermal paths and time constants. If the ambient temperature is changing, the resonator and the thermistor will be at different temperatures. This results in an error in the compensation which is manifested as "apparent hysteresis".[3] The identical effect occurs if the resonator is measured in a passive system such as a "pi" network/vector-voltmeter test set since the temperature sensor is, as in a TCXO, separated from the resonator.

In this paper hysteresis data will be presented on resonators and oscillators. The dual-c-mode self-temperature sensing technique was used, in which, temperature inaccuracies are eliminated since the resonator and the thermometer are one and the same.[4] Measurements were made using both a pi-network/vector-voltmeter system[5] and several dual-mode oscillators.[6]

Experimental Scope

As reference 2 states, thermal hysteresis is not well understood. The phenomenon had to be better defined so that possible mechanisms could be identified. To that end 35 resonators from six different manufacturers were measured. There were units from two manufacturing lots from two of the suppliers and units from one lot from three of the manufacturers. The sixth manufacturer was the supplier of ceramic flatpack enclosed units.[7] The total number of f-T curves analyzed for this paper is 720. The

US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT

data shown are representative of the conclusions drawn from all of the curves.

f-T Profile

All of the f-T cycles were performed in a quasi-static manner. The temperature setting was changed and then, after a pre-set period of time, referred to as the "soak" interval, the frequency measurement was initiated . The standard temperature was 3.2°C and the standard "soak" interval was 4 minutes which resulted in a 12 hour f-T cycle. The longest soak interval used was 60 minutes. The standard temperature cycle was -55°C to +85°C and return.

Experimental Configuration

Figure 2 is a block diagram of the experimental configuration. Ten thick film "pi" networks were switched to a common vector-voltmeter and synthesizer. The three dual-mode oscillators were switched to two counters, one for the fundamental mode and one for the third overtone. The oscillator measurements implemented the true self-temperature sensing technique in that the measurements of the two modes were made simultaneously. In the "pi" network measurements, the modes were measured sequentially with a time separation of less then 5 seconds.

f-T Repeatability

There are two aspects to the compensation problem. The first is f-T repeatability. Figure 3 shows data from four different temperature cycles. In all cases a 6^{th} order polynomial has been subtracted from the raw frequencies to simulate the fractional frequency output from a compensated oscillator. The four temperature cycles repeat fairly well as contrasted with figure 4. (The resonator in figure 4 was removed from one "pi" network and inserted in another so it is not surprising that the f-T curve did not repeat.) The abscissa in figures 3 and 4, and the ones to follow, are labeled with three scales. The top two scales are the thermometric beat frequency[3] in both Hertz and ppm. The bottom scale is the temperature limits.

The repeatability of the f-T characteristic, as well as its smoothness is still an open area for investigation. This report is concerned only with the temperature dependent thermal hysteresis as defined above.

Thermal Hysteresis

In spite of the fact that the f-T repeatability is relatively poor in Figure 4, the temperature dependant thermal hysteresis for both cycles, plotted as a function of the thermometric beat frequency in figure 5, is virtually identical. The hysteresis curves that follow show tempera-

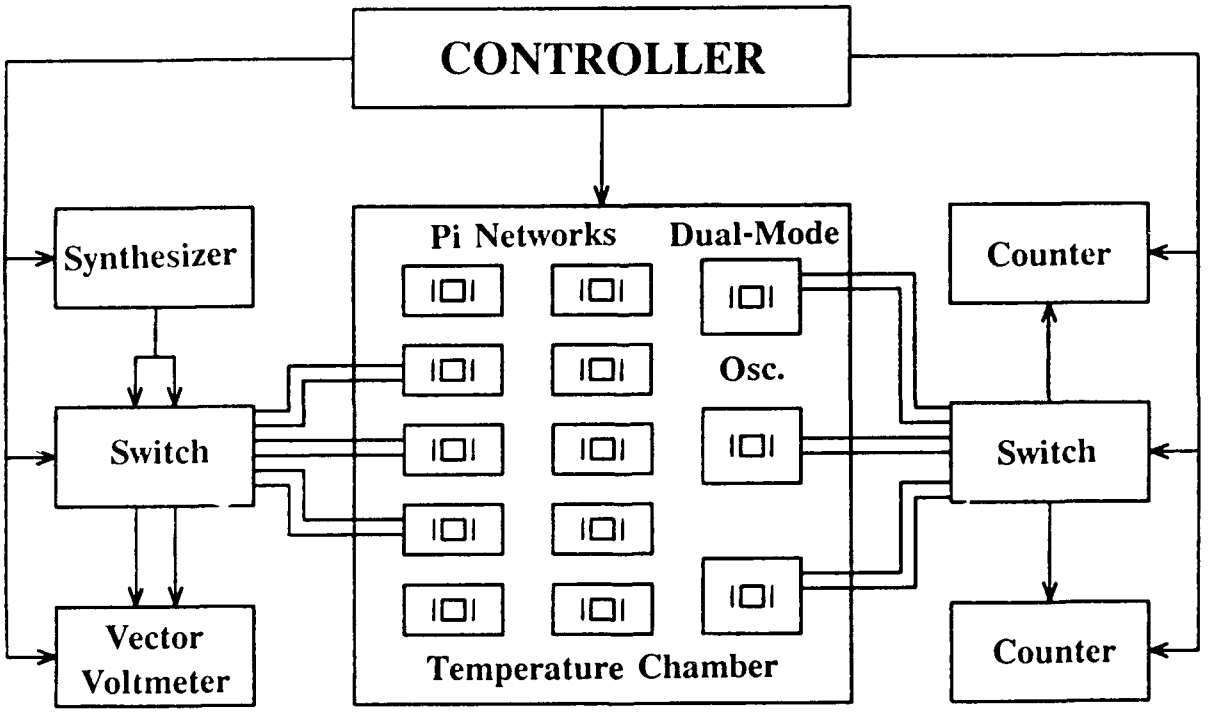


Figure 2 - Experimental configuration.

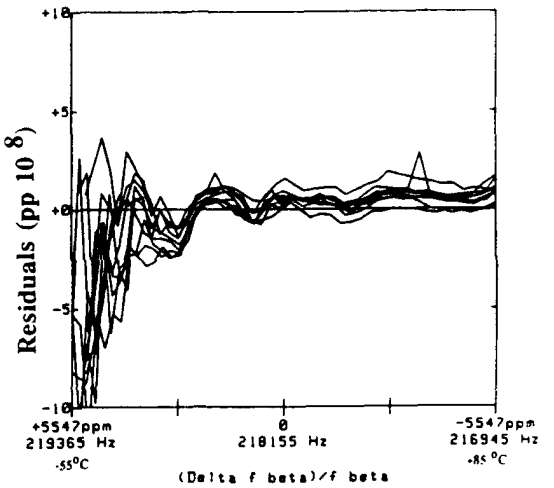


Figure 3 - f-T repeatability (pi network).

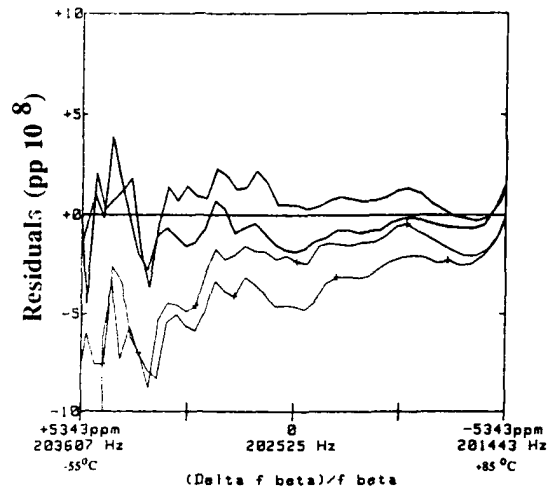


Figure 4 - f-T repeatability (pi network).

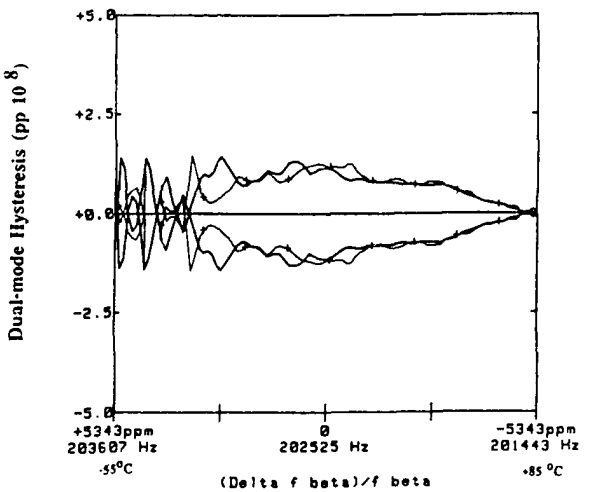


Figure 5 - Hysteresis of device in Fig. 4.

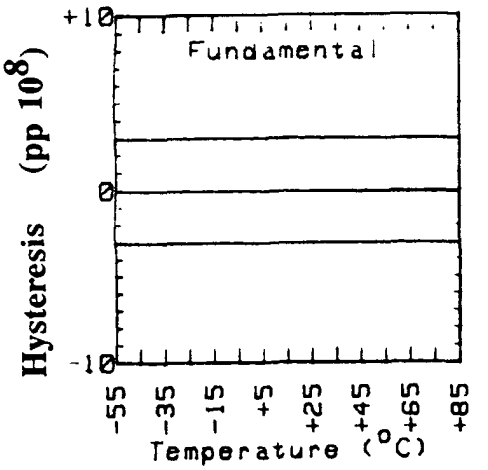


Figure 6 - Fundamental hysteresis for simulation.

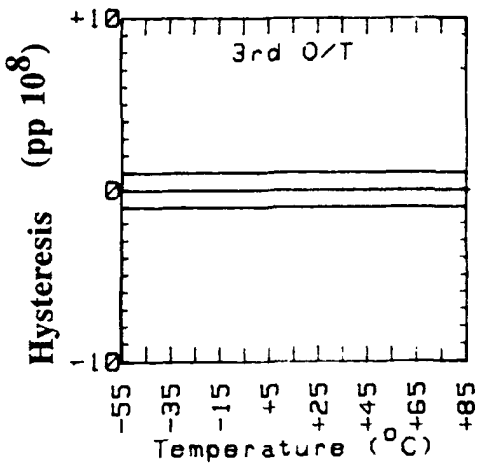


Figure 7 - Third overtone hysteresis for simulation.

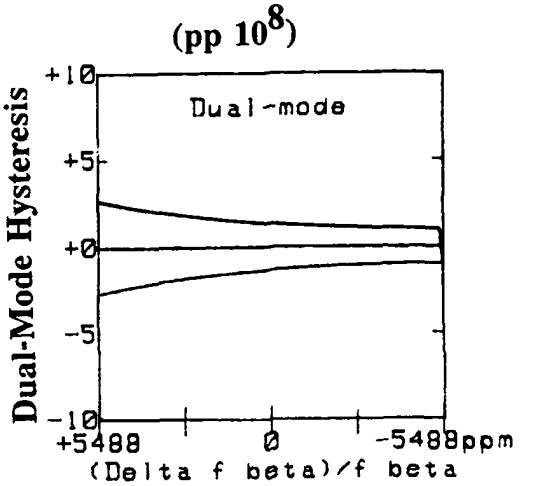


Figure 8 - Dual-mode hysteresis from simulation.

ture dependent thermal hysteresis. This is the first work which allows direct comparison of temperature dependent thermal hysteresis behavior because thermal lag and imperfect compensation have been eliminated.

Thermometric Beat Frequency and its Consequences

The thermometric beat frequency is defined[3] as

$$f_\beta = 3f_1 - f_3, \qquad (1)$$

where f_1 is the frequency of the fundamental mode, f_3 is the frequency of the third overtone and f_β is the thermometric beat frequency.

A consequence of equation 1 is that the dual-mode hysteresis of the fundamental mode and of the third overtone are identical.[8] If H_1 is the hysteresis of the fundamental and H_3 is the hysteresis of the third overtone, then the dual-mode hysteresis, $H(f_\beta)$ is given by

$$H(f_\beta) = 3\frac{df_1}{df_\beta}H_3 - \frac{df_3}{df_\beta}H_1. \qquad (2)$$

To reiterate, the dual-mode hysteresis is measured because measurement of the fundamental mode hysteresis alone or third overtone hysteresis alone is usually impossible due to thermal lag causing "apparent hysteresis."

A consequence of equation 1 is that 3 times the slope of the fundamental mode and the slope of the third overtone with respect to f_β always differ by 1 Hz/Hz. A full derivation and discussion of equation 2 can be found in reference 8.

Equation 2 shows that distortion and mixing of the hysteresis occurs due to the dual-mode technique. For example, if the hysteresis of the fundamental mode was $\pm 3 \times 10^{-8}$ and the hysteresis of the third overtone was $\pm 1 \times 10^{-8}$ over the entire temperature range, as shown in figures 6 and 7, the dual-mode hysteresis would look as shown in figure 8. In this example the flare at the low temperature end is due to the slope of the f-T curve being largest there, not due to the fact that the hysteresis is largest there. Even though it is difficult to compare dual-mode data with conventional data, the curves are self consistent, that is, dual-mode hysteresis measured under various conditions can be compared to determine the effect of various parameters.

Parameters Investigated

Six parameters were investigated:

1. Hysteresis repeatability (no change in parameters)
2. Manufacturer and manufacturing lots
3. Temperature range and sequence
4. Drive current
5. "Soak" interval
6. Measurement technique, i.e., "pi" network or oscillator

Hysteresis Repeatability

An important consideration in studying hysteresis is the stability of the phenomenon. If the hysteresis were different each time one measured a device, it would be very difficult to determine the effect of measurement parameters or process variations. It has been reported that hysteresis decreases with repeated f-T cycles, especially between the first and subsequent cycles.[9] Figures 9 and 10 show several examples of repeated cycles. It can be seen that if the parameters are not changed, the hysteresis repeats extremely well. In figure 10, even the anomalous structure repeats itself. This anomalous structure will take on an interesting connotation when the units are grouped by manufacturing lots. The large fluctuations at the low temperature end are due to the non-simultaneity of the "pi" network measurements coupled with the rapid temperature variations caused by liquid nitrogen cooling in the temperature chamber.

Manufacturer and Manufacturing Lots

In figures 11 to 16 hysteresis data is shown for all of the units from 6 different manufacturing lots. Figures 12 and 13 show the results for two lots from manufacturer B and figures 15 and 16 show the results for two lots from manufacturer D. Figures 11 and 14 show the results for lots from manufacturer A and C respectively. Only single unit lots were studied from the other two suppliers.

It is quite evident that each manufacturing lot has a

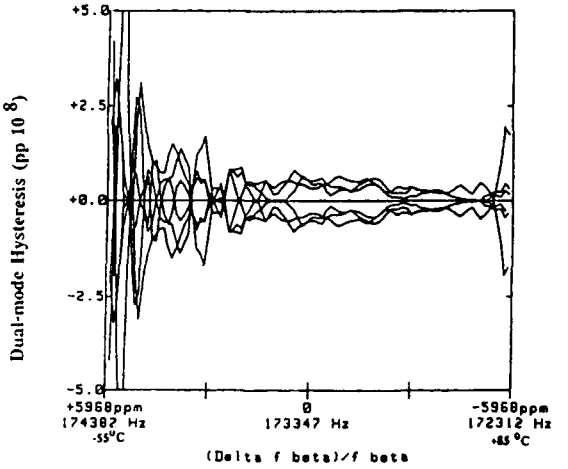


Figure 9 - Hysteresis repeatability.

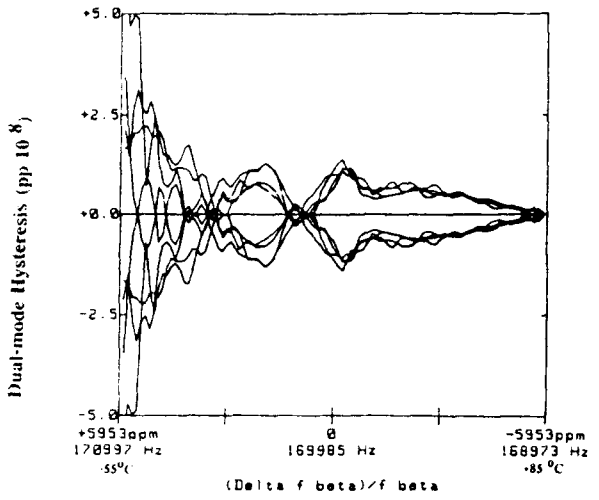


Figure 10 - Hysteresis repeatability.

signature, and that the signature is different for each lot even from the same manufacturer with ostensibly identically processed units. It is interesting that manufacturer B fabricated both the best and the worst units, although they were of different designs. It is difficult to develop a simple model for mechanisms which cause the unusual shape of the hysteresis in units shown in figure 16.

The "sign" of the hysteresis is also related (albeit loosely) to manufacturing lot. There were five "signs" observed:

1. frequency during increasing temperature > frequency during decreasing temperature
2. frequency during increasing temperature < frequency during decreasing temperature
3. a crossover in the middle of the range with the frequency during increasing temperature > frequency during decreasing temperature at the high temperature limit (the reverse at the low temperature limit)
4. the reverse of case 3
5. indistinguishable pattern due to the noise being larger than the hysteresis (only occurred in two of the devices with very low hysteresis)

The results for the manufacturing lots are

Lot from figure 11 : all case 4
Lot from figure 12 : 1 case 4, 1 case 1
Lot from figure 13 : 2 case 4, 1 case 1
Lot from figure 14 : 1 case 4, 2 case 1
Lot from figure 15 : all case 3
Lot from figure 16 : both case 1
Ceramic flatpacks : 9 case 4, 1 case 1, 2 case 5

"Apparent hysteresis" would manifest itself as case 3 or case 4. The "sign" of the hysteresis does not change when the sequence of temperatures is reversed. It appears,

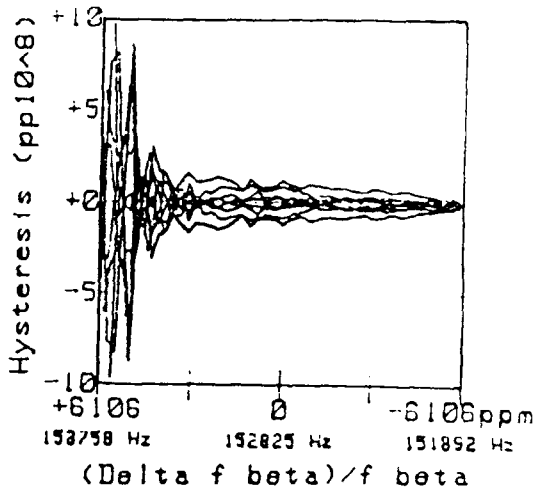


Figure 11 - Manufacturer A, 4 units, 2 point mount, metal enclosure.

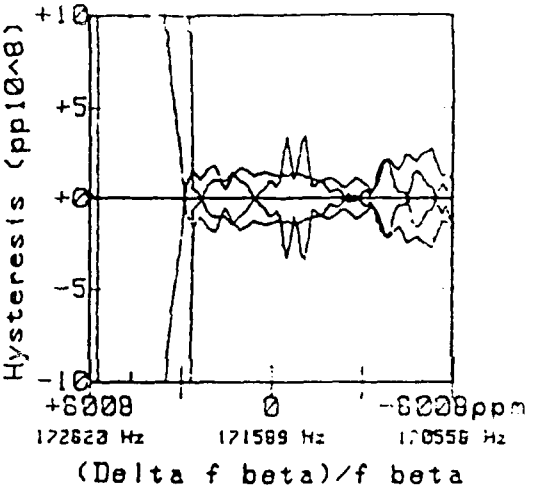


Figure 12 - Manufacturer B, lot 1, 2 units, 4 point mount, metal enclosure.

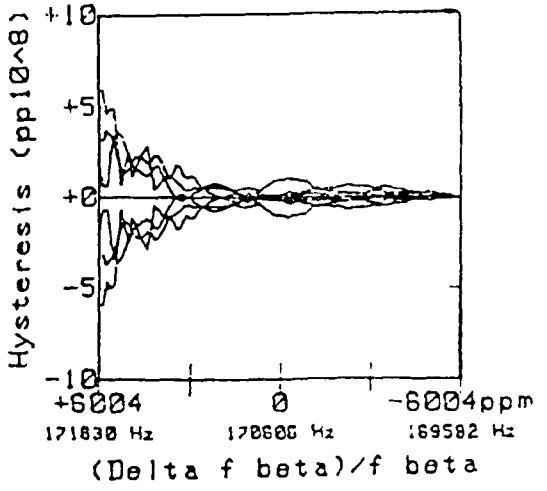


Figure 13 - Manufacturer B, lot 2, 3 units, 2 point mount, glass enclosure.

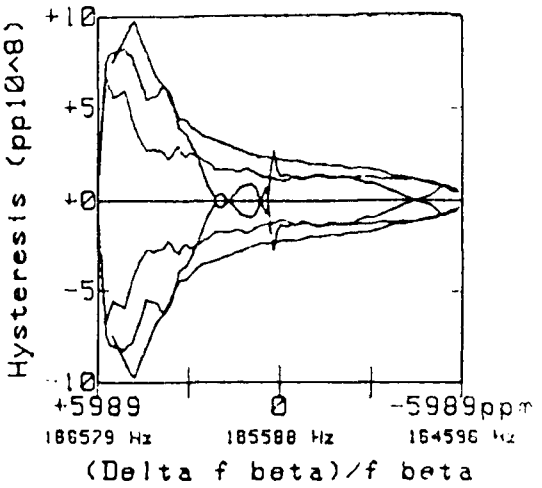


Figure 14 - Manufacturer C, 3 units, 2 point mount, metal enclosure.

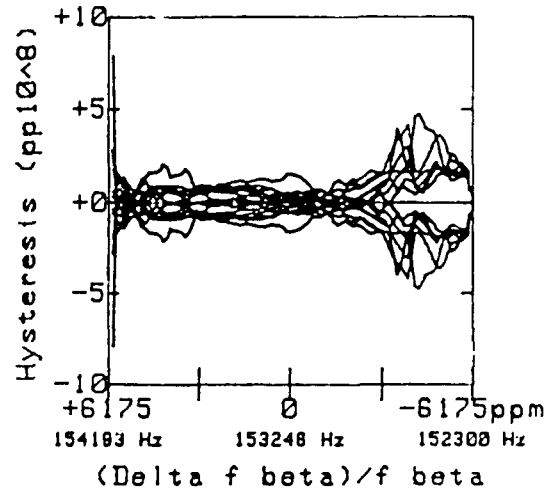


Figure 15 - Manufacturer D, lot 1, 8 units, 4 point mount, metal enclosure.

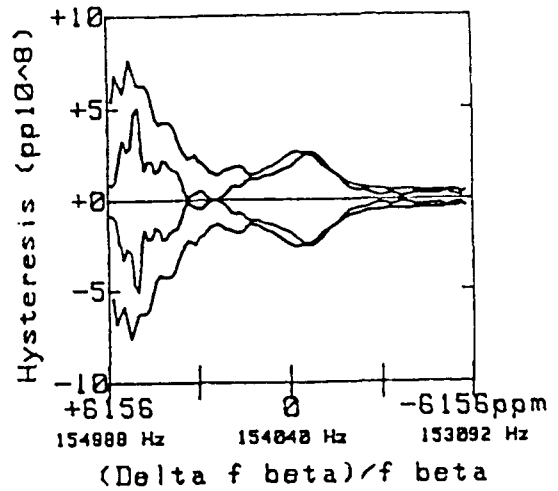


Figure 16 - Manufacturer D, lot 2, 2 units, 4 point mount, metal enclosure.

therefore, that the crossover in the hysteresis is not due to thermal lag or the non-simultaneity of the measurement.

Sequence of Temperatures and Temperature Range

To investigate the effects of thermal history, several f-T cycles were performed with the same temperature limits but in different sequence. The three sequences used were

1. -55°C → +85°C → -55°C
2. +85°C → -55°C → +85°C
3. +25°C → -55°C → +85°C → +25°C

A typical result is shown in figure 17. The different sequences, with the limits identical, made no difference in the thermal hysteresis.

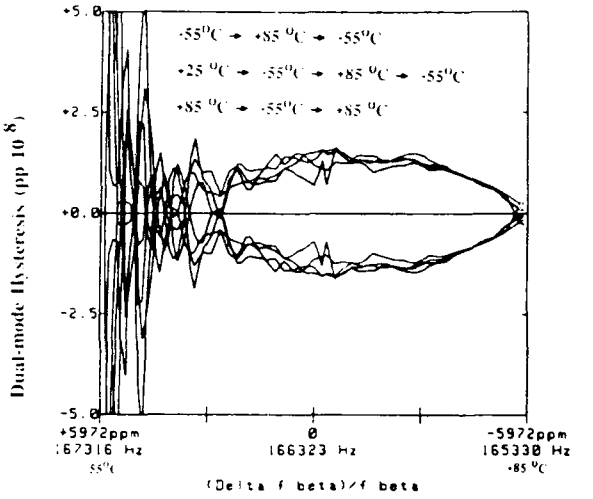


Figure 17 - Temperature sequence.

Several f-T cycles were also performed using th e low T → high T → low T sequence but with the limits changed. Figures 18 - 20 show three resonators, with each resonator cycled over several different temperature ranges. Temperature extreme is the one parameter which made a significant difference in the thermal hysteresis. It appears that hysteresis is dependent on the highest temperature seen during the temperature cycle. The cycles with +85°C as the upper limit have the same hysteresis but the cycles with the upper temperature limit lower than +85°C have lower thermal hysteresis. Since the sequence in this experiment was low T to high T and return, this experiment does not measure the effect of the low temperature limit.

Drive Current

The effect of drive level on the thermal hysteresis was also studied. Figures 21 and 22 show two resonators with the drive on the fundamental mode nominally 1 ma (10μW)

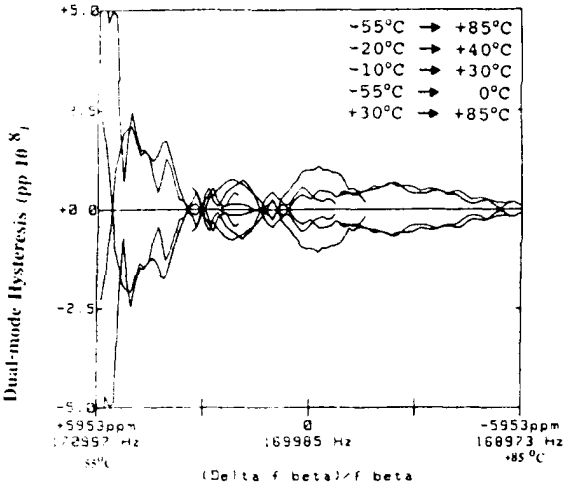


Figure 18 - Various temperature cycle limits.

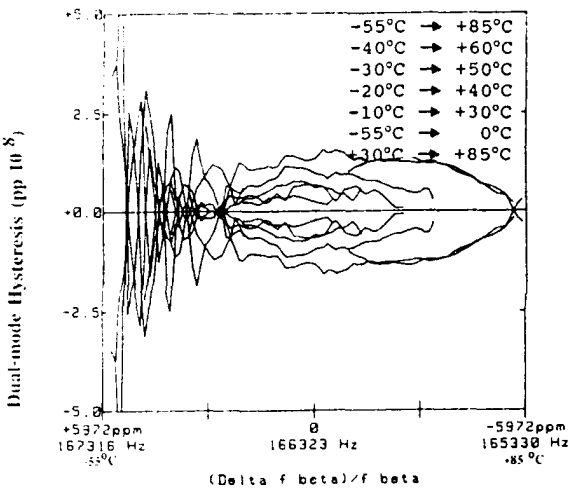


Figure 19 - Various temperature cycle limits.

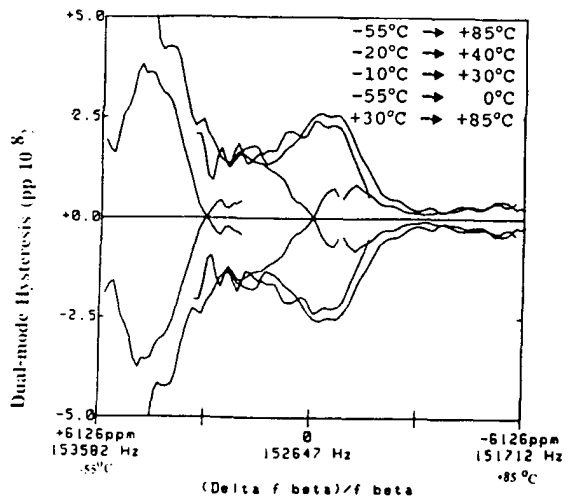


Figure 20 - Various temperature cycle limits.

for the first f-T cycle and 2 ma (40μW) for the second cycle. The drive on the third overtone was 0.3 ma (7.2μW) for the first and 0.6 ma (29μW) for the second. The drive current, at least at the levels reported here and for the 18 resonators measured, did not affect the thermal hysteresis.

"Soak" Interval

The "soak" interval, i.e., the time interval between temperature set and frequency measurement,was increased from 4 minutes to 8 minutes and then to 60 minutes. The f-T cycle for the 60 minute "soak" interval took about 5 days. If thermal transients are contributing to the thermal hysteresis, the change in "soak" interval should give some indication. For most of the units, especially the better performers, the slower cycle made almost no difference. In figure 23 the slower cycle is a little less noisy and possibly a little better at low temperatures, but the essential behavior is unchanged. In figure 24 there is virtually no difference between the two cycles. Figure 25 shows an example of a unit from the one manufacturer which was extremely sensitive to "soak" interval. This points out that there is most likely more than one mechanism responsible.

Measurement Technique

All of the preceding measurements were performed with the resonators in a passive "pi" network. Figure 26 is the thermal hysteresis of a resonator measured both in a "pi" network and a dual-mode oscillator. This technique enables one to separate the contribution of the oscillator circuit from that of the resonator. The contribution of the oscillator circuit, in this example, is very large. The design of the oscillator was later modified by reducing an inductor value in the fundamental mode circuit (and increasing the appropriate capacitance values.)[10] The same resonator

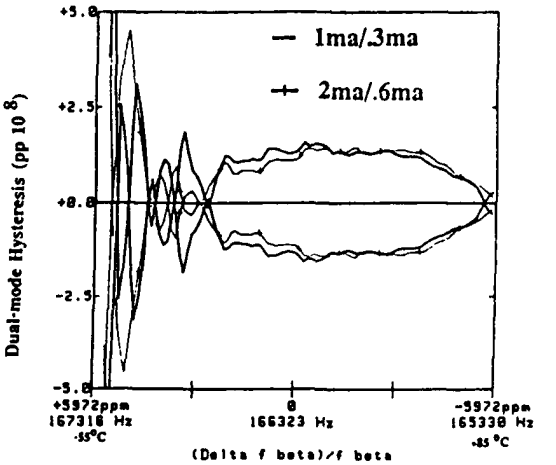


Figure 21 - Drive current.

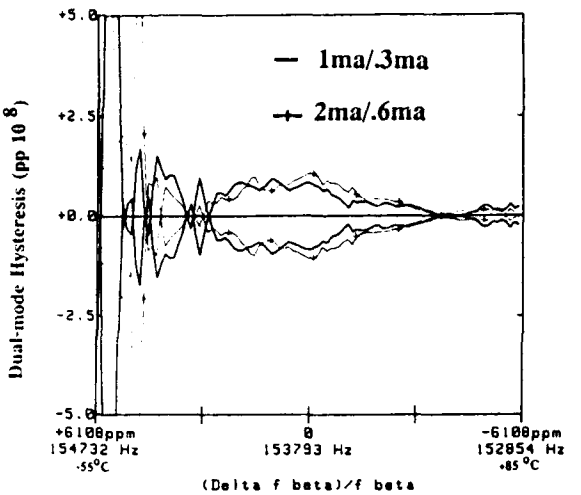


Figure 22 - Drive current.

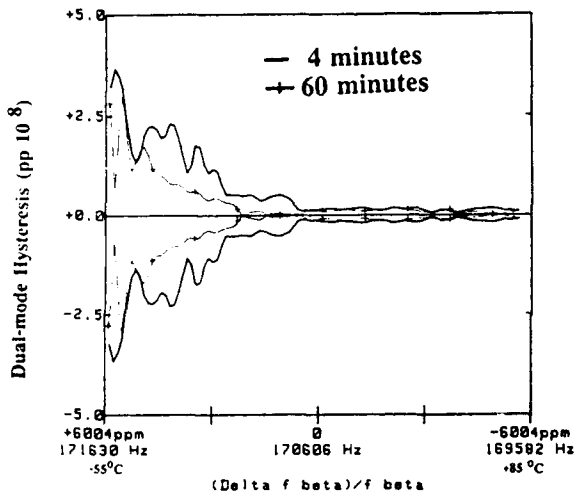


Figure 23 - 4 minute and 1 hour "soak" intervals.

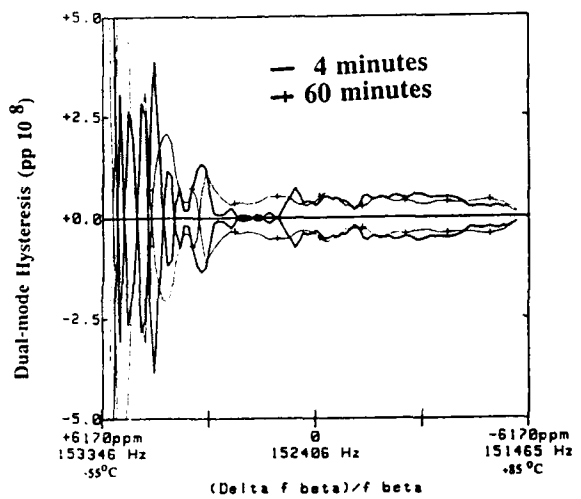


Figure 24 - 4 minute and 1 hour "soak" intervals.

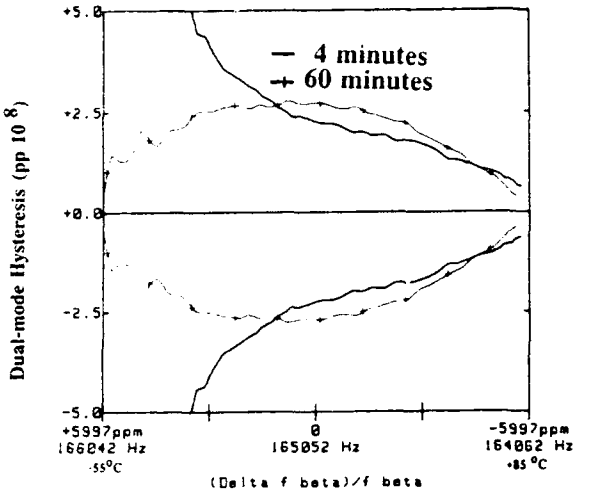


Figure 25 - 4 minute and 1 hour "soak" intervals.

measured in both oscillators is shown in figure 27. Forease of comparison the second oscillator is compared to the "pi" network measurement in figure 28. The redesigned oscillator is almost as good as the resonator in the "pi" network at the higher temperatures. The oscillator is better at the low temperatures, most likely due to the non-simultaneity of the measurement in the "pi" network, or there may be some offsetting phenomena.

Conclusions

The conclusions which can be derived from this work are

1. The fact that resonator hysteresis is a repeatable property for a given temperature cycle will allow easier observation of systematic effects in hysteresis experiments.

2. The observation that manufacturing lots have signatures implies that hysteresis is not an inherent property but is process dependent.

3. Since temperature excursion is a factor, it is not feasible to model hysteresis behavior to improve compensation.

4. Drive current is not a factor in thermal hysteresis, at least at the levels reported here.

5. Resonator hysteresis is not always worse at low temperatures.

6. Oscillator components may be a contributor to hysteresis. A low hysteresis resonator is necessary but not sufficient to insure a low hysteresis oscillator.

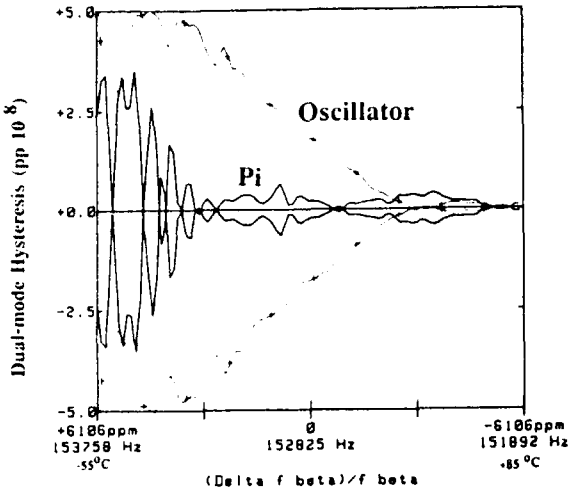


Figure 26 - Pi network and dual-mode oscillator A.

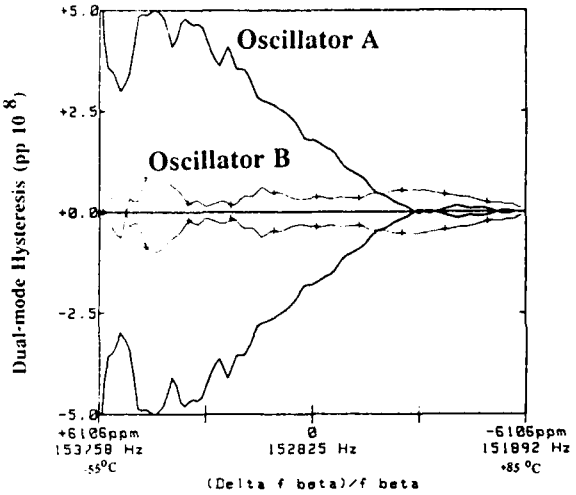


Figure 27 - Dual-mode oscillator A and B.

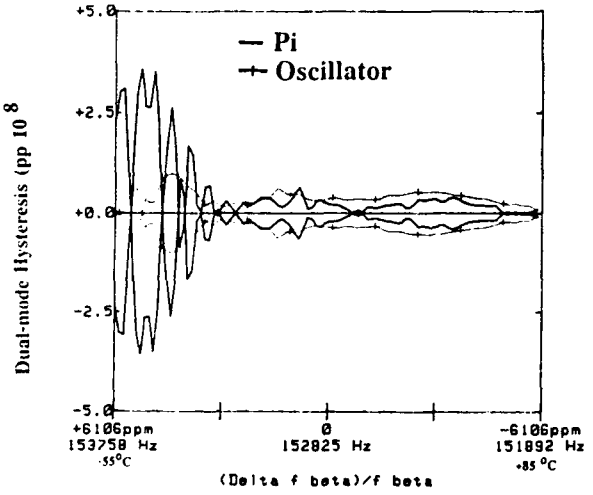


Figure 28 - Pi-network and dual-mode oscillator B.

REFERENCES

1. MIL-O-55310, "General Specification for Crystal Oscillators," Military Specification, copies available from Naval Publications & Form Center, 5801 Tabor Avenue, Philadelphia, PA 19120.

2. J. A. Kusters and J. R. Vig, "Thermal Hysteresis in Quartz Resonators - A Review," Proceedings of the 44th Annual Symposium on Frequency Control, 1990.

3. R. L. Filler, "Measurement and Analysis of Thermal Hysteresis in Resonators and TCXO's," Proceedings of the 42nd Annual Symposium on Frequency Control, 1988.

4. S. S. Schodowski, "Resonator Self-Temperature Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989; and United States Patent #4,872,765 dated Oct. 10,1989.

5. IEC-444, "Measurement of Quartz Crystal Unit Parameters by Zero Phase Technique in a π-Network," International Electrotechnical Commission, copies available from American National Standards Inst. (ANSI), 1430 Broadway, New York, NY 10018.

6. A. Benjaminson and S. Stallings, "A Microcomputer-Compensated Oscillator Using a Dual-Mode Resonator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

7. R. L. Filler, L. J. Keres, T. M. Snowden and J. R. Vig, "Ceramic Flatpack Enclosed AT and SC-Cut Resonators," Proceedings of the IEEE Ultrasonic Symposium, 1980.

8. R. Filler and J. Vig, "Resonators for the Microcomputer Compensated Crystal Oscillator," Proceedings of the 43rd Annual Symposium on Frequency Control, 1989.

9. H. J. Förster, "Thermal Hysteresis of AT + SC-Cut Quartz Crystal Resonators; Automated Measurement Method and Results," Proceedings of the 32nd Annual Symposium on Frequency Control, 1978.

10. A. Benjaminson (private communication), 1990.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ORIENTATIONAL DEPENDENCE OF "TRUE" SC-CUTS

Errol P. EerNisse
Quartztronics, Inc., Salt Lake City, UT

John A. Kusters
Hewlett-Packard Co., Santa Clara, CA

Introduction

The SC-cut, or Stress Compensated-cut, was originally arrived at by two different theoretical approaches and one experimental approach [1-4]. Although the two theoretical models did not treat the exact experimental configuration, all of the early work led to choices of angles for the doubly-rotated cut that were close together. The reasons for this have become clear over the years as other theoretical work has been carried out on the SC-cut. [5,6] Basically, of the six possible initial stress components set up in the quartz resonator blank by various influences such as thermal transients, electrode stress, or radiation deposition, the two in-plane longitudinal components are dominant. This circumstance is the result of the planar geometry of a crystal blank, which leads to planar stress patterns.

Figure 1 outlines the predicted loci of several interesting effects in alpha-quartz. Shown are the locus of the zero first-order temperature coefficient [11], the predictions by Holland for the locus where thermal transient compensation occurs [2], and the locations where stress compensation occurs [3,6]. Also shown is the first experimental confirmation of the theory [4].

Stress compensation occurs in the SC-cut because the shift in resonant frequency is equal, but opposite in sign, for the two in-plane longitudinal stress components.

Thermal transient compensation in the early work [4] was exhibited in a geometry where the dominant stress components were also the two in-plane longitudinal stresses. Many manufacturers now offer "SC-cuts" in their product line that provide marked improvement in thermal transient performance over the singly-rotated AT-cut.[7] Few of them actually obtain the full benefit of thermal transient compensation. This is because the initial stress patterns set up by thermal transients are specific to the mount and packaging configuration, which varies between manufacturers. Theoretical work also exists that suggests even the choice of electrode thickness and material can affect thermal transient compensation.[8]

Recently, an attempt was made to use the crystallographic orientation of a two-point mount SC-cut for a four-point packaging configuration. It was necessary to change the crystallographic orientation for the four-point mount configuration in order to obtain equivalent thermal transient compensation between the two packaging approaches. The fact that a crystallographic orientation change was necessary and the qualitative basis for why it was necessary are the subject of this paper. An experimental matrix method for defining the best crystallographic orientation for a specific mounting and packaging configuration is presented.

CH2818-3/90/0000-185 $1.00 © 1990 IEEE

Theoretical Background

Figure 2 shows a quartz resonator blank with the axes system used herein and the three in-plane stress components T_{xx}, T_{zz}, and T_{xz}. In the case of homogeneous stress throughout the active region of the thickness shear mode, the fractional frequency shift $\Delta f/f_o$ is given by:

$$\Delta f/f_o = K_{xx}T_{xx}+K_{zz}T_{zz}+K_{xz}T_{xz} \quad (1)$$

Here, K_{xx}, K_{zz}, and K_{xz} are calculated using the perturbation formula of Tiersten [9] and the third-order elastic constants of quartz as measured by Thurston, et. al. [10]. Figure 2 shows how K_{xx} and K_{zz} vary with the rotation angle φ of a doubly-rotated plate $(YXwl)\varphi,\theta$, where θ is selected for temperature compensation or the desired turnover temperature (θ is around 34.3°).

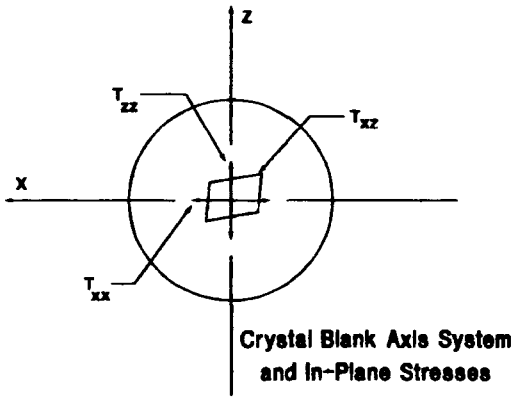


Figure 2. A quartz resonator blank with the axes convention used and representations of the three in-plane stresses.

Note in Figure 3 that around φ=22.3°, K_{xx}=-K_{zz}. This has been emphasized in Figure 3 by also plotting $K_{xx}+K_{zz}$. As one can see in Figure 3, these frequency shift coefficients are not a strong function of φ, so considerable benefit can be obtained by choosing φ near 22°.

In contrast, Figure 4 shows K_{xz} as a function of φ near φ=22°. Note in Figure 4 that K_{xz} has a considerable value for all the φ values where $K_{xx}+K_{zz}\approx 0$. If the heat flow in a thermal transient is not symmetric about the blank center in the plane of the resonator, T_{xz} will $\neq 0$. In this case, there will need to be a corresponding shift in the φ angle to cause all three terms in Eq. 1 to compensate one another. Thus the necessity of small modifications to φ as one changes mounting and packaging if the full benefit of the SC-cut is desired.

In most fabricated quartz crystals, mounting stresses developed during crystal fabrication, non-symmetric electrode stress, other fabrication-related stresses, *and heat-flow into the crystal blank*, may cause three additional non-planar stresses to be present. These are the other longitudinal stress, T_{yy}, and the two remaining shear stresses, T_{xy} and T_{yz}. These stresses are associated with the stress coefficients K_{yy}, K_{xy}, and K_{yz}, respectively. In the presence of a thermal gradient in any region in the crystal, full thermal transient compensation occurs if, and only if, the algebraic sum of the products of the existing stresses and their respective stress coefficients, is zero, i.e.,

$$\Delta f/f_o = K_{xx}T_{xx}+K_{yy}T_{yy}+K_{zz}T_{zz} +K_{xz}T_{xz}+K_{xy}T_{xy}+K_{yz}T_{yz} \quad (2)$$
$$= 0.$$

Inherent in this work is a basic assumption that static stress compensation is necessary, but not sufficient to ensure the maximum performance from

a quartz resonator. Only when all of the dynamic effects due to heating (or cooling) a crystal compensate, does stress and thermal transient compensation occur.

Experimental Evidence for φ Shifts

Figure 5 is from an earlier work in which a comparison was made between conventional cuts and the SC-cut. [12] This figure illustrates the thermal overshoot that occurs in all AT-cut and BT-cut oscillators, and in many other oscillators that are advertised as "SC-cut" oscillators. In this case, the oven power cuts back at about 6 minutes. A properly designed SC-cut essentially approaches its equilibrium temperature at the same time. Other cuts do not, as exhibited by the AT-cut resonator shown.

A key point is that thermal transient behavior may not be due solely to the crystal. Other factors, such as aging or transient effects in the oven control thermistor, frequency trimmer capacitors, inductors in the direct frequency control loop, may all contribute to a transient effect.

Another factor is the coupling between the crystal resonator blank and the oven mass. Asymmetric heat-flow during heating (or cooling) will give rise to stress in the blank. Changing heat-flow conditions may make it impossible to find an orientation with a true thermal transient compensation. Attention must be given to the method of blank attachment, lead conditions, crystal case material, residual (or back-filled) gasses in the case, and the methods of applying heat to the crystal case. These are summarized in Table I.

Thermal Transient Compensation

Factors Affecting Performance

- Type of mount
- Number of mounting leads
- Package configuration
- Electrode mass imbalance
- Electrode Material
- Quality of quartz

Table I.

Experimental Methods for Determining the Optimum φ Angle

Table II summarizes the common methods of testing for thermal transient performance.

Thermal Transient Testing

Experimental Methods

- Thermal Pulse Testing
- Thermal Scan Testing
- Oscillator Warmup Characteristics
- Pulsed Laser (Morley)

Table II.

Of the factors shown, the pulsed laser method [13] must be done in an evacuated system with the crystal case removed.

Thermal step testing has been interesting theoretically [1,2,8] as it is relatively easy to model. As ther-

mal step testing requires the creation of an instantaneous step change in the temperature of the surface of the crystal, it is virtually impossible to do in practice.

Thermal scan testing [4] is illustrated in Figure 6. Here, a hypothetical crystal is scanned in temperature through its turnover point. The turnover point presents a unique, and identifiably reference mark during the testing of the resonator. Under static conditions, i.e., the temperature rate-of-change is much slower than the thermal response time of the resonator, the middle line of Figure 6 is obtained.

As the rate-of-change of temperature is increased, two factors happen. The first is a translation of the apparent turnover point in temperature. This is related to the thermal transient time and represents a thermal lag in the system response.

The second is a frequency shift which occurs at all temperatures. For the crystal data shown, an increasing temperature scan causes the frequency to increase, while a decreasing temperature scan causes the frequency to decrease. The actual sign of frequency change with temperature change depends upon the crystallographic orientation and all of the other factors mentioned above. The key experimental factor is the maximum (or minimum for the other turnover point) frequency value observed, independent of the temperature at which it was observed.

Of importance to the analysis of the experimental data is that, to first order, the amount of frequency offset at the turnover point is directly proportional to the temperature rate-of-change.

The measure of thermal transient performance is therefore the frequency difference between the measured extrema of the frequency-temperature curve, normalized to the rate-of-change of temperature.

The final method outlined is to carefully plot the oscillator warm-up characteristics. This is the final measure of the thermal transient performance. However, as pointed out earlier, other effects due to oscillator components may dominate the crystal resonator response. Only when all other effects have been minimized, or eliminated, can useful information be obtained using the oscillator method.

A useful process for systematically evaluating resonator designs for optimum performance is shown in Table III. This methodology has been used in several companies to eventually produce crystals, and crystal-controlled oscillators, that maximize the advantages of the SC-cut

As the data from a single crystal may not be reliable, it is important that the method outlined in Table III be done with enough samples (at least 5 per set) to have statistical confidence in the final results.

Conclusions

In general, quartz resonators designed as SC-cut devices, do not always perform as expected. To achieve the maximum benefits, both static stress compensation for planar effects, and thermal transient compensation for planar and non-planar thermal effects are required. The methods for testing resonators and choosing the correct φ angle were presented.

References

1. R. Holland, "Nonuniformly Heated Anisotropic Plates: 1. Mechanical Distortion and Relaxation," *IEEE Trans. Sonics and Ultrasonics*, vol SU-21, p. 171, July, 1974.

2. R. Holland, "Nonuniformly Heated Anisotropic Plates: II. Frequency Transients in AT and BT Quartz Plates," *1974 Ultrasonics Symposium Proceedings*, IEEE Cat. #74 CHO 896-1 SU, p. 593.

3. E. P. EerNisse, "Quartz Resonator Frequency Shifts Arising From Electrode Stress," *Proc. 29th Annual Symposium on Frequency Control*, p. 1, 1975.

4. J. A. Kusters, "Transient Thermal Compensation for Quartz Resonators," *IEEE Trans. Sonics and Ultrasonics*, Vol. SU-23, p. 273, July, 1976.

5. E. P. EerNisse, "Calculations on the Stress Compensated (SC-Cut) Quartz Resonator," *Proc. 30th Annual Symposium on Frequency Control*, p. 8, 1976.

6. B. K. Sinha, "Stress Compensated Orientations For Thickness-Shear Quartz Resonators," *Proc. 35th Annual Symposium on Frequency Control*, p.213, 1981.

7. R. W. Ward, "Design of High Performance SC Resonators," *Proc. 35th Annual Symposium on Frequency Control*, p. 99, 1981.

8. B. K. Sinha and H. F. Tiersten, "Thermally Generated Transient Frequency Excursions In Doubly-Rotated Quartz Thickness-Mode Resonators," *J. Appl. Phys.*, vol. 55, p. 3337, May 1984.

9. H. F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," *J. Acoust. Soc. Am.*, vol. 64, p. 832, Sept. 1978.

10. R. N. Thurston, H. J. McSkimin, and P. Andreatch, Jr., "Third-Order Elastic Coefficients of Quartz," *J. Appl. Phys.*, vol. 37, p. 267, 1966.

11. R. Bechmann, A. D. Ballato, and T. J. Lukaszek, "Higher-Order Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha-Quartz," *Proc. IRE*, vol. 50, pp. 1812-1822, August 1962.

12. J. A. Kusters, C. A. Adams, H. Yoshida, and J. G. Leach, "TTC's - Further Developmental Results," *Proc. 31st Annual Symposium on Frequency Control*, 1977.

13. P. E. Morley and R. J. Williamson, "Determination of the Optimum Orientation of an SC Cut Resonator using a Pulsed Laser," presented at the 44th Annual Symposium on Frequency Control, May 1990.

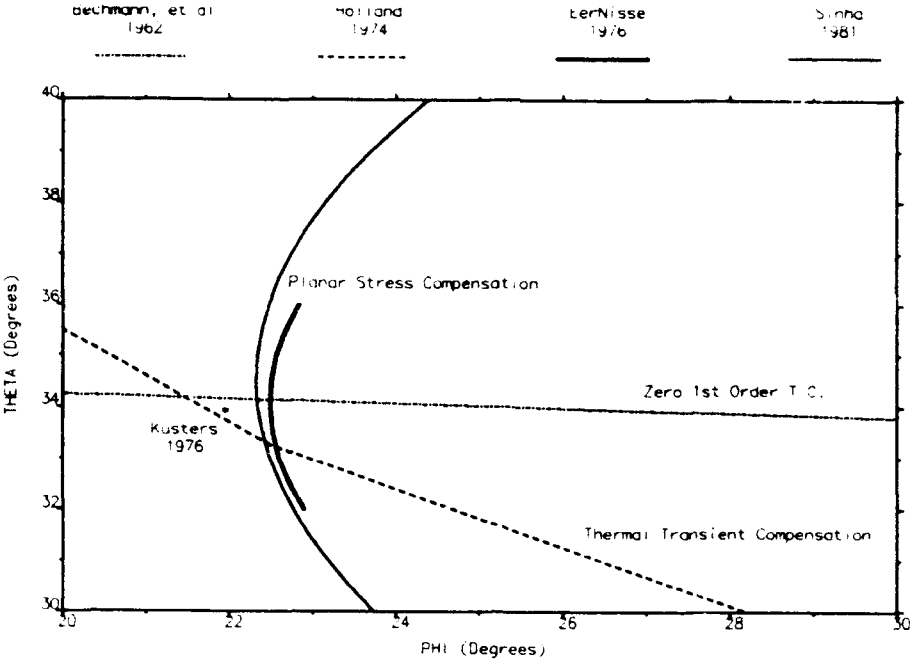


Figure 1. Predicted loci of interest - alpha quartz

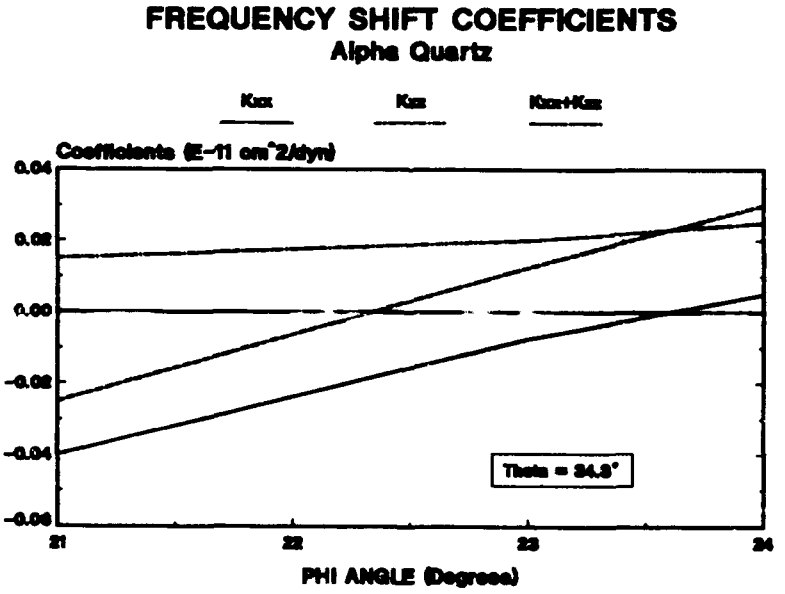


Figure 3. Frequency shift coefficients of the in-planar longitudinal stress components near the SC-cut.

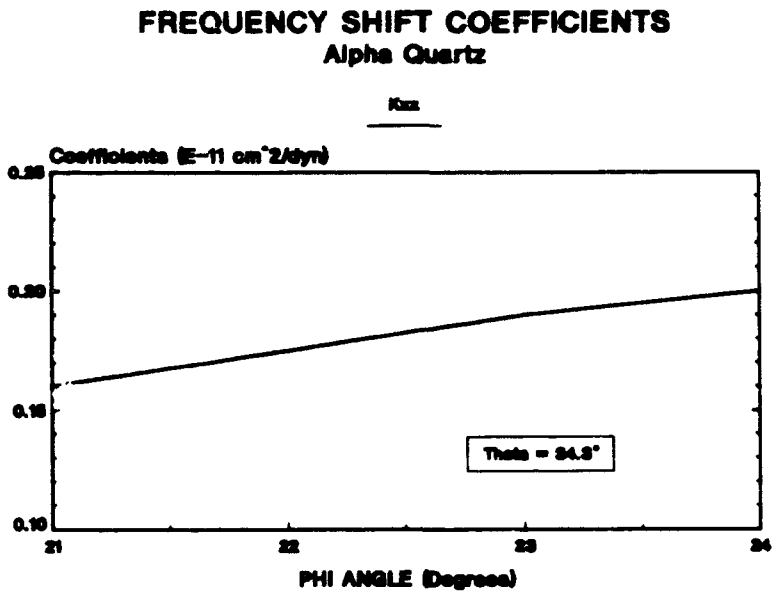


Figure 4. Frequency shift coefficient of the in-planar shear stress near the SC-cut.

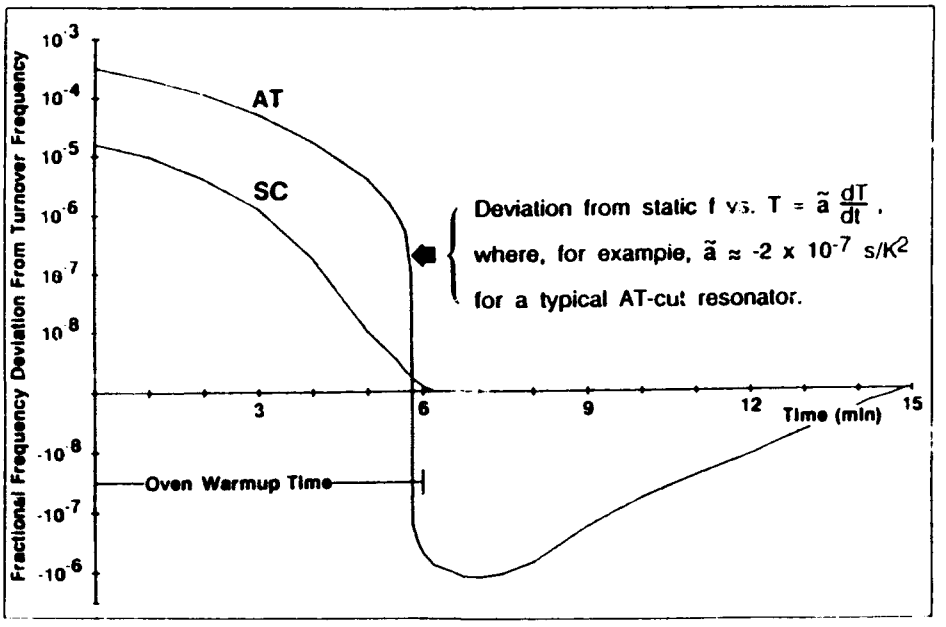


Figure 5. Warm-up characteristics of AT- and SC-cut Resonators

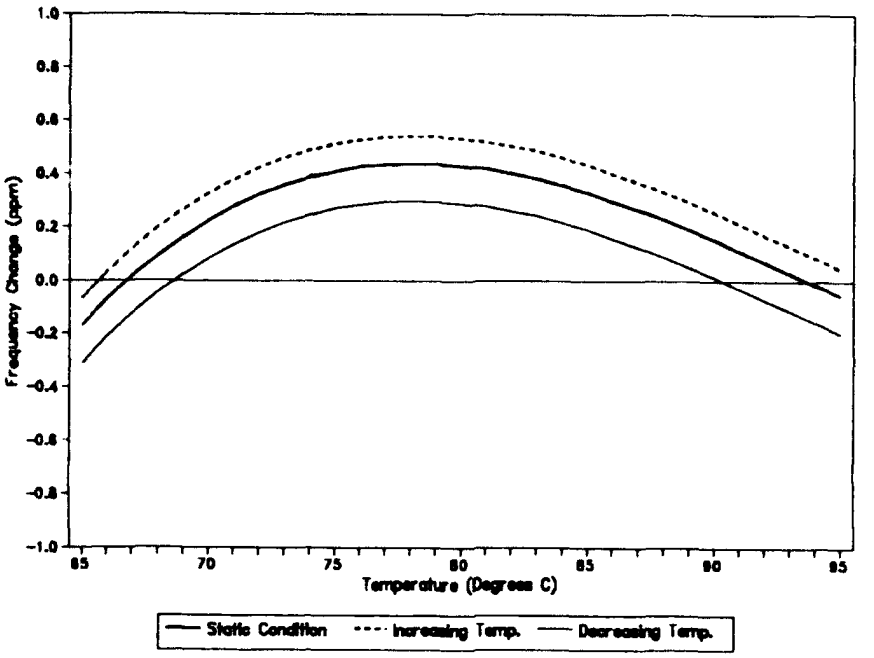


Figure 6. Hypothetical thermal scan data - SC-cut crystal

Experimental Method

Choice of Optimum Phi Angle

- Start with phi = 22°
- Choose theta for desired turnover
- Make a test set – check turnover temperature
- If not OK, correct theta before proceeding
- If OK make 4 more sets with delta phi = 0.6°
- Measure thermal transient response of each set
- Average data in each set discarding obvious data outliers
- Use interpolation on averaged data to compute correct phi
- Make a final test set and test turnover, thermal transient
- If not OK, repeat necessary steps above
- If OK, go to production.

Table III

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DETERMINATION OF THE OPTIMUM ORIENTATION OF AN SC-CUT RESONATOR USING A PULSED LASER

P. E. Morley†, and R. J. Williamson‡

†STC Components, Quartz Crystal Division, Edinburgh Way, Harlow, Essex CM20 2DE, England.

‡STC Technology Ltd, London Road, Harlow Essex CM17 9NA, England

Abstract

The SC-cut resonator, because of its stress compensation characteristics, is very suitable for use in precision oscillators as it has excellent frequency stability with respect to thermal transients. Much theoretical work has been done to determine the optimum crystallographic orientation for stress compensation and the published results are often used in resonator design. From the point of view of the crystal manufacturer, however, it would be more convenient if a rapid method of measuring this force-frequency effect were available. The optimum orientation for any particular resonator design could then be found empirically.

A method has been developed to measure the force-frequency effect in resonators by the use of a pulsed light beam from an excimer laser which is directed onto the vibrating surface of the device. The crystal resonance is maintained by a phase-locked loop with a long time constant, and the phase across the crystal is monitored. The small phase change is measured by the use of a lock-in amplifier which is locked to the repetition rate of the laser. The resulting measurement gives a determination of both the phase and the amplitude of the force-frequency effect.

Several devices have been produced with crystallographic orientations spanning the theoretical value. These were measured using the laser pulsing method, and then the frequency modulation was plotted against angle.

Introduction

The doubly rotated SC-cut was first introduced around 1975 by EerNisse [1,2] as a zero temperature coefficient (ZTC) cut which also exhibited freedom from frequency changes due to in-plane mechanical stresses, such as electrode stress. Contemporary work by Holland [3] predicted a quartz cut which should be frequency compensated for dynamically changing ambient temperature. This was initially called the TS cut (for thermal shock). The crystallographic orientations predicted by the two authors were, using the IEEE standard nomenclature (yxwl) ϕ/θ, (yxwl) 22.5°/34.3° and (yxwl) 22.8°/34.3°, respectively (fig. 1). Later experimental work by Kusters [4] indicated an orientation for thermal transient compensation (the TTC cut) of $(\phi,\theta) = (21.93°,33.93°)$, and this value has been used by many authors in later work [5,6,7]. Clearly, there is a close correlation between the two compensation effects, and, as noted in ref. 4, this is because they are caused by essentially the same internal stresses.

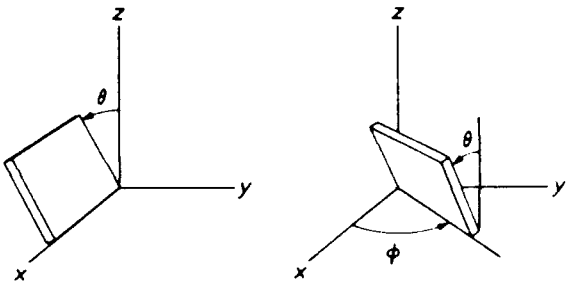


Fig. 1 Angle conventions for singly and doubly rotated quartz plates.

Other cuts have been noted along the AT-SC ZTC locus which exhibit an improvement over the AT-cut for force-frequency related effects, including the IT cut at $\phi = 19.1°$ [8], and a cut, developed for oven operation around 95°C, at $\phi = 23.75°$ [7], which has exhibited low acceleration sensitivity.

Much of the work for the experimental verification of the SC-cut orientation has been performed by direct measurement of the force-frequency coefficient K_f. This has been achieved by the construction of jigs which apply a pair of opposing forces around the periphery of a circular blank, either to apply a compressive effect [9,10] or to produce a bending moment [11]. Variation of the azimuthal angle ψ gives a function, the mean value of which should ideally be zero if the device is compensated for isotropic planar stresses. The thermal transient compensation effect has also been experimentally verified by observing frequency changes due to thermal shock or thermal cycling [4,11,12,13].

CH2818-3/90/0000-193 $1.00 © 1990 IEEE

The work presented here is a development of the work performed at STC Technology [14,15] in which the vibrational distribution of AT-cut resonators was investigated using an excimer laser. This method for mode shape analysis is particularly applicable to thickness-shear modes at high frequencies. The pulsed laser beam is tracked across the resonator surface, resulting in modulation of the crystal frequency which is a function of the position on the quartz plate. It is proposed that the mechanism of modulation is the force-frequency effect, whereby the light-induced heat pulse creates a localized compression in the quartz due to thermal expansion.

The use of a pulsed laser for creating a well-defined stress field on the surface of an SC-cut blank provides a technique for experimental verification of its stress-compensated properties, as theory suggests that the frequency modulation due to such a field should be close to zero if the spot is positioned in the modal centre of the resonator. Results are reported here to illustrate this effect, together with a description of the equipment used for performing the experiments.

Equipment

Laser Optics

The laser used for this work was a Questek 2040 excimer laser. The wavelength of the light emitted by such a laser can be varied by the choice of gases in the laser chamber. In our experiments, argon and fluorine were used, the combination of which produces an emission in the UV at 193 nm. The output energy can be varied between 20 and 200 mJ per 15 ns pulse, with a repetition rate of up to 100 Hz. The output beam is about 15 mm^2. The large beam size enables a very simple means of beam placement as shown in fig. 2. This arrangement was modified in later experiments by removing the second aperture and moving the imaging lens, in order to achieve a higher light intensity at the crystal surface. The spot size was adjusted by changing the distance from the lens to the crystal.

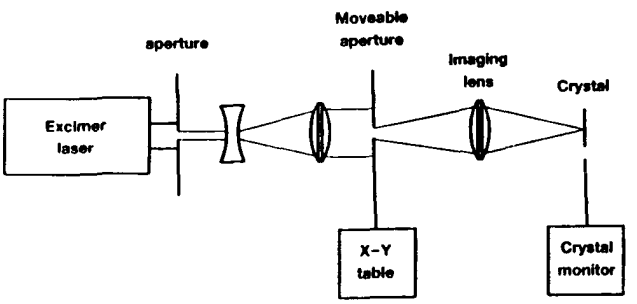


Fig. 2 Laser optics.

Crystal Measurement

The equipment for crystal measurement is shown in fig. 3. The vector voltmeter and low-noise signal generator are used to lock the crystal frequency close to resonance. The phase output of the vector voltmeter could be used to monitor the phase modulation caused by crystal frequency change, but it has been found that a higher sensitivity can be obtained with a double balanced mixer arrangement as shown. This requires a 90° phase shift between the reference and the measurement channels, which is provided by a quadrature hybrid, an Anzac JH-6-4, in place of the conventional power splitter. This considerably simplifies the measurement circuit as it operates over a wide frequency range without circuit modification. The crystal is measured in a simple 100 ohm network made from two SMA coaxial attenuators. Two identical amplifiers raise the signal levels in each measurement arm to about +7 dBm, which maximizes the sensitivity of the HP 10514 double balanced mixer. The output of the mixer is filtered through a 1 MHz low-pass filter to remove the 2.f component, and is then fed into a Stanford Research SR510 lock-in amplifier. This instrument is essentially a narrow band detector which is capable of measuring signals at known audio frequencies which are buried deep in noise. The frequency to which the detector is locked is fed from the source which is used to trigger the laser. The phase at which the measured signal is sampled can be adjusted to maximize the measurement sensitivity, and complex audio filtering is also available.

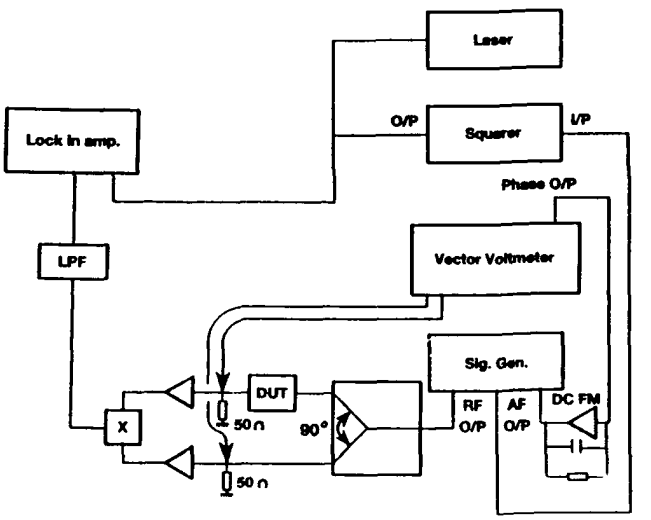


Fig. 3 Crystal measurement circuit.

Theoretical Background

The effect on the resonant frequency of radial edge forces applied to quartz discs was observed over 30 years ago. A force sensitivity coefficient K_f was defined by

Ratajski [9] some years later which relates frequency shift to applied forces by the equation

$$K_f(\psi) = \frac{\delta f}{f_0}\frac{2hD}{F}\frac{1}{N} \tag{1}$$

where $\delta f/f_0$ is the fractional change in frequency due to the application of radial forces F at azimuthal angle ψ on the edge of the plate. N is the frequency constant and $2h$ and D are the thickness and diameter of the plate respectively.

Extensive theoretical work has been carried out previously by EerNisse to determine the nature of K_f [16], and the results of this analysis have been modelled in the form of 4th order trigonometric functions [17]. These coefficients have been used in this work, and the resulting values of K_f are shown graphically in fig. 4.

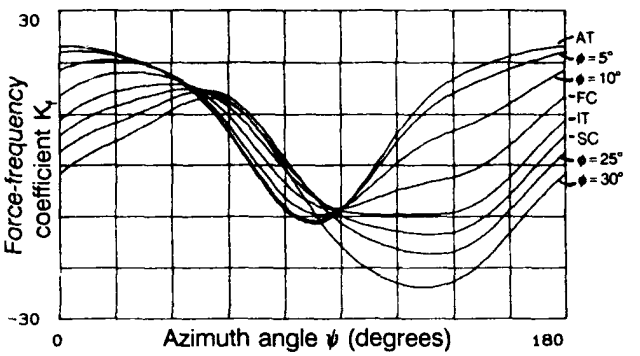


Fig. 4 Force-frequency coefficient vs azimuthal angle ψ.

The application of a laser pulse on the plate surface causes a localized stress related to thermal expansion of the quartz, and can it be considered analogous to the effect of electrode stress, but limited to the area of the spot. It is known that the resultant forces due to thermal expansion are somewhat anisotropic in AT-cut or SC-cut quartz, but this is neglected here, and could be included later. The localization of the stress field clearly results in the induced phase modulation being a function of the vibrational distribution of the thickness mode.

A full analysis of the thickness-shear vibrational modes of doubly rotated contoured quartz resonators has been made by Stevens and Tiersten [18], and this has been incorporated into this work for determination of the form of the mode shape. In Tiersten's nomenclature, the displacement u_{n00}, denoted here as $u(x_1,x_3)$, is given by the Gaussian function

$$u(x_1,x_3) = \exp(-\frac{\alpha_n}{2}x_1^2)\exp(-\frac{\beta_n}{2}x_3^2) \tag{2}$$

where α_n and β_n are defined in ref. 18, and x_1 and x_3 are the coordinates in the plane of the plate after a rotation of $\hat{\beta}_n$, also described in the same reference. The three quantities α_n, β_n and $\hat{\beta}_n$ are functions of overtone number (n) and plate orientation. They have been calculated for various points along the ZTC locus, and their values fitted to 6th order polynomials against the angle ϕ for use in later calculations.

To predict the change in frequency due to changes in the effective elastic constants within a small volume of the quartz disc, Rayleigh's Principle [19] is employed. It states that the time averaged potential energy $<V>$ of a system is equal to the time averaged kinetic energy $<T>$.

Consider a resonator composed of n unit volumes $k=1...n$. The potential energy for the kth unit volume is given by

$$V_k = \frac{1}{2}c_k s_k^2 \tag{3}$$

where c_k is the appropriate elastic modulus and s_k is the strain. For a unit volume,

$$s_k = u_k \sin \omega t \tag{4}$$

$$\text{thus} \quad V = \sum_{k=1}^{n} V_k = \frac{1}{2}\sum_{k=1}^{n} u_k^2 c_k \sin^2 \omega t \tag{5}$$

$$\text{and} \quad \langle V \rangle = \frac{1}{4}\sum_{k=1}^{n} u_k^2 c_k \tag{6}$$

Similarly for the kinetic energy,

$$T_k = \frac{1}{2}\rho \frac{d}{dt}(u_k^2 \sin^2 \omega t) \tag{7}$$

$$\text{thus} \quad T = \frac{1}{2}\rho\omega^2\sum_{k=1}^{n} u_k^2 \cos^2 \omega t \tag{8}$$

$$\text{and} \quad \langle T \rangle = \frac{1}{4}\rho\omega^2\sum_{k=1}^{n} u_k^2 \tag{9}$$

Equating the two energies and rearranging,

$$\omega^2 = \frac{\sum_{k=1}^{n} u_k^2 c_k}{\rho\sum_{k=1}^{n} u_k^2} \tag{10}$$

If a localized change occurs, causing an effective variation of the elastic constants, this can be modelled for a unit volume by considering the jth element. If the elastic modulus for this element changes from c_j to $c_j + \delta c_j$, this results in a modified resonant frequency $(\omega + \delta\omega)$ such that

$$(\omega + \delta\omega)^2 = \frac{\sum_{k=1}^{n} c_k u_k^2 + \delta c_j u_j^2}{\rho\sum_{k=1}^{n} u_k^2} \tag{11}$$

giving $$\frac{\delta\omega}{\omega} = -\frac{\delta c_j u_j^2}{2\omega^2 \rho \sum_{k=1}^{n} u_k^2} \qquad (12)$$

This shows that the change in frequency is proportional to the energy of the thickness mode at the point of laser irradiation. When applied to the full area of the laser spot, this results an expression of the form

$$\frac{\delta f}{f_0} \propto \int_0^{2\pi} K_1(\psi)\, u^2(x_0 + r\cos\psi,\ y_0 + r\sin\psi)\, d\psi \qquad (13)$$

where (x_0,y_0) is the position on the plate of the laser spot centre and r is the spot radius. It is recognized here that this is not a complete solution, since the effect on a constrained quartz cylinder of a radial compression will also result in a circumferential tension component, but this is neglected, and could be the subject of later analysis. However, the evaluation of this function provides a useful insight into the behaviour of devices.

To interpret this expression, numerical evaluations have been made for spot positions covering a 36 mm^2 area of the resonator surface, assuming a spot diameter of 2 mm. The results of these calculations are shown in fig 5a and b for AT-cut (ϕ = 0) and SC-cut (ϕ = 22) respectively. The assumed resonator design for both cases is a 10 MHz 3rd overtone with a 1 dioptre contour. Figure 6a and b illustrates the dependence of the predicted modulation level on laser spot radius, assuming that the laser energy density at the crystal surface is a constant value. This is the case where degradation of the resonator surface is the limiting factor for the light intensity. It shows that if the device is flooded with light, the modulation level should be low, and also indicates an optimum radius for maximum sensitivity.

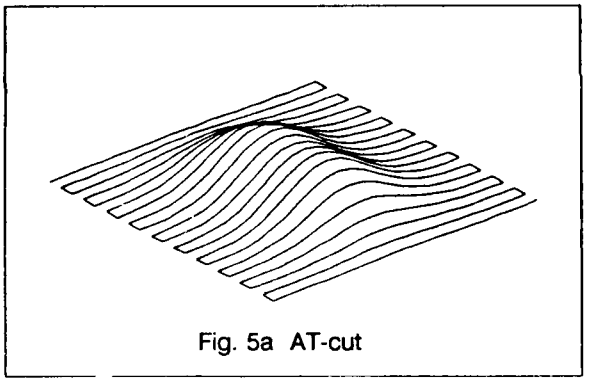


Fig. 5a AT-cut

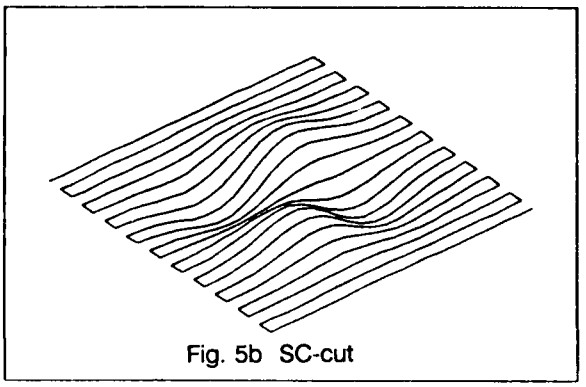


Fig. 5b SC-cut

Fig. 5 Theoretical plots of laser-induced modulation across resonator surface.

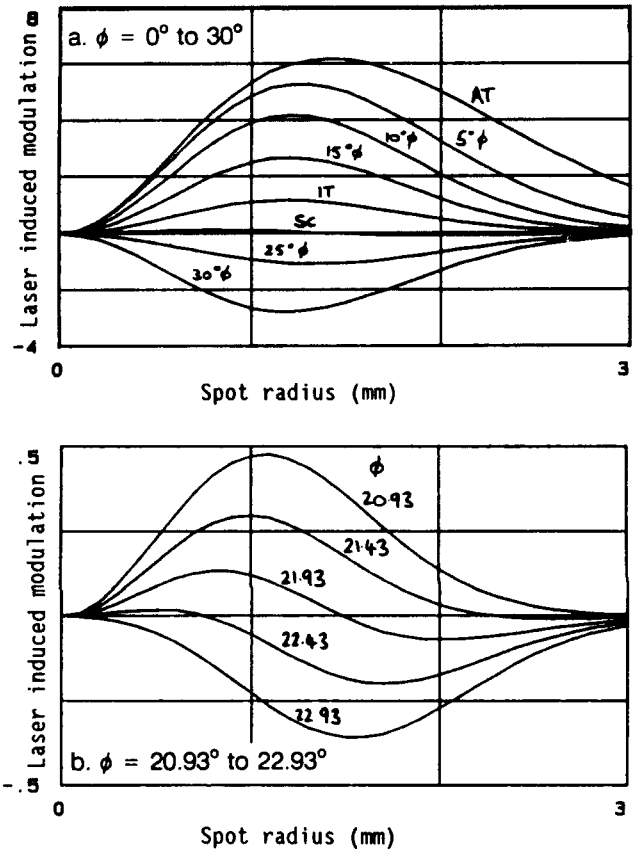


Fig. 6 Theoretical plot of modulation against laser spot radius for various ϕ orientations along the AT-SC ZTC locus.

Experimental Outline

Resonator Designs

Six batches of resonators were fabricated for this exercise, 20 of each design: one batch of AT-cut, and five batches straddling the previously published value for the SC-cut of ϕ = 21°56'. These five batches were cut at 30' intervals, giving a total range in ϕ of 2°.

All of the resonators were mounted on four point mount cold weld TO8 bases with the mounts aligned along the x'' and z'' axes. The designs were 10 MHz 3rd overtone with a 1 dioptre plano-convex geometry, and a

blank diameter of 10 mm. The θ angles were chosen to give turnover temperatures at between 30°C to 40°C to avoid frequency drift with temperature, whilst anticipating a general warming effect due to laser (which was later corroborated by the measurements). The orientation of the batches was verified by measurement of a sample over temperature, and by measurement of all of the resonators for the b-mode to c-mode frequency ratio, giving an accurate determination of the value of ϕ.

Measurement Conditions

For the purposes of these experiments, the laser spot diameter was set to approximately 2 mm. The laser energy was set to 50 mJ per pulse, and the primary aperture was about 10 mm in diameter, giving a fluence at the crystal surface of 40 mJ/cm^2. This was a compromise between obtaining the required modulation level and the possibility of electrode damage. For smaller spot sizes, the light intensity required for a reasonable signal level resulted in scintillation effects on the crystal surface. With larger spot sizes, even the maximum laser power output was insufficient to obtain the required modulation level whilst retaining adequate beam symmetry. A pulse repetition rate of 10 Hz was used for most of the tests. Again, this was a compromise, since, for lower frequencies, $1/f$ noise became a problem, whereas for higher frequencies, the decay time of the crystal phase became a significant proportion of the time between pulses, making the phase setting of the lock-in amplifier a critical variable. With a pulse rate of 10 Hz, the lock-in amplifier phase was always set to 0°, giving a consistent measure of the laser-induced phase modulation.

The maximum value of K_t is seen from fig. 4 to occur at a ψ orientation of about 35° for the SC-cut. The maxima of the phase modulation should therefore occur along a line at 35° from the x'' axis. To simplify the fixturing, and also bearing in mind the other errors involved in the measurement, the crystals were actually mounted such that the x-y table caused the laser spot to traverse the blanks at 45° from the x'' and z'' axes.

For every resonator which was measured, the modal centre was found first, by traversing the laser spot in each direction until the saddle point was found. In the case of the AT-cut the centre of resonance was defined by a local maximum. The beam was then moved in 1 mm steps across the full diameter of the blank in the horizontal direction, and after stabilization, a phase reading was taken at each point. After returning to the centre of the resonator, a second traverse was made in the vertical direction, and readings were taken using the same method.

Experimental Results

The results obtained for the six batches are shown graphically in figs. 7 to 12. The horizontal and vertical axes are superimposed on the same graphs, clearly showing the saddle points. It should be noted that the curves have been shifted in the spot position axis so that the central maxima or minima are all coincident on zero. The actual variation in the saddle point position from unit to unit was found to be up to 2.5 mm. The apparent variation in the maxima and minima either side of the saddle point in both amplitude and position is probably partially caused by a spread at mounting in the azimuthal angle ψ and by positional inaccuracies in the equipment. Figure 13 summarizes the variation in laser-induced modulation for the five groups around the nominal SC-cut orientation, showing the variation in modulation level at the saddle point against ϕ for each crystal, where the orientation was inferred by measurement of f_b/f_c.

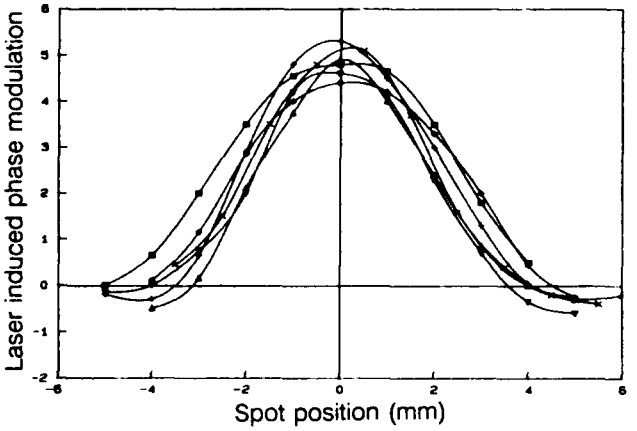


Fig. 7 Plot of laser-induced modulation against spot position for group AT1 (AT-cut).

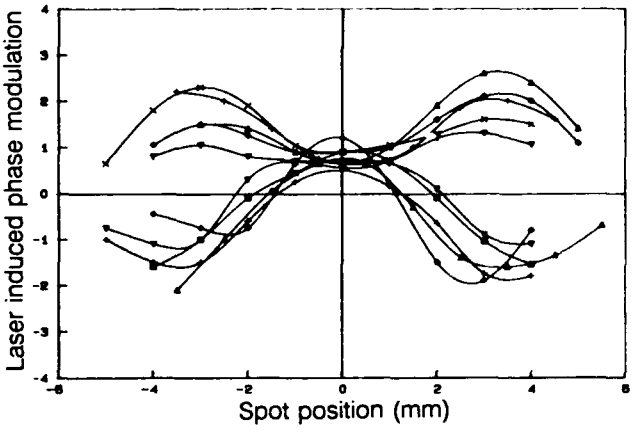


Fig. 8 Plot of laser-induced modulation against spot position for group RT4 (ϕ = 20°56').

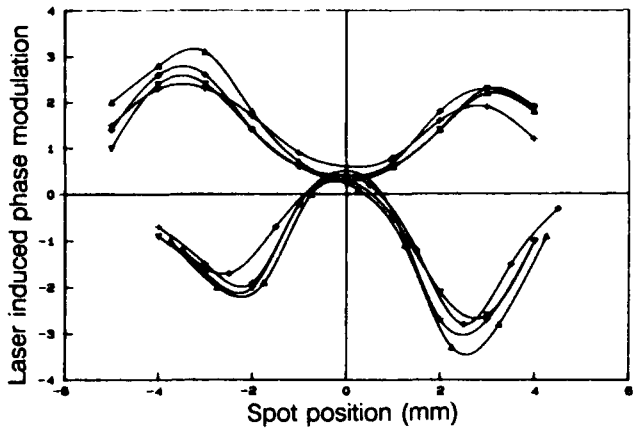


Fig. 9 Plot of laser-induced modulation against spot position for group RT5 (ϕ = 21°26').

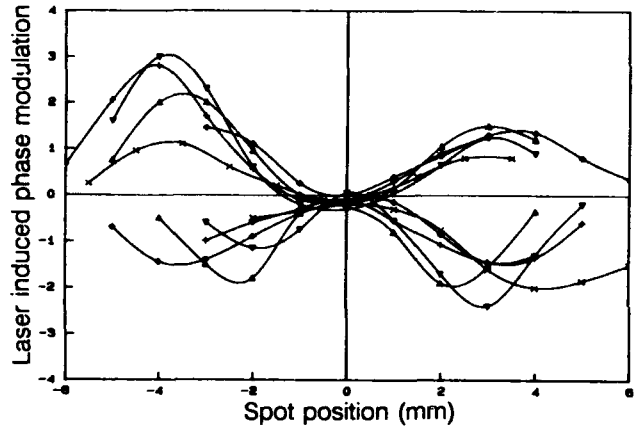


Fig. 10 Plot of laser-induced modulation against spot position for group RT1 (ϕ = 21°56').

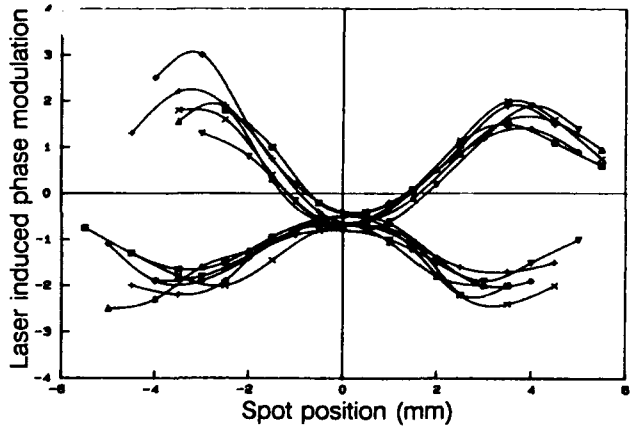


Fig. 11 Plot of laser-induced modulation against spot position for group RT3 (ϕ = 22°26').

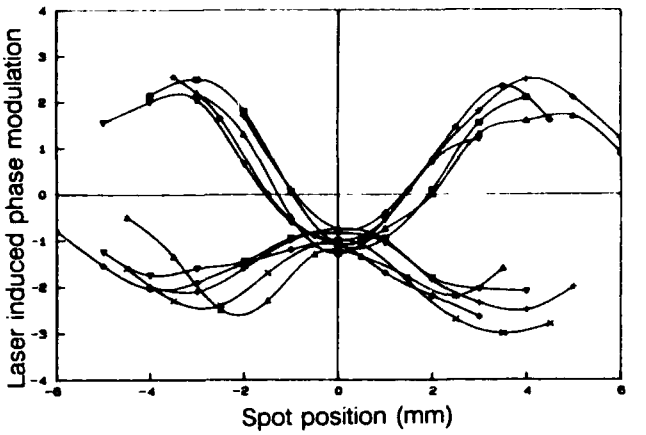


Fig. 12 Plot of laser-induced modulation against spot position for group RT2 (ϕ = 22°56').

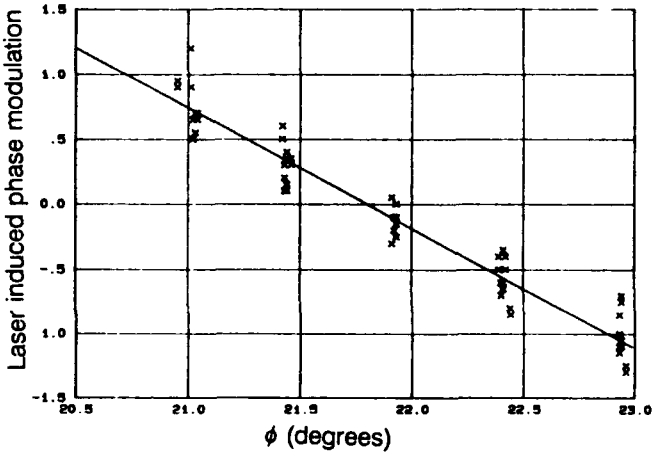


Fig. 13 Distribution of observed modulation at saddle point against ϕ (as inferred by ratio of mode frequencies).

The plot of fig. 13 suggests an orientation of (ϕ,θ) = (21.8°,34.1°) which is compensated for stresses induced by the pulsed laser beam. However, all of the resonators measured here were nominally mounted along the same crystallographic axes. During the course of the measurements, some crystals were rotated by 90° about the plate normal, thus transposing the vertical and horizontal axes, and these were then remeasured. The phase modulation readings were consistently slightly higher in value than those shown in fig. 13, which in turn would predict a slightly higher value for the optimum value of ϕ. This anomaly was considered to be due to irregularities in the intensity of the laser spot, and could be greatly improved by modification of the laser optics. This would be an area of interest in future work.

Conclusions

The method described here clearly demonstrates the mechanism of stress compensation in SC-cut quartz resonators. The modulation induced by the laser has been shown to exhibit a saddle point at the modal centre of the resonator, and the modulation level at this saddle point has been shown to be strongly dependent on the ϕ angle.

For the 10 MHz 3rd overtone designs used here, the results demonstrate the existence of an orientation along the AT-SC locus at $(\phi,\theta) = (21.8°,34.1°)$ which is compensated for stresses introduced by the pulsed thermal energy from the laser. However, the energy distribution of the laser spot is known to be non-uniform, causing a small systematic error in the results. Compensating for this error gives an optimum value of ϕ which is extremely close that obtained by Kusters [4] (21.93°), although more experimental work would be required for a high confidence level for this value, preferably using an optical system which generates a beam with an isotropic energy distribution. Such a system could be created using a simple incandescent light source.

The technique clearly illustrates asymmetries in the mounting and optical processing of resonators, and it could be used to optimize these parameters. EerNisse *et al.* have shown [20] that mode shape and mode location greatly influence the acceleration sensitivity of SC-cut crystals, and have used selective mass loading to improve their characteristics. The laser method could be used, perhaps in conjunction with the selective deposition technique, as a tool for determining the degree of asymmetry. This could be incorporated into a process for optimization of acceleration sensitivity.

Work for the future could include improvement and simplification of the optics, after which further tests would be carried out on resonators of different designs, including other overtone modes and different plate geometries.

Acknowledgements

We would like to take this opportunity to thank Dave Cox for making the resonators used for this work and Reg Spencer for cutting and preparing the blanks.

References

1. E. P. EerNisse, "Quartz resonator frequency shifts arising from electrode stress", 29th Annual Symposium on Frequency Control, 1975, pp. 1-4.

2. E. P. EerNisse, "Calculations on the stress compensated (SC-cut) quartz resonator", 30th Annual Symposium on Frequency Control, 1976, pp. 8-11.

3. R. Holland, "Nonuniformly heated anisotropic plates 1. Mechanical distortion and relaxation", IEEE Transactions on Sonics and Ultrasonics, July 1974, pp. 171-178.

4. J. A. Kusters, "Transient thermal compensation for quartz resonators", IEEE Transactions on Sonics and Ultrasonics, July 1976, pp. 273-276.

5. J. A. Kusters, "The SC cut - An overview", Proceedings 1981 Ultrasonics Symposium, pp. 402-408.

6. R. W. Ward, "Design of high performance SC resonators", 35th Annual Symposium on Frequency Control, pp. 99-103.

7. A. W. Warner, B. Goldfrank and J. Tsaclas, "Further development in SC cut crystal resonator technology", 36th Annual Symposium on Frequency Control, 1982, pp. 208-214.

8. A. Ballato, E. P. EerNisse and T. Lukaszek, "The force-frequency effect in doubly rotated quartz resonators", 31st Annual Symposium on Frequency Control, 1977, pp. 8-16.

9. J. M. Ratajski, "The force sensitivity of AT-cut quartz crystals", IBM Journal of Research and Development, Jan. 1968, pp. 33-49.

10. A. Ballato, E. P. EerNisse and T. J. Lukaszek, "Experimental verification of stress compensation in the SC cut", Proc. 1978 Ultrason. Symp., pp. 144-147.

11. J. C. Brice and W. S. Metcalf, "Quartz-crystal resonators using an unconventional cut", Philips Technical Review, Vol. 40, 1982, No. 1.

12. J. A. Kusters and J. Leach, "Further experimental data on stress and thermal gradient compensated crystals", Proceedings of the IEEE , Vol. 65, Feb. 1977, pp. 282-284.

13. J. A. Kusters, C. Adams, H. Yoshida and J. Leach, "TTC's - Further developmental results", 31st Annual Symposium on Frequency Control, 1977, pp. 3-7.

14. R. J. Williamson, "Plotting of the vibrational distribution of a TS quartz resonator using a pulsed laser", 1989 European Frequency and Time Forum, pp. 22-27.

15. R J. Williamson, "Non-destructive means for determining mode shape in thickness-shear resonators", this proceedings.

16. E. P. EerNisse, "Variational calculation of force-frequency constants of doubly rotated quartz resonators", IEEE Trans. Sonics and Ultrasonics, May 1978, pp. 132-137.

17. M. Mizan and A. Ballato, "The stress coefficient of frequency of quartz plate resonators", 37th Annual Symposium on Frequency Control, 1983, pp. 194-199.

18. D. S. Stevens and H. F. Tiersten, "An analysis of doubly-rotated contoured quartz crystal resonators", 39th Annual Symposium on Frequency Control, 1985, pp. 436-447.

19. Lord Rayleigh, "Theory of Sound", 2nd revised edition, vol. 1, section 88, reprinted by Dover, New York.

20. E. P. EerNisse, R. W. Ward and O. L. Wood, "Acceleration-induced frequency shifts in quartz resonators", 43rd Annual Symposium on Frequency Control, 1989, pp. 388-395.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

CONTAMINATION CONTROL METHODS - A REVIEW OF RECENT PROGRESS

John R. Vig
U.S. Army Electronics Technology and Devices Laboratory, LABCOM
Fort Monmouth, NJ 07703

Abstract

Many of the cleaning issues and processes used in the semiconductor industry are also applicable to the frequency control industry. The amount of contamination control research in the semiconductor industry far outweighs that in the frequency control industry; the frequency control industry can learn much from this semiconductor-related research.

This paper reviews a selection of papers from the First International Symposium on Cleaning Technology in Semiconductor Device Manufacturing (held in October 1989) which are applicable to frequency control device processing. The topics of the 43 papers presented included wet cleaning technologies, dry cleaning technologies, particles and airborne contaminants, and the characterization of cleaning.

Introduction

Contamination control is essential during frequency control device manufacturing - as it is during semiconductor device manufacturing. In the semiconductor industry, it is widely recognized that wafer cleaning plays a major role in achieving high production yields and device reliabilities. Changes in cleaning technology are being driven by progress towards higher density devices in integrated circuits. Environmental and occupational health regulations are also requiring changes. Problems have been identified with the use of liquid chemicals in wafer processing, as is discussed below. A totally dry process may, therefore, be essential for the high-yield processing of future generations of integrated circuits.

The reason for this paper is that many of the cleaning issues and processes in the semiconductor industry are applicable to the frequency control industry as well. The amount of contamination control research in the semiconductor industry far outweighs that in the frequency control industry. Whereas in the past the primary concern in the semiconductor industry was inorganic contamination - while in the frequency control industry it was (adsorption-desorption of) organic contamination - today both types of contaminants are of serious concern in both industries. In quartz resonators and filters, organic contamination can cause aging, noise, and processing yield problems. Inorganic contamination (especially particles and metal flakes) can cause nonlinearities (such as high starting resistance and intermodulation), noise, and yield problems.

The First International Symposium on Cleaning Technology in Semiconductor Device Manufacturing[1] was held in October 1989, in Hollywood, Florida, as part of the 176th Meeting of the Electrochemical Society. The topics of the 43 papers presented included wet cleaning technologies, dry cleaning technologies, particles and airborne contaminants, and the characterization of cleaning. Reviewed here are a selection of papers that are applicable to frequency control device processing.

Review and Discussion

The usual contaminant of concern in clean rooms has been particles. Until recently, little or no attention has been paid to organic vapors. Typical clean room air handling systems filter the particles but not the organic vapors. In "Airborne Concentrations of Organic Vapors and Their Surface Accumulations on Wafers During Processing in Clean Rooms," by A. J. Muller et al.[2] of AT&T, measurements are reported of the airborne concentrations and surface arrival rates of selected organic species at three manufacturing locations, two rural and one urban. The total airborne concentrations were found to be > 100 $\mu g/m^3$ at all three sampling locations. This contrasts with an estimated particulate mass concentration of 20 ng/m^3 in a class 100 clean room, i.e., the mass of organics per unit volume of air is about 5,000 times greater than that of particles. The results are consistent with a previous study (reference 1 of Muller, et al.) which examined the airborne concentrations of semi-volatile organic compounds (10 to 10^{-4} torr vapor pressure range) in "electronic equipment rooms," and found typical values of 30 $\mu g/m^3$, and values as high as 1,000 $\mu g/m^3$ when construction was in progress.

The deposition velocities of organic molecules are higher than those of the typical clean room particles, due

US GOVERNMENT WORK IS NOT PROTECTED
BY US COPYRIGHT

to the organics' much higher diffusivity in air. The arrival rate, i.e., the mass deposition rate, of organic contamination at initially clean surfaces will greatly exceed the arrival rate for particles. Species with less than about 10^{-2} torr vapor pressure can be expected to condense on surfaces. Highly volatile species, such as methanol and acetone, are also likely to be present in clean rooms, and will quickly produce monolayer coverage on clean surfaces, but their high volatility will prevent accumulation much beyond a monolayer.

A wide variety of organics were identified, as is shown in Table 1. The average concentrations and arrival rates for several selected compounds are shown in Table 2. The total surface arrival rates for all the organic species that were quantified was typically > 2 pg/cm^{2}sec, and was sometimes > 20 pg/cm^{2}sec. To see the implications of such arrival rates for a quartz resonator, let us assume an arrival rate of 10 pg/cm^{2}sec, a sticking coefficient of 0.5, a 10 MHz fundamental mode AT-cut resonator, and a 1 cm^2 electrode area. The mass of the active area is then 44 mg, and $\Delta f/f = \Delta m/m \approx 2 \times 10^{-10}$ per second $\approx 2 \times 10^{-5}$ per day. Therefore, for the highest stability, the time that resonators are left exposed to the atmosphere, even in a clean room, ought to be minimized.

Table 1. Organic Contaminants in Clean Rooms

Trichloroethane	Cellosolve acetate
Trichloroethylene	Hexamethylcyclotrisiloxane
Tetrachloroethylene	Octamethylcyclotetrasiloxane
Dichlorobenzenes	3-Methylhexane
Benzene	Trimethylhexane
Toluene	Dimethylcyclohexane
Ethylbenzene	Tetrachloroethane
Xylenes	Propylbenzene
C_3-Alkylbenzenes	3-Ethyltoluene
Methylcyclohexane	4-Ethyltoluene
n-Nonane (Nonane)	1,3,5 Trimethylbenzene
n-Decane (Decane)	2-Ethyltoluene
n-Undecane	1,2,4 Trimethylbenzene
C_{12}-C_{16} n-alkanes	Butyl acetate

Werner Kern,[3] Lam Research Corp., reviewed "The Evolution of Silicon Wafer Cleaning Technology." "Wafer cleaning chemistry has remained essentially unchanged in the past 25 years...what has changed is its implementation..." The cleaning methods have been based on a two step process, Standard Clean 1, commonly referred to as SC-1, followed by Standard Clean 2, SC-2. Both are oxidizing and complexing treatments with hydrogen peroxide solutions. SC-1 is an alkaline mixture; SC-2 is an acidic mixture. SC-1 removes organic films by oxidative breakdown and dissolution, and thereby exposes surfaces for the concurrent or subsequent decontamination reactions. SC-2 removes alkali ions and cations, and metallic contaminants. SC-1 also etches thermally grown SiO_2, at a rate of about 0.1 nm/min; SC-2 does not etch such an oxide.

Table 2. Concentrations and Arrival Rates of Organic Contaminants

Compound	Arrival Rate (At Charcoal) μg/(cm^{2}sec)	Concentration (Active Samplers) μg/cm^3	Ambient Mass Transfer Coefficient * cm/sec
Toluene	3.0 - 5.3 x 10^{-7}	0.5 - 8.9 x 10^{-6}	0.006 - 0.65
Ethylbenzene	0.7 - 1.3 x 10^{-7}	0.2 - 3.7 x 10^{-6}	0.02 - 0.43
Butyl acetate	8.9 - 88 x 10^{-7}	2.6 - 88 x 10^{-6}	0.08 - 2.4
Cellosolve acetate	4.3 - 10.2 x 10^{-7}	0.6 - 4.2 x 10^{-6}	0.1 - 1.67
Nonane	0.6 - 1.91 x 10^{-7}	0.18 - 4.7 x 10^{-6}	0.14 - 0.44
Decane	1.3 - 2.6 x 10^{-7}	0.36 - 11.9 x 10^{-6}	0.014 - 0.46

*Mass Transfer Coefficient = Arrival Rate/Ambient Concentration

SC-1 is a solution of 5 parts H_2O, 1 part H_2O_2, and 1 part NH_4OH. SC-2 consists of 6 parts H_2O, 1 part H_2O_2, and 1 part HCL. The water is ultrafiltered and deionized; the H_2O_2 is 30%, high purity and unstabilized; the NH_4OH is 29% electronic grade; and the HCL is 37% electronic grade. The processing temperatures are 75°C to 80°C.

The implementation methods have evolved from simple immersion techniques through the incorporation of megasonic techniques for particle removal, to completely enclosed and automated systems in which the wafers are kept stationary during the entire cleaning, rinsing and drying processes. The process fluids flow sequentially and continuously over stationary wafers loaded in cassettes. The recontamination encountered when wafers are pulled out from a liquid is thereby avoided. The megasonic cleaning process differs from the usual ultrasonic process in that it uses sonic waves of 850-900 kHz frequency rather that the 20-80 kHz used in ultrasonic cleaners. The pressure waves travel through the liquid parallel to the wafer surfaces. The pressure waves' impact can remove particles as small as 0.3 μm diameter from surfaces.

Kern also discussed rinsing and drying methods, and the criticality of the rinsing and drying steps. If these steps are not done correctly, the wafers become recontaminated. Rinsing is often done with flowing high-purity, ultrafiltered, high-resistivity DI water, at room temperature. Megasonic rinsing, and centrifugal spray rinsing can be advantageous. Rinsing in a closed system has the advantage that the wafers are not removed between the cleaning, rinsing and drying steps. Wafer drying must be done by physical removal of the water rather than by allowing the water to evaporate. Spin drying is the most widely used technique. Hot forced air drying, capillary drying, and solvent vapor drying, usually with isopropyl alcohol or nonflammable azeotropic mixtures which displace the water, are also used.

As device geometries become smaller and smaller, the cleaning levels required are becoming more and more difficult to achieve with wet cleaning techniques. The particles contained in liquid chemicals, and the ability of liquids to clean within small openings become significant issues. The trend in particle contamination requirements[4] in the integrated circuit fabrication is summarized in Table 3. The generally accepted assumption is that maximum permissible defect size is 50% of the linewidth. Therefore, for integrated circuits such as the 16 MB DRAMs, particles as small as ≈ 0.1 μm in diameter will cause problems.

Table 3. Evolution of Particle Control Requirements

	1985	1988	1990	1993
DRAM (bits)	256K	1M	4M	16M
Linewidths (μm)	2.0	1.0	0.5	0.25
Max. Defect Size (μm)	1.0	0.5	0.25	0.125

In a "Review of Particle Control Methods During Wet Chemical Cleaning of Silicon Wafers," V. B. Menon, Sematech, and R. P. Donovan,[5] Research Triangle Inst., discussed the purity of cleaning chemicals, cleaning process chemistries and how the various wet cleaning methods can add particles to surfaces, the effectiveness of megasonic cleaning methods, and particle contamination during rinsing and drying processes. In Table 4, particle concentrations of "low particulate grade" bottled chemicals are compared with those of filtered and recirculated chemicals. Continuous filtration can produce significantly lower particle concentrations. While it takes more than an hour to reach the steady state levels, after only 6 to 10 minutes, the concentrations are below that found in bottled chemicals. HF solutions are among those that are known to add particles to wafers. Fig. 1 shows the particle concentration for a buffered HF solution as a function of recirculation/filtration time. It took several hours for particle levels to stabilize.

Table 4. Particles in Recirculated vs. Bottled Chemicals - FSI ChemFill™ Recirculation System vs. CleanRoom™ Low Particulate Grade Chemicals (1 Gallon Bottles) From Ashland Chemical Company, Columbus, OH

Operating Time (Hrs.)	Particles/ml > 0.5 um in H_2SO_4	NH_4OH	H_2O_2	HCl
0.1	27	28	0.06	0.02
0.5	10	4.3	0.04	0.01
1.0	7.0	1.9	0.03	0.01
2.0	4.6	0.85	0.03	0.01
4.0	3.1	0.38	0.02	<0.01
6.0	2.4	0.24	0.02	<0.01
8.0	2.0	0.17	0.02	<0.01
Bottled Chemicals Specification	60	500	100	40

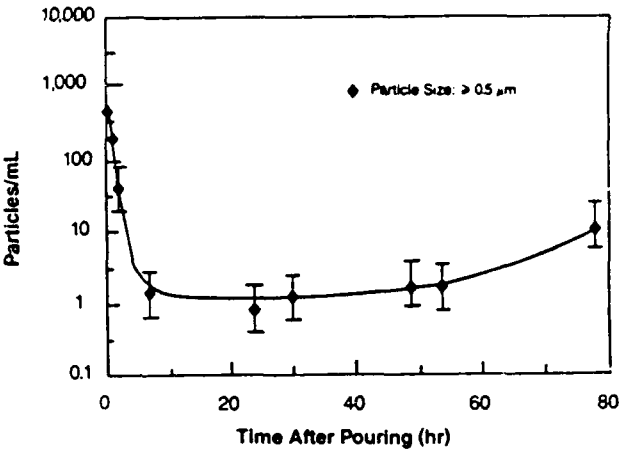


Figure 1. Particles in Buffered HF vs. Recirculation Time

The rinsing and drying steps are major sources of particulate contamination. Cascade overflow rinsing was found to be superior to methods such as quick-dump with top-spray rinsing. The moving parts of spin dryers can be sources of particles, however, a good spin dryer can add fewer than 10 particles on a 6-inch diameter wafer. Isopropyl alcohol (IPA) vapor dryers have the potential for avoiding the particle generation problems of spin dryers;

however, the particles added in IPA vapor dryers so far have not been found to be significantly lower than those added in a well functioning spin dryer.

An effective particle control strategy requires: 1) stringent control of chemical purity, 2) an understanding of the wafer-chemical interactions, 3) a rational rinse and dry protocol, and 4) monitoring of particles in the incoming chemicals, the process tanks, the rinse tank, the dryer, and on the wafers.

In "Wet Chemical Processing in a Trench," C. M. Tipton and R. A. Bowling[6] of Texas Instruments, Inc. showed that etching and cleaning liquids can penetrate very small openings via capillary forces. This work is relevant to the etching of, and cleaning in, the etch channels of quartz. The trenches investigated are comparable in size to the diameters of etch channels, i.e., on the order of 1 μm. Failure to etch was observed only at aspect ratios (depth:width) of greater than 450:1. Within a trench, liquids with higher surface tensions more easily displace liquids with lower surface tensions. For example, on thermal SiO_2 surfaces, DI water (which was found to have a 41° contact angle) will displace 5% HF (16° contact angle) and 0.5% buffered HF (31° contact angle). However, since DI water is known to have a very low (≈ 4°) contact angle on clean quartz, rinsing of the etchant from etch channels presents a problem. The surface tension of etchants can be lowered with surfactants. The use of surfactants, therefore, is advantageous not only because it can result in more uniform etching, as was shown previously,[7,8] but also because it enhances the rinsing of etchants from the etch channels.

In "Ultra Clean Ice Scrubber Cleaning With Jetting Fine Ice Particles," T. Ohmori et al.,[9] Mitsubishi Electric Corp. and Taiyo Sanso Co., describe a method of removing particles, and films of grease and oil from surfaces. The contaminants are removed with a jet of fine, ultraclean ice particles. Particles of 30 μm to 300 μm diameter are formed by spraying ultrapure water into cold (-80°C to -150°C) nitrogen gas that is formed when liquid nitrogen is sprayed into a freezing chamber. The ice particles are sprayed through a jet nozzle onto the wafer surfaces, as is shown in Fig. 2. Table 5 shows the comparative test results for particle cleaning methods. Compared are the rates of removal of standard 0.322 μm diameter polystyrene particles from silicon wafers.

Fig. 3 shows an explanation of the cleaning mechanisms. Particles are removed due to the impact of the ice particles, including the action of the ice particles melting momentarily upon impact. The molten and refrozen ice particles can incorporate particulate contaminants which are removed by the impact of subsequent ice particles. Organic films can become brittle when the ice particles cool the substrate. The adhesion between the film and the substrate is weakened when the substrate contracts. The impact of ice particles can then break up and remove the film. The effectiveness depends on the melting point of the film. Films with melting points above -10°C, which includes fingerprints, were successfully removed. Ice scrubber cleaning removed fingerprints and oil inks more effectively than did trichloroethylene or trichlorotrifluoroethane.

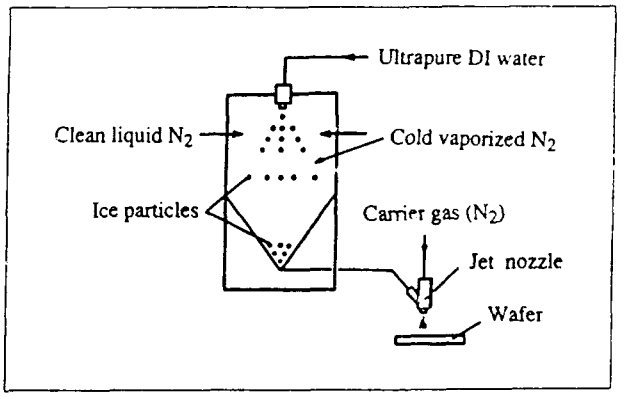


Figure 2. Ice Scrubber Cleaning

Table 5. Comparison of Particle Cleaning Methods

Cleaning Methods	Removal Rate (%)	Condition	
Ice Scrubber	97.6	Pressure 3kg/cm² Angle 80°	Time 30sec
	94.0	Pressure 2kg/cm² Angle 80°	Time 30sec
Megasonic	95.0	Frequency 950kHz	Time 10min
Brush Scrubber	87.4	Brush-pressure 0.8kg/cm²	Time 40sec
High-Pressure Water	84.4	Water-Pressure 100kg/cm²	Time 40sec
Ultrasonic	83.9	Frequency 27kHz	Time 15min
Dry Ice Scrubber	68.9		Time 30sec

In "UV/Ozone Cleaning of Surfaces," J. R. Vig[10] reviewed a dry cleaning method that can rapidly produce clean surfaces, in air or in a vacuum system, at ambient temperatures. Placing properly precleaned surfaces within a few millimeters of an ozone-producing ultraviolet light source can produce near-atomically clean surfaces in about one minute. In addition to reference 10, details of this

method can also be found in the Proceedings of the 28th (1974) and 29th (1975) Annual Symposium on Frequency Control (pp. 96-108 and pp. 220-229, respectively).

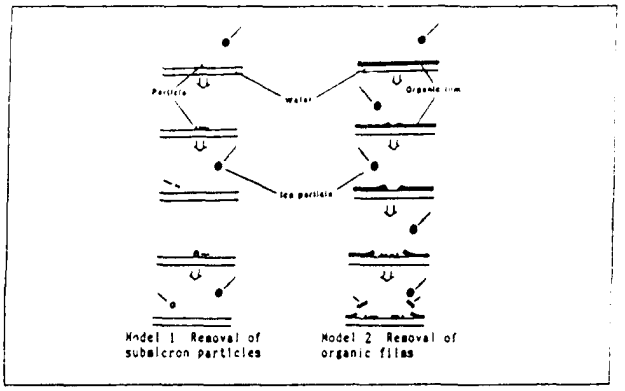


Figure 3. Ice Scrubber Cleaning Mechanism

In "UV-Enhanced Dry Cleaning of Silicon Wafers," T. Ito et al., Fujitsu Laboratories Ltd., describe a dry cleaning method which uses UV-excited high-purity chlorine for the removal of inorganic materials, such as metals, from silicon wafers. Chlorine radicals produced by the UV light react with metal contaminants, resulting in volatile chlorine compounds. The technique also etches the silicon wafer surface by means of a process in which the chlorine radicals diffuse through the thin native oxide and react with the silicon to form volatile silicon chloride compounds, such as $SiCl_2$ and $SiCl_4$. The combination of UV/ozone and UV/chlorine may form the basis of a completely dry cleaning process. The combination may also be useful for the final cleaning and frequency adjusting of quartz resonators. The questions of whether or not the chlorine radicals diffuse into the quartz surface, and if yes, whether or not vacuum baking is capable of removing such chlorine, would need to be answered first.

Large quantities of hydrofluoric acid are used in the manufacture of silicon semiconductor devices. The cleanliness and the purity of the HF used directly impacts device yield and reliability. The practice of using ultrapure HF for a short time then discarding it creates costly consumption and disposal problems. J. Davison et al., Athens Corp. and Texas Instruments, in "The Use of Reprocessed HF in Semiconductor Quartz and Wafer Cleaning Operations," describe a reprocessor that was developed to recycle and repurify the HF solutions used in semiconductor quartz and wafer cleaning operations. Ion exchange and filtration are used to remove ionic and particulate contaminants. The reprocessed HF is ultrapure.

Conclusion

Now that the contamination control requirements of the semiconductor industry have become similar to that of the frequency control industry, the frequency control industry can benefit from the vastly greater resources being expended on contamination control in the semiconductor industry.

Acknowledgement

The author thanks Dr. Jerzy Ruzyllo of Pennsylvania State University, who was co-chairman of the First International Symposium on Cleaning Technology in Semiconductor Device Manufacturing, for providing preprints of the papers reviewed here, and for helpful discussions.

References

[1] Proceedings of the First International Symposium on Cleaning Technology in Semiconductor Device Manufacturing, Semiconductor Cleaning Technology / 1989, Proceedings Volume 90-9, The Electrochemical Society, Inc., 10 South Main St., Pennington NJ, 08534-2896, 1990.

[2] A. J. Muller, et al. "Airborne Concentrations of Organic Vapors and Their Surface Accumulations on Wafers During Processing in Clean Rooms," ibid., pp. 204-211.

[3] Werner Kern, "The Evolution of Silicon Wafer Cleaning Technology," ibid., pp. 3-19. This paper also appeared in the J. Electrochem. Soc., Vol. 137, pp. 1887-1892, 1990.

[4] W. C. Krusell, and D. I. Golland, "Cleaning Technologies for High Volume Production of Silicon Wafers," ibid., pp. 23-32.

[5] V. B. Menon and R. P. Donovan, "Review of Particle Control Methods During Wet Chemical Cleaning of Silicon Wafers," ibid., pp. 167-181.

[6] C. M. Tipton and R. A. Bowling, "Wet Chemical Processing in a Trench," ibid., pp. 33-42.

[7] R. J. Brandmayr and J. R. Vig, "Chemical Polishing in Etching Solutions That Contain Surfactants," Proc. 39th Ann. Symp. on Frequency Control, pp. 276-281, 1985.

[8] R. J. Brandmayr and J. R. Vig, "Further Results on the Use of Surfactants in Chemically Polishing Quartz Crystals," Proc. 40th Ann. Symp. on Frequency Control, pp. 86-90, 1986.

[9] T. Ohmori, et al., "Ultra Clean Ice Scrubber Cleaning With Jetting Fine Ice Particles," see ref. 1 above, pp. 182-191.

[10] J. R. Vig, "UV/Ozone Cleaning of Surfaces," ibid., pp. 105-113. This paper is a condensed version of a book chapter of the same title and author, in Treatise on Clean Surface Technology, Vol. 1, K. L. Mittal, ed., pp. 1-26, Plenum Press, NY, 1987.

[11] T. Ito, et al., "UV-Enhanced Dry Cleaning of Silicon Wafers," see ref. 1 above, pp. 114-120.

[12] J. Davison, et al., "The Use of Reprocessed HF in Semiconductor Quartz and Wafer Cleaning Operations," ibid., pp. 83-91.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

RESISTIVITY, ADHESIVE STRENGTH, AND RESIDUAL STRESS MEASUREMENTS OF THIN FILM METALLIZATIONS ON SINGLE CRYSTAL QUARTZ*

P. T. Vianco, W. R. Conley, and J. K. G. Panitz
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185

ABSTRACT

Resistivity, adhesive strength, and residual stress measurements were made of the thin films, 450 Å Cr/1800 Å Au and 450 Å Mo/1800 Å Au deposited on optically polished, z-plate single crystal quartz surfaces. The films were analyzed after deposition and following annealing at 380°C or 450°C for 30 min in either air or vacuum. The resistivity of the Cr/Au film increased by 1090% and 1780% after vacuum annealing at 380°C and 450°C, respectively. These increases were reduced when 1000 Å of Ni or Mo were introduced as barrier layers between the Cr and Au to prevent interdiffusion of the two layers. The resistivity of the Mo/Au films remained unchanged after all annealing exposures. The Cr/Au, Cr/Mo/Au, and Cr/Ni/Au films had adhesion strengths of 41 to 70 MPa in both the as-deposited and annealed conditions. The adhesive strength of the Mo/Au metallization was 23 to 31 MPa except after the 450°C air anneal, where the value dropped to zero. The residual stress of the as-deposited Cr/Au film was 170 MPa (tensile) but became compressive at −120 to −250 MPa after annealing. The Mo/Au metallization was deposited with a zero stress value that also became compressive at −140 to −180 MPa.

INTRODUCTION

Single crystal quartz has been used in numerous applications beyond the long-established role as a resonator for oscillator circuits. Innovative designs of the resonant cavity have given rise to sensors based upon predictive changes of the resonant frequency caused by exposure to external environments such as high and low temperatures or applied stresses. The result is a new generation of components which measure thermal fluctuations, acceleration,and rate of rotation. The manufacturing processes, as well as the user environments, expose these devices to many harsh conditions such as elevated temperature. Although the substrate material, quartz, is stable over extended temperature ranges, the mechanical and electrical properties of the thin film metallizations used to provide the electric field are not necessarily so, resulting in possible changes to the resonator performance.

The thin film electrodes used on piezoelectric resonators typically contain at least two layers. The first is a deposit of 100 to 500 Å of metals such as Cr or Mo which promote adhesion between the substrate and the second, more highly conductive layer which is usually Au of 1000 to 2000 Å thick. Exposing these multilayer films to elevated temperature during subsequent processing and use of the device can cause interdiffusion of the layers. Interdiffusion potentially changes the resistivity, adhesion, and residual stress of the films which can degrade the device performance. This investigation examined the resistivity, adhesion, and residual stress of thin films deposited on single crystal quartz substrates.

Experimental Procedures

A. Candidate films and heat treatments. The two metallizations of principal interest were: *(1)* 450 Å of Cr (on the quartz) with 1800 Å of Au and *(2)* 450 Å of Mo (on the quartz) followed by 1800 Å of Au. Included in this investigation were variations to the Cr/Au films in which barrier layers of 1000 Å Ni or 1000 Å Mo were introduced between the Cr and Au layers. Also examined was a pure Au film, 1800 Å thick.

The heat treatment parameters were selected to coincide with secondary processing steps common to sensor assembly and use. A matrix of annealing conditions was created from temperatures of 380°C and 450°C, a time period of 30 min, and atmospheres of lab air or vacuum ($\leq 2\times10^{-6}$ torr).

* This work was performed at Sandia National Laboratories under United States Department of Energy contract DE-AC04-76DP00789.

CH2818-3/90/0000-207 $1.00 © 1990 IEEE

B. Deposition parameters. The films were deposited by electron beam evaporation. A multi-source turret hearth permitted the sequential deposition of multiple layers without breaking vacuum. Deposition rates for *(1)* Mo, 5 ± 1 Å /sec; *(2)* Cr, 5 ± 1 Å /sec; *(3)* Ni, 4 ± 1 Å /sec; and *(4)* Au, 7 ± 1 Å /sec were selected for their stability, ease of control, and minimal heat loading to the substrates. The background pressure was maintained below 5×10^{-6} torr during all depositions. Final film thicknesses were determined by contact surface profilometry of partially shadowed fused silica plates.

The substrates for the resistivity measurements were preheated in the vacuum chamber to 200°C for 60 min with this temperature being maintained during the deposition. The substrates used to acquire the adhesion strength and residual stress data were not heated.

All substrates were cleaned by *(a)* degreasing in Freon vapor, *(b)* ultrasonic agitation in acetone, *(c)* ultrasonic agitation in ethanol, and *(d)* then blown dry with nitrogen.

C. Resistivity measurements. The substrates were single crystal quartz plates measuring 26.2×24.4×0.152 mm. The surfaces on which the films were deposited were normal to the z-direction (0001) and optically polished. The metal films were deposited through an aperture (shadow) mask to form 8 lines on the quartz blank, each measuring 10.2 mm long and 1.27 mm wide. The geometry of the films and substrate appear in Fig. 1.

Two wafers were deposited with each metallization type in the same run. The resistivities of the 8 strips on each wafer were measured; an average value was calculated for each group of 4 strips. The wafers were then broken in half so as to separate each of the groups of 4 strips. Each group was then exposed to one of the 4 annealing treatments. This procedure ensured that the data was taken from films deposited in the same run. Furthermore, it allowed for a precise comparison between the as-deposited resistivity and the value after heat treatment.

The resistance of each film was measured by the four-point probe technique. To insure electrical contact to the film while preventing mechanical damage, osmium tipped, light load probes were used. The two outer probes, which supplied a direct current of 45.3 mA, were 7.6 mm apart. The voltage probes were spaced 5.1 mm apart, collinear with the current probes. The resistivity was calculated using the expression $R\times A/L$, where R is the measured resistance (ohms), A is the cross sectional area of the strip (thickness times 1.27×10^{-3}; m^2), and L is the separation between the voltage probes (5.1×10^{-3} m). The data spread quoted with each average value denotes plus-or-minus one standard deviation of the results.

D. Adhesive strength measurements. The substrates used for these tests were quartz plates measuring 17.8×17.8×1.60 mm. The surface on which the films were deposited was optically polished and oriented perpendicular to the z-direction. A total of 15 substrates were coated with each metallization type. Three of the substrates were used to test the film in the as-deposited condition. The remaining substrates were separated into 4 groups of 3 samples. Each group was subject to one of the 4 heat treatments from the matrix outlined in section A.

The adhesive strength was measured by the Sebastian method, which is diagrammed in Fig. 2. An aluminum pin with a flared base 2.5 mm in diameter was bonded to the film surface with an epoxy which was cured by heating the unit in an air oven at 150°C for 60 min. The sample was then placed into the tester, which pulled the pin from the substrate. The force measured by a load cell combined with the footprint of the pin determined the stress (normal to the film) at failure which is displayed on the instrument. The failure mode of the bonds was noted to validate the reading. Typically, three tests were performed on each sample for a maximum of 9 tests per exposure condition. Resistivity and Auger depth profile analysis of as-deposited films subjected to the epoxy curing cycle demonstrated no significant change to either property. These results were used to infer that the epoxy curing cycle did not change the film character. The values presented in this report are the average stresses; the accompanying error range represents plus-or-minus one standard deviation of the data.

E. Residual stress measurements. The substrates were single crystal quartz beams measuring 76×13×0.79 mm with the broad surfaces perpendicular to the z-direction. Both sides of the extended surfaces were optically polished.

The magnitude of the residual stress was calculated from the curvature of the beam measured in a large optics Michelson interferometer built at Sandia National Laboratories [1]. Digital image analysis and the specially designed

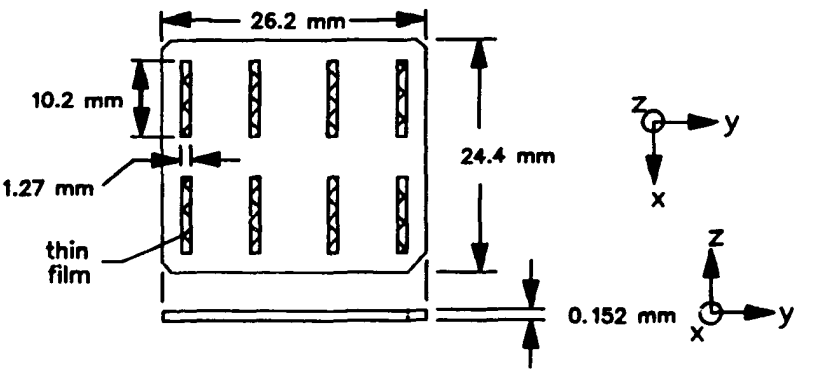


Figure 1. Substrate and metallization geometries for resistivity test samples

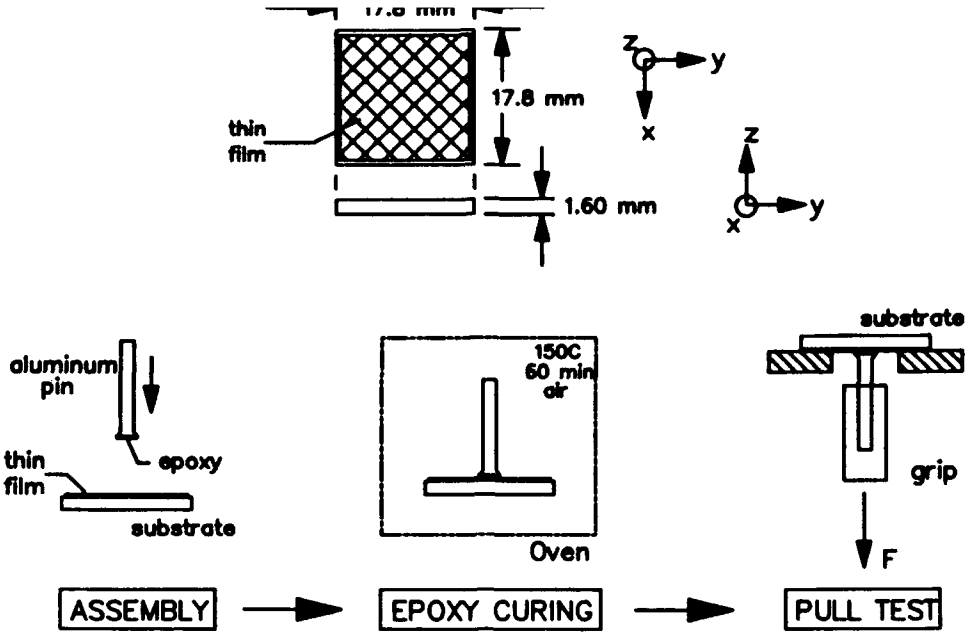


Figure 2. Substrate geometry and assembly procedure for adhesive test samples.

hardware made it possible to measure very small radii of curvature (on the order of 1 km). Small changes in curvature resulting from thin films deposited on relatively thick substrates allowed for the calculation of the film's residual stress without the need to know the film's elastic modulus. The equation used to calculate the film stress is given below (for details, see the analytical analysis in Reference 1):

$$\sigma_f = E_s t_f/(6R_s)[(t_s/t_f)^2 - t_s/t_f + 6(E_f/E_s)(y_f/t_f)] \quad (1)$$

where t_f is the film thickness, t_s is the substrate thickness, E_f is the film modulus, E_s is the substrate modulus, R_s is the radius of curvature at the film/substrate interface, and y_f is the position in the film (at which the stress is being calculated) as measured from the film center plane towards the film/substrate interface. The maximum stress will be at the interface; that is where y_f/t_f is 0.5. A very thin film and a thick substrate cause the third term in the square brackets of equation (1) to be negligible with respect to the other two terms. Therefore, it is not necessary to know the film modulus in order to calculate the film stress. The modulus of the quartz beam was calculated from available data and confirmed in the same apparatus by applying known forces to the beam and recording the resulting curvature.

A layer of 1800 Å of Au was deposited on one surface of the beam to form the reflecting layer from which to measure the curvature. The substrate was then heated at 450°C for 1 hour to stabilize the Au film. The thin film of interest was then deposited on the other side. A total of 4 beams were deposited with each metallization at the same time. An average curvature and error term based upon one standard deviation were calculated from the 4 beams.

F. Auger depth profile measurements. Auger depth profiles were made of each of the films in the as-deposited condition as well as after each of the annealing treatments. Deposition of the films was at 200°C. Sputtering was performed by an argon ion beam. Surface sputtering followed by signal detection provided a (qualitative) chemistry profile of the film layer.

Results and Discussion

A. Resistivity. Shown in Table 1 are the as-deposited resistivity data from two separate deposition runs of each metallization. These values correspond to measured resistances of approximately 0.5 to 0.6 ohms over the line geometries described in Fig. 1. An estimate of the percent error introduced by the measurement technique is approximately 6%. Therefore, the different values of resistivity for each film indicated a real variation between the two separate runs. The effect of deposition temperature on film resistivity was investigated by evaporating the films at ambient temperature (<30°C), measuring the resistivity after coating, annealing the films at 200°C for 30 min, and finally performing a second resistivity measurement. The results are outlined in Table 2. It was apparent that the run-to-run variations of resistivity exemplified in Table 1 were comparable to the change in resistivity due to *(1)* changing the substrate temperature from ambient to 200°C during deposition, or *(2)* by the low temperature annealing (200°C, 30 min) of the film deposited at ambient temperature.

The percent resistivity changes as a result f the heat treatments are presented in Table 3. The results of the Auger depth profiles will also be discussed. The Au and Mo/Au films showed resistivity decreases for all heat treatments. No interdiffusion between the Mo and Au layers was observed. This observation is confirmed by the phase diagram of the Au-Mo couple [2], which shows only the terminal solid solutions of *(1)* maximum 1.25 at.% Mo in Au

and (2) maximum 0.4 at.% Au in Mo for a simple eutectic without intermediate intermetallic compounds for the annealing temperatures. The trend of decreasing resistivity with annealing was probably caused by changes to the microstructure of the films [3,4,5]. The Mo/Au film surface retained a golden color except after annealing at 450°C in air, where some blistering caused the surface to darken slightly (Fig. 3a). The blistering was initiated with air entering the film and oxidizing the Mo layer. Certain oxides of Mo (e.g., MoO_3) are volatile at 450°C [6]. Vapor formation caused the Au layer to burst and form blisters. Trace Mo was observed on the film surface by Auger analysis; suggesting that the analyzed area included blisters.

The discontinuities introduced into the film by the blisters, did not affect the resistivity. However, experiments in which such films were annealed in air at 500°C for 180 min showed large resistivity increases as the blisters grew in size (Fig. 3b).

The Cr/Au film experienced large resistivity increases of 1130% and 1090% after annealing at 380°C for 30 min in air and vacuum, respectively. Those values climbed to 2240% and 1780%, respectively, when the heat treating temperature was increased to 450°C. This general increase of resistivity was caused by the diffusion of Cr into Au as demonstrated by a comparison of the Auger depth profiles shown in Fig. 4. The low resistivity of the Au film, 2.3×10^{-6} ohm-cm [7], was severely degraded by the Cr, which has a resistivity of 36.6×10^{-6} ohm-cm.

The air annealed films had slightly higher resistivity increases as opposed to the vacuum annealed specimens. Oxidation of the Cr within the Au layer was not extensive as observed by the Auger analysis. Therefore, it appears that the oxidation of the Cr at the film surface increased the amount of Cr entering the Au layer. The surfaces of the films developed a gray tint indicative of Cr and which darkened as the annealing temperature was increased for

Table 1. Resistivity of metallizations deposited at 200°C in two separate runs

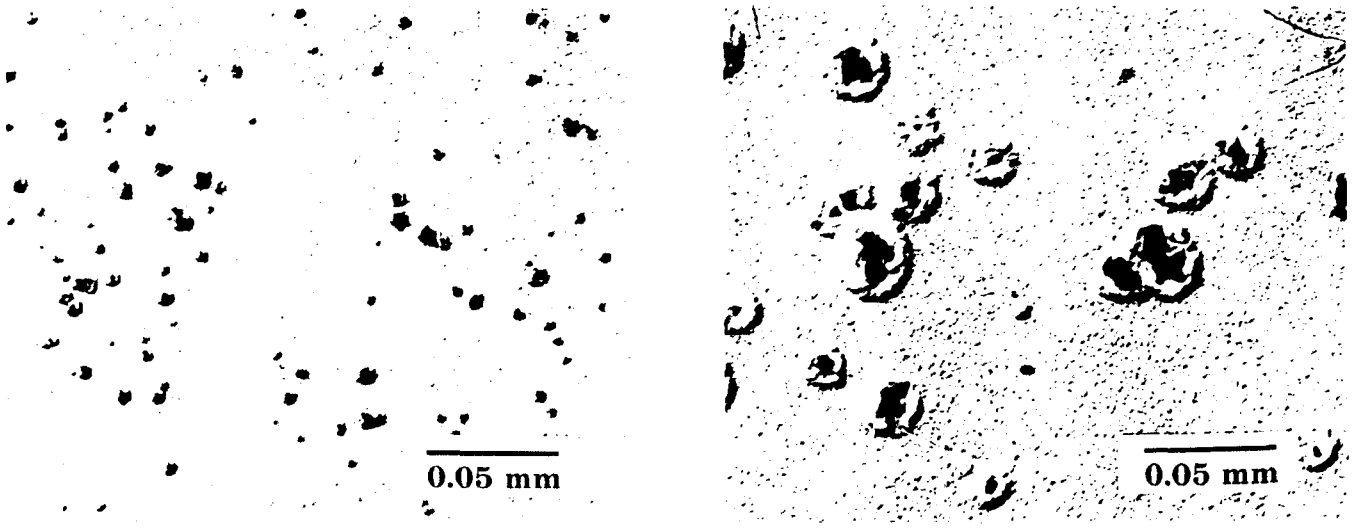
Film	Resistivity $\times10^{-8}\Omega$m 200C Deposit Temp.	
1800 Å Au	2.39±0.02,	4.53±0.04
450 Å Mo/1800 Å Au	4.04±0.05,	3.98±0.03
450 Å Cr/1800 Å Au	3.59±0.04,	5.64±0.10
450 Å Cr/1000 Å Mo/1800 Å Au	5.24±0.03,	6.50±0.05
450 Å Cr/1000 Å Ni/1800 Å Au	5.44±0.03,	5.28±0.04

Table 2. Resistivity of metallizations deposited at room temperature in two separate runs and then annealed in air at 200°C for 30 min

Film	Resistivity $\times10^{-8}\Omega$m R.T. Dep. → 200C, 30 min., air	
1800 Å Au	3.81±0.03	3.31±0.03
450 Å Mo/1800 Å Au	4.24±0.02	3.90±0.02
450 Å Cr/1800 Å Au	4.71±0.02	5.10±0.10
450 Å Cr/1000 Å Mo/1800 Å Au	7.59±0.04	6.95±0.07
450 Å Cr/1000 Å Ni/1800 Å Au	6.33±0.07	5.28±0.06

Table 3. Percent change of resistivity of metallizations deposited at 200°C

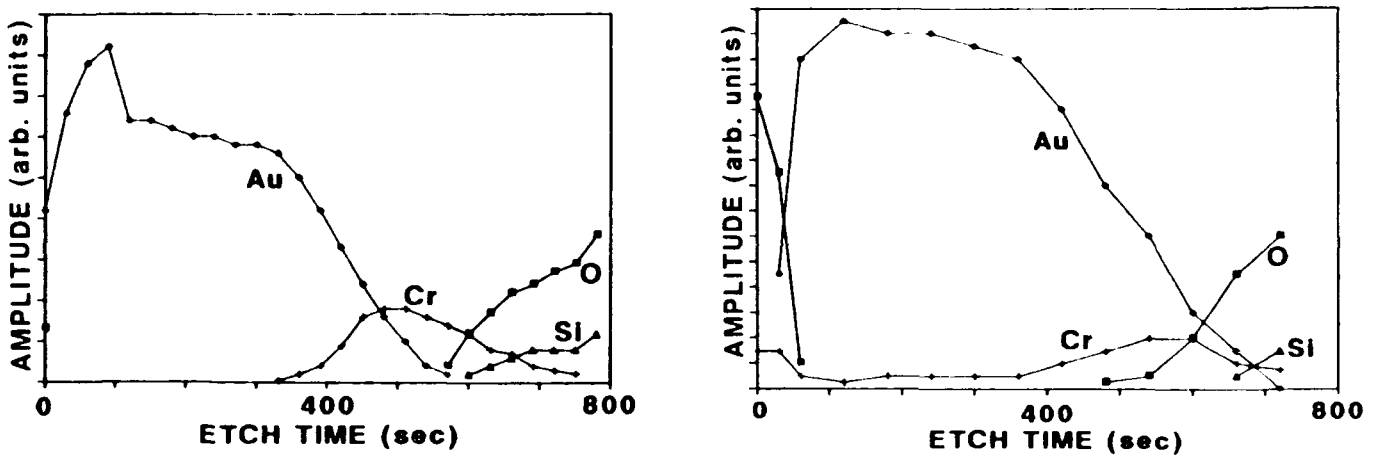
Film	Percent Change (%) 380C, 30 min air	vac.	450C, 30 min air	vac.
1800 Å Au	−2.4	−2.6	−8.4	−8.8
450 Å Mo/1800 Å Au	−3.8	−9.0	−9.4	−9.4
450 Å Cr/1800 Å Au	1130	1090	2240	1780
450 Å Cr/1000 Å Mo/1800 Å Au	−4.3	−9.4	992	914
450 Å Cr/1000 Å Ni/1800 Å Au	416	38.0	715	496



(a) at 450°C for 30 min

(b) at 500°C for 180 min

Figure 3. Optical micrographs of blisters on Mo/Au films annealed in air



(a) after deposition

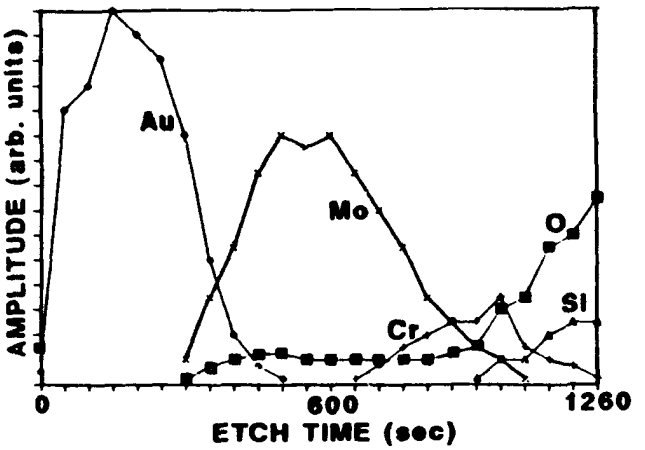
(b) following annealing in air at 450°C for 30 min

Figure 4. Auger depth profile of the Cr/Au film

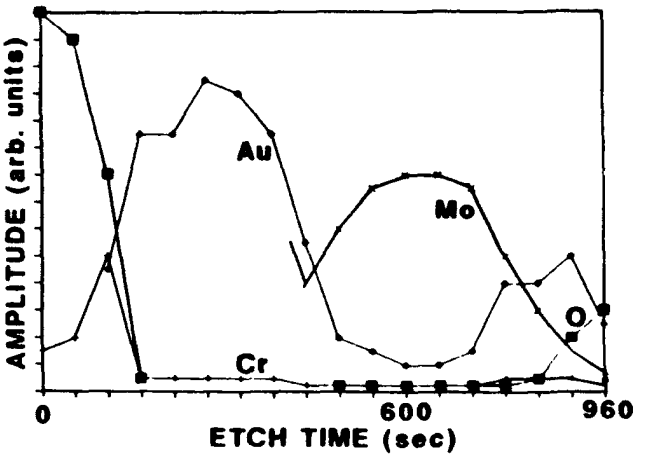
both air and vacuum conditions.

For the Cr/Mo/Au film, resistivity decreases were observed after annealing at 380°C. The values following the air and vacuum anneals at this temperature were nearly the same as those observed for the Mo/Au film. The Cr/Mo/Au film behaved similarly to the Mo/Au because the Mo layer prohibited the interdiffusion of the Cr and Au layers. However, raising the annealing temperature to 450°C caused the Mo barrier to break down. Shown in Fig. 5 is the Auger depth profile of *(a)* the as-deposited Cr/Mo/Au film and *(b)* that of the film after annealing at 450°C in air. These profiles clearly show the movement of both Au and Cr past the Mo layer. No interdiffusion was observed in the Mo/Au films at either of the annealing temperatures. Therefore, it appears that the Cr moved through the Mo barrier first, thereby providing a path by which the Au passed through the Mo barrier back towards the quartz. The Cr-Mo phase diagram [8] shows a continuous solution without intermediate phases and a (calculated) miscibility gap below 880°C which may permit solutions of 5 at.% Mo in Cr and 9 at.% Cr in Mo to form. This was apparently the case for the Cr-Mo couple.

The slightly higher resistivity change of air versus vacuum annealing (450°C) was not caused by oxidation of Cr within the Au layer. The oxide signal in Fig. 5b follows the Mo signal with a magnitude similar to the as-deposited case (Fig. 5a). A change to the appearance of the film was noted only after annealing at 450°C at which point a gray tint was observed on the surface. This gray tint was probably due to the oxidation of Cr and a roughening of the surface. Blistering was not observed. Therefore, an enhancement of the amount of Cr in the Au layer after air annealing

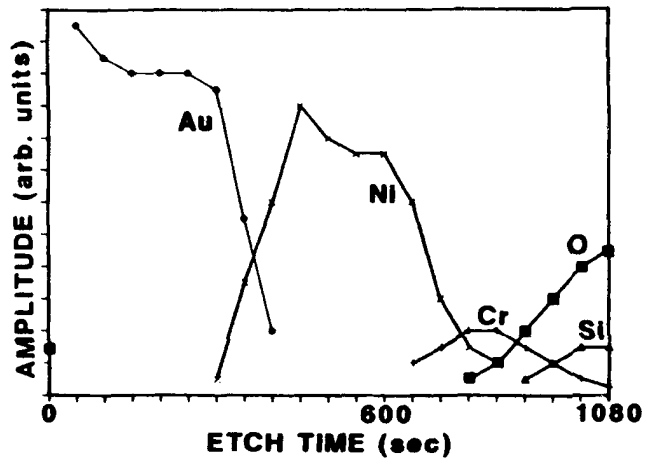


(a) after deposition

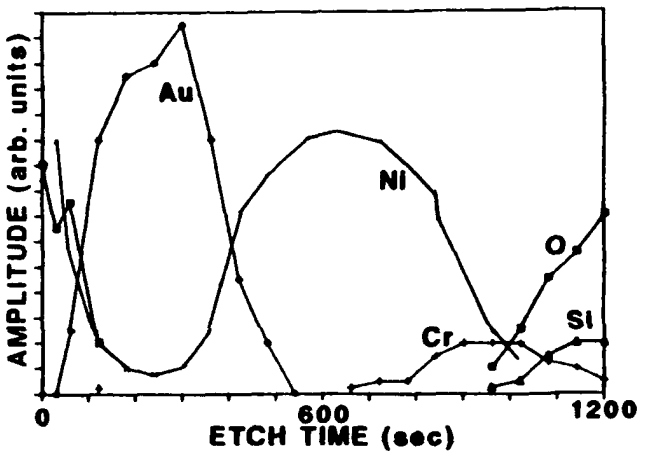


(b) following annealing in air at 450°C for 30 min

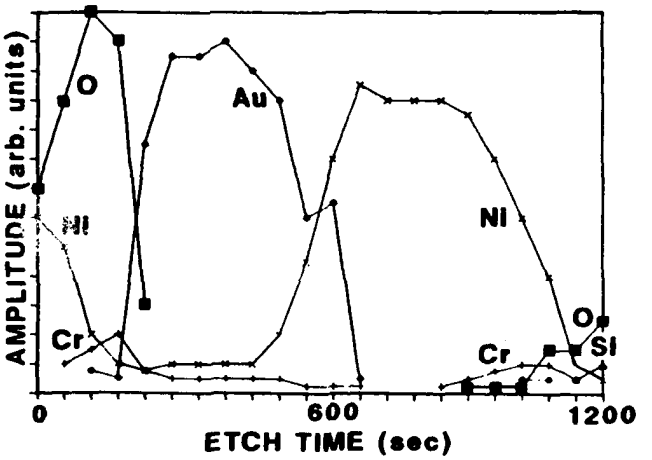
Figure 5. Auger depth profile of the Cr/Mo/Au film



(a) after deposition



(b) following annealing in air at 380°C for 30 min



(c) after annealing in air at 450°C for 30 min

Figure 6. Auger depth profile of Cr/Ni/Au film

versus vacuum annealing caused the higher resistivity changes of the former case

The Ni layer proved to be a poor barrier between the Cr and the Au. Resistivity increased under all annealing conditions. The Auger depth profiles representing the as-deposited film; the film annealed at 380°C in air; and the profile of the film following a 450°C air heat treatment are shown in Fig. 6. The resistivity increases at 380°C were due largely to the interdiffusion of Ni into the Au layer. At 450°C, the Cr layer which was relatively inactive at 380°C, alloyed with the newly formed Ni-Au layer and caused 715% and 496% resistivity increases for air and vacuum annealing, respectively. The resistivity values were higher after air annealing than when done in vacuum for both temperatures. Oxidation of the Ni and Cr components following air annealing was observed at the film surface but not in the film interior. Therefore, the oxidation of Ni and/or Cr on the surface enhanced the diffusion of either species (depending on the temperature) into the Au layer, causing the larger resistivity increases for air versus vacuum annealing.

Air annealing caused the film surface to develop an orange hue at 380°C which darkened to a brownish color at 450°C. The various colors were caused by the oxidation of the Ni and Cr layers on the surface.

In conclusion, the Mo/Au film resulted in no increase to the resistivity under the prescribed heat treatments. Mo also prevented a resistivity increase in the Cr/Mo/Au films by acting as a barrier to the interdiffusion of Cr and Au. The Mo barrier became ineffective at the 450°C annealing temperature. Replacing Mo with Ni resulted in extensive interdiffusion between Ni and Au at 380°C as well as between Cr and the Ni-Au layer at 450°C. Although the resistivity of the Cr/Ni/Au films increased in all cases, the increases were less than those observed with the Cr/Au films. Surface oxidation-enhanced diffusion of Cr or Ni components caused slight increases to the resistivity values of the air annealed samples as compared to the vacuum annealed films.

B. Adhesive strength tests. The adhesion test results appear in Table 4. A particular pull test was considered a valid datum if *(1)* the machine failed to pull the pin off of the specimen prior to reaching the maximum load, *(2)* fracture took place in the epoxy film at a stress in excess of 9000 psi (which is approximately 90% of the load capacity of the machine), or *(3)* the film separated from the substrate at any stress level.

The 1800 Å Au film did not adhere to the quartz substrate in the as-deposited state or following any of the heat treatments. These tests were made to demonstrate that reported adhesion values are not anomalously high due to epoxy reaching the substrate through defects (pinholes) in the metallizations.

The Mo/Au film showed an as-deposited strength of 25 ± 17 MPa. The strength value remained unchanged after all of the heat treatments except the 450°C air anneal in which the adhesion stress dropped to zero. The blistering of the film resulted in the oxidation of the underlying Mo layer (beyond the actual open areas). The powdery character of remaining Mo oxide provided inadequate adhesion between the Au layer and the quartz.

The Cr/Au films showed an as-deposited adhesive strength of 48 ± 10 MPa. The value increased to the machine limit of 70 MPa after annealing at 380°C in both air and vacuum. The values dropped slightly following the 450°C annealing treatments. These results clearly indicate that in spite of the large degree of intermixing between the Cr and Au layers, the Cr/Au alloy film retained exceptional adhesive strength.

The Cr/Mo/Au had an as-deposited adhesive strength of 56 ± 10 MPa, which was statistically similar to that of the Cr/Au film. This was expected because in both cases, the Cr layer was providing the adhesion. The values remained effectively unchanged for films after each of the heat treatments.

The results of the Cr/Mo/Au system have also resulted in some added insight into the Cr/Au system. For example, the Mo layer was an effective barrier to the interdiffusion of Cr and Au at the 380°C annealing temperature. Yet the adhesive strength of the annealed Cr/Au films exceeded those of the Cr/Mo/Au films. Therefore, since the Au component of the alloyed Cr-Au film is not expected to add any adhesive strength, the elevated temperature appears to enhance the adhesion between the Cr component of the *intermixed Cr/Au film* and the quartz substrate. At 450°C, the Mo barrier of the Cr/Mo/Au film broke down and the Cr andAu intermixed. The resulting

Table 4. Adhesive strength of metallizations deposited at room temperature and annealed

Film	Adhesive Strength (MPa)				
	As-Depos.	380C, 30 min air	380C, 30 min vac.	450C, 30 min air	450C, 30 min vac.
1800 Å Au	0	0	0	0	0
450 Å Mo/1800 Å Au	25±17	29±13	31±10	0	23±14
450 Å Cr/1800 Å Au	48±10	>70*	>70*	61±12	68±7.0
450 Å Cr/1000 Å Mo/1800 Å Au	56±10	57±12	41±23	64±10	56±10
450 Å Cr/1000 Å Ni/1800 Å Au	65±3.0	65±8.0	66±4.0	69±2.0	69±4.0

* Machine limit.

adhesion was not significantly different from that of the Cr/Au films after similar annealing.

The adhesive strength data of the Cr/Ni/Au films exhibited no change from the as-deposited value of 65 ± 3 MPa in spite of the interdiffusion observed at the annealing temperatures. The as-deposited strength was significantly larger than that of the Cr/Au films. This observation clearly indicated that even with similar layers (Cr) next to the substrate, the adhesive strength of a multilayer film is affected by all of the individual layers.

In conclusion, the Cr/Au, Cr/Ni/Au, and Cr/Mo/Au films exhibited excellent adhesive strengths in the as-deposited condition as well as after each of the heat treatments. The Mo/Au metallization had generally lower strength values with severe degradation to the film adherence after air annealing at 450°C due to oxidation and loss of the Mo underlayer by blistering.

C. Residual stress measurements. The residual stress data of the Mo/Au and Cr/Au films are listed in Table 5. The Mo/Au metallization had practically no stress in the as-deposited condition while the Cr/Au film had a tensile stress of 170 ± 30 MPa. For both cases, however, the film stresses became compressive upon annealing.

The Cr/Au data clearly showed that the residual stress magnitudes were larger following air annealing than after heat treating in vacuum; the difference was relatively small at 380°C and larger at 450°C. Assuming that the residual stress changes were caused by interdiffusion between the Cr and Au, then the residual stress magnitudes indicate that mixing was larger during air annealing than after vacuum heat treatment, thereby confirming the explanation of surface oxidation enhanced diffusion of Cr into Au used to explain a similar trend in the resistivity results noted earlier. However, the substrate curvature may also reflect a volume expansion at the film surface due to oxidation which would appear as a greater compressive residual stress than noted after vacuum annealing. Further tests are needed to determine how air annealing enhanced the film stress as compared to vacuum heat treatment.

No appreciable difference in the magnitude of the residual stress was observed for the Mo/Au film after air annealing at 450°C, in spite of the severe structural degradation.

An important result of these tests on the Cr/Au film is that although they were deposited in tension, after heat treatment they became compressive. This implies that at some point, the stress became zero. Electron beam deposition techniques offer very little control of the properties ofthe film, in particular the residual stresses. The data in Table 5 show that annealing programs can be implemented to control the film stress for optimum device performance.

Summary

1. The resistivity and adhesive strengths were measured for four multilayer films which were electron beam deposited on optically polished, z-plate single crystal quartz surfaces. Measurements were also made after the films had been vacuum and air annealed at 380°C or 450°C for 30 min.
2. The 450 Å Cr/1800 Å Au films exhibited extensive interdiffusion during all heat treatments. Resistivity increases of 1090% and 1780% resulted from vacuum annealing at 380°C and 450°C, respectively. These increases were reduced significantly when barrier layers of Mo and Ni were introduced between the Cr and Au.
3. The 450 Å Cr/1800 Å Au, 450 Å Cr/1000 Å Ni/1800 Å Au, and 450 Å Cr/1000 Å Mo/1800 Å Au films showed adhesive strengths in the range of 41 to 70 MPa in both the as-deposited and annealed conditions. The adhesive strength of the Mo/Au film was in the range of 23 to 31 MPa except after the 450°C, air anneal where the value dropped to zero.
4. The residual stress of the as-deposited 450 Å Cr/1800 Å Au was tensile at 170 ± 30 MPa and became compressive after each heat treatment. The residual stress of the 450 Å Mo/1800 Å Au film was approximately zero and became compressive following heat treatment. These results indicate the ability to control film stresses by thermal annealing.

ACKNOWLEDGMENTS

The authors wish to thank C. H. Sifford and J. A. Romero (now at Motorola, Inc.) for their meticulous process and measurement techniques. This work was performed at Sandia National Laboratories and supported by the United States Department of Energy under contract DE-AC04-76DP00789.

Table 5. Residual stress of metallizations deposited at room temperature and annealed

Film	Film Residual Stress (MPa)				
	As-Depos.	380C, 30 min air	380C, 30 min vac.	450C, 30 min air	450C, 30 min vac.
450 Å Cr/1800 Å Au	170 (T) ±30	120 (C) ±30	83 (C) ±60	250 (C) ±30	180 (C) ±30
450 Å Mo/1800 Å Au	7 (T) ±40	*	*	180 (C) ±30	140 (C) ±30

REFERENCES

[1] R. E. Cuthrell, F. P. Gerstle and D. M. Mattox, "Large Optics Michelson Interferometer for Measuring Residual Stresses in Films of Unknown Elastic Modulus," *Proc. Symp. Cer. Thin and Thick Films*, Amer. Cer. Soc, 1989.

[2] *Binary Phase Diagrams*, ed. T. Massalski. Metals Park, OH: ASM Inter., 1986, pp. 283-284.

[3] D. M. Mattox, "Adhesion and Surface Preparation" in *Deposition Technologies for Films and Coatings*, 1982, pp. 63-82.

[4] A. F. Jankowski and T. O. Wilford, "Grain Size Variation in Coatings," *Journal of Metals*, vol. 39, pp. 28-30, June 1987.

[5] C. van Opdorp and M. J. Verkerk, "Stress Distribution in Released Vacuum-Deposited Aluminum Films," *J. Appl. Phys.*, vol. 63, pp. 1518-1525, March 1988.

[6] O. Kubaschewski and C. Alcock, *Metallurgical Thermochemistry*, 5th ed., NY: Pergamon, 1979, p. 366.

[7] G. F. Carter, *Principles of Physical and Chemical Metallurgy*. Metals Park, OH: ASM, 1979, p. 102.

[8] op. cit. [1], pp. 836-837.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A PARAMETRIC STUDY OF THE VARIABLES INVOLVED IN QUARTZ GROWTH

Gary R. Johnson, Robert A. Irvine, and Jonathan W. Foise

Sawyer Research Products
35400 Lakeland Blvd.
Eastlake, Ohio 44095

ABSTRACT

As part of a research effort with the U.S. Army LABCOM, we have conducted a series of parametric experiments to examine several process variables and their effect upon the production of quartz crystals by hydrothermal growth. The experiments were designed with the aid of a computer software package which allowed for the examination of not only the effects of individual variables, but the interactions among them. Some of the advantages and disadvantages of designed experiments are discussed.

As the requirements for electronic devices become more critical, so too do the requirements placed upon the materials from which they are produced. If the crystal growing industry is to meet these requirements, a thorough understanding of the growth process is needed.

As part of a research effort with the U.S. Army Laboratory Command, we have been conducting a research project to develop the technology to produce high purity, low defect cultured quartz, so as to meet the future materials requirements of the electronics industry. The particular material goals are:

Table 1

Etch channel density:	$<10/cm^2$
Inclusion density:	<10/bar
Impurity concentrations (ppb)	
Al:	<200
Li:	<300
Na:	<500
K :	< 40
Fe:	<100
Q (3500/3800 cm^{-1}):	$>2.5 \times 10^6$
Strain:	none
Fringe distortion:	<0.05 RMS

Phase one of this research was designed to determine the limits of existing sodium carbonate technology. In particular, this effort involved the use of a noble metal liner and high purity inputs in a production autoclave.[1] Through the use of these techniques we were able to achieve the following:

Table 2

Etch channel density:	$<86/cm^2$
Inclusion density:	<420/bar
Impurity concentrations (ppb)	
Al:	700
Li:	300
Na:	1640
K :	300
Fe:	1800
Q (3500/3800 cm^{-1}):	$>2.5 \times 10^6$
Strain:	variable
Fringe distortion:	variable

Phase two of this research effort was to establish the interrelations among a variety of process variables and the quality of the quartz produced. This would allow an optimization of the process. In particular, it was hoped to study the interactions among various variables.

The first part of this second phase examined the interrelations among baffle design, temperature difference (between the dissolving and growing chambers) and power in production scale autoclaves.[2] We were able to show the linear relationship between main heater power and temperature difference (TD) and its dependence upon baffle design.

To study the interrelationships among the process variables and the quartz quality, a series of parametric experiments were designed through the use of a commercial experimental design and statistical analysis software package called

CH2818/3/90/0000-216 $1.00 © 1990 IEEE

Expert in a Chip (EChip™).[3] This package uses a technique known as robust product design. This technique allows variables to be adjusted simultaneously, thus decreasing numbers of experiments to be run.[4,5]

Based upon the results of the baffle-TD-power experiments and our production experience, the following variables were chosen:

Table 3

Variable	Range
Growing temperature	339° - 355°
Initial TD	3° - 6° C
Baffle	17% single hole vs. 7.5% perforated
Solution molarity	0.80 - 0.85M Na_2CO_3
Lithium additive	0 - 0.05 M Li^+
Hydrothermal etch temperature	-15° - 0° C
Quartz supply tape	Lascas vs. broken crystals

The original experimental design included the study of these seven variables at three levels (quadratic design) for a total of 46 trial runs (forty-one trials plus five replicates). With an estimated 80 day run length, plus turn-around and inspection time, it was not possible to fit this many experimental runs into our normal production schedule. The design was modified so that only growing temperature, which was felt to be the most critical variable, would be examined at three levels and that the rest would be examined at two. We would also reduce the number of replicates from five to three. This reduced the number of runs to twenty-three.

In addition to the variables listed above, the seed rack was designed such that a variety of seed variables could be examined including width (41 mm X, 0° rotation around X and 66 mm X, 1.5°), strain, etch, location in the rack, and horizontal vs. vertical positioning. However, the geometry of the rack remained unchanged from run to run to eliminate possible effects on fluid flow.

Runs were conducted over a period of about a year, ending in October 1989. A Z-target of 36 mm was selected, so that LABCOM requested material suitable for 15 mm SC cuts could be produced. This gave a target of 80 days run time, with runs ranging from 62 to 99 days to achieve target.

Bars from these runs were inspected for dimensions, weight, mechanical and microscopic defects and were analyzed for infrared alpha, impurity levels, and etch channel densities.

These results were then entered into the EChip program. EChip presents its analysis in two ways. First, it reports which input variables (and for 3-level and higher variables, interactions between input variables) have a significant effect upon the responses. Second, it generates a response surface to show the particular effect of those variables on a given response variable.

Figure 1 shows a response surface for the fraction of bars with cracks for the 66 mm X seeds as a function of growing temperature and supply type (1 is broken cultured crystal, 0 is natural quartz lascas). A number of characteristics of EChip's response surfaces can be pointed out. The response surface represents an empirical model based upon a regression analysis of the actual data. There is some extrapolation found in such a model. The limits of the actual data region are represented by the "convex hull" which is shown on the graph by the region marked in "*".[6]

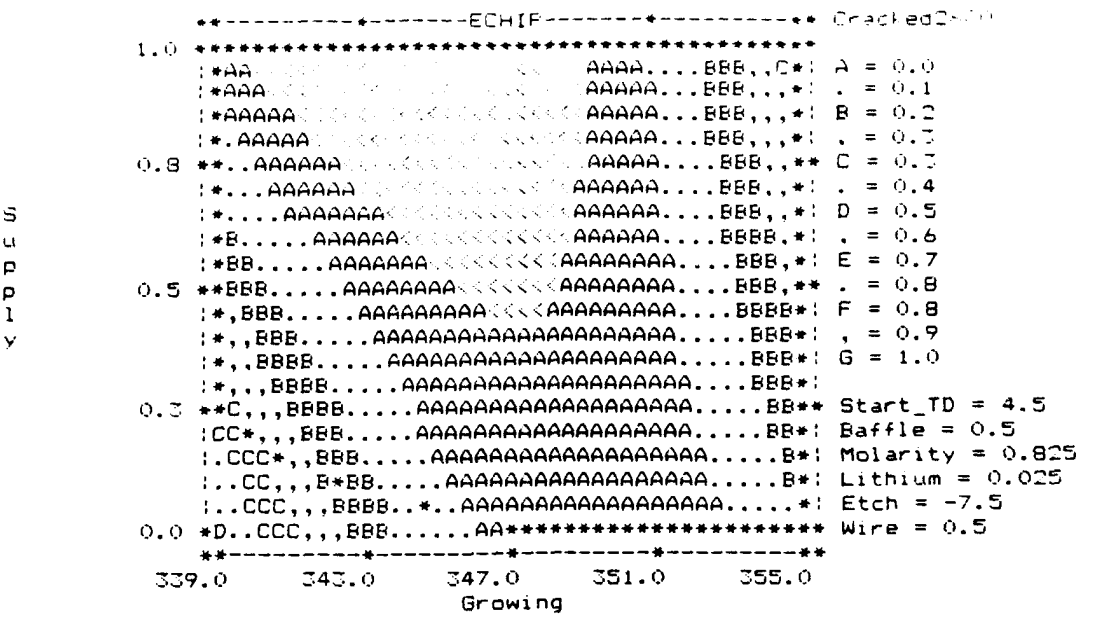
One must also take care in confusing the model with a prediction of actual values. Examining figure 1, one can find a region where the fraction with cracking is less than the lowest experimental value, represented by the "<"s. Since the lowest value is zero, this obviously does not predict an actual value for cracking you would get by growing quartz under those conditions. However, one can use the model to predict this as the region with the lowest response value.

The response surface is a two dimensional representation of a three dimensional surface, analogous to a geographical topological map. We have developed a technique for converting EChip's two dimensional graphs into three dimensions by converting these graphs into spreadsheet data and then using Lotus 1-2-3™ and 3-D Graphics™ to produce three dimensional response surfaces.[7] Figure 2 shows such a representation of the data from figure 1. In this figure, it is easy to see that the highest levels of cracking occur for bars grown at a high growing temperature from broken crystal or low growing temperature from lascas. One can also see the interaction between the variables.

EChip calculates the precision of the response surface based upon the fit of the regression analysis, the reproducibility of the replicate runs and how far outside of the convex hull the surface extrapolated. The width of the contour intervals of the response surface is representative of this precision. For example, as one extrapolates outside of the region of the actual data, the precision decreases and the contour lines broaden. Figure 3 shows a graph of the response surface for inclusions (number/cc, 100 - 150 microns size) as a function of growing

Figure 1

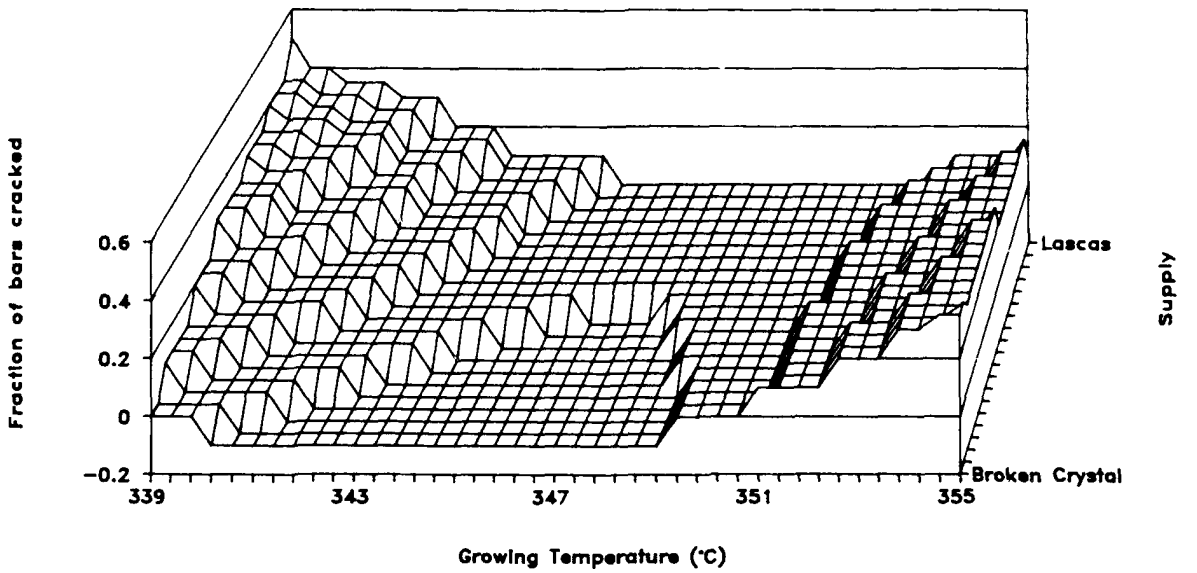
EChip Response Surface for Cracking



EChip generated response surface for tne fraction of bars withcracks as a function of growing temperature and supply type (1 is broken cultured crystal, 0 is natural quartz (lascas)).

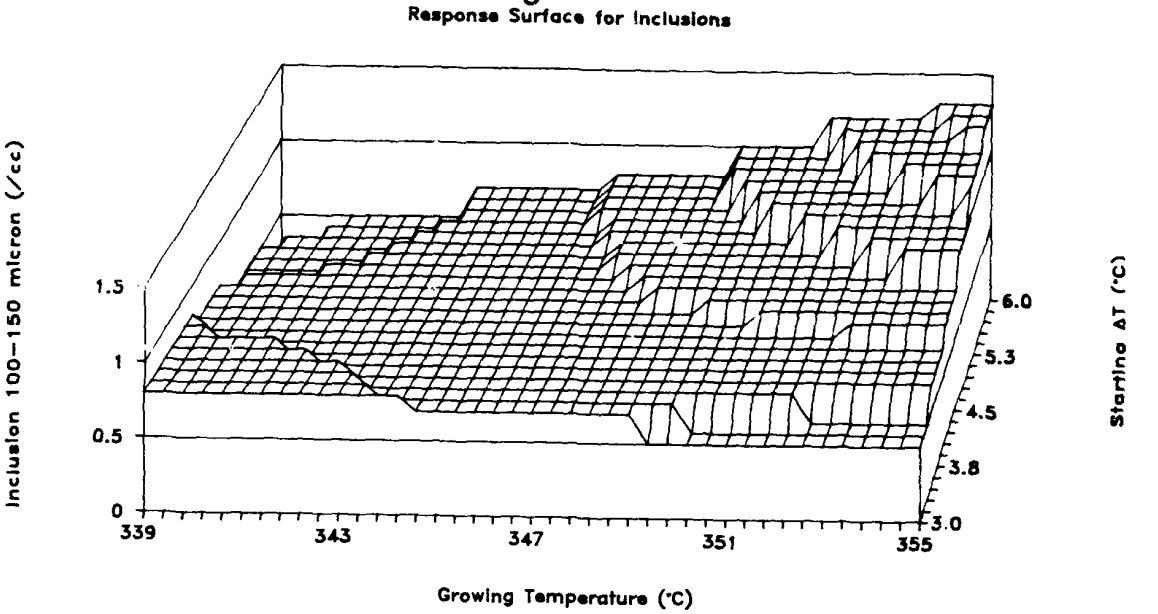
Figure 2

Response Surface for Cracking



Three dimensional representation of the response surface for the fraction of bars with cracks as a function of growing temperature and supply type.

Figure 3



Response surface for inclusions as a function of growing temperature and initial TD.

temperature and initial TD. Since this plot shows data for the middle of the convex hull, the broadness of the contours in the middle region of the plot is probably due to a lack of reproducibility in the data.

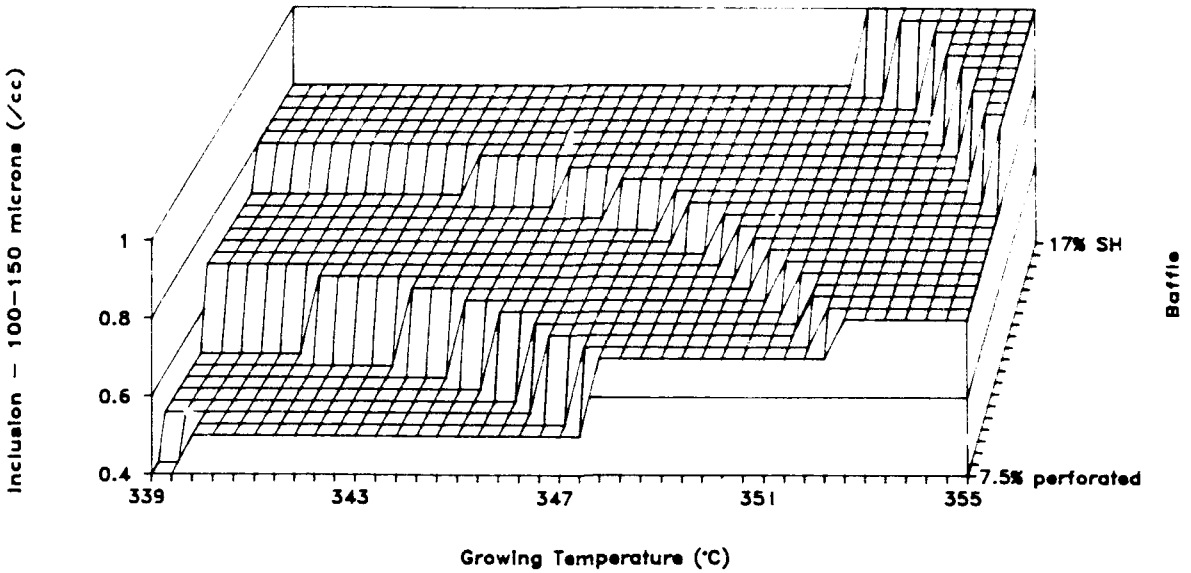
One of EChip's limitations is in its inability to distinguish between continuous and categorical variables; all variables must be treated as continuous variables. One way to handle this is by rethinking how you look at a variable. In figure 1, supply type could be though of as broken crystals vs. lascas and thus as a categorical variable. However, one can also think of it as a continuous variable from 100% lascas, through a mixture of the two, to 100% broken crystal. In fact, this reflects actual production practices. Figure 4 shows the response surface for inclusions (number/cc, 100 - 150 microns size) as a function of growing temperature and baffle type. Here too, baffle type has been thought of not as perforated vs. single hole, but as a continuous variation in the number of holes. The limitation here is this does not reflect actual practice and may not even reflect the actual behavior of the system. One must therefore be careful in interpreting the meaning of baffle values on this contour plot between the two endpoints.

Another limitation of EChip is its inability to examine correlations between response variables; for example, between inclusion densities and cracking. However, this can be overcome by the use of other statistical methods. For example, scatter graphs in conjunction with linear regression analysis can be used to look at interrelations between response variables. Figure 5 shows a scatter diagram for the concentration of aluminum in a sample bar from each run vs. the concentration of lithium, with a linear regression fit of the data. One can also use more traditional statistical tests, such as the student t-test, to supplement EChip. One must remember with all these methods that they generate empirical models, it is up to the researcher to create the theory, if desired.

One last consideration is that EChip is designed to be used as a "blackbox". While this is a great advantage for the statistical novice, one must be careful to use the program within its stated limitations. EChip does include some "whistle-blowing" abilities which alert the user to potential statistical problems which may require expert help.

Figure 4

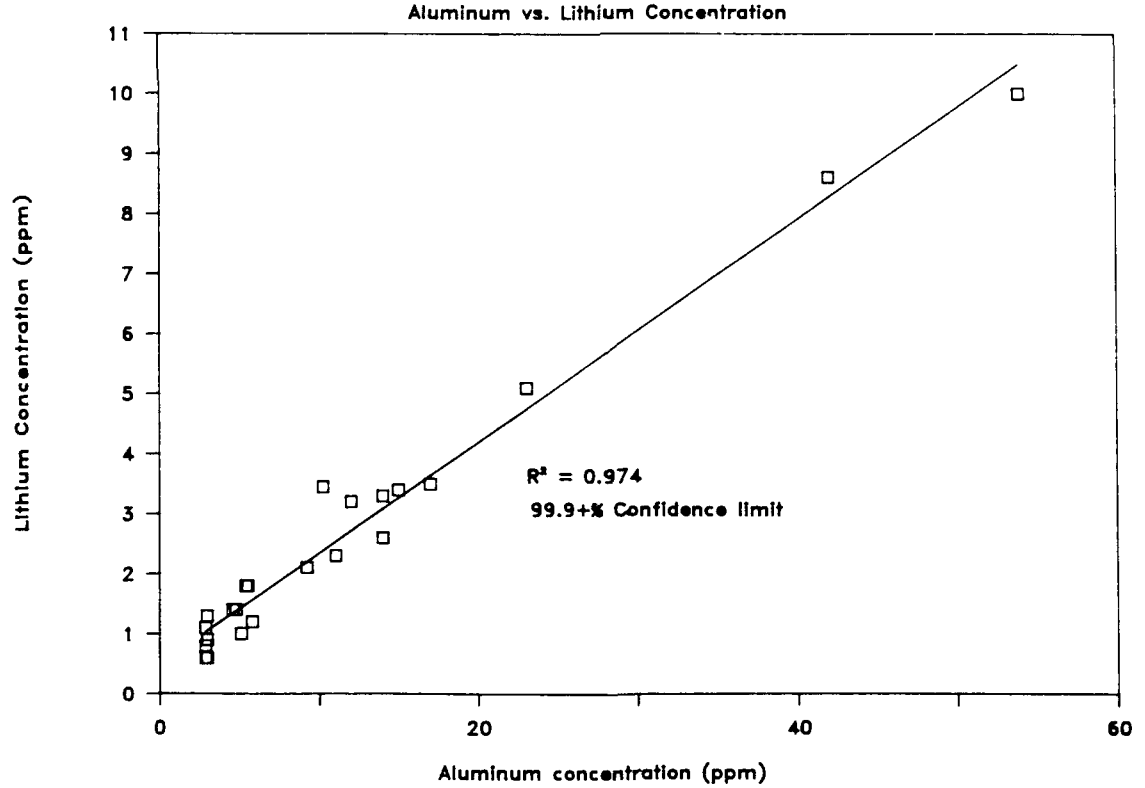
Response Surface for Inclusions



Response surface for inclusions as a function of growing temperature and baffle type.

Figure 5

Aluminum vs. Lithium Concentration



Scatter diagram of the concentration of aluminum versus the concentration of lithium in a bar. A linear regression fit of the data is also shown.

There were some limitations due to the design we chose as well. By looking at a variable at only two levels, instead of three, one is no longer able to look for curvature in that variable. It also means that one cannot look at interactions between two level variables, only between two and three level variables and among third and higher order variables.

These design limitations can be eased to some extent by following EChip's normal experimental sequence.[4,5,6] Table 4 gives some typical examples.

Table 4

*Screening trials: many variables at 2-levels

7 variables / 2 levels - 14 trials + replicates

12 variables / 2 levels - 19 trials + replicates

*Response trials: significant variables at 3-levels

4 variables / 3 levels - 20 trials + replicates

6 variables / 3 levels - 33 trials + replicates

*Detail/confirmation trials:

2 variables / 4 levels - 13 trials + replicates

However, the cost of such a detailed study is in the number of experiments required (44 to 75 in the examples above). For a batch process with a turn-around time of 100 days, such as quartz growth, such a study would be very expensive and very time consuming.

In spite of all these limitations, designed experiments have turned out to be very useful to us. From these 23 experimental runs we have discovered 73 different interactions and relationships among the seven input variables studied and the various responses examined, using EChip and the other statistical methods mentioned. In addition are the negative responses, the relationships and interactions which were believed to exist but have been disproved. We are now in the process of using this round of experiments as screening trials for further experiments to examine the details of particular responses.

References

[1] R. Irvine, J. Foise, E. Leeson, G. Johnson, "Recent Experiments in a Silver Lined Autoclave", Proc. 42nd Annual Freq. Control Symp., pp. 127-137, 1988.

[2] G. Johnson, R. Irvine, J. Foise, "Experimental Determination of the Relationship Among Baffle, Temperature Difference and Power for the Hydrothermal Growth of Quartz", Proc. 43rd Annual Freq. Control Symp., pp. 447-458, 1989.

[3] Expert in a Chip (EChip)™ - Expert in a Chip, Inc., 518 Holly Knoll Rd., Hockessin, DE 19707.

[4] T.A. Donnelly, "Robust Product Design", Machine Design, Oct. 8, 1987.

[5] T. Donnelly, "Desktop software for simultaneous optimization of multiple responses", American Laboratory, Vol. 21 (9), pp. 100-107, Sept. 1989.

[6] B. Wheeler, EChip Course Text. Expert in a Chip, Inc., Delaware, 1987.

[7] 3-D Graphics™ - Intex Solutions, Inc., 161 Highland Ave., Needham, MA 02194.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A STUDY OF THE ELECTRODIFFUSION PROCESS IN QUARTZ

D. W. HART, JENNIFER FRANK, DOWLAN SMITH, and J. J. MARTIN,
Department of Physics
Oklahoma State University
Stillwater, OK 74078-0444

and

J. G. GUALTIERI
U.S. Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

ABSTRACT

Electrodiffusion (sweeping) is a high-temperature process used to replace the interstitial alkalis present in as-grown quartz with hydrogen. Sweeping affects point defects such as the substitutional aluminum with its associated interstitial alkali and the extended dislocation networks with their precipitated impurities that form the etch channels. Quartz is commercially swept to improve the radiation hardness and to reduce the formation of the etch channels. We have been investigating the electric current versus time curves observed during sweeping runs as possible indicators of the completeness of the electrodiffusion process. Hanson reported the appearance of a current peak or plateau near 250°C during the warm-up portion of the sweeping process when the electric field was initially applied at room temperature. He also reported that this peak was not present in samples that had previously been swept. We have tracked both the growth of the Al-OH center and the decay of the Al-Na center with the sample current versus sweeping time. We find that the current peak or plateau disappears when the sample is heated to 400°C and immediately cooled back to room temperature. However, no reduction in the Al-Na centers or production of the Al-OH centers match the disappearance of the peak. Instead, the Al-Na center is slowly replaced by the Al-OH after the sample has been held for some time at temperatures near 500°C. Since hydrogen is much less mobile than the alkalis, this replacement causes a reduction of the sweeping current. It appears that the appearance of a steady current during a time interval might be used as a signature for the completeness of the electrodiffusion process. We find that when the current remains steady to $\pm$2% over an eight to ten hour interval that the sample passes the OSU test for sweeping effectiveness. It does appear, however, that the tendency to form etch channels reduces more slowly than the rate of replacement of the alkalis at point defects.

INTRODUCTION

Electrodiffusion (sweeping) is commercially employed to reduce the formation of etch channels during processing [1] and to improve radiation hardness [2]. The process removes neither the point defects nor the dislocation loops. Instead the improved performance of swept quartz is caused by the modification of the defects. For example, substitutional aluminum with an associated interstitial alkali is a common point defect. Sweeping in air (with water vapor present) or in hydrogen removes the alkali from the quartz and replaces them with a hydrogen. A similar replacement process must also take place with the impurities trapped in the dislocation networks so that they are much less chemically active after sweeping. Verhoogen [3] carried out the first deliberate alkali sweeping experiments during his study of ionic diffusion in quartz. King [4] was the first to apply the sweeping process to quartz later used for resonator studies. The process for sweeping specific alkalis into quartz was developed by Fraser [5] who clearly demonstrated that the 53-K acoustic loss peak was due to the Al-Na center. Kats [6] studied the effects of alkali and hydrogen sweeping on the infrared spectra of both natural and cultured quartz. Recently, Martin [7] reviewed current electrodiffusion results.

At the present time, several methods are used to determine if the sweeping process is complete for point defects. Of these, the electron-spin resonance, ESR, test developed by Halliburton, *et al*. [8] and the measurement of the high temperature series resistance of a resonator developed by Koehler [9] are probably the most reliable indicators. The ESR test requires the removal of a test sample from the swept bar and an elaborate radiation procedure. The resistance test requires the fabrication of a resonator blank. No test other than actually etching samples is available for etch channel reduction by sweeping. Hanson [10] reported the appearance of a current peak or plateau near 250°C during the warm-up portion of the sweeping process when the electric field was initially applied at room temperature. He found that this peak was absent in samples that previously had been swept and suggested that it could be used as a sweeping test. Tests for sweeping effectiveness that do not require elaborate sample preparation or that can be performed during the electrodiffusion process would clearly be useful. We report here an investigation of the effects of sweeping temperature and sweeping time on the replacement of the interstitial alkalis by hydrogen and on the reduction of etch channel formation. Our results suggest that the appearance of a constant current during the sweeping run is a simple reliable indicator of sweeping completeness

US GOVERNMENT WORK IS NOT PROTECTED BY
US COPYRIGHT

for point defects. The study also shows that infrared measurements of the Al-OH absorption can also be used as a indicator.

EXPERIMENTAL PROCEDURE

Cultured quartz was used exclusively for this investigation. Except for the sample designated HA-A all samples were pure Z-growth material. Both full sized lumbered bars and smaller infrared or 5 Mz 5th overtone AT-cut blanks were used for the sweeping tests. The HA-A samples were resonator blanks taken from an R-growth stone selected for its 60-70 ppm aluminum content. The full bars were air-swept at LABCOM using the system described by Gualtieri, et al. [11] The smaller samples were H-swept at OSU using the system described by Martin. [7] Both systems were controlled by HP Series 80 microcomputers. Au/Cr or Au electrodes were vapor deposited on cleaned polished surfaces. The samples were then placed between steel or stainless steel contacts in the sweeping furnace. Electric fields between 800 V/cm and 2000 V/cm were applied at room temperature. The computer recorded the sample current and temperature during the sweeping run. Temperature-ramped sweeping was performed [11], i.e., the furnace was slowly brought up to the sweeping temperature, held there for the selected time, and then ramped back to room temperature.

A Perkin-Elmer model 1/60 Fourier transform infrared spectrophotometer was used to monitor the room-temperature OH-vibration spectrum before and after air-sweeping the full bars. Low temperature IR absorption scans were made on the smaller samples using a Beckman Model 4240 dispersive spectrophotometer. These samples were held in a liquid nitrogen cryostat that had CaF_2 windows. The reduction of the Al-Na centers in the resonator blanks caused by sweeping was determined by measuring the acoustic loss over the 8 K to 100 K temperature range. A set of adjacent AT-cut sections from one bar were swept for different times. After etching in a hot saturated ammonium bifluoride solution etch channel counts were made using an Olympus Nomarski microscope.

RESULTS AND DISCUSSION

Figure 1 shows the current density, j, and temperature versus time curves for an initial air-sweep at 300°C. The sample was a Y-bar designated A-GC-3539 with Au/Cr electrodes. An electric field of 2000 V/cm was used. The peak labeled A is the initial peak first observed by Hanson [10] when the sample is warmed with the field on. The second peak, marked B, represents the replacement of alkalis by hydrogen. The current density is small for this 300°C run and continues to decrease throughout the entire run. The current density reduction over the eight hour period before cool-down was 15 %. The current density . Curve A in Fig. 2 shows the room temperature infrared spectrum of this sample prior to the air-sweeping runs. The sharp band at 3585 cm^{-1} and the broader band near 3410 cm^{-1} are the OH-related growth-defect centers. Curve B shows the spectrum obtained after the initial 300°C air-sweep. No significant Al-OH band at 3380 cm^{-1} was produced by this low temperature run. The run caused a 50% reduction in the 3585 cm^{-1} growth-defect band. The non-porous Au/Cr electrodes inhibit the introduction hydrogen into the sample during sweeping. Gualtieri and co-workers [12,13] have extensively studied this effect.

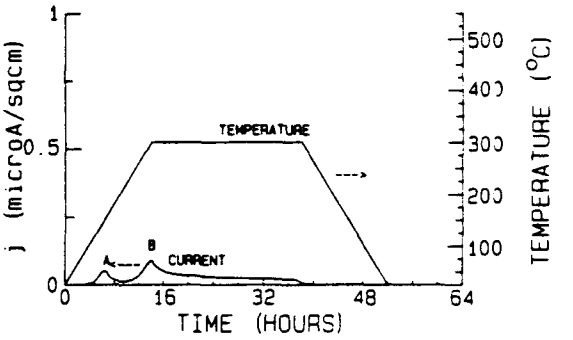


Figure 1. The current and temperature versus time curves for the initial 300°C air sweep on sample A-GC-3539 are shown.

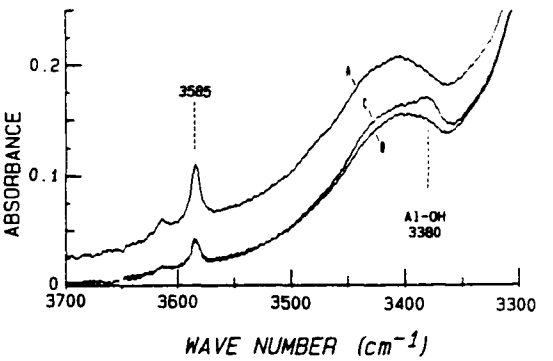


Figure 2. The room temperature IR spectra for sample A-GC-3539 are shown for the as-received condition, A; after the initial 300°C sweep, B; and after a subsequent 400°C sweep, C. Only the second sweep produced the Al-OH band.

Figure 3 shows the current density and temperature versus time curves for the subsequent 400°C air-sweeping run on sample A-GC-3539. The initial current peak, A, is nearly gone. The much larger peak B and the overall larger current indicate that hydrogen exchange is more efficient at this higher temperature. We note that j is still decreasing with time at the end of the 26 hr run. THe current density reduction, over the eight hour period before cooldown, was 4%. The 3380 cm^{-1} band in curve C of Fig. 2 shows that the Al-OH centers are being produced by the air-sweeping. Interestingly, the growth-defect band does not seem to recover.

The initial 300°C air-sweep on sample A-GC-3539 did not introduce any Al-OH centers. As shown in the subsequent 400°C run the current peak was removed by the first sweep. The second run did introduce some Al-OH centers and the sample can be

considered as partially swept. The current was still decreasing with time at the end of both runs. A third sweep at 500° C produced a current density reduction, during the last eight hours before cooldown, of 2%. The IR spectrum after this sweep retraces curve C in Fig. 2.

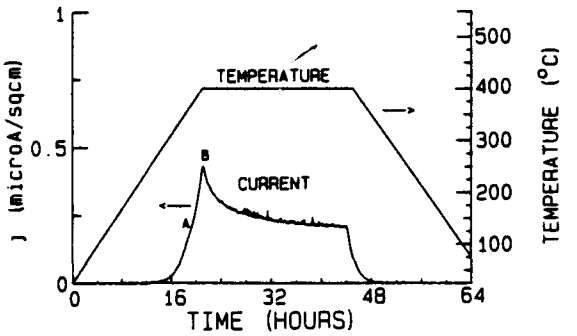


Figure 3. The current density, j, and temperature are shown as functions of time for the subsequent 400°C run on sample A-GC-3539. Current peak A is missing.

The current peak or plateau observed on warm-up was investigated by carrying out a series of H-sweeping runs in which a test sample as ramped up to a selected temperature and then immediately cooled back to room temperature. The sample was then cycled repeatedly to successively higher temperatures. The curves on the far left side of Fig. 4 shows the current and temperature data collected during a cycle to 248°C on sample R4 from bar MC1-7. The curves offset to 5 hr are for a subsequent cycle to 320°C. The curves starting at 10 hr are for a full 12 hr 490°C sweep made on the control sample R1. The initial current peak is clearly evident in the data for the control sample. The dashed lines in Fig. 5 show the current density, j, versus temperature data collected during the warm-up portion of each of the T-cycles. These curves map out the current peak for the control sample as shown by the solid line in Fig. 5. The dashed curves also show that the current peak is removed by sweeping at low temperatures for very short times.

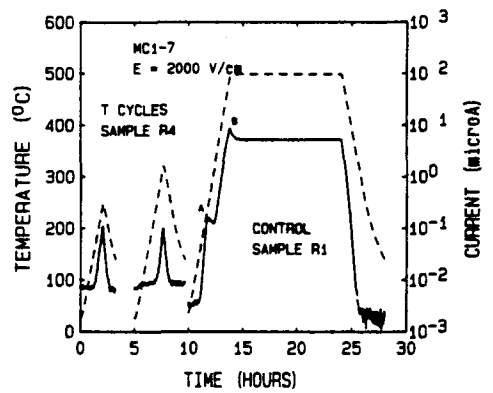


Figure 4. The temperature and current curves for several sweeping cycles on sample R4 and for the full sweep on the control sample R1 are shown.

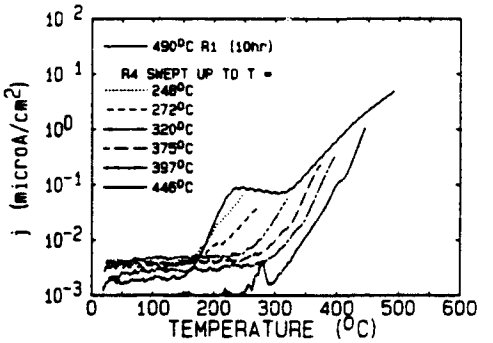


Figure. 5. The dashed curves show the current density, j, versus temperature curves for the warm-up portion of the sweeping cycles. They "map out" the current peak observed for the control sample.

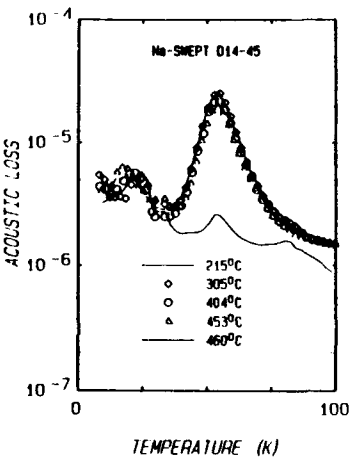


Figure 6. The 53 K Al-Na acoustic loss peak is shown after the resonator blank was swept while being cycled to the indicated temperature. The Al-Na does not go out until the sample is cycled to temperatures well above the current peak.

A 5 MHz 5th overtone AT-cut blank from D14-45 Premium Q bar was Na-swept using our standard process. [14] It was run through a series of sweeping cycles similar to those described above. The acoustic loss versus temperature curves measured after a number of the cycles are shown in Fig. 6. No decrease in the Al-Na loss peak was observed until the sample was cycled above 450°C. The initial height of the 53 K loss peak [15] agrees with the approximately 1 ppm aluminum concentration observed in D14-45 quartz. As shown below, the Al-OH is not effectively produced until the sample has spent time at temperatures well above that at which the initial current peak is removed. The conversion of the Al-Na and presumably Al-Li centers into Al-OH does not begin until the sample is cycled through temperatures well above

the point where the initial current peak is removed. Therefore, the current peak is not associated with the aluminum center and does not reliably indicate that the sample is fully swept.

The H-sweeping of several Z-plate samples at a fixed temperature of 490°C was investigated. After each sweeping run the low temperature infrared spectrum was taken on each sample. Figure 1 shows the spectra taken on sample SQB1 in the as-received condition, after it was swept for approximately ten minutes, and after a cumulative sweeping of 41 hours. The two bands at 3300 and 3200 cm^{-1} are intrinsic Si-O overtone vibrations [6]. Only the OH-related growth-defect bands are present in the as-received condition. Taking the spectrum at 80 K sharpens the 3585 cm^{-1} band shown in Fig. 2 and shifts it to 3581 cm^{-1}; the broad 3410 cm^{-1} shown in Fig 2 is split into sharper bands at 3431 and 3398 cm^{-1}. After H-sweeping the sample for ten minutes strong Al-OH bands at 3361 and 3306 cm^{-1} are seen in the low temperature spectrum. The sample was successively swept and the Al-OH concentration measured by low temperature IR absorption after each run. The upper curve in Fig. 1 shows the spectrum obtained after a cumulative sweeping time of 41 hours. Figure 8 shows the growth of the Al-OH concentration versus cumulative H-sweeping time for samples SQB1 with 1.5 ppm Al and for sample HA-A with 60 ppm Al. We find that when the Al-OH concentration as measured by infrared absorption techniques remains constant with cumulative sweeping time that the sample is fully swept according to the OSU ESR test for sweeping effectiveness. Because of its lower sensitivity the infrared test is not as reliable for low aluminum samples as the ESR test.

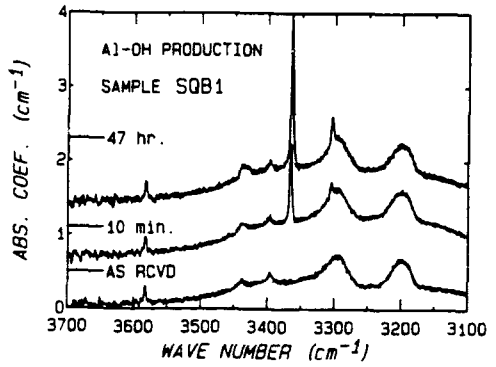


Figure 1. The low temperature infrared absorption spectra for sample SQB1 taken in the as-received condition and after cumulative H-sweeping times of ten minutes and 41 hours are shown.

The upper portion of Fig. 9 shows the fractional current change, $\Delta I/I_{end}$, and the lower portions shows temperature fluctuations, ΔT, measured during a H-sweeping run on sample SQB1. I_{end} was the average current at the end of the 30 hour run. The 1.6 C° peak-peak temperature oscillations were due to a slightly mistuned controller. These temperature oscillations cause a 6% peak-peak oscillation in the sweeping current. After averaging out the short term temperature

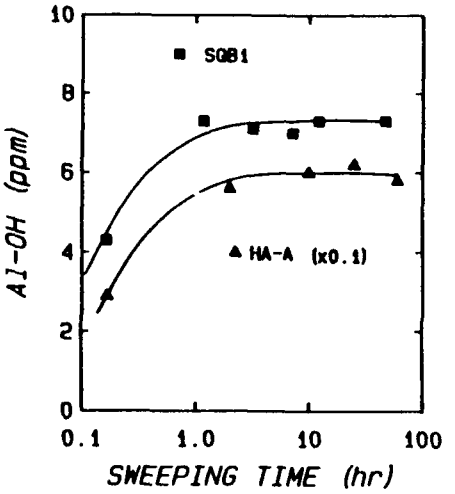


Figure 8. The growth of the Al-OH center with cumulative sweeping time is shown for samples SQB1 and HA-A.

oscillations, $\Delta I/I_{end}$ is seen to be constant to better than ±2% over the last 20 hours of the run. The oscillations were removed on later runs by carefully retuning the temperature controller. As reported above, the OSU ESR test showed that this sample was fully swept. We consistently find that when the sweeping current is constant to within ±2% over the last 8-10 hours of the run that the samples are fully swept. Therefore, it appears that the observation of a constant current (assuming the temperature is steady) is a good indicator that the electrodiffusion is complete for the point defects.

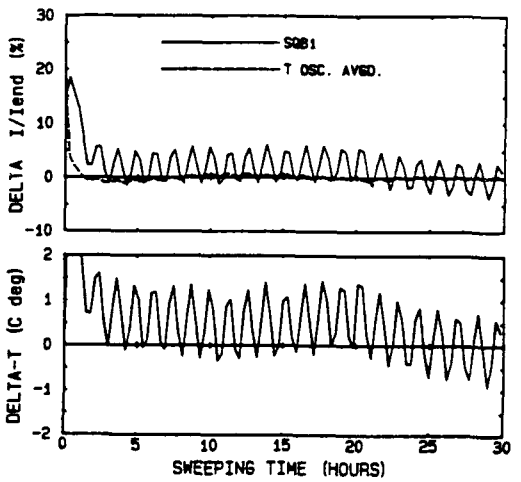


Figure 9. The fractional current change and the temperature deviation observed during a 30 hour H-sweeping run on sample SQB1 are shown.

Electrodiffusion also affects the dislocation networks responsible for the formation of the etch channels. A set of adjacent AT-cut plates were taken from our Premium Q bar PQ-C. Au electrodes were vapor deposited on the polished surfaces and the plates were H-swept at 490°C for different times. Figure 10 shows the etch channel density measured on the unswept control sample and on samples swept for times ranging from 10 minutes to 20 hours. The current was steady to ±2% for the last 10 hours of the 20 hour run. The etch channel density appears to decrease more slowly than the Al-OH production reported above. Previous reports [11,16,17] showed that air-sweeping of full bars reduced the etch channel density to less than 10 channels/cm^2. Since the channels tend to lie along the Z-direction [18] sweeping AT-cut plates is probably less efficient than sweeping a full bar where the full electric field is along Z.

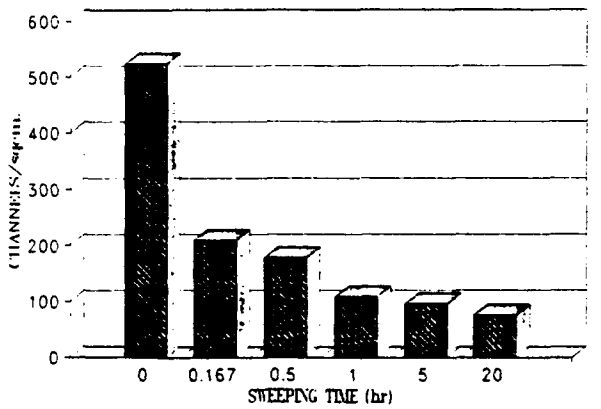


Figure. 10. The reduction of the etch channel density versus H-sweeping time is shown for a series of AT-cut samples from bar PQ-C.

SUMMARY AND CONCLUSIONS

The initial current peak observed during the warm-up portion of a temperature ramped sweeping run disappears when the sample is swept at 300°C; however, the sample cannot be considered as swept since no Al-OH was produced. Al-OH was produced by a subsequent temperature ramped sweep at 400°C. Subsequent tests showed that the low temperature current density peak disappears upon heating to near 300°C with immediate cooling back to room temperature. No Al-Na conversion to Al-OH was observed during such cycles until the sample was taken above 450°C. Therefore, the disappearance of the initial low temperature peak cannot be used as a reliable indicator of sweeping.

The Al-OH conversion was measured as a function of sweeping time for runs at 490°C. We found that the Al-OH concentration was stable after the samples had been swept for a few hours. For these samples the fractional change in the sweeping current was less than ±2% and the OSU ESR test showed them to be fully swept. Therefore, the observation of a stable current can be used as an "on-line" indicator that the electrodiffusion process for point defects is complete. The etch channel density continued to decrease for longer sweeping times than were needed to produce a steady Al-OH concentration. Thus, the modification of the dislocation networks proceeds more slowly than that of point defects.

ACKNOWLEDGEMETS

The authors thank L. E. Halliburton and Ben Zapata for performing the ESR tests of sweeping effectiveness. An early portion of the work at OSU was sponsored by the Solid State Sciences Division, Rome Air Development Command, USAF. Jennifer Frank was a particpant in the NSF Research Experience for Undergraduates Program. Dowlan Smith was sponsored by the OSU Center for Energy Research.

REFERENCES

[1] J. R. Vig, J. W. LeBus, and R. L. Filler, "Chemically polished quartz," in Proc. 31st Ann. Symposium on Freq. Control, NTIS, May 1977, pp. 113-143.

[2] J. C. King and D. R. Koehler, "Radiation effects on resonators," in Precision Frequency Control, E.A. Gerber and A. Ballato, eds. Orlando, FL: Academic Press, 1985, vol. 1, pp147-159.

[3] J. Verhoogen, "Ionic diffusion and the properties of quartz," Am. Mineral., vol. 37, pp. 637-655, 1952.

[4] J. C. King, "Electrolysis of synthetic quartz; Effect upon resonator performance," in Proc. IEE(London), vol. B109, Suppl. 22, pp295-301, 1962.

[5] D. B. Fraser, "Anelastic effects of alkali ions in crystalline quartz," J. Appl. Phys. , vol 35, pp. 2913-2918, 1964.

[6] A. Kats, "Hydrogen in alpha-quartz," Phillips Res. Repts., vol. 17, pp. 132-279, 1962.

[7] J. J. Martin, "Electrodiffusion (sweeping) of ions in quartz-a review," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-35, pp. 288-296, 1988.

[8] L. E. Halliburton, N. Kuomvakalis, 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.

[9] D. R. Koehler, "Radiation induced conductivity and high temperture Q changes in quartz resonators," in Proc. 35th Ann. Symposium on Frequency Control., EIA, pp.322-328, 1981.

[10] W. P. Hanson, "Probable ion signature in quartz electrodiffusion data," Proc. 38th Ann. Symposium on Frequency Control,. IEEE, pp. 38-41, May 1984.

[11] J. Gualtieri, "The influence of temperture and electric field effects the etch channel density in swept-cultured quartz," Proc. 39th Ann. Symposium on Freq. Control, IEEE, pp. 247-254, 1985.

[12] J. Gualtieri, L. Calderon, and R. Lareau, "Update on possible electrode mechanisms in the sweeping of alpha quartz," IEEE, pp. 509-516, 1989.

[13] J. Gualtieri, "Nonuniformities in the sweeping of quartz,", Proc. Ann. Symposium on Freq. Control, IEEE, in press.

[14] A. R. Lopez, J. D. West, and J. J. Martin, "Electrodiffusion of ions in alpha-quartz," Mat. Res. Soc. Symp. Proc., vol. 60, pp. 451-457, 1986.
[15] J. J. Martin, A. Lopez, A. F. Armington, and J. Balascio, "Evaluation of resonators fabricated from high quality quartz," Proc. 42nd Ann. Symposium on Freq. Control, IEEE, pp162-168, 1988.
[16] J. F. Balascio and A. F. Armington, "Developmental results for the production of high quality quartz," Proc. 40th Ann. Symposium on Freq. Control, IEEE, pp. 70-75, 1986.
[17] J. R. Hunt and R. C. Smythe, "Chemically milled VHF and UHF AT-cut resonators," Proc. 39th Ann. Symposium on Frequency Control, IEEE, pp. 292-300, 1985.
[18] G. Johnson and R. Irvine, "Etch channels in single crystal cultured quartz," IEEE, Proc. 41st Ann. Symposium on Freq. Control, pp. 175-182, 1987.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NONUNIFORMITIES IN THE AIR-SWEEPING OF QUARTZ

J. Gualtieri, J. Kosinski, and R. A. Murray

US Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

ABSTRACT

Electrode metal-quartz contact, metal porosity, and metal-quartz interfacial diffusion of hydrogen-species are required for efficient and uniform hydrogen introduction during air-sweeping [1]. Nonuniform H-introductioncan be indicated by the co-introduction of color centers. Color-center introduction during air sweeping leads to nonuniform sweeping along the Z-axis [2] similar to the nonuniformity found with vacuum sweeping [3].

Using infrared absorption and scanning electron microscopic techniques, we have found that color-center introduction into quartz is influenced by the porosity of evaporated metal electrodes [1]. It appears that thermal-stress relief of electrode-metal films, during sweeping, normally causes void formation (porosity). If the porosity is extensive, H-introduction predominates and color-center introduction mechanisms are suppressed.

Samples swept with evaporated Au/Cr electrodes relying on thermal stress-induced porosity sometimes displayed sweeping nonuniformities. However, nonuniformities were not found when samples were swept using evaporated Au/Cr electrodes containing periodic stripe openings. Electrodes with stripe openings were fabricated using photolithographic techniques. The stripe openings provided sufficient electrode-quartz-water vapor (3-phase) regions to ensure uniform sweeping.

Uniform sweeping was also obtained using magnetron-sputtered amorphous Y-Ba-Cu-O films. These electrically conducting oxide films apparently allow diffusion of hydrogen without the need of porosity or electrode openings to assist in the H-indiffusion sweeping process.

The high-temperature annealing phase of the sweeping process will sometimes produce a light-scattering haze in the sample. The haze occurs mainly in unswept regions of the quartz. Thus, the occurrence of haze can be another demonstration of nonuniform sweeping. Haze was found not to depend on electrode effects, absorption in the near IR (due to molecular water), or to absorption in the far IR (due to precipitation of silicates as suggested by Cohen [4]). Instead, haze was found to occur in regions of the quartz showing strong OH-absorption [5], and appears to be associated with lithium [6].

INTRODUCTION

It is well known that the sweeping of quartz not only results in an improvement in mechanical Q, but also improves the resonator stability in a radiation environment [7] and reduces the etch-channel density of etch-processed quartz. Previous investigators have indicated that vacuum sweeping can be nonuniform and the use of resonators processed in this manner may lead to unpredictable results [3].

Nonuniform air-sweeping of quartz is associated with the nonporosity of evaporated metal electrodes [1]. That is, openings in the metallization are required when hydrogen diffusion through the metal electrode is too slow. The nonuniformity along the Z-axis (see Fig. 1 and Figs. 9 and 10 of [2]) is associated with the development of color centers when Au/Cr, V, and Al-electrodes are used. These materials form metal-silicide compounds that inhibit the metal/quartz interfacial (lateral) diffusion of H-species [1] prior to the diffusion of hydrogen into the quartz. During sweeping, aluminum compensated by electron-holes (Al-h+), develops much more slowly than aluminum compensated by hydrogen

US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT

(Al-OH) [8]. This contributes to nonuniform sweeping along the Z-axis [3].

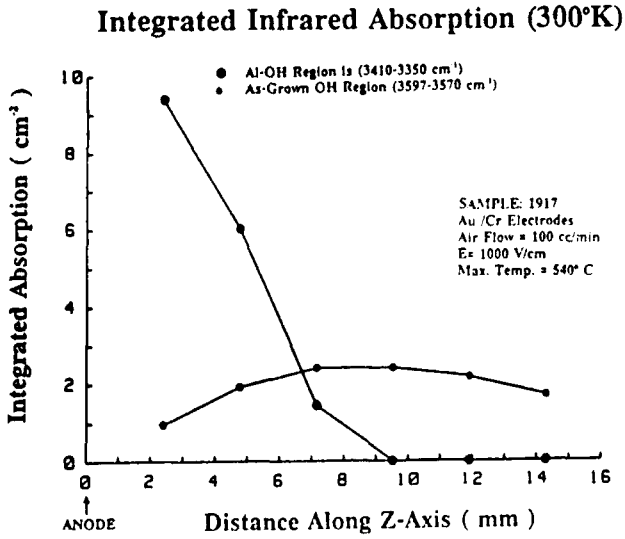


Fig. 1. The integrated absorption of a series of infrared scans. The IR beam was unpolarized and in the X-direction. The Al-OH band was nonuniform along the Z-axis. The as-grown OH band is depleted in the region near the anode.

Light-scattering haze regions, unrelated to electrode effects, sometimes are observed in swept samples. The haze occurs in regions of the quartz not containing Al-OH or Al-h+ (color centers). Thus, haze can be another demonstration of nonuniformly swept quartz. Attempts to relate the occurrence of haze with absorption in the near infrared, associated with molecular water content in quartz or absorption in the far infrared, associated with the existence of metal-silicate compounds, as suggested by Cohen [4], have not been successful. Instead, very large differences have been found in the OH-vibrational spectral region (3700 cm-1 to 3100 cm-1) [5] and the association of lithium seems implicated.

The goal of the research reported in this paper was to increase our understanding of: 1) the mechanism for electrode void formation before and during sweeping that leads to nonuniform H-sweeping; 2) the conditions that restrict H-introduction and allow nonuniformities to develop; and 3) the techniques and electrode types available to circumvent the restriction of uniform hydrogen indiffusion.

EXPERIMENTAL METHODS

Samples of cultured quartz were used in this investigation. Swept samples were lumbered into Y-bars after seed removal [9]. The Z-surfaces of these bars were lapped, polished, and cleaned as outlined previously [1,10]. For electrodes, evaporated Au/Cr films were applied to the polished and cleaned Z-surfaces of the lumbered bars. In some cases, open regions (stripes) were photolithographically etched into the metal film (anode region). Temperature-ramped sweeping (maximum temperature 537° C) was carried out in an atmosphere of flowing laboratory air (flow rate = 100 cc/min) [11]. An electric field of 1000 V/cm was used in all experiments.

A Perkin-Elmer model 1760 Fourier Transform Infrared Spectrophotometer (FTIR), having a wavenumber scan range of 4000 to 450 cm^{-1}, was used to obtain room-temperature OH-vibrational spectra, difference spectra, and deconvoluted spectra of haze regions of the quartz. Optical microscopic techniques were used to study voided electrode regions and to image interior features of swept bars. Scanning electron microscopic (SEM) techniques were used to study the microstructure of metallized surfaces.

Color center development was observed visually when lower quality quartz was swept [12]. The coloration can be detected, using magnification, by the apparent light-brown coloration of inclusions, when high quality quartz is swept. Alternatively, the detection of weak color-center absorption may be accomplished using visible lasers as sources [13].

Films of Y-Ba-Cu-O were grown by magnetron sputtering from a single stoichiometric $YBa_2Cu_3O_7$ target. The films were grown at room temperature in flowing Argon (flow rate = 35 cc/min) on the Z-surfaces of the quartz Y-bar substrates [14]. The films were annealed, in flowing-air atmosphere, as part of the temperature-ramped sweeping process (maximum temperature = 500° C). Rutherford backscattering spectroscopy (RBS) [2] was used to evaluate the possible diffusion of the electrode constituents into the quartz.

Secondary ion mass spectrometry (SIMS) was used to verify the presence of H-species at the electrode/quartz interface [1].

EXPERIMENTAL OBSERVATIONS

Air-sweeping nonuniformities have been found to be related to electrode effects [1,9]. Microscopic examination of Au/Cr anode surfaces after sweeping reveals the formation of numerous voided areas. In some instances, it is clear that the voids in the electrodes are surrounded by thicker metal regions (see Fig. 2). When voiding occurs, the color centers are bleached

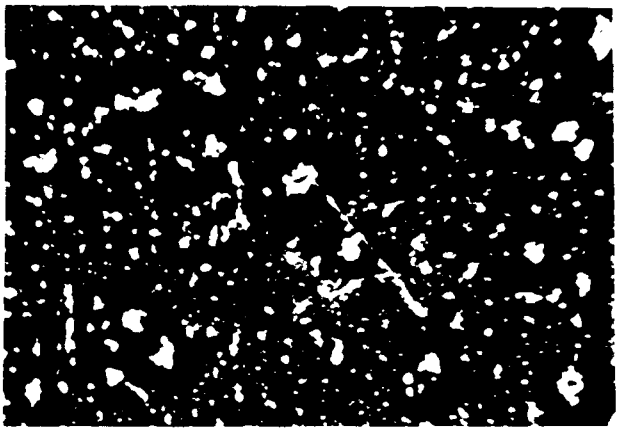
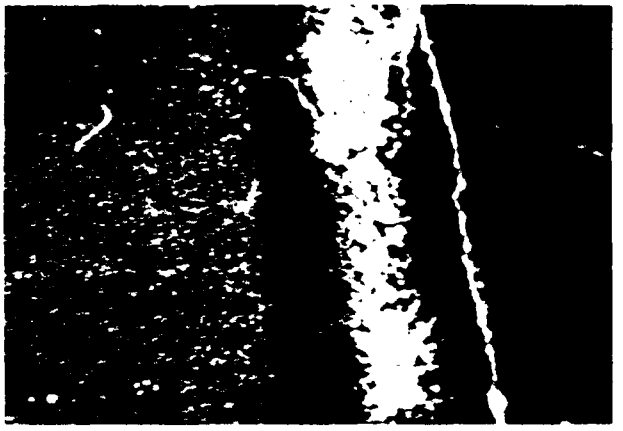


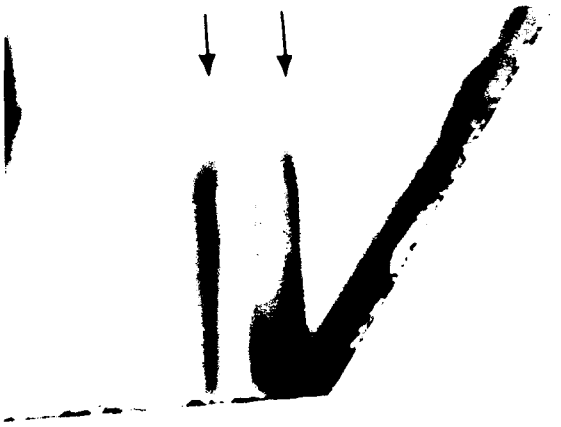
Fig. 2. Photomicrograph of a Au/Cr anode surface after sweeping. Voided areas (white) are surrounded by thicker metal (dark) regions. This porosity develops during sweeping. 200 X.

by the indiffusion of hydrogen. If voiding is extensive, formation of Al-OH predominates and sweeping is uniform. If the voiding and interfacial diffusion of hydrogen is inhibited, e.g. by the formation of Cr-silicide compounds [1], then color centers develop and the sweeping is nonuniform.

When Au/Cr electrodes were partially immersion overplated with gold, unbleached color-center regions (see Fig. 3) could be related to the particle size of the gold anode above the unbleached region. In the unbleached regions, the particle size was small. In the adjacent (bleached) regions the particle size was about 10 times larger (see Fig. 4).



(a)

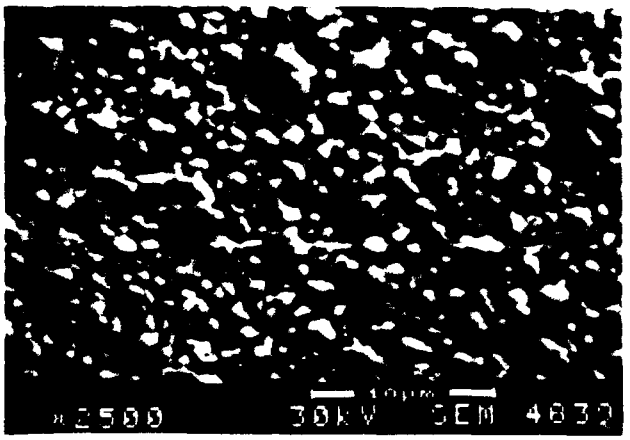


(b)

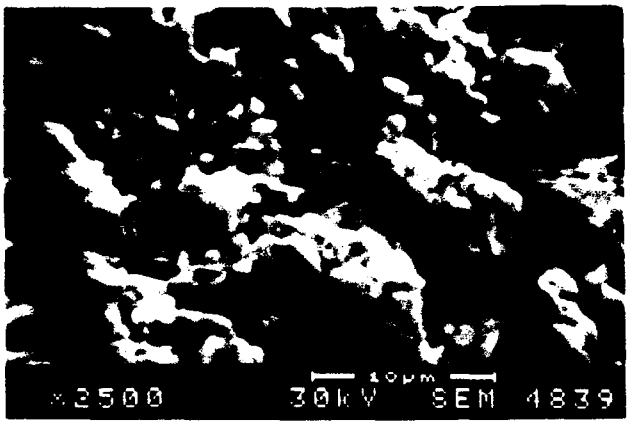
Fig. 3. Photomicrographs of (a) a Au/Cr anode surface after sweeping. This Z-axis view is through the quartz from the cathode side. All dark areas to the left of the crystal edge are color centers. This view shows two vertical (unbleached) color-center areas, the one on the right coincides with a 35^{o} cut that terminates the bar. 15 X. (b) This X-axis view shows the two vertical (unbleached) color-center areas, indicated by arrows, and the 35^{o} termination. (Anode area, Z-axis is vertical.) 7.5X.

Voided areas were only found in the large particle size region. The smaller particle size regions were unintentionally created during the overplating procedure. Overplating was used to increase the metallization thickness to allow parallel-gap welding of gold-plated molybdenum lead-in ribbon to the metallization. The smaller particle size regions coincided with the meniscus

region of the partially immersed cathode (for the purpose of overplating).



(a)

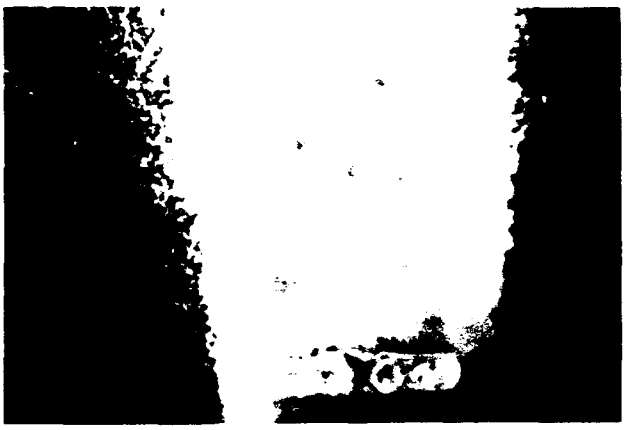


(b)

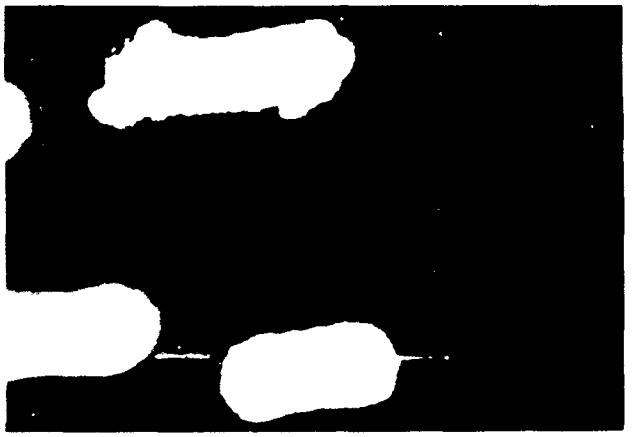
Fig. 4. SEM photographs of the Au/Cr electrode (a) above one of the unbleached color-center regions described in Fig. 3. The metal particle size is about 1 μm. 2500 X. (b) an adjacent electrode region which has developed porosity that has lead to H-indiffusion and bleaching of color centers. The metal particle size is about 10 μm. 2500 X.

Welded-ribbon contact areas of metal electrodes have been observed to void extensively, leading to increased H-indiffusion. Regions beneath the welded ribbon on Au/Cr metallizations showed nearly complete bleaching of color centers (see Fig. 5(a)). The welded-ribbon contact areas were found to contain a very high density of tiny openings (see Fig. 5(b)).

Using Au/Cr striped electrodes (see Fig. 6) and lower quality quartz, it was found that hydrogen compensation of Al develops uniformly along the Z-axis (see Fig. 7). Color centers have been found to form beneath portions of the electrode not striped. Edge voiding (see Fig. 8) was found to occur on thinner metal stripes (25 μm width).



(a)



(b)

Fig. 5. Photomicrograph of a Au/Cr anode surface after sweeping (Z-axis views). (a) This view (through the quartz from the cathode side), reveals color centers (dark areas). The light wide vertical band marks the area where a gold-plated molybdenum ribbon was welded to the electrode. This area shows complete bleaching of color centers. When the ribbon was peeled back, metal under the weld spots adhered to the ribbon, exposing several open-quartz areas. 16.5 X. (b) The same view as in (a) using backlighting to reveal porosity. The contact area contains a high density of tiny openings. The noncontact area to the right has fewer and larger openings. 37.5 X.

Uniform sweeping, without any appearance of color-center formation, also was obtained using electrically conducting amorphous Y-Ba-Cu-O film electrodes (see Fig. 9).

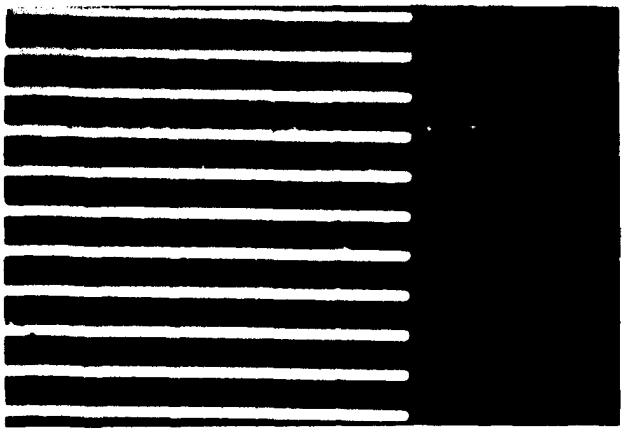


Fig. 6. Photomicrograph of a striped Au/Cr anode surface after sweeping. The metal stripe width (dark) is 130 μm, the open-quartz stripe width (light) is 65 μm. Color centers were found to develop beneath the unstriped metal-electrode area on the right. 26.5 X.

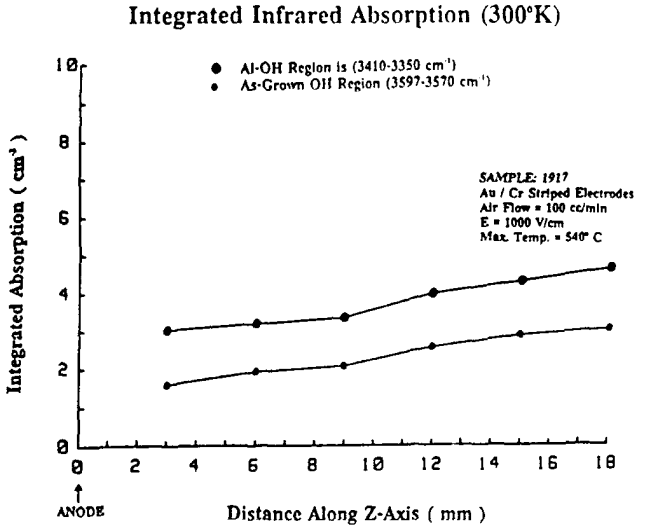


Fig. 7. The integrated absorption of a series of infrared scans (similar to Fig. 1) using Au/Cr striped electrodes. Both as-grown OH and Al-OH bands are fairly uniform, increasing toward the cathode (seed) side of the bar.

DISCUSSION (NONUNIFORM SWEEPING)

Quartz sweeping is affected by gas adsorption [1,2] by the voided metal film which, in turn, is highly dependent on changes in the structure of the film during annealing [15]. These changes occur when sweeping is performed in controlled atmospheres, such as forming gas, as well as in air. The development of voids using air-sweeping and Au/Cr electrodes is probably enhanced by the oxidation behavior of Cr. Chrome oxide nodules and ridges develop at 400° C to relieve stress, providing a high density of mass transport paths [16]. This stress-relief transition could be related to the onset of void formation and the noticeable increase in H-introduction above 400° C [1,17].

Openings in metal-film electrodes can exist after evaporation of the metal, resulting from surface contamination and probably stress voiding. The existence of internal stresses in films deposited on a substrate is well known. Thermal stresses result from a difference in the thermal expansion coefficients of the film and the substrate. Intrinsic stresses also exist, arising from impurities in the film and structural modifications occurring during film growth, in amorphous to crystalline transitions, and further growth of crystal films during high-temperature annealing processes.

Fig. 8. Photomicrograph of a Au/Cr striped anode, backlighted to reveal voids. The dark stripes are metal, the white stripes are quartz. The thin (25 μm) metal stripes are only voided at their edges. 200 X.

Stresses in the Au film of the Au/Cr/quartz config-

uration are tensile before annealing and the stresses can be as large as 2-3 X 10^9 dyn/cm^2 [18,19]. Annealing the films tends to expand the metal more than the substrate, resulting in a changeover at higher temperatures to compressive stress [20]. The film relieves stress by undergoing internal and surface modifications, such as particle growth, void formation, and surface diffusion [20]. Since the atomic mobility is high, during the extended time and elevated temperature of sweeping, the gold film can relax by a stress-relief-vacancy-creep mechanism [21] causing formation of voids surrounded by thicker metal areas (see Fig. 2). Analysis of self diffusion under such conditions is extremely complicated and depends on vacancy source and sink geometry.

The fact that metal voids form predominantly along the narrow metal stripe edges, where stresses are expected to be greatest [22], suggests that a stress-induced mechanism is involved in the nucleation of voids in metal films used for air-sweeping quartz.

The creation of voids beneath the welded-ribbon contact areas could result from the combination of a simple load on the metal (applied compressive stress) and the intrinsic tensile stress of the film. It is well known that when tensile and compressive stresses coexist in metal films, vacancy diffusion can occur from boundaries in tension to boundaries in compression [21] causing metal voids at elevated temperatures [23].

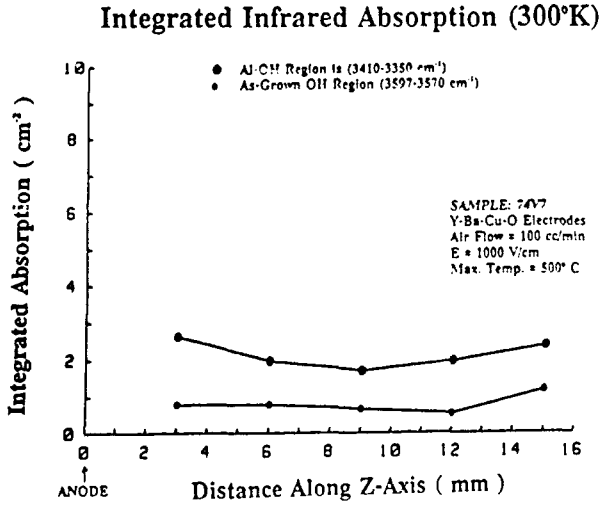


Fig. 9. The integrated absorption of a series of infrared scans (similar to Fig. 1), using Y-Ba-Cu-O electrodes. Both as-grown OH and Al-OH bands are uniform across the bar.

Nonuniform sweeping is sometimes revealed by the occasional formation of light-scattering haze in samples subjected to high-temperature annealing. The haze occurs in unswept regions of the quartz. The formation of haze does not depend on sweeping or electrode effects. However, voids in the electrode will result in the development of clear striae along the Z-axis, indicative of the formation of Al-OH along a cylindrical column of the quartz. When haze occurs near these cylindrical columns, the haze region surrounds the clear region forming a denser cylindrical shell of haze, which in 2-dimensions resembles a smoke ring (see Fig. 2(c) of [1]). This suggests that the formation of Al-OH, along the Z-axis, has influenced the source of the haze to diffuse laterally (in directions other than parallel to Z).

Infrared spectroscopic differences between haze and clear (unswept) regions were investigated. No differences were found in the near infrared region (5500 - 3800 cm^{-1}) which could be attributed to the presence of molecular water [24,25] (see Fig. 10).

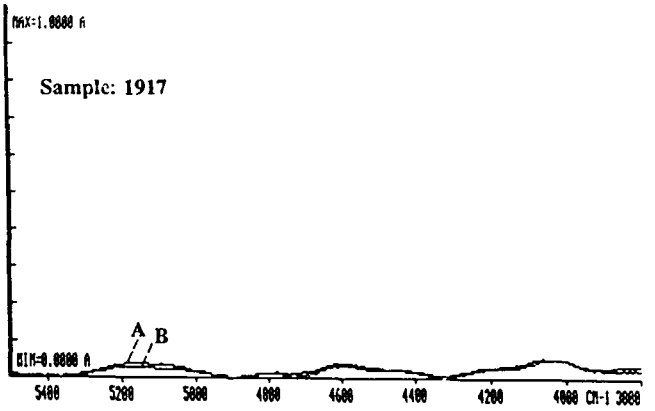


Fig. 10. Infrared (300° K) absorption spectra of quartz. The IR beam was along Y (4.2 cm thick, A = clear path, B = haze path). No differences were found between haze and clear (unswept) paths in the 5500 - 3800 cm-1 region .

Great differences were found in the OH-absorption (3700 - 3100 cm^{-1}) region, in agreement with Iwasaki [5]. Using deconvoluted difference spectroscopy to reveal individual bands, the principal difference OH-bands can be related to the presence of lithium [6]. Also, there is a close relationship between the IR bands associated with haze regions found in our swept-cultured samples and

the IR bands associated with light-scattering "cluster inclusions" found in natural quartz [26] (see Fig. 11). No differences were found in the far infrared region (1000 - 450 cm^{-1}) associated with the precipitation of silicates such as $LiSiO_3$ [27-29] or $LiAlSiO_4$, as suggested by Cohen [4] (see Fig. 12).

Low temperature IR spectroscopy would aid in more accurate identification and correlation of the individual bands. Also, localized infrared modes associated with the presence of cations might be detected at wavenumbers lower than 450 cm^{-1}, in the 200-400 cm^{-1} region [30]. The source of the haze, its character and its ability to diffuse laterally remain puzzling.

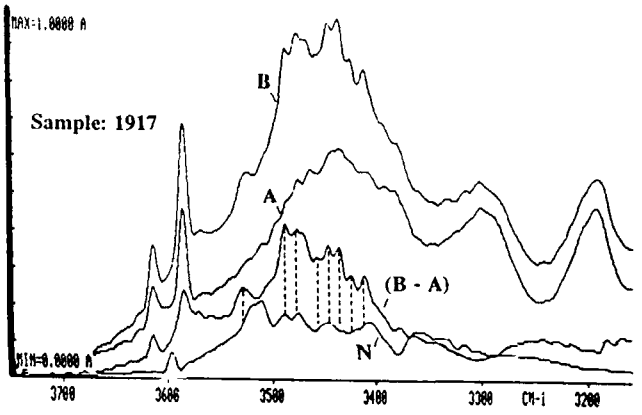


Fig. 11 Deconvoluted absorption spectra of quartz. The IR beam was along Y (4.2 cm thick, A = clear path, B = haze path). The difference spectrum (B-A) is plotted and the principal difference OH-bands are indicated. They are centered on: 3525(Li), 3486, 3476(Li), 3455(Na), 3444(Li), 3435, 3423, and 3411 cm^{-1}. The correlation between (B-A) and a similarly generated natural sample (N) (1.7 cm thick, IR beam perpendicular to Z).

DISCUSSION (UNIFORM SWEEPING)

Au/Cr Films Containing Stripe Openings:

Since stress-induced surface migration of metal to create sufficient voids for uniform sweeping is difficult to control, periodic openings (stripes) were etched in Au/Cr metallizations. The openings provided sufficient electrode-quartz-watervapor (3-phase) regions to achieve uniform H-indiffusion during sweeping (see Fig. 7). The formation of color centers beneath those portions of the Au/Cr electrode not striped, provides proof of the efficacy of this method of sweeping quartz.

The metal-stripe width (see Fig. 6) was designed to be less than twice the apparent Au/Cr-quartz interfacial H-diffusion length found during the sweeping process. An estimate of the interfacial H-diffusivity for Au/Cr electrodes was previously determined to be approximately 7×10^{-9} cm^2/s [1]. This was calculated using a sweeping time of 36 hours with the temperature above 400° C. The radius of the interfacially diffused area was found to be 0.03 cm [1]. The metal stripe width was made less than twice this radius (since hydrogen can diffuse inward interfacially from either stripe edge).

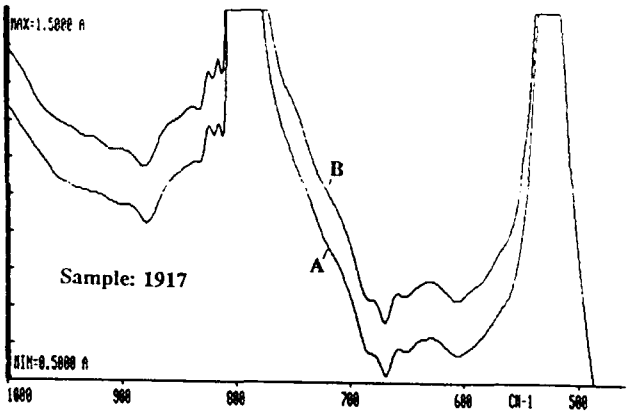


Fig. 12 Infrared absorption spectra of quartz. No differences were found between clear (A) and haze (B) paths in the 1000 - 450 cm-1 region (clear spectrum displaced). The IR beam was along Z (5mm thick).

Hydrogen indiffusion in open-quartz regions adjacent to a Au/Cr boundary was previously reported and discussed [1] (see Fig. 10 of [1]). For this reason, it is possible that hydrogen diffuses into the quartz through the open-quartz regions between the metal stripes. Since the H-introduction region of open-quartz surface was found to be approximately 0.06 cm wide [1], the open-stripe width was made less than 0.06 cm to ensure uniform sweeping.

Y-Ba-Cu-O Films:

Under the deposition and annealing conditions used here, the Y-Ba-Cu-O films are amorphous and remain amorphous if the annealing temperature remains at or below 500° C [31]. The 1.75 μm thick films were transparent and copper colored. The amorphous to crystalline transition begins above 500° C and annealing above 550° C causes grain growth, voiding, and tearing of the films [31]. However, sweeping at temperatures at or below 500° C ensures that the film will remain amorphous. Voiding or tearing of the films was not detected at sweeping temperatures at or below 500° C (see Fig. 13).

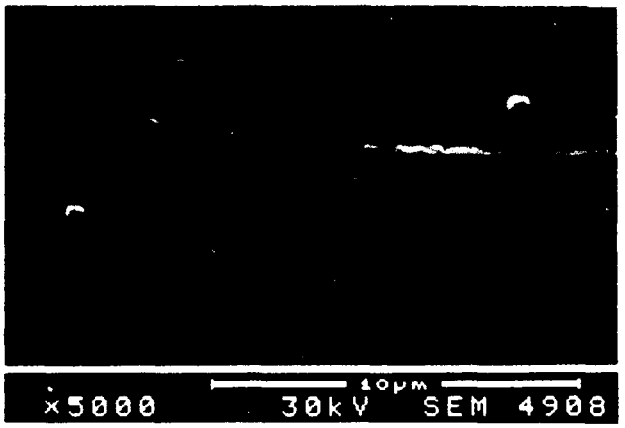


Fig. 13. SEM photograph of a Y-Ba-Cu-O film electrode on quartz (after sweeping). Disregarding random film inhomogeneities, voiding of the film was not detected using magnifications as high as 5,000 X.

Since uniform indiffusion of hydrogen, to form Al-OH, was observed without any appearance of electron-hole compensation (color centers) of Al (see Fig. 9); it was assumed that sufficient hydrogen, from the water vapor in the flowing-air atmosphere, had diffused through the film. Therefore, voiding of the electrode and interfacial diffusion of hydrogen, necessary in the case of Au/Cr electrodes (and a source of nonuniform sweeping) is not required for the Y-Ba-Cu-O electrodes. It is known that both oxygen and hydrogen readily diffuse into the Y-Ba-Cu-O crystalline phase [32] and that $YBa_2Cu_3O_7$ in powdered crystalline form reacts very strongly with water [33].

Not much is known about the amorphous phase of Y-Ba-Cu-O, but it is suspected that films deposited at room temperature without a reactive oxygen atmosphere will be subject to hydrolysis [34]. SIMS techniques have shown that hydrogen is present in the Y-Ba-Cu-O/quartz interface after sweeping. RBS has revealed that no Y, Ba, or Cu metal constituents have diffused into the quartz.

CONCLUSIONS

Stress-induced voiding of evaporated metal electrodes normally provides openings (electrode-quartz-water vapor regions) that allow the indiffusion of hydrogen for sweeping. The voiding is influenced by metal particle size and the application of compressive stresses.

Since the extensive stress-induced voiding necessary for uniform sweeping is difficult to control, sufficient open electrode regions can be created artificially to ensure uniform H-sweeping.

Uniform H-sweeping can be achieved by using conducting oxide electrodes that permit the diffusion of hydrogen without the need for electrode porosity.

ACKNOWLEDGEMENTS

The authors thank Mr. William Washington for preparation of samples, Dr. George McLane and Mr. Melvin Wade for metallizations, Ms. Mary Saunders and Mr. Armand Balekdjian for the photolithographic work, Dr. Robert Pffefer and Mr. Chris Wrenn for the RBS work, Dr. Richard Lareau for the SIMS work, Mr. Donald Eckart for the SEM work, and Dr. Joel Martin of Oklahoma State University for helpful discussions.

REFERENCES

[1] J. G. Gualtieri, L. Calderon, and R. T. Lareau, "Update on possible electrode mechanisms in the sweeping of alpha quartz," Proc. 43rd Annual Frequency Control Symposium (AFCS), IEEE, pp. 509-516, 1989.

[2] J. G. Gualtieri, "Possible mechanisms for the introduction of hydrogen into alpha quartz during sweeping," Proc. 42nd AFCS, pp. 155-161, 1988.

[3] H. G. Lipson, A. Kahan, and J. O'Connor, "Aluminum and hydroxide defect centers in vacuum swept quartz," Proc. 37th AFCS, pp. 169-176, 1983.

[4] A. J. Cohen, "Substitutional and interstitial aluminum impurity in quartz, structure and color center interrelationships," J. Phys. Chem. Solids, Vol. 13, pp. 321-325, 1960.

[5] F. Iwasaki, "Hydrogen bonded OH in synthetic quartz," Jap. Jour. Appl. Phys. Vol. 19, No.7, pp. 1247-1256, 1980.

[6] A. Kats, "Hydrogen in Alpha-Quartz," Philips Res. Repts. vol. 17, pp 133-195 and 260-272, 1962.

[7] B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckwalter, "Quartz crystal radiation effects," IEEE Trans. Nucl. Sci., Vol. NS-17, pp. 217-221, 1970.

[8] G. B. Krefft, "Effects of high-temperature electrolysis on the coloration characteristics and OH-absorption bands in alpha quartz," Radiation Effects Vol. 26, pp. 249-259, 1975.

[9] J. G. Gualtieri and J. R. Vig, "Sweeping and irradiation studies in quartz," Proc. 38th Ann. Freq. Control Symp., IEEE, May 1984, pp. 42-49.

[10] J. G. Gualtieri and D. W. Eckart, "The influence of surface finish and metallization on electrode electromigration in alpha quartz during sweeping," Proc. 40th. Ann. Freq. Control Symp., IEEE, May 1986, pp. 115-120.

[11] J. G. Gualtieri, "The influence of temperature and electric field on the etch-channel density in swept - cultured quartz," Proc. 39th. AFCS, pp. 247-254, 1985.

[12] J. G. Gualtieri, R. T. Lareau, and D. W. Eckart, "Electrode Effects in the Sweeping of Alpha Quartz," to be published in IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, Sep., 1990.

[13] H. G. Lipson, "Infrared and laser spectroscopic characterization of aluminum defects in cultured quartz," Proc. 40th AFCS, pp.63-69, 1986.

[14] G. McLane, R. Pffefer, W. Savin, C. Wrenn, "Triode magnetron sputtered superconducting Y-Ba-Cu-O thin films," Proc. Army Workshop on High Tc Superconductors, May 23-25, 1989, Redstone Arsenal, Alabama. Doc# GAC/AC PR-89-02, published by IIT Research Institute, 10 West 35th St., Chicago, IL 60616.

[15] R. Aberman and R. Koch, "Internal stress of thin silver and gold films and its dependence on gas adsorption," Thin Solid Films, Vol. 62, pp. 195-208, 1979.

[16] D. J. Young and M. Cohen, "Oxidation behavior of chromium between 300° and 600° C," J. Electrochem. Soc., Vol. 124, pp. 769-774, 1977.

[17] D. W. Hart, Jennifer Frank, Dowlan Smith, J. J. Martin, and J. G. Gualtieri "A Study of the Electrodiffusion Process in Quartz," Proc. 44th AFCS, 1990.

[18] K. L. Chopra, Thin Film Phenomena, Ch V, McGraw-Hill Book Co., New York, 1969.

[19] E. Klokholm and B. S. Berry, "Intrinsic stress in evaporated metal films," J. Electrochem. Soc., Vol. 115, No. 8, pp. 823-826, 1968.

[20] A. Christou and H. Day, "Structure and thermal stability of sputtered Au-Ta films," J. Appl. Phys., Vol. 44, No. 8, pp. 3386-3393, 1973.

[21] A. H. Cottrell, An Introduction to Metallurgy, pp. 355-358 and p. 843, St. Martins Press, New York, 1967.

[22] S. K. Groothius and W. H. Schroen, "Stress related failures causing open metallization," Proc. 25th International Reliability Physics Symposium (IRPS), pp. 1-8, 1987.

[23] J. T. Yue, W. P. Funsten, and R. V. Taylor, "Stress induced voids in aluminum interconnects during IC processing," Proc. 23rd IRPS, pp. 126-137, 1985.

[24] P. Cordier, J. C. Doukhan, J. Bignon, and X. Buisson, "Growth defects in cultured quartz, a microscopic investigation," Proc. 43rd. AFCS, pp. 459-469, 1989.

[25] R. D. Aines, S. H. Kirby, and G. R. Rossman,

"Hydrogen speciation in synthetic quartz," Physics and Chemistry of Minerals, Vol. 11, pp. 204-212, 1984.

[26] Clifford Frondel, Dana's system of mineralogy, VIII, silica minerals, p. 238, John Wiley and Sons, Inc, New York, 1962.

[27] F. A. Miller and C. H. Wilkins, "Infrared spectra and characteristic frequencies of inorganic ions," Anal. Chem., Vol. 24, pp. 1253-1294, 1952.

[28] G. W. Arnold, "Near-surface nucleation and crystallization of ion-implanted lithia-alumina-silica glass," J. Appl. Phys., Vol. 46, No. 10, pp. 4466-4473, 1975.

[29] A. I. Berezhnoi and Y. M. Polukhin, "Formationof crystalline centers in light sensitive glasses," Inorg. Mater. Vol. 3, pp. 880-884, 1967.

[30] F. Gervais, A Blin, D. Massiot, and J. P. Coutures, "Infrared reflectivity spectroscopy of silicate glasses," Jour. Non-Cryst. Solids, Vol. 89, pp. 384-401, 1987.

[31] S. I. Shah, "Annealingstudies of $Yba_2Cu_3O_{7-x}$ thin films, "Appl. Phys. Lett., Vol. 53 (7), pp. 612-614, 1988.

[32] H. Wiesmann, D. H. Chen, R. L. Sabatini, J. Hurst, J. Ochab, and M. W. Ruckman, "Effect of stoichiometry and post-annealing conditions on magnetron sputtered films of $Yba_2Cu_3O_7$ on cubic zirconia," J. Appl. Phys., Vol. 65, pp. 1644-1647, 1989.

[33] R. L. Barns and R. A. Laudise, "Stability of superconducting $YBa_2Cu_3O_7$ in the presence of water," Appl. Phys. Lett., Vol. 51 (17), pp. 1373-1375, 1987.

[34] Private Communication (Mar '90), with Dr. H. Wiesmann, Dept. of Applied Science, Brookhaven National Laboratory, Upton, New York 11973.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PLASMA ETCHING OF MONOCRYSTAL QUARTZ

L. Spassov, E. Yossifov, D. Yankov

Institute of Solid State Physics, Bulgarian Academy of Sciences, Sofia, Bulgaria

H.-U. Poll, S. Schreiter

Technical University, Karl-Marx-Stadt, DDR

Abstract

We worked out the plasma etching method as well as some special equipment aiming the etching of monocrystal quartz. For this purpose the influence of different plasma parameters on the etching rate and the surface profile were studied. Quartz wafers were etched with a high rate (over 0.65 μm/min) as a result of which thin quartz membranes were formed. On such membranes aluminium electrodes were deposited and resonance systems at 42.950 MHz were obtained.

1. Introduction

The production of high frequency piezoelectric devicesis is limited by the level of technology and equipment for the mechanical preparation of quartz wafers. By using modern mechanical methods a thickness of 80 μm can be attained, which corresponds to about 20 MHz fundamental resonant frequency. Lately this method has been considered to be on the limits of its efficiency.

Therefore a new method for the preparation of thin wafers and membranes is needed [1,2]. These membranes serve as a basis for development of new types of sensors for low pressure measurement, too [3].

Thin membranes can be produced by etching methods. The rate of wet etching processes is determined by the crystal cut. Surface roughness is increased in the process of wet etching because minute defects of the surface provide for etch grooves which are limited by etch stable crystal planes. These disadvantages can be avoided by using plasma etching [4,5].

The high energy of etch active particles produced in a low pressure glow discharge exceeds the energy of all chemical bonds between the substrate molecules in such a degree that the etch rate becomes independent of the crystal cut.

2. Experiments

At the Technical University of Karl-Marx-Stadt, DDR, and the Institute of Solid State Physics of Bulgarian Academy of Sciences the plasma etching method with a high rate is established and applied in the production of thin quartz membranes and the investigation of the influence of quartz wafers orientation on the etch rate.

The quartz wafers were etched in a special plasma etching device developed at TU - Karl-Marx-Stadt. The plasma reactor consists of a planar electrode and an asymmetric grounded electrode in a vacuum vessel of a conventional high vacuum equip-

CH2818-3/90/0000-238 $1.00 © 1990 IEEE

ment (fig. 1)

The discharge was maintained at a high frequency (27 MHz), electrical power density about 5 W/cm² and pressure from 10 to 100 Pa. Etch gas was CF_4. The etch configuration provided for a high self-bias voltage and in this way high energy of the etch active ions. The achieved rate was from 0.2 to 1.3 μm/min, depending on the plasma parameters.

3. Results

3.1. Etching of quartz wafers with different cuts

The different crystal orientations of the wafers were of X, Y, Z and AT-cut. The wafers surface were obtained through lapping with different abrasives or polishing.

The etch rate was measured by means of a microbalance with a sensitivity of 10 mg. Precise estimation of the geometrical dimensions of wafers with the help of microscopic techniques results in values for the volume with errors smaller than 2%.

Etch rate is expressed by the relation:

$$r = \frac{\Delta d}{\Delta t},$$

where Δd is change of the thickness and Δt - the etch time.

The change of thickness can not be measured directly with high accuracy because of the roughness of the wafers. Therefore, at first we estimate the geometrical dimensions, which allow us to calculate the average mass density. Figure 2 demonstrates an exactly linear change with mean slope $\frac{\Delta m}{\Delta t} = \alpha$ and initial mass m_0. Than we calculate the etch rate:

$$r = \frac{1}{A} \cdot \frac{\Delta m}{\Delta t} = \frac{d}{m_0} \cdot \alpha,$$

where d is thickness and A - area of the surface.

Errors in the measured etch rate can be results only of the thickness measurement because the weighing error is small. Within a small error the rate is independent of the crystal cut. The surface roughness of wafers causes a greater influence on the rate. The difference between the etching rates of the two types of wafers (lapped and polished) is negligible - about 2.8%. Perhaps the little bit higher etch rate connected with rough wafers is due to the larger effective surface.

3.2. Etching of quartz wafers with a different initial surface roughness and under different conditions in the etching process

Figure 3 shows the topography as well as some typical profiles of etched wafers with etching depth - 30, 60, and 100 μm. For comparison we also give the same characteristics of the lapped wafers. With the increase of the etching depth the dimensions of the etching forms change from 1-2 μm to 10-15 μm and respectively Ra - from 0.18 to 0.28 μm.

The next figure 4 demonstrates the topography of the surface of a polished wafer, which has been etched to a depth of 100 μm. The etching figures in the peripheral zone (fig. 4a) and the central part of the wafer (fig. 4b) have forms, similar to the surface mechanical defects that appear in the process of lapping and polishing of the wafers (i.e. break-offs in the periphery and scratches). Most of the surface of the etched wafer is equally etched and smooth.

On figure 5 a comparison is shown between

profiles of the surface relief of a lapped (a) and polished (b) wafers before (1) and after (2) etching at rate 1.24 μm/min and etch depth 100 μm. The roughness of the lapped wafer which we measured before and after etching is respectively 0.12 μm and 0.22 μm. The roughness of the polished wafer is one and the same before and after etching - Ra = 0.002 μm.

Figure 6 shows the dependence of the etching rate Ve on the pressure and the self-bias voltage V.

We etched wafers that have undergone different mechanical operations with different abrasives. The way the wafers have been mechanically treated does not influence the etching rate which is constant with an allowance of 3%.

3.3. Preparation of a resonator system on thin quartz membranes

On quartz wafers AT-cut with a thickness 150 μm and diameter 8 mm which have been lapped with abrasive 3 μm, as a result of etching membranes were constructed. The membranes were situated in the central part of the wafers. On such membranes aluminium electrodes were deposited, whose diameter was 2 mm and thickness - 800 A (fig. 7). These resonance systems were fixed with conductive glue to standard holders for miniature resonators M2P (fig. 8). The measured electrical parameters are the following:

- resonance frequency - 42.957 MHz
- dynamic resistance - 105 Ω
- spectral characteristic is shown in figure 9.

4. Conclusions

Plasma etching method and corresponding equipment with a high rate etching are established and applied for production of thin quartz membranes. The influence of the orientation of the wafers and of different plasma parameters on the etching rate and the quality of the surface has been investigated.

In order to improve the electrical parameters of the resonance systems the following must be done:

- the surfaces of the wafers should be well polished before etching and should be plane-parallel;
- the parameters of the plasma process and the construction of the masks ought to be optimized in view of manufacturing membranes whose surface characteristics and plane-parallelity are similar to those of the initial wafers.

5. Acknowledgements

The authors are very grateful to A. Stratieva and D. Mikhailov for preparing of the substrats and to R. Zakharieva of Interquartz - Sofia for the assistance in the creation of resonators.

6. References

[1] Бронников В.Г. и др. "Результаты исследований монолитных кварцевых фильтров на обратных мезаструктур" - Электрон. техника. Серия радиодетали и радиокомпоненты. Вып. 2(51) 1983

[2] Качальников и др. "Реализации высокочастотных пьезоэлектрических резонаторов на основе обратных мезаструктур" - Электрон.

техника. Серия радиодетали и радиокомпоненты, вып. 3(52) 1983

[3] Frisch et all. "Vibrating quartz diaphragm pressure sensor", patent USA 4, 479, 070 - Oct. 1984

[4] Poll H.-U. u.a. "Die Glimmentladung in perfluorierten Gasen zum Plasmaatzen und zur Schichtabscheidung". Wissenschaftliche Schriften-reihe der Technischen Hochschule, Karl-Marx Stadt, 4.1985

[5] Киреев В.Д. и др. "Плазмохимическое и ионно-химическое травление микроструктур". Москва, "Радио и связ", 1983

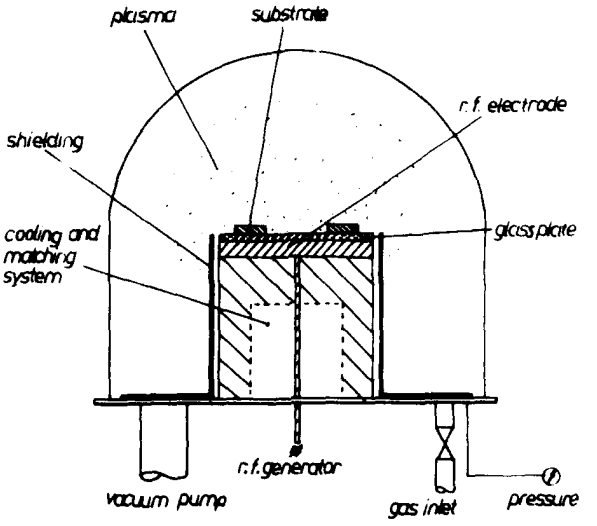


Fig.1. Schematic representation of the used asymmetruc high rate plasma etch arrangement

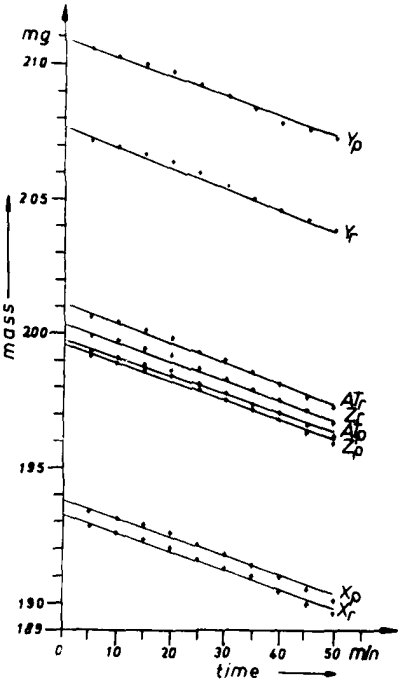


Fig.2. Mass variation with the time as a result of plasma etching r - lappeds surface, p - polished surface

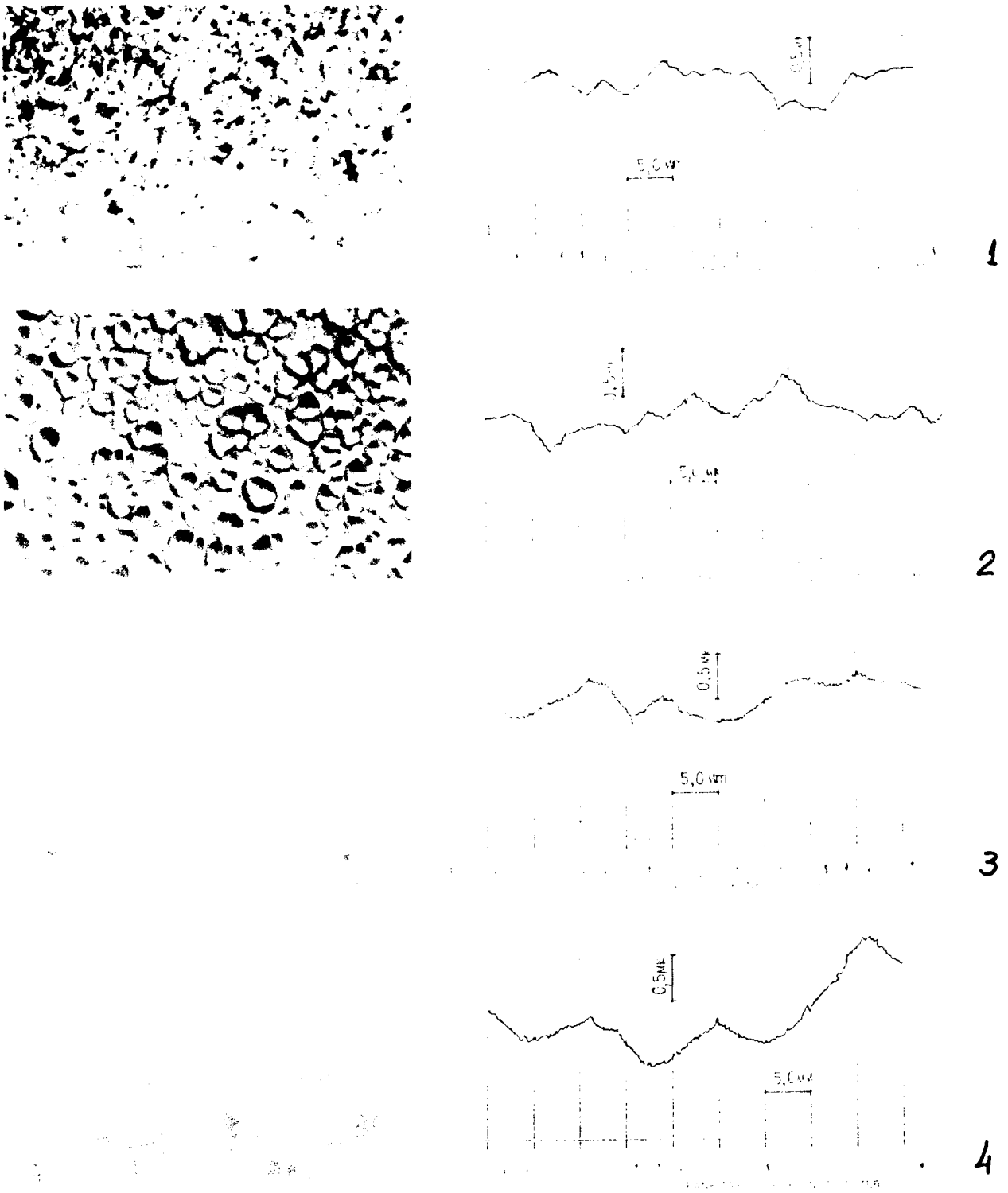


Fig.7. Surface characteristics of etch wafers

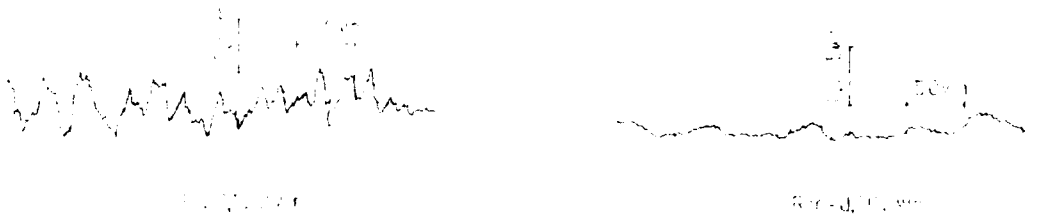
1 - Lapped surface

2 - Surface at an etch depth 20 μm

3 - Surface at an etch depth 60 μm

4 - Surface at an etch depth 100 μm

A B



1

2

A B

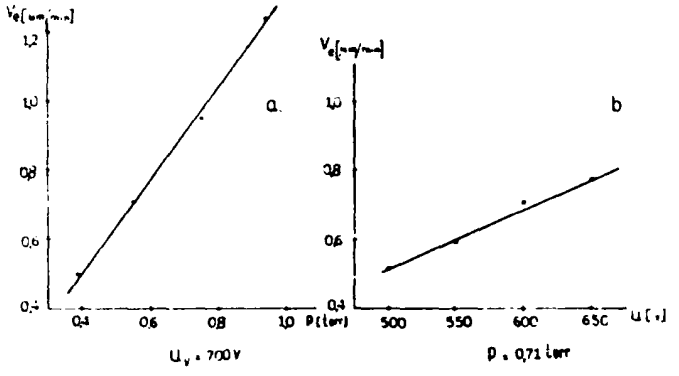


Fig.6. Dependance of the etching rate Ve on the pressure P and self-bias voltage V

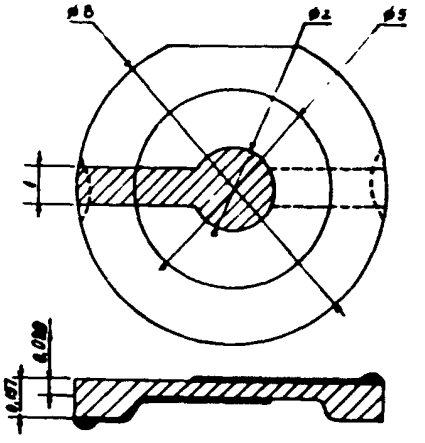


Fig.7. Diagram of the resonance system

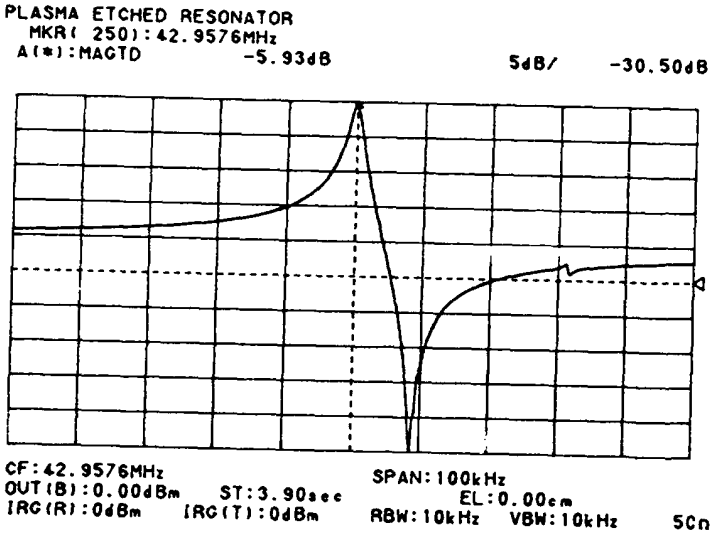


Fig.9. Spectral characteristic of the resonator

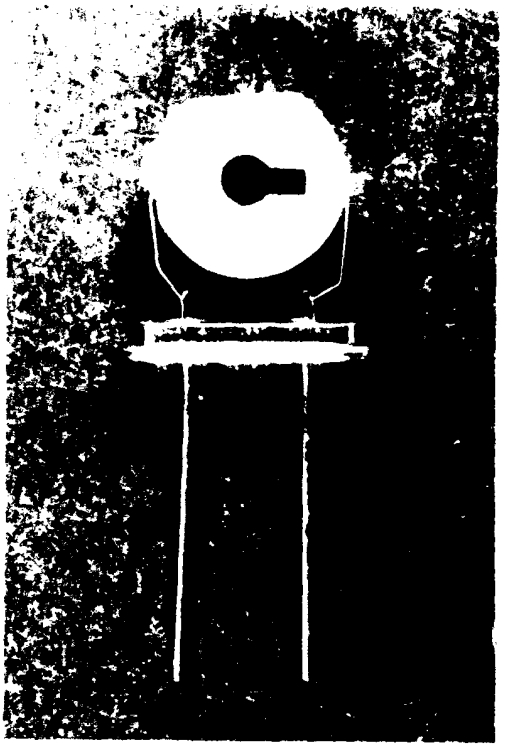


Fig.8. General view of the resonator

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AUTOMATIC FREQUENCY CONTROL IN CHEMICAL ETCHING OF QUARTZ CRYSTAL BLANKS

Franz L. Sauerland

Transat Corp.
31000 Bainbridge Road
Solon, Ohio 44139

ABSTRACT

This paper describes a method and system for measuring the resonance frequency of piezoelectric resonators in conductive fluids. One of its applications is in the chemical etching of quartz crystal blanks, where it can be used to monitor and control either the etch rate of the etchant or the etching of blanks to a targeted thickness or frequency.

The system comprises a "Crystal Etch Monitor" in conjunction with special electrodes that are designed to take account and advantage of the conductive nature of the etchant. It can etch a load of blanks or a "monitor" blank different from the etch load but immersed in the same etchant. When the blank reaches a predetermined target, an etch-termination signal is triggered that can be used to alert an operator or to initiate action for automatic etch termination.

BACKGROUND

During etching, the blank's thickness is reduced by

(1) $\Delta T = K (F2-F1) / F1 \times F2$,

where F1 and F2 are the blank frequencies (in MHz) at the start and end of etching, respectively, and the constant K has a value of 1000. In industry parlance, Delta T is frequently referred to as "Delta F Square" or "F Square" (F^2).

If the instantaneous blank frequency F can be measured, then etching to a target Delta T can be controlled by monitoring F (starting with its initial value F1) and by terminating the process when F reaches the target value F2, which can be obtained from equation (1).

The etch rate is a function of the concentration and temperature of the etchant and the surface finish of the blank. Fig.1 illustrates the effect of blank surface. It is copied from Ref.1 and shows that the rate is constant for polished or pre-etched blanks but varies for lapped blanks during the initial etch.

ELECTRODE

The Crystal Etch Monitor (CEM) measures the blank frequency during etching by means of special electrodes located in close vicinity to the blank. One of the various suitable electrode configurations exposes only one surface of the blank to the etchant and allows only one individual blank to be measured at one time. This blank will henceforth be called BLANK in order to distinguish it from other blanks that may be contained in the same etch tank.

Fig. 2 shows one type of BLANK-electrode assembly immersed in etch fluid F in a tank T. On one side of the BLANK B is a rf-conductive electrode E in close vicinity of or in direct contact with the blank surface. It is insulated from the etchant by an insulator I and connected to terminal T1 via an insulated cable C1. On the other side of the blank, the (conductive) etchant acts as another electrode that is connected to terminal T2 via conductor C and insulated cable C2.

Fig. 3 shows another type of BLANK-electrode assembly. In addition to the basic components of Fig. 2 it comprises a spring SP that presses the BLANK against a seal S.

CH2818-3/90/0000-246 $1.00 © 1990 IEEE

MONITOR BLANK

The BLANK may be

a) a single blank to be etched to target,
b) a representative of an etch load,
c) a monitor blank that can be used for repeated etching and may have a frequency different from the etch load.

In case a), the CEM can control the etch process according to equation (1).

In case b), the CEM can control the process according to equation (1) if the value for K is doubled to account for the fact that the BLANK is etched on one side only, i. e. at half the rate of the etch load.

In case c), the monitor etch rate is constant, since the monitor will normally be pre-etched from use over repeated etch cycles. If the etch load is also pre-etched or polished, etching can be controlled as in case b). However, if the etch load has a lapped surface, etching can be reliably controlled ONLY if there is a defined and repeatable correlation between the thickness removal of the monitor and the thickness removal of the etch load. This correlation is discussed below under "Experimental Results".

MEASUREMENT CONSIDERATIONS

The objective is to determine etch removal in terms of frequency measurements. Since the frequency changes during etching, it must be measured either continuously or fast and frequently.

It remains to be determined which frequency to measure. In normal blank and resonator measurements, it is the series resonance frequency that is of interest. However, etch measurement conditions are different from normal conditions because of the immersed BLANK's very low values for both Q and Factor-of-Merit M. As a result, not only is it impossible to determine the exact series resonance via a single-point measurement, but even the traditional multipoint measurements fail because the admittance/impedance curves are no longer circles in the complex plane. This is due to the fact that the Co-reactance can no longer be considered constant over the spectrum of interest, and it is evidenced by the fact that the factors interrelating the characteristic resonator frequencies become imaginary for low M (Ref. 2).

Fortunately it is not necessary to measure the exact series resonance frequency if there is another characteristic BLANK frequency that can be measured with good repeatability. This is apparent from equation (1), which depends on frequency DIFFERENCE, i. e. requires only relative accuracy of frequency measurement.

The characteristic frequency chosen for the CEM is the frequency of the inflection point of the slope of the BLANK's impedance vs. frequency curve between series resonance and anti-resonance. This frequency is bounded within relatively narrow limits and can be measured with good repeatability.

Another consideration is the accuracy of Delta T measurements in terms of the frequency resolution. The present CEM has a 5-digit floating point frequency display, i. e. its worst resolution is 100 ppm. This means that the accuracy will be poor for small values of Delta T. For instance, the maximum tolerance will be 10% if the frequency difference is 1000 ppm. The corresponding Delta T can be obtained from equation (1) (for K=2000 and (F2-F1) = .001) as

$$\text{(2)} \qquad \text{Delta T} = 2/\ F2 \ \text{(in MHz)}$$

Equation (2) shows that for a given Delta T and limited frequency resolution, the measurement accuracy increases with increasing BLANK frequency. From this point of view, a high monitor frequency is desirable. On the other hand, a low frequency (thick BLANK) would be desirable for maximum longevity of the monitor.

SYSTEM CONNECTION AND OPERATION

Fig. 4 shows a diagram of the etch control system, including the CEM and the electrode/BLANK arrangement of Fig. 2. Terminals T1 and T2 are connected to a measurement probe P that is connected via coaxial cable to the CEM, which provides two output signals: S1 triggers when the BLANK frequency equals the target frequency; S2 can be set to trigger at a predetermined frequency below the target frequency and can be used to initiate anticipatory action (such as alerting the operator) before etch termination.

As mentioned before, the CEM can be used to monitor the etch rate of an etchant or to control the etching of one or more blanks to a predetermined target. In the latter case, the basic operating steps are:

1. Enter target.
2. At the start of etching, push "START".
3. Upon reaching target, the CEM switches to STOP and triggers an etch-termination signal.

Since some users may be accustomed to etching and targeting in terms of Delta T and others in terms of frequency, the CEM offers the choice between both modes, as follows:

Delta T Targeting
The operator enters the target Delta T for the etch load. At the start of etching, the CEM measures the BLANK's start frequency F1, computes F2 from equation (2), and terminates etching when the BLANK frequency reaches F2.

Frequency Targeting
The operator enters start frequency and target frequency of the etch load. From this, the CEM computes Delta T via equation (1). At the start of etching, the CEM measures the BLANK's start frequency F1, computes F2 from equation (2), and terminates etching when the BLANK frequency reaches F2.

EXPERIMENTAL RESULTS

1. Monitor
For now, 10 MHz blanks with 11.5 mm diameter were selected for monitors and packaged similar to Fig. 2, using materials that withstand the harsh etching environment. If this monitor is etched to half its thickness, it can monitor a total etch load removal of about 100 (F^2), provided it is free of etch channels.

2. Monitor Etch Rate Correlation
Measurements were made to test and establish the relationship between thickness removal of monitor and etch load. These measurements - based on an early CEM prototype and a 10 MHz monitor - are summarized in Table 1 and explained as follows:

Four different groups of lapped crystals were tested. Each group comprised 3 subgroups of 5 crystals that were preselected for narrow frequency spread. Each subgroup was etched separately.

The table lists the following data for each group: Surface Finish (in micron); Thickness Removal of monitor (in F^2); "Work Removal", defined as thickness removal of etch load (in F^2); "Absolute Repeatability", defined as the total spread (in F^2) readings for all finished crystals; "Maximum Error on Work", defined as the maximum error (in KHz) for frequency readings for all finished crystals.

In all cases the etch load thickness removal is larger than the monitor removal. The difference becomes larger for coarser surface finish, as is to be expected according to Fig. 1.

In all cases, the "maximum frequency error" is less than 600 ppm.

In the two plano-plano cases with thickness removal of .444 F^2 or more, the "Absolute Repeatability" is better than 5%, while for the smaller thickness removal of .219 F^2 it increases to 8.6%. The lower accuracy for lower Delta T is to be expected from equation (2). According to this equation, a 10 MHz BLANK can have a resolution-dependent error of up to 10% for a Delta T of 0.2.

All results were affected by an instability in monitor frequency readings of +/-2 KHz. For the 10 MHz monitor this corresponds to an instability of +/-200 ppm or +/- .02 F^2. This instability has been reduced to about +/-50 ppm in the present CEM.

The same monitor was used for all measurements. It was continuously submerged in etchant for several operating shifts without deterioration or response change.

CONCLUSIONS

Obviously, more work is needed to optimize the system and to gain experience with it as a production tool. However, several promising conclusions can be drawn from these two main results:

1) The system permits reliable blank frequency readings with a repeatability better than 500 ppm.

2) For given thickness removal and blank surface, there is a repeatable correlation between the thickness removals of monitor blank and etch load.

From 1) we can derive these conclusions:

1a) The system can be used to etch to a predetermined frequency.

1b) The system can be used to monitor the etch rate of an etchant.

1c) The rate monitor system per 1b) can likely be expanded to control the etch rate of an etchant - for instance by feedback adjustment of one or more of the variables affecting the etch rate.

1d) A system per 1c) can be used for etching to a predetermined thickness removal, since the etch time for given etch conditions (etch rate, blank surface, Delta T) is predictable.

1e) A system per 1c) could also be used to control the etch rate of etchants that, while causing thickness removal in quartz (or other piezoelectric resonators), are used for etching materials other than quartz, such as silicon.

From 2) we can conclude:

2a) Etch control using a BLANK that is a representative of the etch load is feasible and can be effected in terms of equation (1).

2b) Etch control using a monitor blank is feasible and can be effected in terms of empirical correlations between the thickness removals of monitor and etch load. These correlations depend on blank surface and target Delta T and can be stored in written or computer memory.

For small values of Delta T, its measurement accuracy is limited by the resolution of the frequency measurement. This can be alleviated in two ways:

a) by increasing the resolution;

b) by controlling etch time and etch rate (as in 1d), while basing the etch rate measurement on frequency readings taken at steps sufficiently larger than the frequency resolution.

CEM SPECIFICATIONS

Freq. Range	1 - 95 MHz *
Delta T Range	0.001 - 1000
Display	Freq. or Delta T
Computer Interface	RS 232

* While the CEM has measurement capabilities covering this range, it has so far only been applied to etching of fundamental monitor blanks from about 4 to 25 MHz and to measuring third overtones to about 45 MHz.

ACKNOWLEDGMENTS

The system described is the result of teamwork of Transat engineers . The practical evaluation was helped greatly through cooperation with J. Kent and E. Hughes of Innovative Frequency Control Products, who supplied the "Experimental Results" described above.

REFERENCES

1. DESIGN AND I..LEMENTATION OF AN ETCH SYSTEM FOR PRODUCTION USE
D. Ang, Proc. FCS 1978

2. THE PIEZOELECTRIC CRYSTAL UNIT - DEFINITIONS AND METHODS OF MEASUREMENT
E. Hafner, Proc. IEEE, Feb. 1969

3. FREQUENCY MEASUREMENT OF PIEZOELECTRIC RESONATORS IN ELECTRICALLY CONDUCTIVE FLUIDS
F. Sauerland, U.S. Patent Application 7-478245 and Foreign Applications

4. CRYSTAL ETCH MONITOR
Transat Corp., Preliminary Data Sheet

GROUP #	DESCRIPTION	SURFACE FINISH (u)	MONITOR REMOVAL (F^2)	WORK REMOVAL (F^2)	ABSOLUTE REPEAT (F^2)	MAX ERROR ON WORK (KHz)
1	1.2288 MHz BICONVEX	12	.140	1.388	.16	.230
2	11.8 MHz PLANO-PLANO	5	.140	.9949	.05	6.9
3	31.0 MHz PLANO-PLANO	3	.100	.444	.019	18.45
4	31.0 MHz PLANO-PLANO	3	.050	.219	.019	18.6

TABLE 1

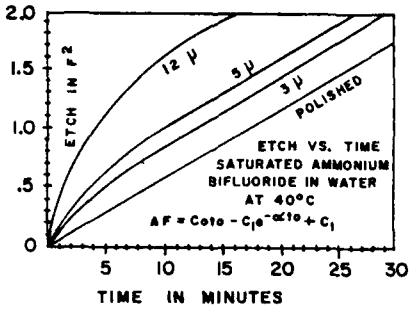


FIG. 1

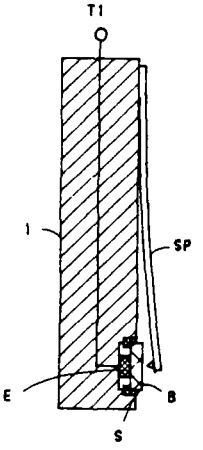


FIG. 3

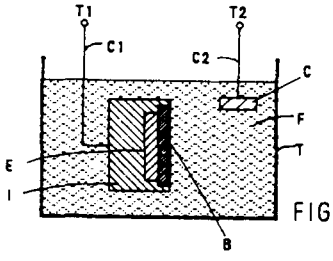


FIG. 2

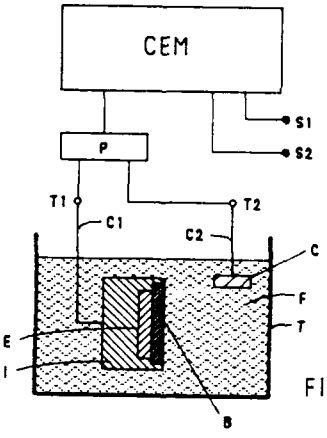


FIG. 4

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A COMPARISON OF LASER-INDUCED TWINNING OF QUARTZ BASED ON CHARACTERISTICS AND SUPPLIER

K.L. Blisnuk, J.G. Gualtieri and R.A. Murray

USALABCOM, ET & D Laboratory, Fort Monmouth, N.J. 07703-5000

Abstract

AT-cut samples were irradiated at room temperature using a focused CO_2 laser. For both swept and unswept material, it was found that irradiance levels necessary for twin nucleation did not correlate with $\alpha(3500\ cm^{-1})$ in the range $0.028 < \alpha < 0.189$.

Surface twins, forming on AT-cut plates, remained approximately the size of the irradiated area. After repeated exposures, the twinned area formed figures resembling the calculated ferrobielastic stress diagram for AT-cuts.[1] Twinned areas were more irregular in high etch channel regions. SC-cut twins showed a tendency to extend, in a direction close to the z-axis, towards the edge of the sample in only one direction. This result differs with the more symmetric stress diagram calculated for the SC-cut.

Introduction

The quality of quartz material, including susceptibility to twinning, is an important concern in many applications. Dauphiné twins can form from the normal processing of a quartz crystal, e.g., cutting quartz bars can produce surface twins in the plane of the cut.[2] Dauphiné twinning in quartz can be induced by internal stresses created by mechanical deformation or thermal expansion produced by the absorption of laser radiation.

Electrical, or Dauphiné, twinning in quartz results when silicon-oxygen atomic groups are shifted through distances of about 0.3 angstroms. Since the movement is the same order of magnitude as thermal vibrations at room temperature, it does not break Si-O bonds.[3] The structures of an interpenetrating twin and the untwinned area can be brought into coincidence if one is rotated through 180 degrees about the c-axis,[1] as shown in Figure 1. Twinning can occur when the quartz is subjected to localized heating which induces internal stresses. If the stresses are high enough, the quartz will twin in that area.

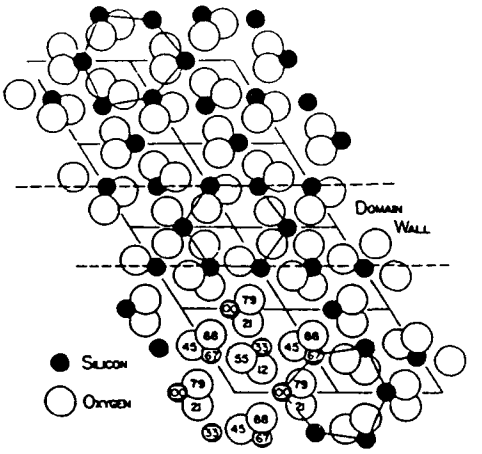


Figure 1. Dauphiné twin formation and untwinned region (from Anderson, et. al. (Ref. 1)).

The presence of twinning can degrade the performance of quartz resonators.[4] It has been asserted by some that high Q quartz twins easier than low Q material. The Hewlett Packard Company's quartz resonator and transducer operation specifies a Q_{IR} of no less than 2.2 million and no greater than 2.6 million. Although a Q_{IR} greater than 2.6 million would be desirable for their applications, experience with their major supplier of quartz has taught them that when $Q_{IR} > 2.6$ million, the incidence of twinning increases.[5] Frondel[6] also noted that quartz of relatively pure quality twins more readily, however, no quantitative evidence was given. Bandyopadhyay and Merkle[7] studied the influence of impurities and defects on laser-induced damage in quartz. Their results indicated that OH^- concentrations (used to determine Q_{IR}) did not influence damage resistance using either single- or multiple-pulse irradiation.

US GOVERNMENT WORK IS NOT PROTECTED
US COPYRIGHT

The objective was to determine whether the Q_{IR} or other properties of cultured quartz influences twinning and if the tendency to twin can be correlated with supplier. Twinning was induced in quartz by the absorption of CO_2 laser radiation following work done by Anderson et al.[1] This method of twinning quartz is efficient and reproducible. The samples were irradiated at room temperature.

Experimental Method

Lumbered Y-bars were used for producing AT-cut plates. The absorption coefficient α(3500 cm^{-1}), using EIA Standard EIA-477-1, was measured to determine Q_{IR}-values before further processing was done. One-half of each bar was swept and AT-cut plates were cut from each bar and lapped mechanically to a 1 μm polish. The plates were chemically polished in a saturated N_2H_4F:HF:H_2O solution for 1-2 hours at 70°C, removing about 13 μm from each surface. The z-axis was determined optically and the thickness of each sample was measured. A sketch of surface features was also made to assist in identification of each piece. The α's ranged from 0.028 to 0.189, which, using equation (10) of Reference (8), corresponds to a Q range of 2.55 million to 0.63 million. The sample's etch-channel densities ranged from 0 to 200 /cm^2.

A diagram of the laboratory setup is shown in Figure 2. An Apollo flowing-gas CO_2 laser was used. The laser beam was focused down from its 7 mm width to 1 mm to increase the power density. A beamsplitter was used to take a portion of the unfocused beam (17%) and direct it to a thermopile detector for power measurement. Samples were irradiated for one second, then the samples were chemically etched in 2:1 HF:H_2O solution to reveal the twins, if any had been produced. If no twins were found, the process was repeated at a higher beam power. The Q_{IR}, quartz cut and type (whether it was swept or unswept and how it was classified according to its growth), etch channel density (ρ) in the area of the twin, surface finish, irradiation levels used and quartz supplier were recorded. A summary of the quartz material used is shown in Table I. SC-cut blanks were used to determine twin shape only, as the blanks were already processed by the supplier without including a precise analysis of Q_{IR}. The as received SC-blanks were too thin to allow an accurate Q_{IR} determination and so, no further analysis was performed.

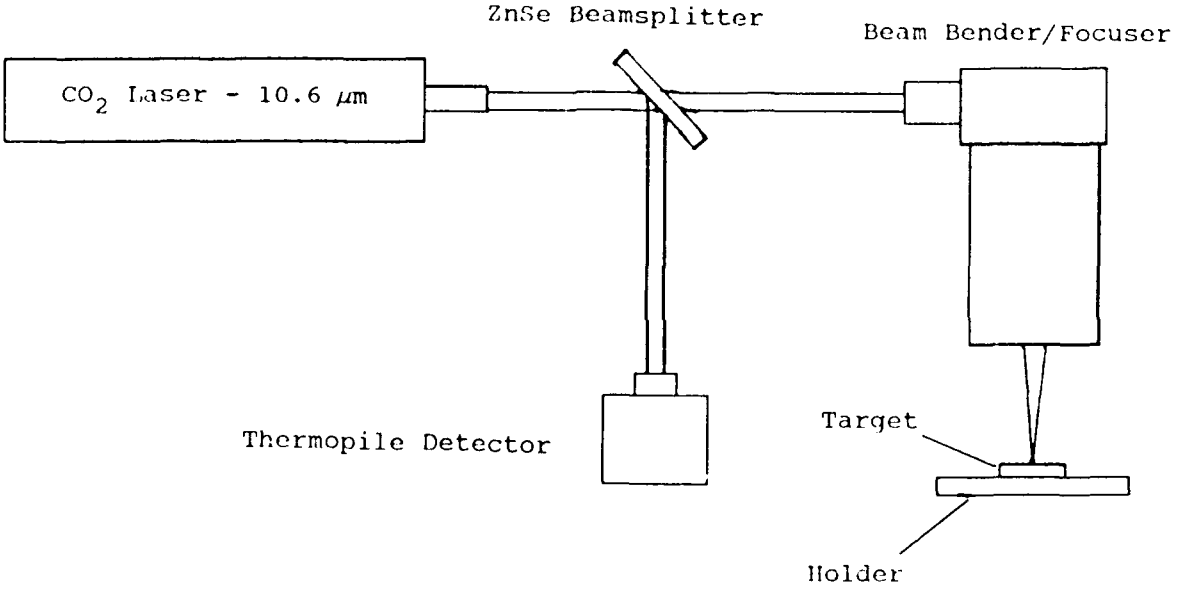


Figure 2. Laboratory setup for laser irradiation of quartz plates.

Experimental Results

The twin nucleation data obtained were divided into separate supplier files of swept/unswept groups. They were then compared to determine if relations between α and the power density (threshold irradiance required to twin), etch channel density or twin size could be found. Etch channel densities, twin sizes and irradiances were also compared.

Figure 3 shows a graph of threshold irradiance vs. α. For the range of α's studied, there is no relationship between α and the amount of energy required to initiate twinning. Comparing α to twin size, as in Figure 4, again, shows no correlation. This supports the findings of Bandyopadhyay and Merkle, who found that factors related to Q_{IR} do not influence quartz laser damage behavior.

	AT SWEPT	AT UNSWEPT	SC SWEPT	SC UNSWEPT
SUPPLIER A	15 Q=0.7, 1.9*	15 Q=0.6, 1.8		
SUPPLIER B	5 Q=1.3	10 Q=1.3, 1.4		
SUPPLIER C	10 Q=1.2, 2.4	29 Q=0.9, 1.2, 1.3, 1.8, 2.6	18 Q>2.0	17 Q>2.0
TOTAL	30	54	18	17

* All Q values given are times 10^6.

Table I. Sample categories. A summation of the types of samples irradiated. SC-cuts were irradiated only to determine the shape of their twins.

Unswept quartz from all three suppliers showed a common tendency for the twin size to increase with increasing threshold irradiance (see Figure 5). The same comparison for swept material (see Figure 6) indicates no relationship between the two parameters for any supplier. This change of the twinning behavior due to sweeping has not been reported before. Figures 7 and 8 compare twin size vs. etch channel density for the Q ranges shown. Figure 7 shows an increase in twin size as etch channel density increases for supplier C only. Figure 8, which compares the same parameters with low-Q data, shows no such relationship for any supplier. With the exception of data for supplier C in Fig. 7, data in both Fig. 7 and 8 indicate no correlation between the two parameters. Comparison of etch channel densities to the twinning irradiances, as shown by Figures 9 and 10, did not show any correlation.

The size and shape of the twins formed on AT- and SC-cuts were also studied. Figure 11(a) shows the predicted ferrobielastic stress diagram for AT-cuts,[1] where $\theta = 35.25^\circ$. AT-cuts formed twins on the surface of the sample, which remained confined to the area under direct irradiation. After repeated irradiations, these twins nucleated distinct lobes corresponding to the preferred twinning directions shown in Figure 11(a) (solid lines). The dashed lines show the anti-twinning directions. The figure was calculated from ferrobielastic stress theory, and is based on the AT-cut calculations of Anderson, et al.[1] They found the angular dependence of ferrobielastic switching stress to be:

$$[3\sin^2\psi - \cos^2\psi\sin^2\theta]\cos^2\psi\sin\theta\cos\theta.$$

Figures 11(b) and 11(c) compare the twins formed on AT-cuts having two distinct regions. One region contained a large amount of etch-channels and the other area was etch-channel free. The shape of the twins in each region show different features. Figure 11(b) shows a twin formed in the area containing no etch-channels. It has straighter boundary lines than the twin formed in the region with etch-channels (Figure 11(c)). The twinned area in Figure 11(c) has jagged boundaries except near the "border" between the two regions. There the twin has straight boundaries, similar to those of the twin in Figure 11(b).

The ferrobielastic stress diagram for the SC-cut is shown in Figure 12(a), where $\phi = 22^\circ$ and $\theta = 33.9^\circ$. Following Anderson, et al.[1], the calculated angular dependence of the ferrobielastic switching stress for the SC-cut was found to be:

$$[3\cos^2\phi\sin^2\psi + 3\sin^2\phi\sin^2\theta\cos^2\psi - 8\cos\phi\sin\psi\sin\phi\sin\theta\cos\psi - \sin^2\phi\sin^2\psi$$

$$- \cos^2\phi\sin^2\theta\cos^2\psi][\sin\phi\sin\psi + \cos\phi\sin\theta\cos\psi]\cos\theta\cos\psi.$$

When irradiated at room temperature (as the AT-cut samples were), the SC-cut showed a tendency to form one dominant lobe which extended to the edge of the sample (Figure 12(b)), and three smaller lobes extending from the center of irradiation. The diameter of this sample is about 14 mm. Figure 12(c) shows the twin nucleated on an SC-cut blank irradiated at a temperature of 400° C. This twin shows complete formation in all lobes, as the stress needed to induce twinning decreases with increasing temperature.[1]

Conclusions

Correlations between α and twin size or α and twin nucleation irradiance were not found for the quartz samples used in this study. This is an important negative result because it does not support ideas currently held in the crystal industry. Taken with the work of Bandyopadhyay and Merkle,[7] who found no correlation between OH levels and laser damage in quartz, this may indicate that factors other than Q_{IR} are involved in the ease with which quartz twins. Further examination of higher Q_{IR} quartz material needs to be done, since the maximum Q_{IR} of the samples in this study measured only 2.6 million.

A correlation between twin size and twin nucleation threshold irradiance for all three suppliers was found for unswept quartz, but not for swept quartz. This has not been reported in the literature. Possibly the interstitial Li and Na atoms in the unswept quartz react to the laser radiation, or distort the lattice, so that the twin sizes were larger for samples with higher nucleation thresholds. When these Li and Na lattice sites in the quartz were replaced with H, as happens during sweeping, the correlation between twin size and twin nucleation threshold irradiance vanishes. Only slight differences between swept and unswept samples are seen in the irradiance vs. α and in the twin size vs. α, so this behavior is difficult to explain. It is possible that the levels of interstitial alkali atoms in quartz affects twinning more than the Q_{IR} which measures OH.

A correlation was also found between twin size and etch channel density for high Q_{IR} samples from supplier C. The twin size of the samples increased with etch channel density. This was not found in samples from the other two suppliers.

Supplier C samples with a Q_{IR} of about 1.8 x 10^6 required an unusual amount of energy before they would twin, producing noticeably large twins which penetrated through the sample. Again, the other two suppliers did not show this behavior at comparable Q_{IR}-values. Low-Q_{IR} samples formed small surface twins, confined to the region under irradiation. This characteristic may make it possible to "label" some quartz with surface twins identifying the material as to type, cut, supplier, etc. Quartz material from supplier C showed greater differences from the other suppliers in irradiance levels and twin sizes formed. This may be due to irregularities in the limited number of samples tested, and not indicative of supplier C material as a whole.

A secondary ion mass spectroscopy (SIMS) scan was performed on a twinned SC-cut sample. Compositional differences between twinned areas and untwinned areas showed no differences. Further work will be done on other samples and comparisons made between suppliers to see if any compositional trends develop.

Acknowledgements

We wish to thank W. Washington and R. Brandmayr for their suggestions and help in the processing and chemical polishing of samples. Thanks are also due R. Lareau and the microanalysis group in ETDL for the SIMS analysis.

Bibliography

[1] Anderson, T.L., Newnham, R.E., Cross, L.E. and Laughner, J.W., Laser-induced Twinning in Quartz", **Phys. Stat. Sol. (a) 37**, 1976, pg. 237.

[2] Ushakovskii, V.T., Kashkurov, K.F., Simonov, A.V., Butorin, P.P., "Artificial Twinning in Synthetic Quartz Crystals", **Soviet Physics - Crystallography**, Vol. 12, No. 6 (June 1968), pg. 954.

[3] McLellan, A.G. "The thermodynamic theory of the growth of Dauphiné twinning in quartz under stress", **J. Phys. C: Solid State Phys.**, Vol. 11, 1978, pg. 4665.

[4] Shiau, S.-M. and Newnham, R.E., "The Effect of Radiation on Twinning in Quartz", **Mat. Res. Bull., Vol. 22**, 1987, pg. 227.

[5] Private communication, J.R. Vig, U.S. Army

LABCOM, and J.A. Kusters, Hewlett-Packard Co., June 1990.

(6) Frondel, C. Dana's System of Mineralogy, 7th ed., Vol. III (New York: James Wiley & Sons, Inc.), 1962, pg. 84.

(7) Bandyopadhyay, P.K. and Merkle, L.D., "Laser-induced Damage in Quartz: A Study of the Influence of Impurities and Defects", **J. Appl. Phys., 63** (5), 1 March 1988, pg. 1392.

(8) Brice, J.C. and Cole, A.M., "The Characterization of Synthetic Quartz by Using Infrared Absorption", **Proc. 32nd Annual Frequency Control Symposium,** 1978, pgs. 1-10.

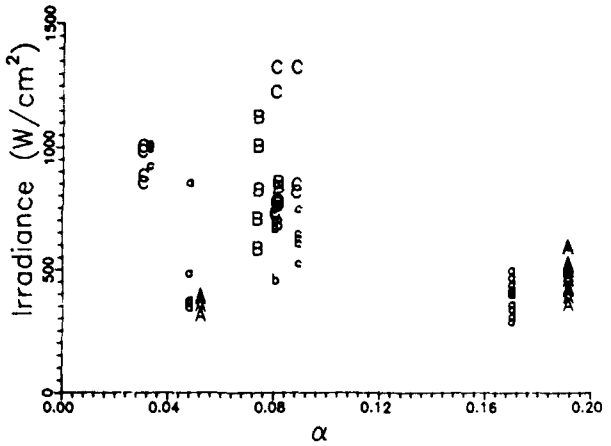


Figure 3. Twin nucleation threshold irradiance vs. α. Uppercase letters indicate unswept material, lowercase is swept material.

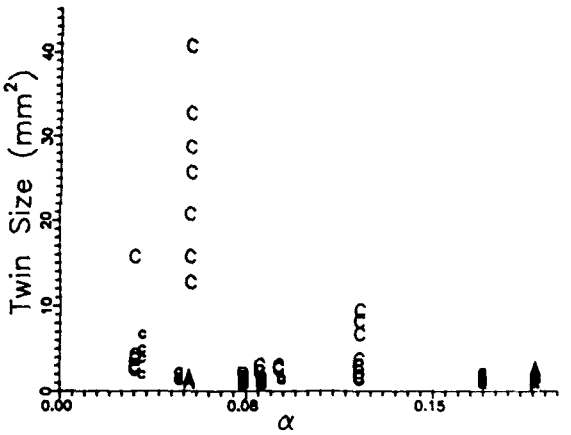


Figure 4. Twin size vs. α.

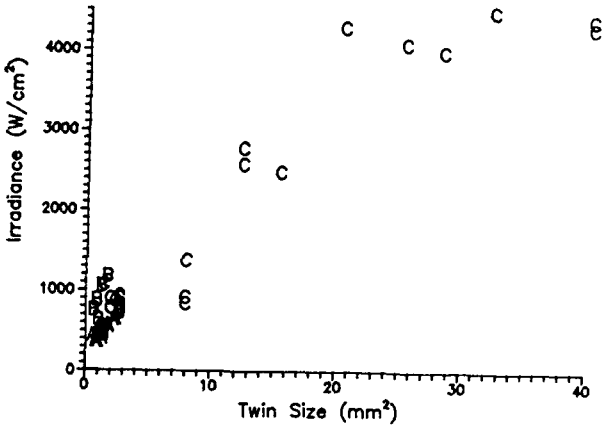


Figure 5. Twin threshold irradiance vs. twin size (unswept material). Q_{IR} varied from 0.63 to 1.9 x 10^6.

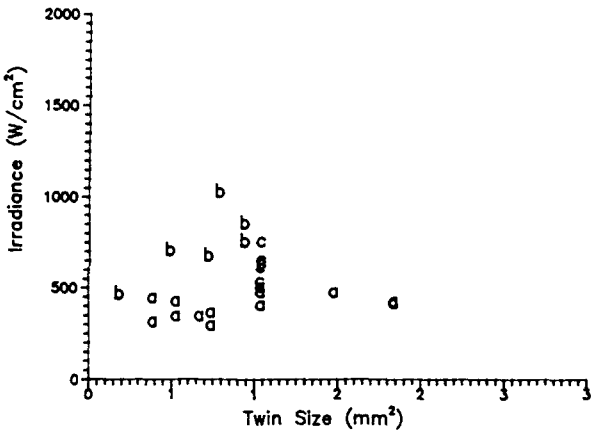


Figure 6. Twin threshold irradiance vs. twin size (swept material). Q_{IR} varied from 0.7 to 1.31 x 10^6.

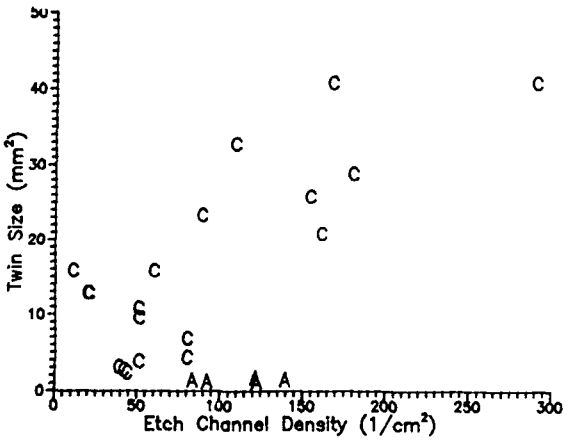


Figure 7. Twin size vs. etch-channel density (unswept material). Q_{IR} varied from 1.8 to 2.6 x 10^6.

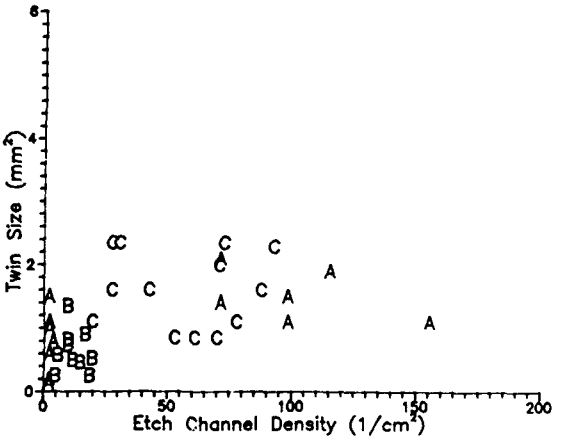


Figure 8. Twin size vs. etch-channel density (unswept material). Q_{IR} varied from 0.63 to 1.4 x 10^6.

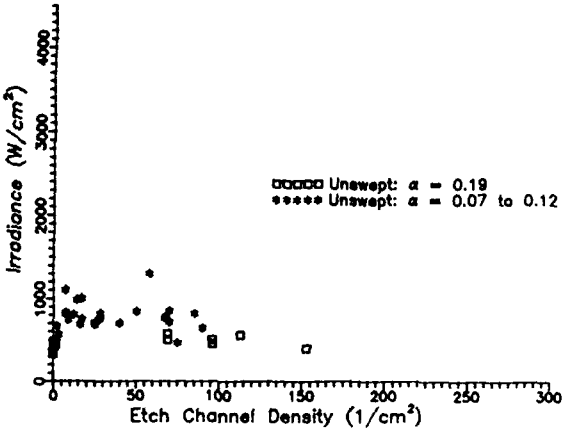


Figure 9. Twin threshold irradiance vs. etch-channel density. Q_{IR} varied from 0.63 to 1.31 x 10^6.

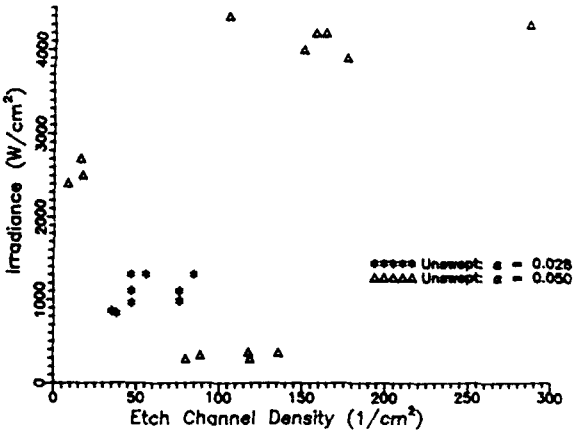
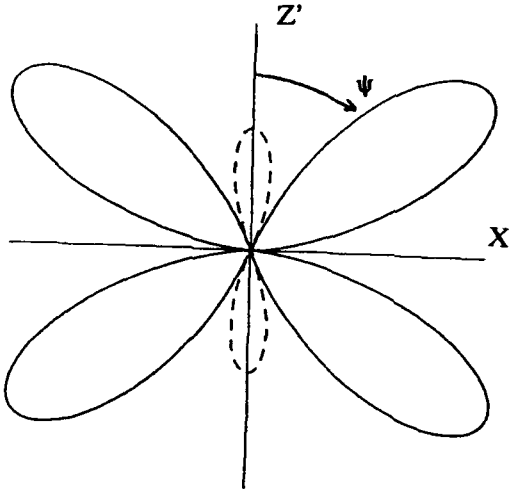
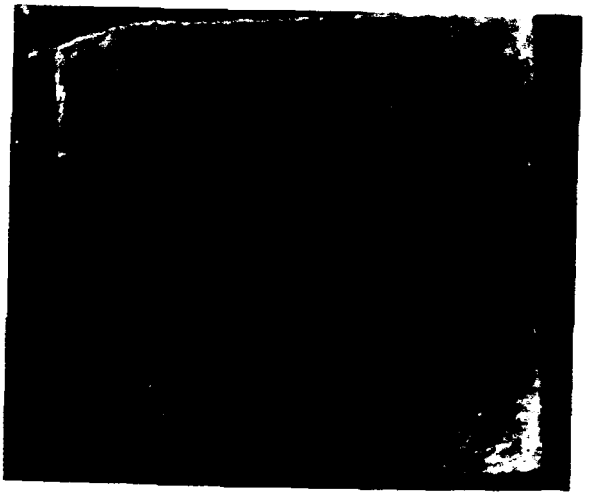


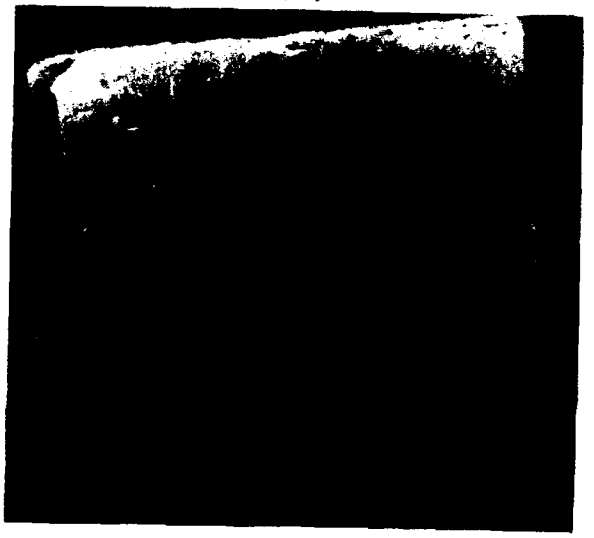
Figure 10. Twin threshold irradiance vs. etch-channel density. Q_{IR} varied from 1.8 to 2.6 x 10^6.



(a)

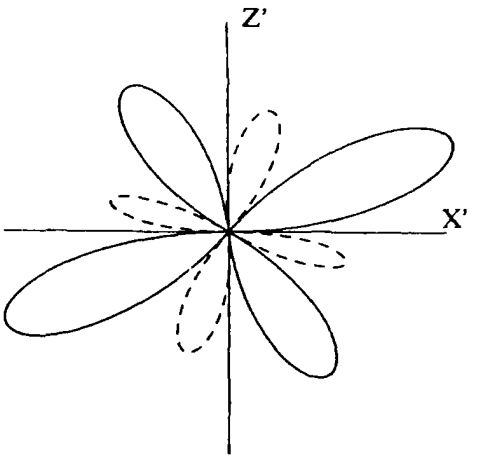


(b)



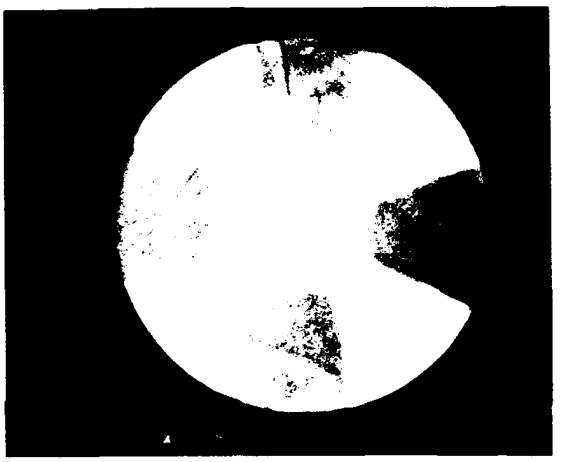
(c)

Figure 11. Twin formation for AT-cut material. (a) Ferrobielastic stress diagram showing preferred twinning directions for the AT-cut. (b) Photograph of an AT-cut plate irradiated in its lower half, which contained etch-channels. Note irregularity of twin except at top, where twin boundary forms angular, straight lines. This area of the plate is the boundary where etch-channels end. (c) Another AT-cut plate with etch-channel distribution as (b). The etch-channel free region was irradiated and the twinned area developed straighter boundaries.



(a)

(b)



(c)

Figure 12. Twin formation for SC-cut material. (a) Ferrobielastic stress diagram showing preferred twinning directions for the SC-cut. (b) SC-cut blank, irradiated at room temperature, shows one lobe of the twin extending to the edge of sample. (c) Different SC-cut blank, irradiated at 400° C, showing a twin better resembling the diagram in (a).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A QUARTZ STANDARD AUTOMATIC FREQUENCY ADJUSTMENT SYSTEM

Marc MOUREY, Serge GALLIOU, Jean-Pascal VALENTIN, Raymond J. BESSON

Ecole Nationale Supérieure de Mécanique et des Microtechniques
La Bouloie - Route de Gray - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

In this paper a system is presented that automatically adjusts the frequency of a quartz standard unit to the frequency of a reference unit. In fact, for some applications, it is necessary to use sources of different nature exactly at the same frequency. As a consequence sources affected by frequency drift must be periodically adjusted. The quartz standard automatic frequency adjustment system (A.F.A.S.) presented here has the following features :

* Automatic starting of adjustment as soon as the signal to be frequency adjusted appears.
* The adjustment value is held between two successive adjustment procedures with a minimum signal degradation.
* Small electronic device easy to add in an ordinary quartz standard unit.

The adjustment is made through a change of the varicap voltage (easy on a quartz standard unit). A phase-locked loop with controlled self-oscillation is used.

The presented A.F.A.S. includes a phasemeter comparing phases between the reference signal and the signal to be frequency adjusted. The output of the phasemeter is shaped so as to trigger a specific numerical block which works as a "without-zero integrator". The without-zero integrator also insures the storage of adjustment as soon as frequencies are close enough. End of frequency adjustment procedure is determined by use of self-oscillation counting.

In this paper, the previous procedures and the schematics used are described in details especially the phase-locked loop system and the self-oscillation.

Results obtained are as follows (at 5 MHz)

frequency difference range : $\Delta f/f = 2\ 10^{-6}$

residual frequency difference after adjustment : $4\ 10^{-12}$.

It is eventually possible to start with frequency differences as large as 10^{-5} but the final resolution then drops down to $3\ 10^{-10}$.

This A.F.A.S. has been successfully used to adjust quartz standard units to a cesium reference.

INTRODUCTION

Time measuring systems or navigational systems often use several time bases which must have sources at the same frequency. Consequently, frequency drifts of such sources must be periodically corrected.

This paper describes a circuit designed to adjust the frequency of a quartz standard unit to the reference unit's.

The main requirements of the automatic frequency adjustment system (AFAS) are first reviewed. In the following sections the built system is described. Some theoretical investigations and a few measurements are presented.

1 - REQUIREMENTS

Sources to be frequency adjusted are quartz oscillators which usually provide an external tuning voltage for that purpose. The automatic frequency adjustment system (A.F.A.S.) uses the monitoring voltage by the means of a well-known phase-locked loop principle. Nevertheless the described solution is closely related to following specific requirements :

- The oscillator controlling voltage must be locked between two successive adjustments

CH2818-3/90/0000-259 $1.00 © 1990 IEEE

within a range of 0 to 10 Volts.

- The oscillator frequency stability must not be perturbed by the A.F.A.S. except during adjustment process. This is particularly important because the A.F.A.S. is to be built in the quartz oscillator unit (and consequently the unit and the A.F.A.S. must have a single supply).

- When pluging in the reference unit, the adjustment process must start automatically.

- The absolute frequency adjustment has to be made with the best accuracy.

Only a few solutions fit these requirements. The main goal involves the use of a control loop with a phase comparator generating the error signal. This last one is the only component able to guarantee a finite (or nul) phase difference yielding frequencies as close as possible.

A presence detector starts up the adjustment process. At the end of that process the tuning voltage is held on without drift by the means of a numerical value. Moreover, in order not to disturb the oscillator output frequency when the reference is disconnected at the end of the process, all other signals vanish.

2 - A.F.A.S. OPERATING

Fig. 1 shows the block diagram of the reference unit connected to the quartz oscillator through the A.F.A.S. including a phase comparator (mixer and low-pass filter) and the processing electronic described below.

2.1/ The phase comparator

Considering noise performances and accuracy an analogical phase comparator is choosen. This is better than numerical solutions for example. When associating it with a proper low-pass filter, its output voltage is almost a linear function of the total phase difference between $[-\pi, 0]$ or $[0, \pi]$ (see Fig. 2). In both intervals the static gain is the same and only the sign changes. That means the looped system will find the interval where the stability condition are achieved.

2.2/ The processing electronic

According to the processing diagram (Fig. 3), as soon as the reference signal is present a coarse adjustment is first triggered. The end of this coarse adjustment starts up a fine adjustment around the retained value of the former. When both measured frequencies are close enough (this will be discussed below) the fine adjustment is stopped. Its value is retained and the supply voltage of unused components is switched off in order to inhibit spurious frequencies.

This simplified sequence may be detailed from the functional diagram (Fig. 4). The reference signal is shaped to trigger a retriggerable monostable multivibrator whose output enables the system working after a general reset (see part 1 Fig. 4). From this time, two counters (Ct1 and Ct2) count "up" or "down" according to the state ot the phase-comparator output. The outputs of each counter are translated into a voltage. One of these analog quantities is divided by a proper value before being added to the other (see parts 4 and 5 Fig. 4). The resulting voltage monitors the V.C.O., i.e. the unit to be frequency adjusted. In such a way, one of the analog quantity to be added to the other is a ratio of that last one. So, both counters are presetted at the same value (zero, for example) without any drawback on the system accuracy.

During all the transient time both counters outputs are equal. When the V.C.O. global phase reaches the reference global phase, a sign change occurs on the phase-comparator output, involving a "up/down" change of the counters. Then, oscillations appear around the zero-phase difference because numerical outputs may only change step by step (this is described in Fig. 5). The number of transitions is recorded by the counter Ct5 which stops the counter Ct1 after five transitions (experiments prove that it is a proper value to be sure the phase is locked around the reference one) while the counter Ct2 is always working. The coarse adjustment value is then locked (DAC1) whereas the fine adjustment voltage becomes free over a range equal to a ratio of the DAC2 full-scale. Five other transitions are still counted by Ct10 before locking Ct 2 that it is to say the fine adjustment value. At the end of the processing sequence many components such as the counter-clock generator, are inhibited.

2.3/ Resolution

The system resolution is obviously related to the counting capability versus the maximum range of frequency adjustment. If the coarse adjustment counter Ct1 is a n_1-bit counter and the fine adjustment counter Ct2 a n_2-bit counter, the resolution may be expressed as :

$$R_c = \frac{(\Delta f_{max}/f) \times N}{2^{n_1} \times 2^{n_2}}$$

where N is the number of coarse-adjustment steps within the fine-adjustment full-scale. We have choosen $n_1 = n_2 = 12$ and a frequency adjustment range of $\Delta f_{max} = 2$ Hz at f = 5 MHz. In this case, the maximum resolution will be $R_c \simeq 2{,}5\ 10^{-14}$ when N = 1. In fact, this calculation is only available for the counting set. In our system with a feedback-loop we already said oscillations take place. So, the system accuracy depends on the oscillation magnitude which may be identified to global phase difference, that is frequency difference.

System oscillations are analysed below.

3 - ANALYSIS OF THE SYSTEM OSCILLATIONS

Fig. 6 shows a block diagram of the feedback control system.

- The mixer and the low-pass filter are gathered in the error detector. When considering a reference signal :

$$E_R = E_{Ro} \sin(\omega_R t + \alpha) = E_{Ro} \sin \theta_R$$

and the VCO output signal :

$$E_S = E_{So} \sin \omega_S t = E_{So} \sin \theta_S,$$

the error detector output has the form :

$$S_p = E(\cos \theta/2 - \sin \theta/2) \text{ where } \theta = \theta_R - \theta_S.$$

Near $\theta = \pm \pi/2$, this may be linearized, thus $\Delta S_p = \pm K_p \Delta\theta$.

- The error detector is followed by a Schmitt trigger which exhibits non linearity because of its hysteresis h. Between mixer and trigger there is an amplifier (which also cuts off high frequencies) and the trigger output is connected to the "up/down" input ot the counters. All these functions may be simplified (see Fig. 7).

Neglecting the higher harmonics (a low pass filter exists in the closed-loop system) the describing function technique may be used. So, if $\Delta\theta = \theta_1 \sin \omega t$ the describing function becomes :

$$\text{For } \theta_1 > \frac{\theta_h}{2}, \quad T(\theta_1) = B(\theta_1)\, e^{j\Psi(\theta_1)}$$

$$\text{with } B(\theta_1) = \frac{4}{\pi\theta_1}$$

$$\text{and } \Psi(\theta_1) = -\text{Arctg} \frac{\theta_h/2}{\theta_1 \sqrt{1 - (\frac{\theta_h}{2\theta_1})^2}}$$

$$(T(\theta_1) = 0 \text{ for } \theta_1 < \theta_h/2$$

h : hysteresis width of the trigger.

θ_1 : phase-difference magnitude.

- The trigger output controls the "up" counting or the "down" counting of a constant frequency F. During the steady state, this means counters integrate the self oscillation signal at a periode of $T = 2\pi/\omega$ with a gain of F. If ΔN is the integer number change of their outputs, that is :

$$\Delta N = \text{Int} \left[F \int W(t)\, dt \right] \quad (F >> 1/T).$$

So the transfer function of such components is $C(p) = K_c/p$ ($K_c = F$) where $p = j\omega$ is the Laplace variable (assuming discrete functions are close to linear functions because of the large capability of the counters and the high value of F).

- Digital to analog converters may be described from : $\Delta U = K_D \Delta N$ with $K_D = U_{Max}/2^n$ for a n-bit to 0-U_{Max} full-scale convertion.

- Experiments proved the V.C.O. transfer function may be written :

$$\frac{\Delta\theta_S}{\Delta U} = \frac{K_V}{p(1 + \tau p)^2}$$

The product of all these transfer functions gives the open-loop transfer function

$$F(j\omega) = T(\theta_1) \cdot \frac{K_p K_c K_D K_V}{(j\omega)^2 (1 + \tau j\omega)^2}$$

$$= R(\theta_1, \omega) + j\, I(\theta_1, \omega)$$

Then the self-oscillation conditions, that is $R(\theta_1, \omega) = 1$ and $I(\theta_1, \omega) = 0$, provide the phase-difference magnitude θ_1 (i.e. the system accuracy) versus the oscillation frequency $f = \omega/2\pi$ for different hysteresis width h.

Fig. 8, Fig. 9 and Fig. 10 sum up those theoretical results.

One component has not be mentioned until now. It is a monostable multivibrator (see Fig. 4) which improves the result in delaying the fine adjustment locking of one quater of the self-oscillation period. In such a way, the final frequency is a mean value of the peak to peak magnitude of the self-oscillation.

4 - EXPERIMENTAL RESULTS

The A.F.A.S. has been tested using a Cesium beam frequency standard as a reference unit.

Table 1 shows the final deviation of both compared frequencies when the unit to be frequency adjusted is a precision quartz crystal oscillator. It depends on the initial frequency difference. The same conclusion may be driven when the A.F.A.S. locks a synthesizer output frequency to the reference, as shown in Table 2.

Initial frequency offset	Final frequency difference
2.10^{-7}	7.5×10^{-12}
6.10^{-8}	5×10^{-12}
2.10^{-9}	4×10^{-12}
5.10^{-10}	4×10^{-12}

Table 1

Initial frequency offset	Final frequency difference
2.10^{-6}	3.10^{-10}
4.10^{-7}	2.10^{-10}
10^{-7}	3.10^{-10}

Table 2

It turns out, that the A.F.A.S. which has been described does not anymore work properly, when both frequencies are initially too far away (in practise 10^{-5}). In such a case it is convenient to use an error detector combining phase and frequency differences (see Fig. 11) but a fine adjustment is no more possible.

CONCLUSION

Our goal performances have been achieved using the described A.F.A.S. The circuit may now be turned into a specific integrated circuit built-in the quartz Xtal oscillator unit.

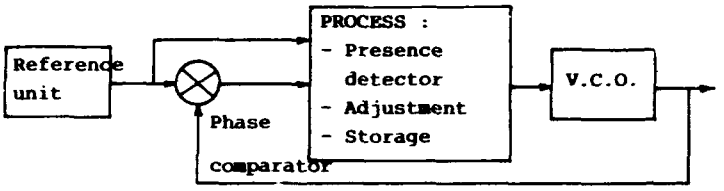


Fig. 1 : Block diagram showing the V.C.O. to be frequency adjusted within the reference frequency.

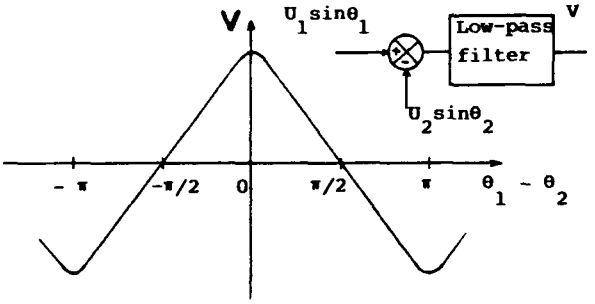


Fig. 2 : Typical response of an analogical phase comparator (mixer + low-pass filter)

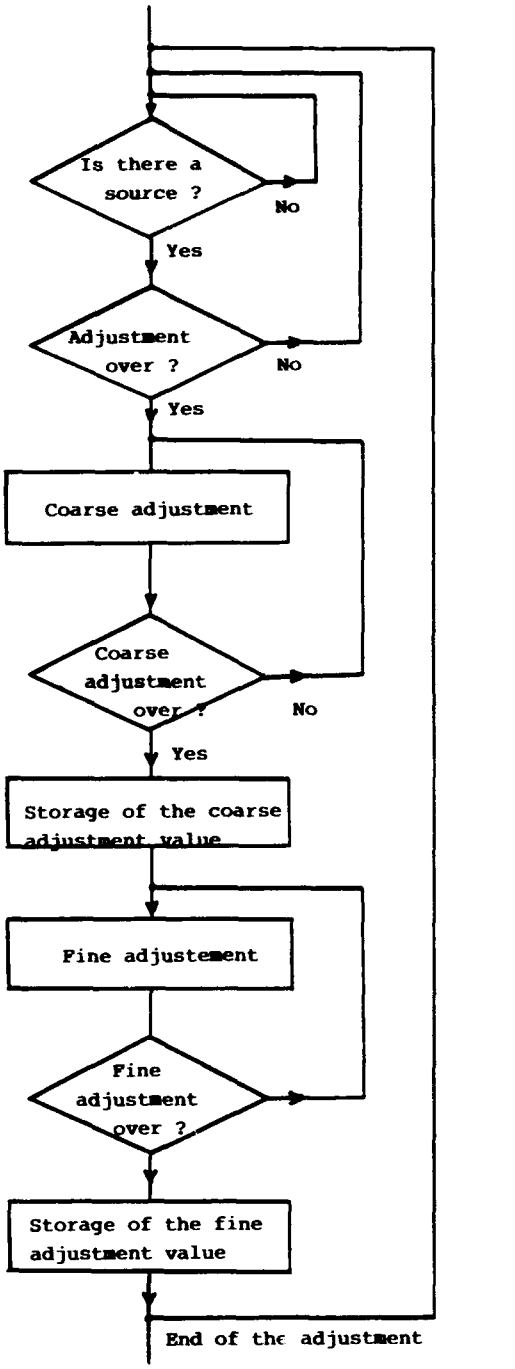


Fig. 3 : The A.F.A.S. processing diagram.

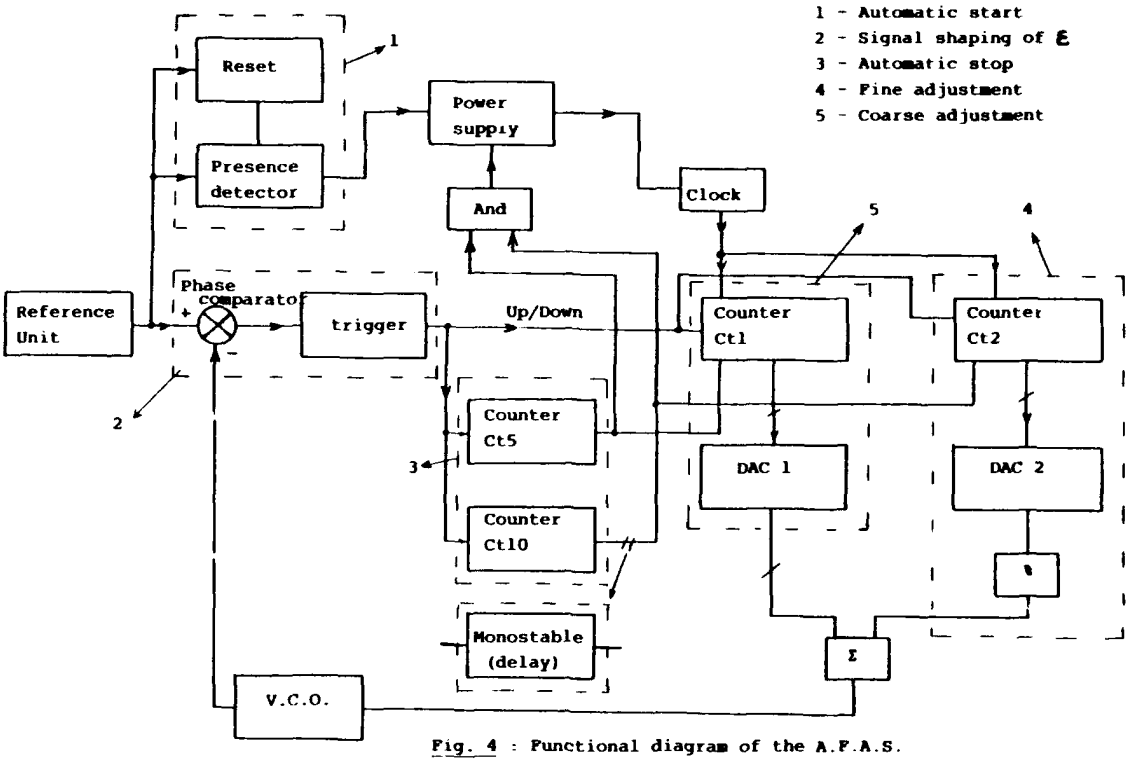


Fig. 4 : Functional diagram of the A.F.A.S.

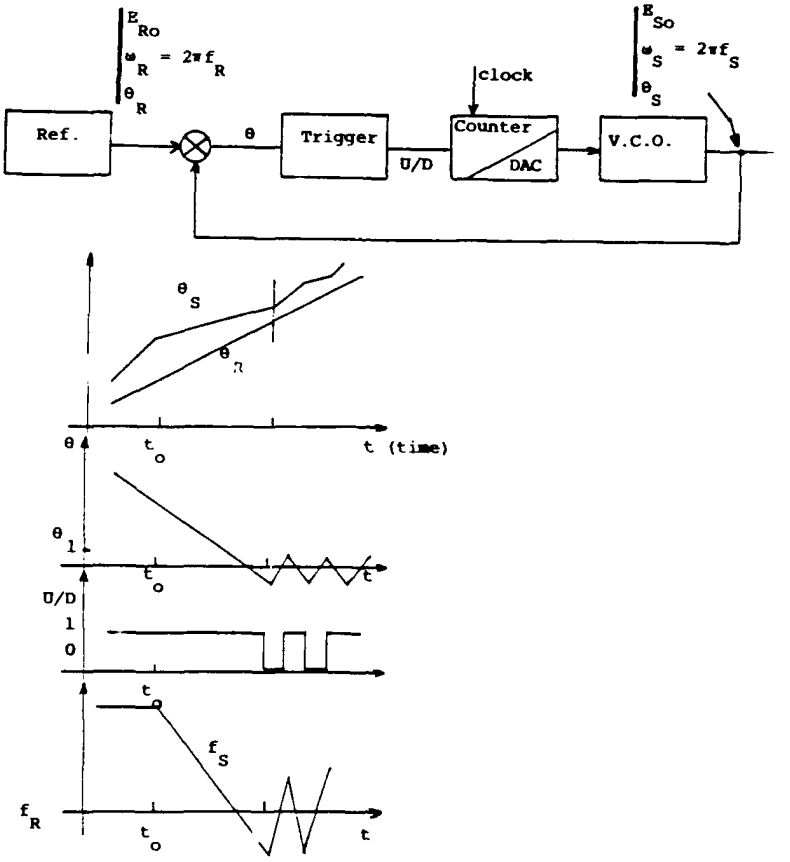


Fig. 5 : During the steady state oscillations occur

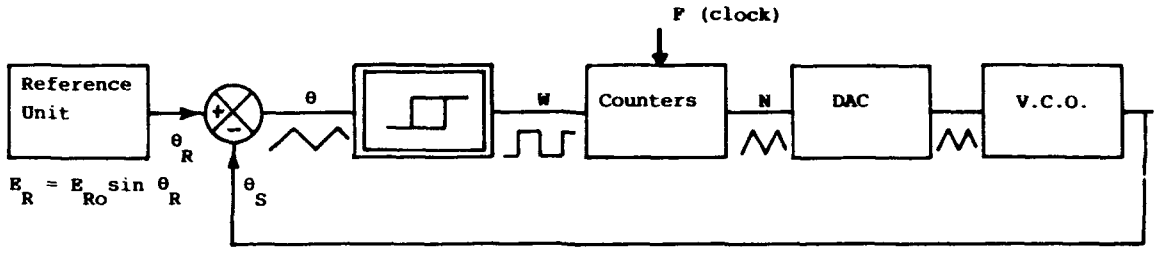


Fig. 6 : The closed loop system.

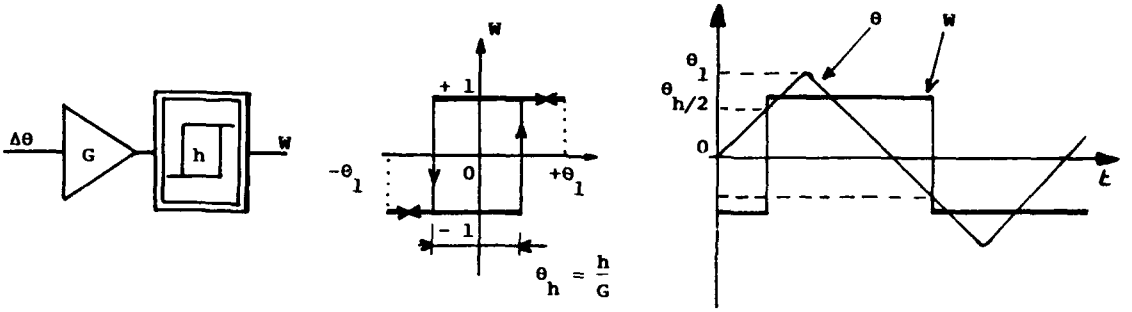


Fig. 7 : When considering the fundamental component of the trigger output, the trigger is a phasor.

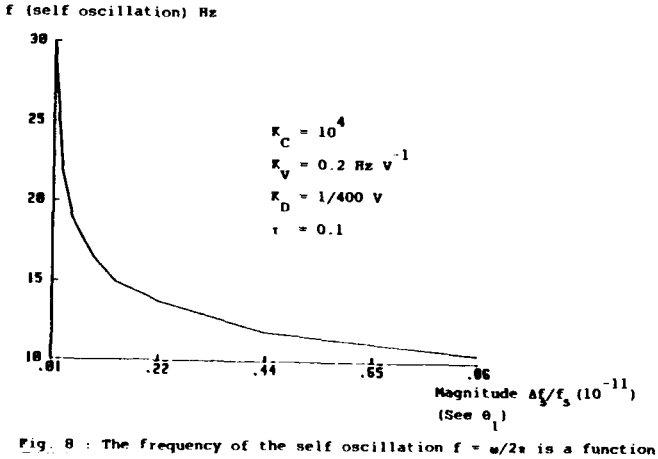


Fig. 8 : The frequency of the self oscillation f = ω/2π is a function of the instantaneous frequency change magnitude (5 MHz).

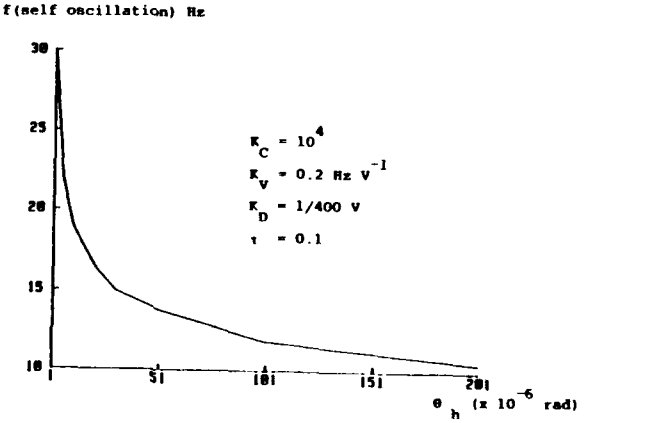


Fig. 9 : Self oscillation frequency changes versus the hysteresis width (i.e. θ_h).

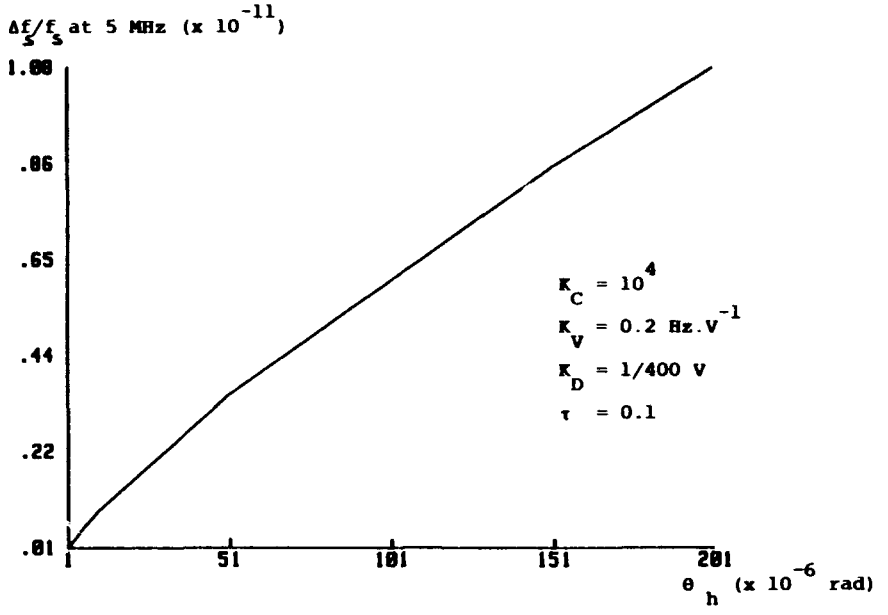


Fig. 10 : Peak magnitude of the frequency change (5 MHz) versus the hysteresis width.

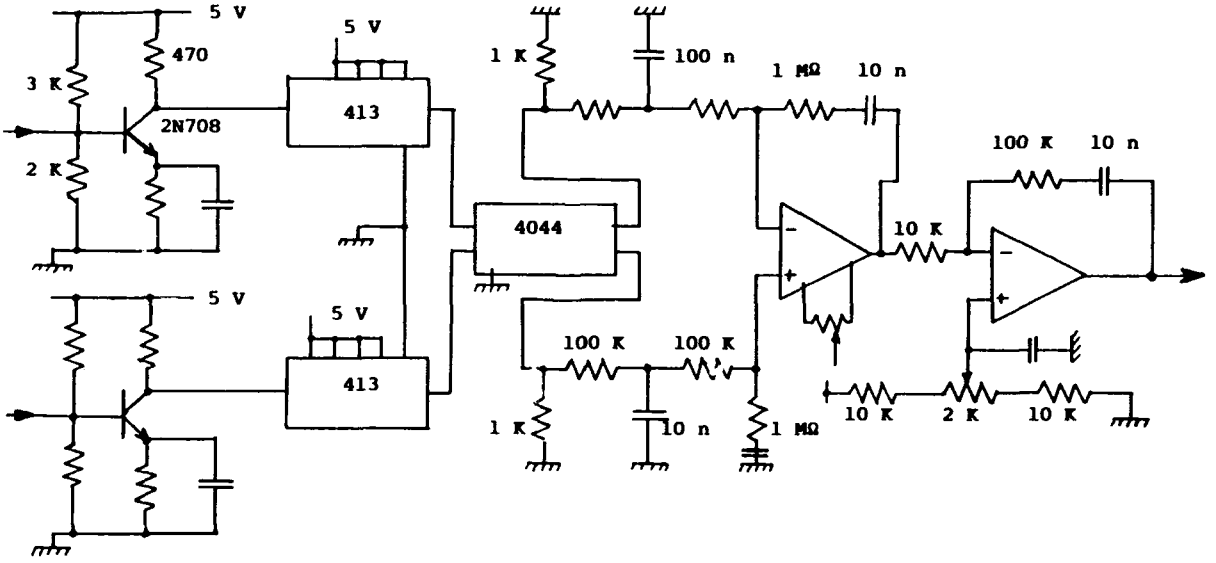


Fig. 11 : The A.F.A.S. phase comparator for oscillator having a frequency at more than 10^{-5} from that of the reference.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FABRICATION CONSIDERATIONS FOR A COUPLED-TRIPLE RESONATOR CRYSTAL

G. E. Roberts, A. G. Staples, S. Toliver

Ericsson GE Mobile Communications, Inc.
Mountain View Road
Lynchburg, Virginia 24502

ABSTRACT - With the increased tendency to use mobile and cellular radio systems to transmit data, requirements on the maximum value and flatness of the group delay in the passband as well as attenuation characteristics in IF filters are becoming more stringent. Therefore, it is becoming very important to be able to design the radio IF filters from the most efficient characteristics possible. This means using bandpass characteristics derived from odd order lowpass prototype characteristics as well as even order. Therefore, the use of a coupled-triple resonator crystal in conjunction with coupled-dual crystals appears to be one possible option.

In this paper, one way of fabricating a coupled-triple resonator crystal is presented. The three resonators are electrically displayed via a lead frame and substrate to be able to monitor their frequencies throughout the fabrication process as well as the coupling between resonators. A specific filter design is presented.

Introduction

Since the crystal filters in most mobile radio applications use coupled-dual resonator crystals, the resulting filters are limited to bandpass realizations derived from even order lowpass prototypes. These realizations are obtained by cascading the coupled-dual crystals and coupling adjacent crystals with an equivalent capacitance to ground. When considering a design relative to a given set of passband and stopband specifications, a 3-pole Chebyshev bandpass characteristic (from a 3rd order lowpass prototype producing a 6-order bandpass characteristic) is often called for instead of a "c-case" 4th order Chebyshev lowpass prototype or even a "b-case" 4th order Chebyshev lowpass prototype [1] even when manufacturing and temperature tolerances of the crystal are factored into the design considerations. Since only coupled-dual crystals are used in the actual filter realizations, the resulting filter is derived from one of the 4th order Chebyshev lowpass prototypes producing a bandpass characteristic of 8th order which is two orders more than required. A similar situation often occurs when a 5-pole Chebyshev bandpass filter could be used instead of a 6-pole filter.

With the increased tendency to use mobile and cellular radio systems to transmit and receive data, requirements on the maximum value and the flatness of receiver IF filter's group delay in the passband as well as attenuation are becoming more stringent. As a result, it is becoming very important to be able to design the radio IF filters to have the most efficient characteristics possible. This means using bandpass characteristics derived from odd order lowpass prototype characteristics as well as even order lowpass characteristics in order to realize a IF crystal filter as the requirements dictate. Therefore, the use of a coupled-triple resonator crystal in conjunction with coupled-dual crystals appears to be one possible option.

However, the initial problem and perhaps the most critical in developing a coupled-triple resonator crystal is how to measure the basic electrical parameters of the crystal such as resonator frequencies and coupling during the fabrication. Since the work described in this paper was begun well before the development of the 4-frequency measurement method [2,3], only the neutralizing inductance method of measurement [4,5] and tuning frequency method presented by Rennick [6] was seriously considered. The neutralizing inductance method was discarded for the same reason cited by Rennick as it is undesirable to use in the fabrication process because measurements must be made in a vapor plating machine. Therefore, the tuning frequency method was chosen with the fabrication process and packaging developed around this measurement method.

Triple resonator theory

Following Rennick, a basic coupled-triple resonator crystal is shown in Fig. 1 with its equivalent circuit shown in Fig. 2 with the behavior of the capacitive-T circuit shown in Fig. 3 in the case of a single crystal equivalent circuit. Since the Q of each resonator is large, it is assumed here that the resistance associated with such resonators can be neglected. An important feature of a coupled-triple resonator is that if the middle terminal pair is shorted as in Fig. 4A, then a nodal equivalent circuit is formed which is in ladder network form as shown in Fig.4B. This nodal equivalent circuit can be transformed to a mesh circuit by simply transforming the elements through either tee of capacitors associated with the end resonators using

$$Z_{new} = 1/(\omega^2 C^2 Z), \qquad (1)$$

where C is the capacitance of the capacitive-T

CH2818-3/90/0000-267 $1.00 © 1990 IEEE

network and Z is the impedance being transformed through the capacitive-T network. Then the $1:L_1/L_3$ transformer is eliminated by setting $L_1=L_3=L$ and all three resonators are made the same size by setting $L_2=L$ also. As in Rennick, the transformation can be seen by considering Fig. 3 by evaluating Eq. (1) at $\omega^2 = \omega_0{}^2 = 1/(LC)$ where L and C are the parallel inductance and capacitance attached to node (A) in Fig. 3A. This transformation is discussed at length by Rennick. The resulting negative capacitances in each mesh in Fig. 5A are combined with the positive capacitances in the same mesh the three positive capacitance C_1, C_2 and C_3 are equated to C for convenience. The resulting network is shown in Fig. 5B where

$$C_{nj} = C_1 M_{nj}/L_1.$$

The basic tuning frequency measurement techniques presented by Rennick amounts to proper handling of a normalized coefficient matrix obtained from the circuit analysis of the particular coupled-resonator equivalent circuit under the condition that the resonator frequencies are in error. In the case of the coupled-triple resonator, this matrix is:

$$\begin{bmatrix} (F-E_1) & K_{12} & 0 \\ K12 & (F-E_2) & K_{23} \\ 0 & K_{23} & (F-E_3) \end{bmatrix} \qquad (2)$$

where $F = (\omega^2 - \omega_0{}^2)/(2\pi \cdot \omega \cdot BW)$, (3)

BW = bandwidth of the filter in Hz,

f_0 = center frequency of the filter,

$K_{mn} = (C\ f_0)/(C_{mn}\ BW)$, (4)
in the narrow band of frequencies around f_0, where $f_0/f \doteq 1$,

C = the positive capacitance in the series branches in Fig. 4A.

C_{mn} = the shunt coupling capacitance between adjacent meshes in Fig. 4B.

The error function, E_i. associated with the i-th loop or i-th resonator in a coupled-triple resonator crystal is generated by changing the impedance across port A, B, or C in Fig.2 from a simple short-circuit condition. If the i-th resonator or port is open-circuited, the capacitance C_{0i} appears it and the error function becomes

$$E_i = (f_0 C_i)/(BW\ C_{0i}). \qquad (5)$$

In the case considered here, $C_i = C$ and $C_{0i} = C_0$.

As pointed out in [6], each mesh in Fig. 5B is required to be resonant at f_0 when all other meshes are open-circuited. In the physical triple-resonator crystal, the resonators cannot be isolated since each one is coupled to the other two on the plate. If one resonator is to the measured physically, the other two then influences the measurement of that resonator. Similarly, a measure of coupling between any two adjacent meshes of Fig. 5B could be determined by opening the third mesh and finding the two short-circuit frequencies of the two adjacent meshes. Rennick defines the difference between these frequencies as coupling. When it is divided by bandwidth, he calls it normalized coupling coefficient or apparent coupling as in

$$K_{mn} = (f_2 - f_1)/(BW)$$

where K_{mn} is the same as that given in Eq. (4).

Likewise, it is pointed out in [6] that the turning frequency of any resonator occurs at a zero of the input impedance to the resonator in question when the other two resonators are open-circuited and thereby mistuned by the C_0. This input impedance may be expressed as

$$Z = H \frac{(\text{short-circuit natural frequencies})}{(\text{open-circuit natural frequencies})} \qquad (8)$$

In our measurement of tuning frequency, there will be three zeros of this input impedance, but two of these will be far removed from the others due to mistuning of the resonators not being measured. The short-circuit natural frequencies of the systems are simply the zeros of the determinant of the matrix given in Expression (2) with all the $E_i = 0$.

To determine the tuning frequency of the first resonator (resonator A)

$$E_1 = 0 \qquad (10a)$$

$$E_m = (f_0 C_m)/BW\ C_0) \qquad (10b)$$

where m is not equal to 1,or

$$E_2 = (f_0 C)/(BW\ C_0) \qquad (10c)$$

and

$$E_3 = (f_0 C)/(BW\ C_0) \qquad (10d)$$

since here $C_2 = C_3$.

The three zeros of interest are obtained by substituting these values of E_i into the matrix in Expression (2) and determining the zeros of the polynomial derived from the determinant of the resulting matrix. This determinant is given by

$$\begin{vmatrix} F & K_{12} & 0 \\ K_{12} & [F-(f_0/BW)(C/C_0)] & K_{23} \\ 0 & K_{23} & [F-(f_0/BW)(C/C_0)] \end{vmatrix} \qquad (11)$$

The zero at the lowest frequency is tuning frequency of the first resonator expressed in terms of F. The resulting frequency which should be measured is then

$$f_m = f_0 + (F_m\ BW)/2. \tag{12}$$

To determine the apparent coupling between resonator one and resonator two, short resonator two also and leave resonator three open, then the determinant equation becomes

$$\begin{vmatrix} F & K_{12} & 0 \\ K_{12} & F & K_{23} \\ 0 & K_{23} & [F-(f_0/BW)(C/C_0)] \end{vmatrix} = 0 \tag{13}$$

Then K_{12} will be the difference between the two lowest frequencies divided by BW or

$$K_{12} = (f_2 - f_1)/BW. \tag{14}$$

The tuning frequencies and apparent coupling coefficients as calculated above are those that should be measured on a good coupled-triple resonator crystal where all of the frequencies and couplings are correct to begin with. If this is not the case, an appropriate plating algorithm will produce the correct frequencies and apparent couplings.

When designing a coupled-resonator crystal filter from a lowpass prototype having normalized element values X_1, X_2, and X_3 for a 3-pole crystal filter, the values of K_{mn} for the above calculations are given by

$$K_{mn}^2 = 1/(X_m X_n), \tag{15}$$

where m = 1,2, and n = 2,3.

We are now in a position to summarize the equations that are to be used for a coupled-triple resonator crystal. Starting from a 3rd order lowpass prototype, Eq. (15) gives

$$K_{12}^2 = 1/(X_1X_2) \tag{16a}$$

$$K_{23}^2 = 1/(X_2X_3) \tag{16b}$$

If all three ports of the coupled-triple resonator are open-circuited as in Fig. 2, the system determinant equation becomes

$$\begin{vmatrix} (F-E_1) & K_{12} & 0 \\ K_{12} & (F-E_2) & K_{23} \\ 0 & K_{23} & (F-E_3) \end{vmatrix} = 0 \tag{17}$$

and the resulting cubic equation in F becomes

$$F^3 + A_1F^2 + A_2F + A_3 = 0$$

where $A_1 = -(E_1+E_2+E_3)$,

$$A_2 = E_1E_2 + E_1E_3 + E_2E_3 - K_{12}^2 - K_{23}^2,$$

$$A_3 = E_1E_2E_3 + E_1K_{23}^2 + E_3K_{12}^2.$$

The simplest case to consider is the coupled-triple resonator crystal with no mistuning at the ends to absorb the C_0 that occurs across those resonators as well as across the middle resonator. The resulting tuning frequencies are calculated as follows, where ATF is the A-resonator tuning frequency, BTF is the B-resonator tuning frequency, and CTF is the C-resonator tuning frequency.

ATF: $E_1=0$
$E_2=(C\ f_0)/(C_0\ BW)$ Terminals B B' open,
$E_3=(C\ f_0)/(C_0\ BW)$ Terminals C C' open,

BTF: $E_1=(C\ f_0)/(C_0\ BW)$ Terminals A A' open,
$E_2=0$,
$E_3=(C\ f_0)/(C_0\ BW)$ Terminals C C' open,

CTF: $E_1=(C\ f_0)/(C_0\ BW)$ Terminals A A' open,
$E_2=(C\ f_0)/(C_0\ BW)$ Terminals B B' open,
$E_3=0$.

Note that the resonator at which the tuning frequency is determined or measured is short-circuited by virtue of the fact that a generator and resistor is placed across it to measure the driving-point impedance. Of course, in each situation above, the values of E_m are substituted into Eq. (17) and the three values of F are found to be F_i where i=1,2,3. To find the actual frequencies, they are substituted into

$$f_i = f_0 + (F_i\ BW)/2, \tag{18}$$

where i=1,2,3. Also note that the lowest value of frequency is the tuning frequency in question. Of course, the value of bandwidth is known from the design information if the values of tuning frequencies are to be calculated. If the coupled-triple resonator is measured, frequencies come out directly.

The apparent coupling between resonators is then determined by shorting two adjacent resonators and opening the other and determining the two short-circuit frequencies which are the two frequencies having lowest value.

Apparent coupling between A and B is determined by substituting

$$E_1 = E_2 = 0, \tag{19a}$$

and

$$E_3 = (C\ f_0)/(C_0\ BW), \tag{19b}$$

into Eq. (17), determining the three values of F, taking the two lowest values F_1 and F_2, taking their difference and K_{12} is calculated as

$$K_{12} = (F_2 - F_1)/BW. \tag{20}$$

The actual frequencies are calculated by Eq. (18).

The apparent coupling between B and C is determined by substituting

$$E_1 = (C\ f_0)/(C_0\ BW), \tag{21a}$$

and

$$E_2 = E_3 = 0, \tag{21b}$$

into Eq. (17) and following the same procedure followed in determining the apparent coupling between A and B.

Of course, to determine the short-circuit natural frequencies of the coupled-triple resonator, Eq. (17) is used with

$$E_1 = E_2 = E_3 = 0,$$

and the frequencies are determined by use of Eq. (18).

Coupled-triple resonator crystal fabrication

The basic consideration in deciding how to fabricate a coupled-triple resonator crystal is how to provide separate electrical access to each of the three resonators as required to measure the resonator tuning frequencies, the apparent coupling between adjacent resonators, and the short-circuit natural frequencies. Therefore, all three terminal pairs in Fig. 2 associated with the resonators must be displayed such that short-circuits, open-circuits or any other predetermined electrical circuit may be selectively connected across the resonator terminal pairs.

The coupled-triple resonator arrangement which will be discussed here was devised by Roberts and Staples [7] and is shown in Figs. 7 and 8. It consists of three acoustically coupled resonators fabricated on a single wafer of quartz. This wafer is physically supported and spaced from a ceramic connection interface substrate by six leads which are formed from a resilient electrical conductor material (e.g., beryllium copper) in order to connect the conductor pattern on the crystal wafer to the conductor pattern on the ceramic substrate as shown in Fig. 7. The conductor pattern on the ceramic substrate is designed to provide separate electrical access to each of the three resonators.

Once the baseplated crystal wafer is attached to the six leads and bonded via conductive cement, for example, the six leads are inserted into the holes in the substrate to the appropriate stop provided by the bend in the formed leads as shown in Fig. 6B. The leads are then cemented to the conductor pattern on the opposite side of the substrate. All three resonator terminal pairs then can be displayed via pads on the substrate.

The terminals associated with the center electrode pair are displayed via pads X and Y in Figs. 7 and 8. The two outside crystal electrode pairs are displayed between pads W and Y, and pads Z and Y. The substrate is then attached to the three leads of a crystal base by attaching pads W and Z which are attached to the two outside pins (forks) of the base while pad Y is attached to the center fork of the base. Therefore, two outside crystal resonators are displayed electrically between their respective outside base leads and ground at the center pin of the base. The center resonator is displayed electrically by probing pad X on the substrate and to the center pin of the base or pad Y.

When the resonators are on their desired frequencies (depending on the finishing process), the crystal middle electrode pair - middle resonator - may be short-circuited without disturbing the crystal itself by connecting pads X and Y on the back of the substrate usually with conducting cement. After the resulting crystal-substrate-base assembly is sealed as in Figs. 7 and 8, the coupled-triple resonator will function in accordance with the connections in Fig. 4A and the equivalent circuit in Fig. 4B.

There are several interesting processes that go into the actual fabrication of the crystal-substrate assembly. The leads which attach the crystal to the substrate were etched out of a sheet of beryllium copper in sets of three leads which are an integral part of a connecting "palm" in a lead frame shown in Fig. 6A. The formation of the bent lead wires is effected using conventional metal forming techniques.

The shape of the individual leads on the lead frame in Fig. 6A are depicted in detail in Figs. 6B and 6C including the first and second bends. In addition, a weakened pre-defined breakpoint is provided along the shank of each individual lead. After the distal ends of the bent lead wires are cemented to the appropriate conductor points on the crystal wafer, the frame palm is broken away from the individual lead wires. A fixture was developed to hold the leads in place which are attached to the crystal wafer in order to protect the wafer while detaching the palm. This then leaves three leads attached to each side of the wafer as in Fig. 7A but without the palms. The resulting six leads are precisely spaced to be inserted into the appropriate "holes" of the ceramic substrate where they are cemented to the conductor pattern on the back side of the substrate. At this point, the substrate serves as a carrier for the crystal wafer and an interconnect device. In addition, the bent leads appear to be extremely good shock absorbers for the crystal wafer.

After the crystal-substrate assembly is attached to the crystal base as in Fig. 7. The crystal-substrate base assembly was inserted into a fixture via the base leads making contact with three leads both mechanically and electrically. The top of the substrate is kept stationary when pad X on the substrate is probed. This fixture allows the monitoring of the three tuning frequencies during processing as well as the apparent coupling between adjacent resonators and

the short-circuit natural frequencies. Once the tuning frequencies are set to their final values and the apparent coupling between adjacent resonators is correct, the shorting connection is made from pad X to Y thereby shorting the middle electrode pairs together. The cover is slid over the entire assembly and sealed to the base, then the coupled-triple resonator is ready for final test.

Specific design example

The first application of a coupled-triple resonator crystal which was investigated was a 21.4 MHz 3-pole filter having a minimum 3 dB passband of 13.5 kHz and a minimum stopband attenuation of 25 dB at 21.4 MHz +- 22 kHz.

A third order Chebyshev normalized lowpass prototype of 0.1 dB ripple as shown in Fig. 9 was chosen to start the design. If a Dishal transformation is performed on this lowpass prototype, it is found that by choosing a ripple bandwidth of 10800 Hz and an inductance of .88280138E-2 H a termination resistance of 580.72727 Ohm results. If an extremely conservative value of Q = 30000 is chosen, then resulting circuit shown in Fig. 10 should produce a theoretical stopband attenuation of 31.0 dB at 21.4 MHz +-22 kHz and a 3 dB bandwidth of 14700 Hz.

Requiring the X and Z dimensions of the resonator electrode to be

$$LX = 1.26\ LZ$$

and using the expression

$$A = 1.9/(6.4516508\ L\ f_0^3(1E-18)),$$

where A = area of each electrode in square inches,

f_0 = center frequency of the filter in Hz,

L = the inductance of each resonator of the filter in henries.

Then the electrode dimensions are determined to be

LZ = .051976130 inches,
LX = .065489925 inches.

Following the procedure outlined in [7],[8], and [9] applied along the Z axis of the crystal, it is found that for electrode LX = .052 inches by LZ= .065 inches and a spacing between electrodes of .015 inches and a synchronous peak separation frequency SPSF of 9927 Hz corresponding to C12= C23=13.506531 pf are produced. The etch frequency of the unelectroded wafer should be 21,563 kHz.

Since the filter in Fig. 10 is symmetrical about the middle resonator, the basic triple-resonator crystal design is accomplished by simply performing the design of a coupled-dual. The actual wafer used was a race track arrangement with diameter of .320 inches and height of .240 inches.

From Eq. (15),

$$K_{12}^2 = K_{23}^2 = 1/(X_1X_2),$$

$$= 1/((1.0315598)(1.1473972)),$$
$$= .8448737264.$$

Therefore,

$$K_{12} = K_{23} = .9191701295.$$

To determine the value of C to be used in the equations for tuning frequencies, the expression

$$C = (C_1C_2)/(C_{12}-C_1)$$

is used where C_1 = .62683208E-14 f and C_{12} = .13506531E-10 f producing C = .627123125E-14 f. A nominal value of C_0 was assumed to be 2.4 pf initially.

Next, the tuning frequencies were calculated. From symmetry, ATF = CTF. Therefore, for ATF and BTF

$$E_1 = 0,$$

$$E_2 = E_3 = (C/C_0)(f_0/BW),$$

$$= (.627123125E-14/2.4E-12)(21.4E6/10800),$$
$$= 5.177636912.$$

The third order polynomial equation in F from (17) becomes

$$F^3-10.3552738\ F^2 + 25.118176S4\ F + 4.374449392=0.$$

Then

$$F = -.16302800\ ,\ 6.1678390\ ,\ 4.3508729,$$

and choosing the lowest value for

$$f_n = f_0 + (F_i\ BW)/2,$$

$$= 21.4E6 + (-.16302800)(10800)/2.$$

Therefore,

ATF = CTF = 21399120 Hz.

Following a similar procedure

BTF = 21398337 Hz.

Likewise, the apparent coupling between the resonators is obtained by the procedure outlined previously from Eqs. 12, 13 and 14 producing

f_1 = 21394653 Hz,
f_2 = 21464467 Hz,

with the apparent coupling being calculated as $f_2 - f_1$ = 9814 Hz. The third frequency occurs at 21428840 Hz and is ignored.

The short-circuit natural frequencies are then

calculated from Eq. 17 with $E_1=E_2=E_3=0$ and become

f_1 = 21392985 Hz,
f_2 = 21.4E6 Hz,
f_3 = 21407015 Hz.

Results

Data at several steps in the fabrication process is given here on a typical samples using a silver-nickel process with a solder-seal package.

	After Baseplate (Hz)	Final Frequency (Hz)
ATF	21465839	21399574
BTF	21452724	21398533
CTF	21461137	21399265

Apparent coupling between A and B resonators at final frequency is given by

$ACAB = f_2 - f_1$,

= 21404075 - 21394021,
= 10054 Hz.

Apparent coupling between C and B resonators is

$ACAB = f_2 - f_1$

= 21404542 - 21393991
= 10551 Hz.

The short-circuit natural frequencies measured while driving the A resonator are given by

f_1 = 21392423
f_2 = 21398789
f_3 = 21407818

The resulting filter data was

Ripple = .40 dB
Upper 3 dB frequency = 21409459 Hz
Lower 3 dB frequency = 21392702 Hz
3 dB Bandwidth = 16757 Hz

Attenuation at f_0 + 22 kHz = 25 dB
Attenuation at f_0 - 22 kHz = 34.2 dB.

First major spurious = -19.0 dB at 21490219 Hz.

Also, the experimental value of C_0 across both the A- and C-resonators in the sealed assembly was 3.1 pf. No data was taken for the value of C_0 across the B-resonator.

A typical plot of the attenuation response is given for a second unit and shown in Fig. 11. Although no process data is available, the short-circuit natural frequencies on the unit are 21391500 Hz, 21397750 Hz, 21407000 Hz. The filter fixture terminations are transformed to 50 ohms. With 50 ohms on the source (or load), a vector impedance meter probing J1 or J2 to ground reads 560 ohms at an angle of -7 degrees in Fig. 12 to produce the response in Fig. 11.

The apparent coupling value between A and B and C and B are given by

ACAB = 21403750 - 21393000 = 10750 Hz

ACCB = 21403250 - 21392750 = 10500 Hz

The actual passband shape of both triple resonator units are similar except for the slight shift in the passband.

Conclusions

Although the initial design called for an apparent coupling between resonators of 9814 Hz, the actual apparent coupling between resonators came in the 10 to 11 kHz range which contributed to the .3 dB extra passband ripple as well as the extra bandwidth. The extra ripple could also be explained by the fact that resonator inductances are probably not exactly what the design calls for. The initial assumption of C_0=2.4 pf put the resonator frequencies of the initial units very close to those required by the filter.

Once the fabrication method was devised, processing of units did not prove to be unusually difficult to build and the substrate-wafer assembly proved extremely convenient to work with. The spurious content and spurious levels were better than expected although they were not a major consideration in the original design. Since the techniques used in the fabrication of these coupled-triple resonators which were over and above those used on a coupled-dual unit can be either carried out in batch form or can be automated, this type of coupled-triple resonator crystal arrangement could be a viable candidate for use in situations where an efficient attenuation or group delay characteristic is needed.

References

[1] Von R. Saal, Der Entwurf von Filtern mit Hilfe des Kataloges normierter Tiefpasse, Telefunken G.M.B.H, Backnang/Wurtt, Western-Germany,1961.

[2] H. J. Peppiatt and G. E. Roberts, "Method of Measuring Parameters of a Crystal Filter," U.S. Patent 4,093,914, June 6, 1978.

[3] G. E. Roberts, "A four frequency process for accurately measuring coupled dual resonator crystals," Proc. 33rd Annu. Symp. on Freq. Control, 1979, pp. 159-165.

[4] G. E. Roberts, "Apparatus for measuring the resonant frequency and coupling coefficient of a plurality of coupled piezoelectric resonators", U.S. Patent 3,963,982, June 15, 1976.

[5] Aristotelis Arvanitis,"Multi-resonator crystal filter having a reduced number of pins and method of fabrication thereof," U.S. Patent 4,342,014, July 27, 1982.

[6] R. E. Rennick, "An equivalent circuit apprach to the design and analysis of monolithic crystal filters," IEEE Trans. on Sonic and Ultrasonics, Vol. SU-20, No. 4, Oct. 1973, pp. 347-357.

[7] G. E. Roberts, and A. G. Staples, "Monolithic crystal filter and method of manufacturing same," U. S. Patent 4,484,158, Nov. 20, 1984.

[8] G. E. Roberts, "The design of monolithic crystal filters," Engineering Notebook, General Electric Mobile Radio Products Department, Vol. 11, May 2, 1978.

[9] W. D. Beaver, "Theory and design principles of the monolithic crystal filter," Ph. D. Dissertation, Lehigh University, Bethleham, Pa., 1967.

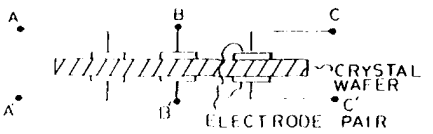
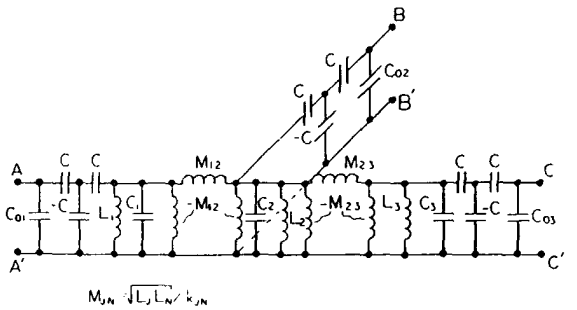


Fig. 1. Sketch of cross-section of coupled-triple resonator crystal with all three resonators open-circuited.


A
B
B'
C
C
C

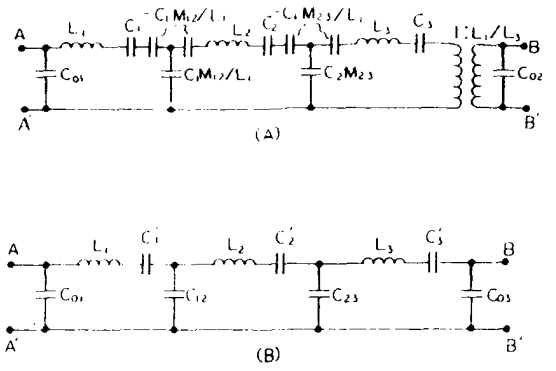
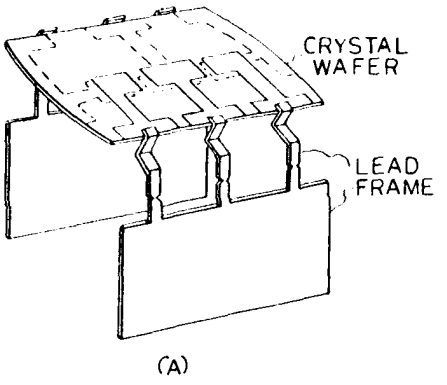
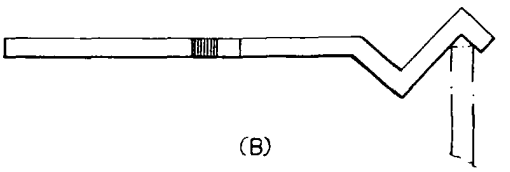


Fig. 5. A) Equivalent circuit of coupled-triple resonator in Fig. 4A after left-side capacitive-T impedance inverter is moved through the network, B) equivalent circuit of coupled-triple resonator after C_1, C_2, and C_3 are set equal to C, the negative and positive series capacitors are combined, and L_1, L_2, and, L_3 are set equal to L.





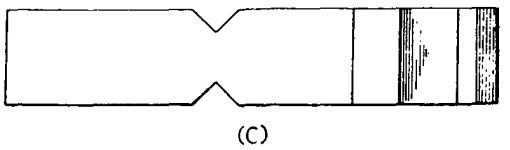


Fig. 6. A) Crystal wafer with baseplated conductor pattern and attached lead frames, B) side view of lead frame fingers and break points, C) top view of lead frame fingers with indicated break point.

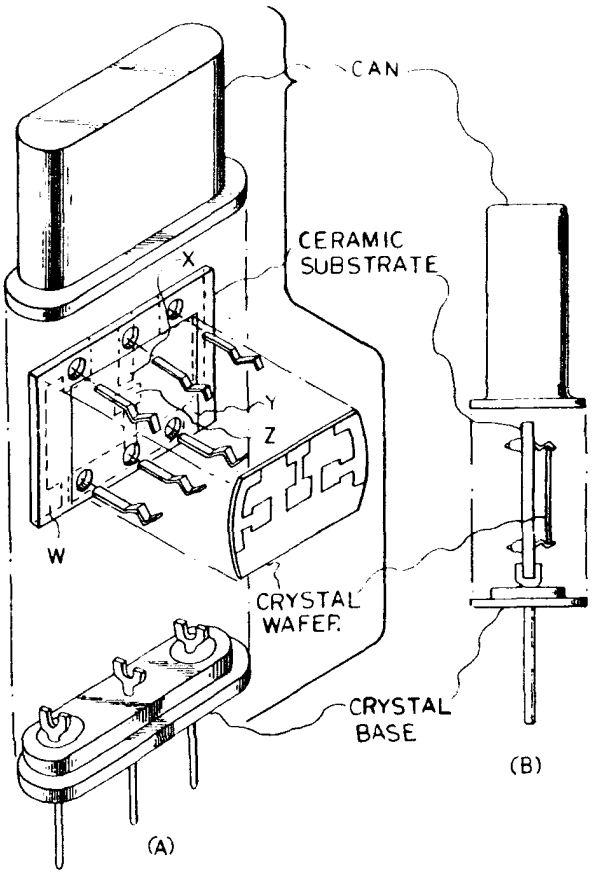


Fig. 7. A) Exploded view of crystal wafer-ceramic substrate-leads-base-cover assembly, B) side view of wafer-substrate-base assembly with cover.

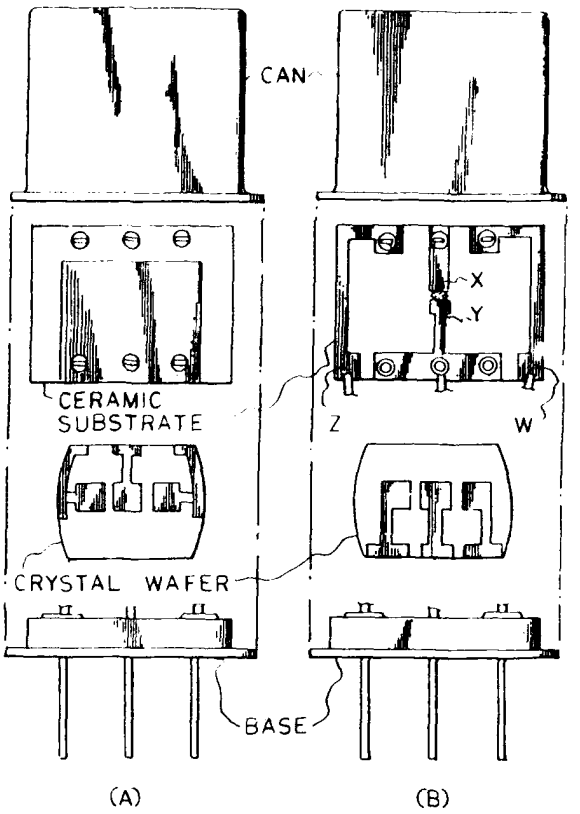


Fig. 8 A) Exploded front view of triple resonator assembly, B) exploded rear view of triple resonator assembly.

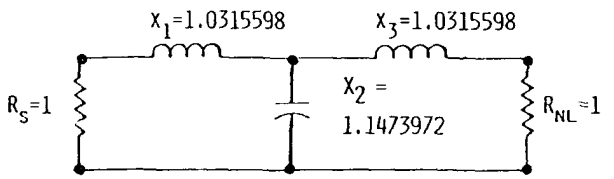
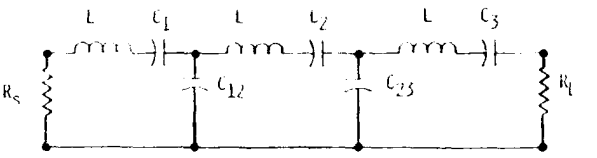


Fig. 9. Equivalent circuit of 3rd-order lowpass prototype.



$R_S = R_L = 580.72727$ Ohms
$L = .88280138E-2$ H.
$C_1 = C_3 = .62683208E-14$ F.
$C_2 = .62712312E-14$ F.
$C_{12} = C_{23} = .13506531E-10$ F.

Fig.10. Bandpass equivalent circuit of 3-pole filter derived from 3rd-order lowpass prototype in Fig.9.

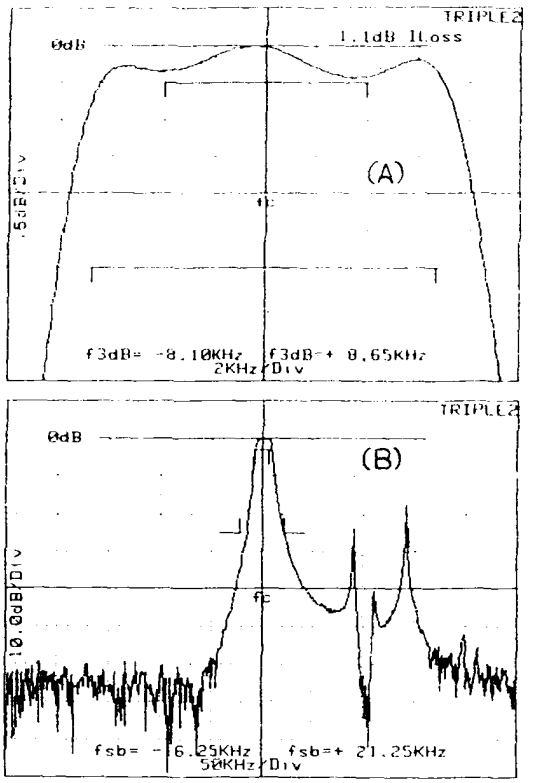


Fig.11. A) Passband attenuation response of 3-pole filter using triple-resonator unit no.2. B)stopband response of the same triple resonator crystal.

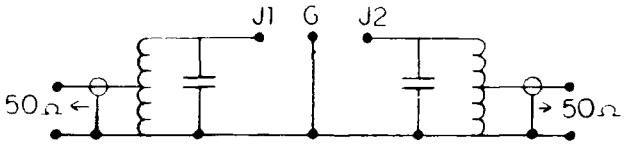


Fig.12. Schematic of circuit used for filter fixture to test triple resonator crystal.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

USAGE OF A QUARTZ RESONATOR AS A PIEZOSORPTION SENSOR FOR MERCURY VAPOURS IN AIR

L. Spassov, D. Yankov, D. Mihailov, R. Zaharieva

Institute of Solid State Physics, Bulgarian Academy of Science, Sofia, Bulgaria

A.N. Mogilevski, A.D. Mayorov

Institute of Geochemistry and Analytical Chemistry, Academy of Science of USSR, Moscow, USSR

Abstract

A quartz resonator with built-in micro heater used as a piezosorption sensor for mercury vapours in the air is reported. The specifical requirements to the resonator and to the heater are noted. With the presented here construction of the sensor a maximal sensitivity of $3 \cdot 10^{-6}$ g/m^3 can be reached. The built-in micro heater allows the regeneration of the sensor under working conditions. The sensor can be used in several devices for control of the environmental pollutions and in other cases in which precise determination of low concentrations of mercury vapours in the air is necessary.

1. Introduction

Recently as a result of the aggravation of the ecological situation working out of a method for measuring the concentration of injurious to health components that are to be found in industrial waste products becomes extremely important. The problem for monitoring of mercury vapours in air and measuring their concentration is one of the aspects of this general task. This problem is a complex one and different ways for its solving have been seeked. Currently there are a number of publications on the piezosorption method for construction of sensors and devices for gas analysis [1]. There is also information about construction of a device for measuring the concentration of mercury vapours in air [2,3,4].

The principle of action of these sensors is based on the dependence of the frequency of the piezoelectric quartz resonator on the change of the weight of the electrodes caused by sorption of mercury vapours. For AT - cut quartz resonator the relationship between the change of the resonance frequency ΔF [Hz] and the mass Δm [g] of the substance adsorbed on the surface of the electrodes is almost linear and is given by [5]:

$$\Delta F = - 2.3 \cdot 10^6 \cdot F^2 \cdot \Delta m / S \qquad (1)$$

where F [MHz] is the resonance frequency of vibration of the piezoelement, and S [cm^2] - the surface of the piezoelement's electrodes.

The main disadvantage of the piezosorption sensor is that it needs regeneration of the quartz resonator after a certain time period of action. For this aim it should be heated in oven approximately to 150°C and simultaneously the oven should be blown with fresh air or other gas. This process needs laboratory conditions for performing the regeneration.

One possibility to solve this problem is

CH2818-3/90/0000-277 $1.00 © 1990 IEEE

creating a quartz resonator with built-in micro heater. This heater allows heating of the quartz resonator up to relative high temperatures to be done easily without its dismounting from the equipment.

2. Realization of the sensor

2.1. Requirements to quartz resonator and to built-in heater

According to the usage of quartz resonators as a sensors for mercury vapours they should cover several requirements. First of all the electrodes of the resonator should adsorbe selectively mercury vapours only and should be stable to other reagents in the air. In order to registrate low concentrations of mercury vapours the resonator should possess a high mass sensitivity. As it was mentioned above the requirement for direct heating of the resonator is important to ease the exploatation of the measuring apparatus. The built-in heater should be with low electric power consumption and at the same time should be able to reach high temperatures. The heater also should be stable to action of mercuty vapours and other gases in the air.

2.2. Construction of the sensor

The task for creating of piezosorption sensor is solved by AT - cut quartz resonator with resonanse frequency about 14 MHz. The dimencions of the resonator are: diameter of the wafer - 14 mm and thickness - about 120 μm. In order to increase the sensitivity the quartz wafer is polished. Thin layers of gold are used as electrodes. The golden layers are at the same time selective towards mercury vapours sorption coating. The diameter of the electrodes is 5 mm which provides a sensitivity about 0.4 Hz/ng calculated by (1). The built-in micro heater is a thin strip situated around the electrodes (Fig. 1) and is made from the same material. The thickness of the electrodes and the heater is about 0.12 μm and they all are deposited in one technological cycle. The resonator is mounted at standard holder with additional output connectors to the heater.

3. Results

3.1. Electrical parameters of the resonator

The typical electrical parameters of a resonator with built-in micro heater are following:

Resonance frequency	-	14.422 MHz
Static capacity	-	8.5 pF
Dynamic resistance	-	14.2 Ω
Dynamic capacity	-	14.4 fF
Dynamic inductivity	-	8.4 mH
Loaded Q	-	$54 \cdot 10^3$

Figure 2 demonstrates the spectral characteristic of the same sample.

Because of the golden electrodes the resonator possesses very high time stability. It was found that the typical value of frequency time instability did not exeed 0.1 Hz per 10 seconds, 0.2 Hz per 2 minutes and 0.5 Hz per 10 minutes.

3.2. Parameters of the micro heater.

In order to determine the necessary electrical conditions to reach the wanted temperature

of 150°C the following experiment was performed. First the resonator was placed in a thermostat with temperature 150°C and its frequency was measured. After that the temperature of the thermostat was lowed down and kept up to 25°C. An electrical voltage was applied on the heater and was gradually increased till the frequency became the same as at 150°C. With the typical value of resistance of the heater about 30 Ω the measured value of the voltage was about 6 V and the current was about 130 mA. In this way the calculated value of the electrical power consumpted by the heater was about 800 mW, which was less than the maximal admissible value for the conductive golden strip of the heater. The usage of a golden heater allows itself regeneration from adsorpted mercury vapours too.

3.3. Parameters of the sensor

Figure 3 shows the block diagram of the experimental apparatus for investigation of the relationship between the quartz resonator frequency and the mercury vapours concentration. The measurement apparatus allows to monitor the frequency of the measuring resonator, as well as the difference between the frequencies of the measuring and the standard generators. The usage of a standard generator made possible the conversion from higher to lower frequencies, so that the latter could be more easily processed further.

For regeneration of the sorption capacity of the resonator we chose a temperature of 150°C at which there are not any relaxation processes in quartz [4]. As it was mentioned above the built-in micro heater ensure this temperature at cited electrical conditions. The desorption time period at this temperature is not more than 30 minutes.

Figure 4 shows the relationships between the frequency response of the measuring quartz resonator and the time period of action of the mercury vapours, which were established at different pressures of the saturated mercury vapours, i.e. they are with different concentrations. These relationships are linear and thus allow us to determine small concentrations of mercury vapours by increasing the time of action.

The concentration of mercury vapours can be determined by investigation of measuring resonator upon the flow rate of deluted mercury vapours throug the chamber. In a particular apparatus [4] the maximal frequency response rate of 50 Hz/min is reached at a flow rate of 400 ml/min.

Figure 5 demonstrates the relationships between the frequency response rate and mercury concentration in the interval from $2 \cdot 10^{-8}$ to $4 \cdot 10^{-7}$ g/m³. The results give the possibility to calculate the detection limit. As seen from Figure 5 the frequency response rate of the measuring resonator at mercury concentration in air $2 \cdot 10^{-8}$ g/m³ is 1 Hz/min.

Being measured for 2 minutes the change of frequency equals 2 Hz. For this time period the frequency instability of the resonator does not exceed 0.2 Hz. In this case the minimal detectable frequency change according to the 3 σ - criterion is approximately 0.6 Hz, while the detection limit corresponding to frequency response rate 0.6 per 2 minutes is $6 \cdot 10^{-9}$ g/m³. Calculated in the same way the detection limit is $3 \cdot 10^{-9}$ g/m³ when being measured for 10 minutes.

The threshold limit value (TLV) of mercury

vapours in air is equal to 10^{-5} g/m^3 according to current standards [6], therefore the above mentioned detection limits correspond to 0.6 and 0.3 times TLV.

4. Conclusions

The results of the investigation work show that the described piezielectric resonator can be used as a sensor for detecting and measuring small concentrations of mercury vapours in air up to the threshold limit value. The built-in micro heater allows the regeneration of the resonator to be done easily without special laboratory conditions. It gives the possibility the represented piezosorption sensor for mercury vapours in air to pass in general use in a great number of practical needs.

5. Acknowledgements

The authors are very grateful to A. Stratieva for preparing of the substrats and to E. Yossifov for the assistance in the creation of resonators.

6. References

[1] Alder I.F., I.I. Meculum, The Analyst, Vol. 108, No. 1291, 1196-1187 (1983)

[2] Scheide E.P., R.B.J. Warner, Am. Ind. Hyg. Assoc. J., Vol. 39 745 (1978)

[3] Scheide E.P., Phys. Teach., Vol. 15(1), 47-51 (1977)

[4] Mogilevski A.N., A.D. Mayorov, N.S. Stroganova, D.B. Stavrovski, I.P. Galkina, L. Spassov, D. Mihailov, R. Zaharieva, To be published

[5] Saubrey G.Z., Z. Physik, Vol. 155, 206 (1959)

[6] State standard of the USSR - GOST 12.3.031 - 83 (in russian)

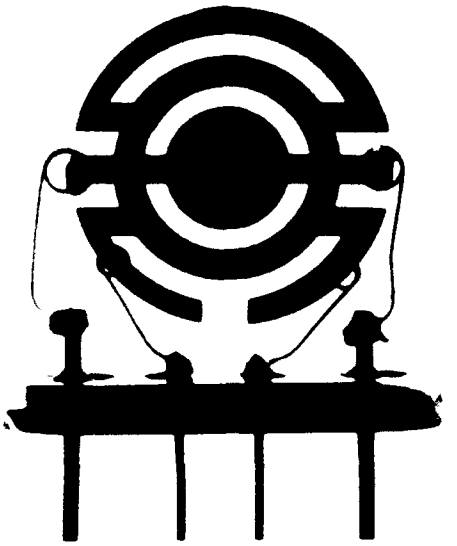


Fig.1. General view of the resonator with built-in heater

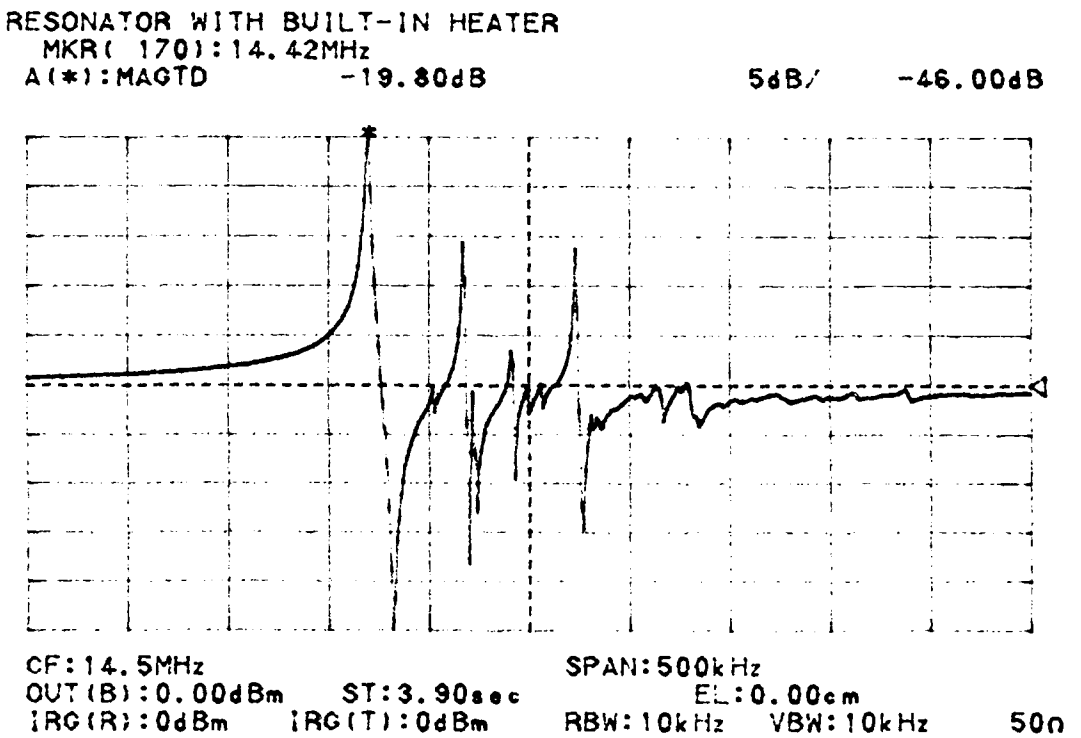


Fig.2. SPECTRAL CHARACTERISTIC OF THE RESONATOR WITH BUILT-IN MICRO HEATER

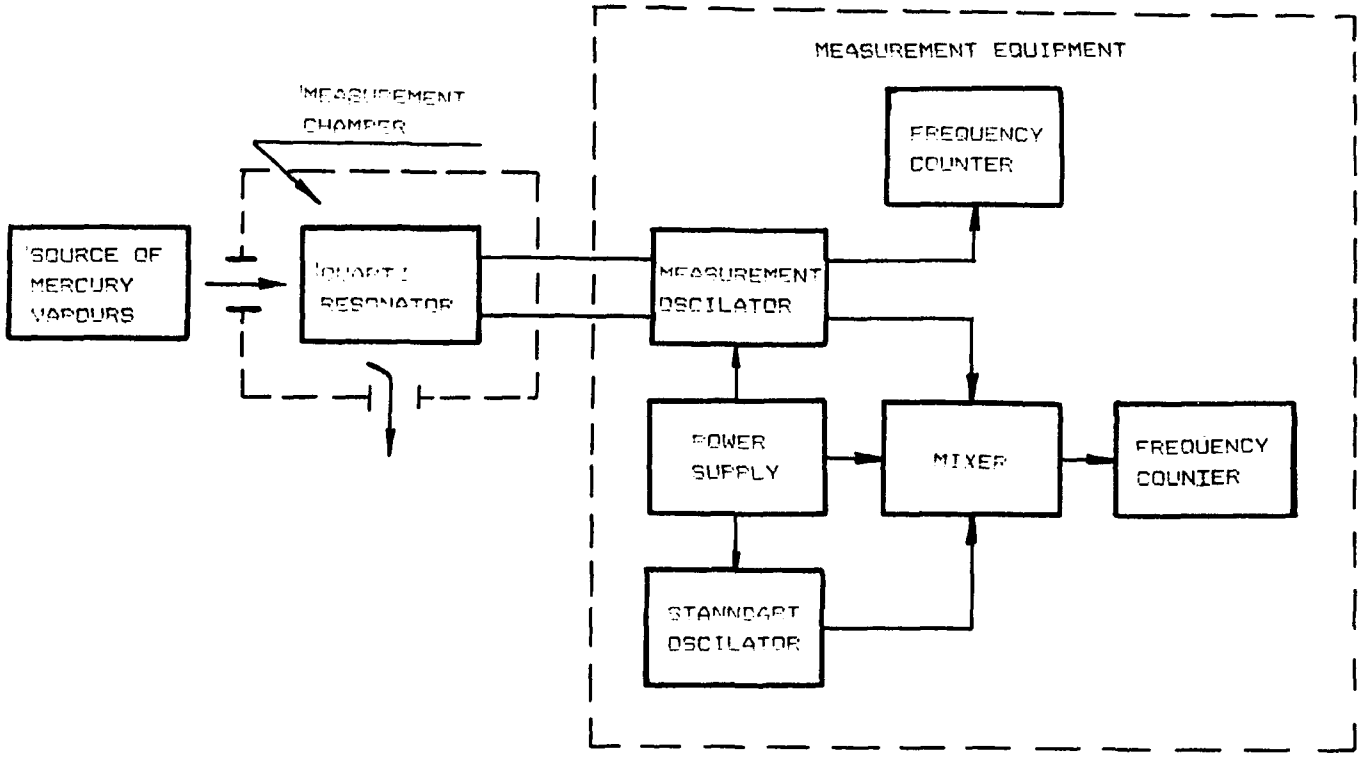


Fig.7. BLOCK DIAGRAM OF EXPERIMENTAL APPARATUS FOR INVESTIGATION OF THE RELATIONSHIP BETWEEN THE QUARTZ RESONATOR'S FREQUENCY AND MERCURY VAPOURS CONCENTRATION.

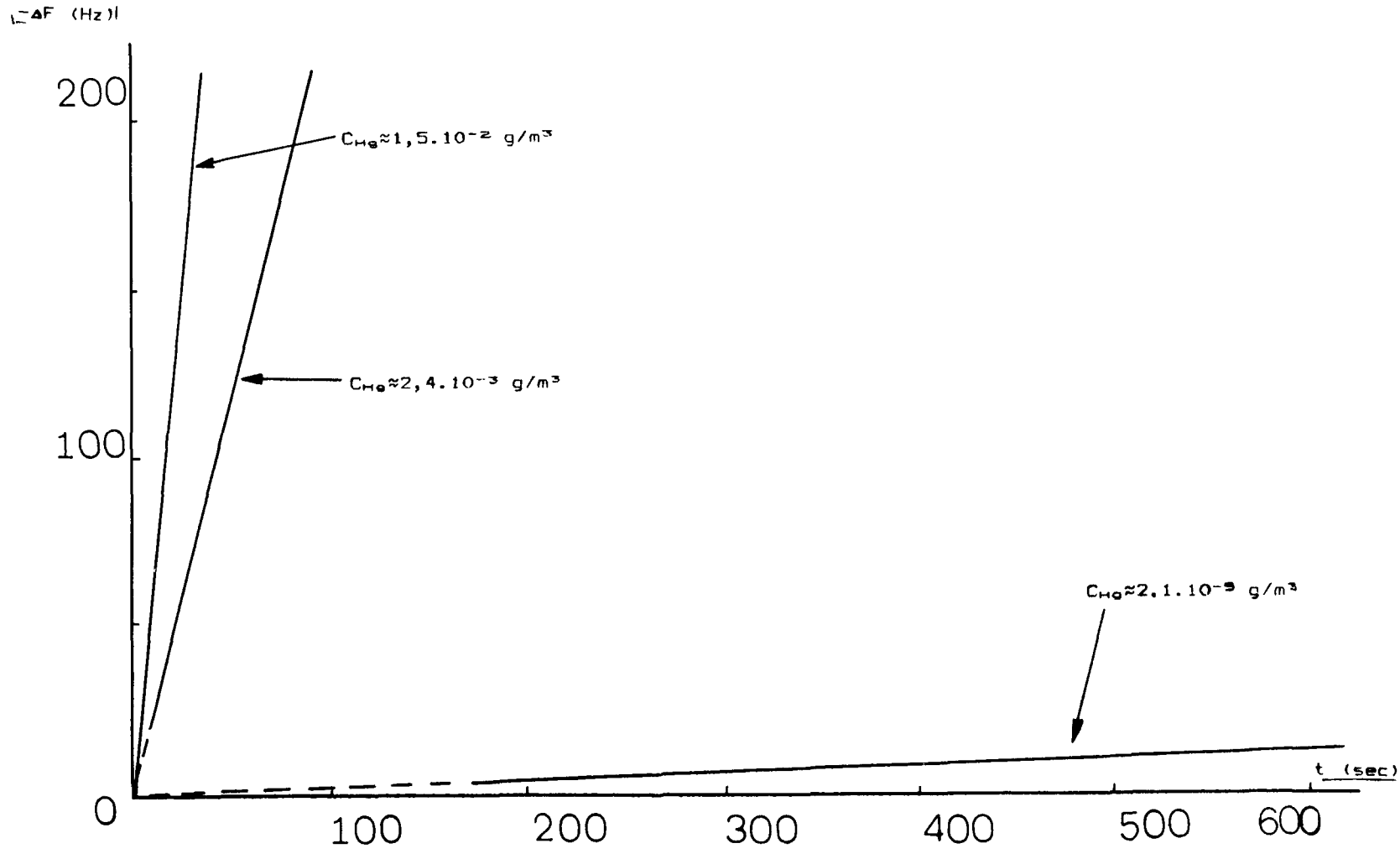


Fig. 1. RELATIONSHIPS BETWEEN THE FREQUENCY RESPONSE AND THE TIME PERIOD OF ACTION AT DIFFERENT CONCENTRATION OF THE MERCURY

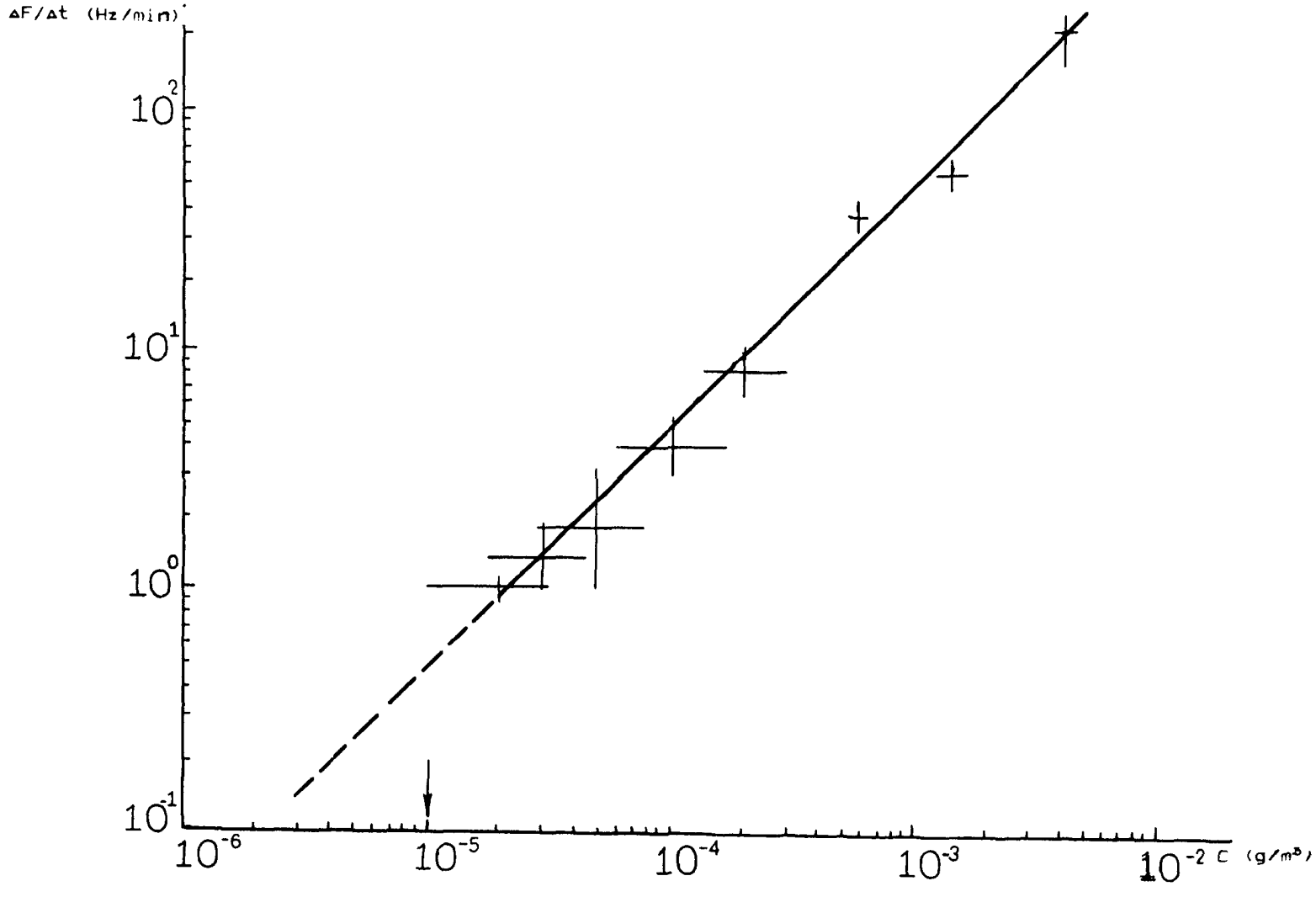


Fig.5. RELATIONSHIPS BETWEEN THE FREQUENCY RESPONSE RATE AND MERCURY CONCENTRATION

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

STATISTICAL PROCESS CONTROL FOR FREQUENCY CONTROL

by
Bill Thomas
Statistical Engineer

Rockwell International Corporation
Collins Defense Communications
855 35th Street NE MS 166-201
Cedar Rapids, Iowa 52498

Abstract

Frequency control like many other processes has an abundance of variables. These variables can be studied using the tools of Statistical Process Control (SPC). This paper will discuss some possibilities of using these tools to improve the accuracy, precision and quality of frequency control. The tools presented are process flow charts, pareto charts, cause and effect charts, histograms, scatter diagrams/regression, control charts and designed experiments in the control of frequency standard processes. Some of the major variables considered will be crystals, thermistors, selastic and workmanship and how to study their relationships to frequency stability. No experimental work has been done to this point. Therefore the paper will discuss a few approaches to studying frequency control using the preventative method of SPC. The major emphasis will be on reducing the variation of the frequency stability by studying variation and being able to sort out the signal (assignable causes) from the noise (random causes). Some specific items included will be when and how to look at the distributions of the data along with when and how to look at the data over time.

Introduction

Increased worldwide competition in the electronics industry has prompted many companies to consider the use of SPC. One of the first areas in the electronic industry was the computer industry. SPC can be applied to product design, process design, manufacturing and/or the providing of services.

SPC is a management philosophy which embraces the methodology of prevention of defects with the goal of improved quality and productivity. When quality is improved it causes improved productivity and reduces cost. Continuous improvement in the product and processes is realized by reducing variation and moving towards the optimum target point of each process variable. This philosophy must encompass the entire company though all levels of management to reap the maximum possible benefits.

Process Flow Charts

This is the first step in SPC implementation. This chart answers the question "what goes on in the process?." It visually describes the sequential process steps and decision points in the process. (Figure 1) In this process flow chart the major steps of manufacturing a frequency control device are displayed. An oven controlled type frequency standard is used as the example.

Pareto Chart

This chart answers the question "what are the major problems in the process?." (Figure 2) It is a bar chart that separates the trivial many problems from the vital few. This helps people understand which problems occur most frequently so they can work those first. [1]

The rule of thumb used is that 20% of the problems create 80% of the impact. Collection of defect types or problem categories is the first step. Next arrange the contributors in bar chart form, from largest contributor down to the smallest contributor on the x axis. The y axis represents the number of occurrences. Figure 2 is an example of major problems in the manufacture of frequency standards.

CH2818-3/90/0000-285 $1.00 © 1990 IEEE

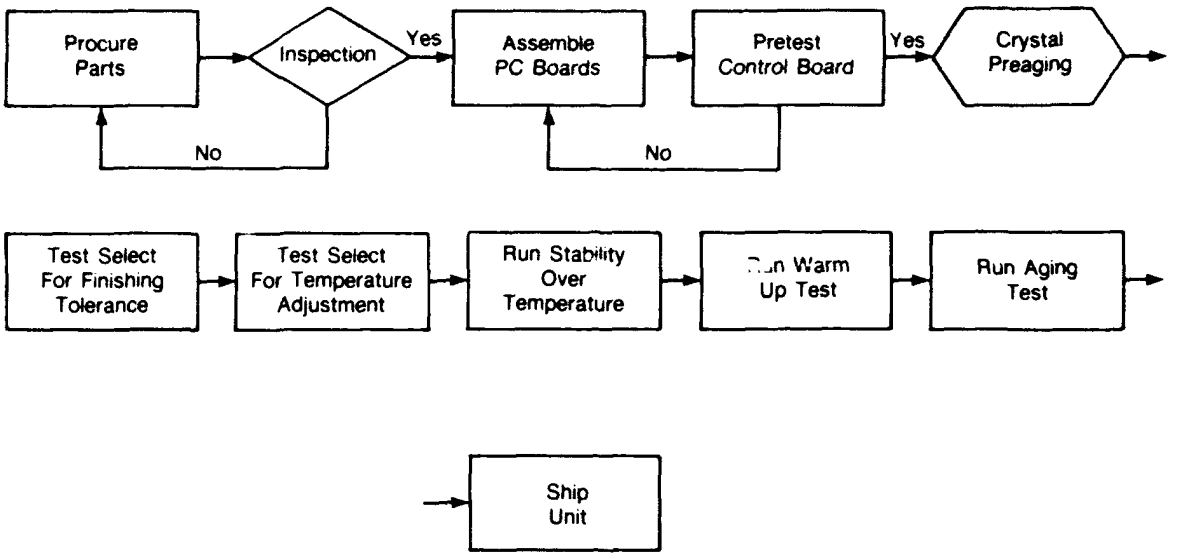


Figure 1. Process Flow Chart.

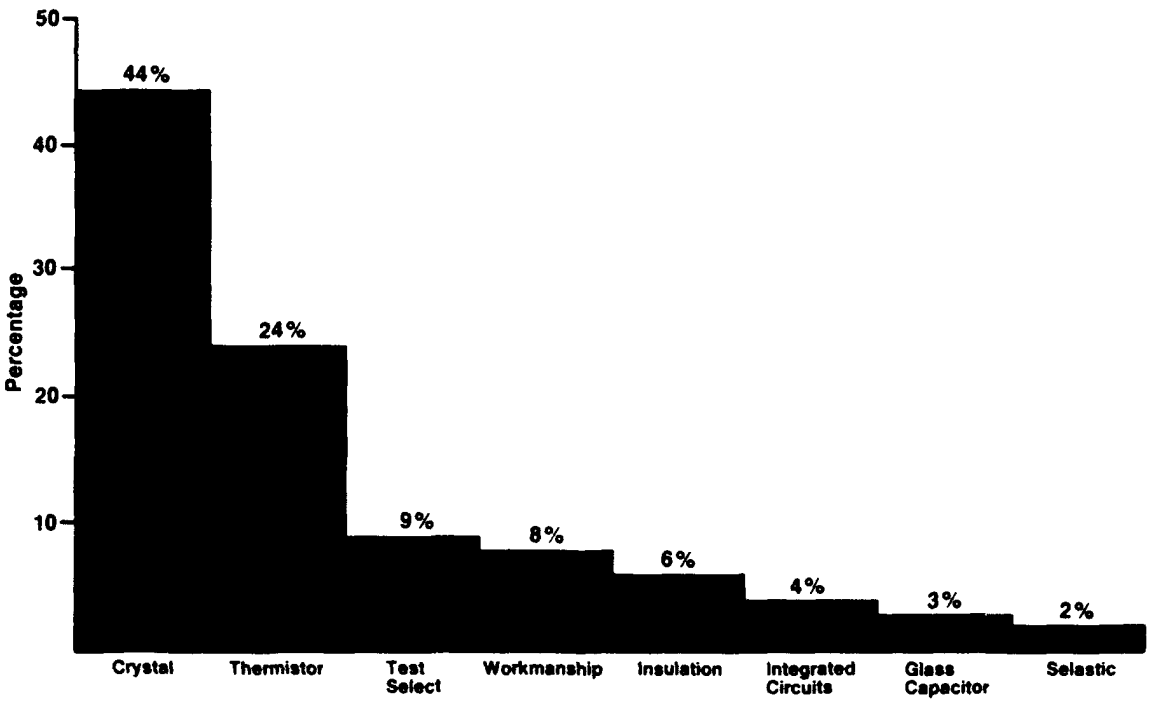


Figure 2. PARETO Chart.

Cause and Effect Chart

The next step is to construct a cause and effect chart.[1]The problem (or defect) is the effect, the possible causes are generated with a "Brainstorming" session. (Figure 3) For the remainder of this paper, when effect is used it is implied that the effect is a major problem. This chart motivates the answer to the question "what are the causes of the major problem?"

To construct a cause and effect chart you simply brainstorm possible causes with the process experts in five major areas (methods, machinery, material, man, and the environment). This has been done for the effect of the crystal in figure 3 since it was the largest problem on the pareto chart (figure 2).

Histogram

The Histogram helps define a possible cause or effect to determine graphically the central tendency, shape, and variation associated with variable (footnote 1) data. It answers the question "what does a look at the past or present data look like?." The data can be compared to a specification to see how the distribution looks in regards to the specification. It is a snapshot of the data and can be very helpful in understanding, correcting and improving a particular cause or effect. It is a special bar chart with the variable data (continuous data) on the x axis and the number or percentage of occurrences on the y axis. It does not take time order into account. [1]

Figure 4 represents the distribution of the frequency stability readings from a frequency standard over the temperature range from 25 degree C to 85 degree C. The shape of any group of data is very important. How it compares to the specifications is also very helpful in improving any process. Figure 5 is another histogram of the crystal turning point temperature for a sample of crystals to describe the shape, central tendency, and variation between crystals. Figure 4 is an effects/outputs. Figure 5 is a possible cause of variation for frequency stability distribution in figure 4.

From figure 4 the assumption that the distribution is normal can be made. This means the data should be consistent and repeatable. Another statistical technique can be used to predict the percentage of frequency standards outside the specifications of 10 megahertz plus or minus 0.05 hz. (10 megahertz has been subtracted from all the data for ease of interpretation and explanation.) The mean of the data is –0.0288 Hz and the standard deviation (a measure of variation) is 0.034 Hz. The percentage outside the specification would be 27.78%. (Figure 4a) This prediction was made with a sample of 80 values and is

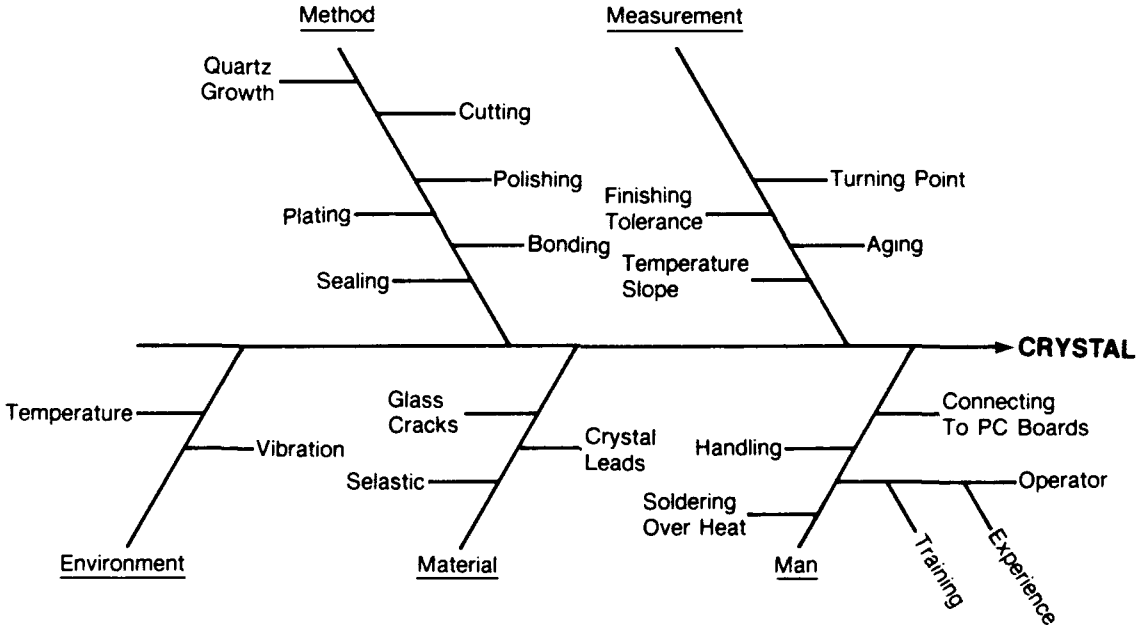


Figure 3. Cause and Effect Chart

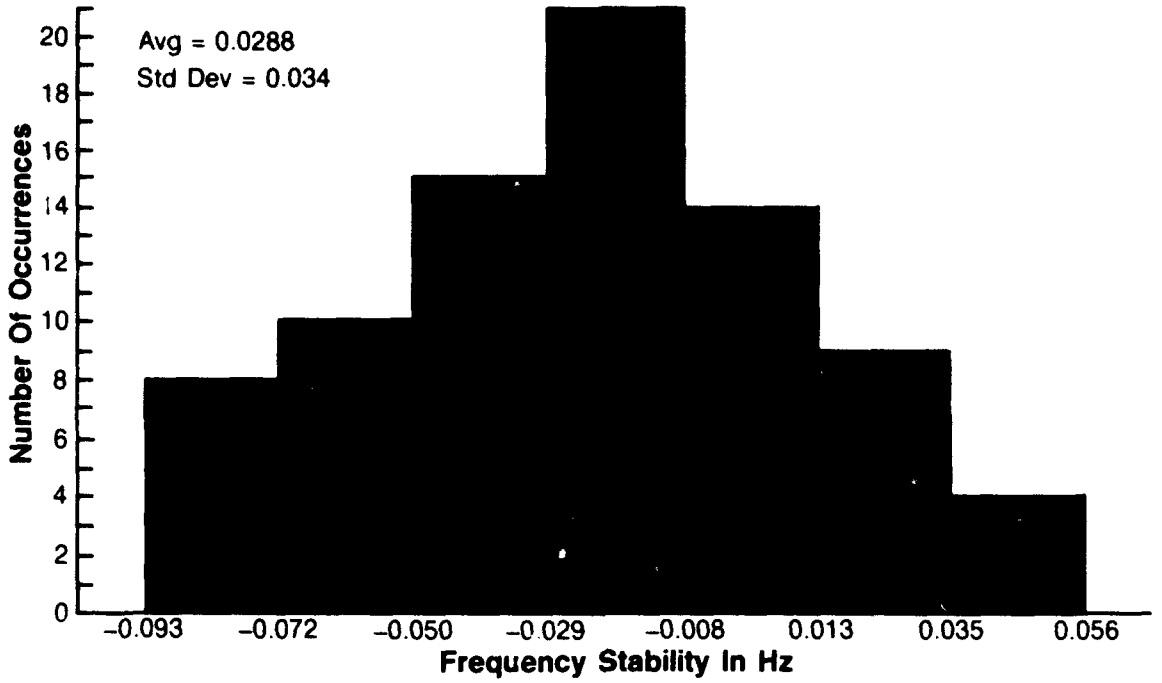


Figure 4. Frequency Stability 25 °C to 85 °C Histogram for 10 MHz. Frequency Std.

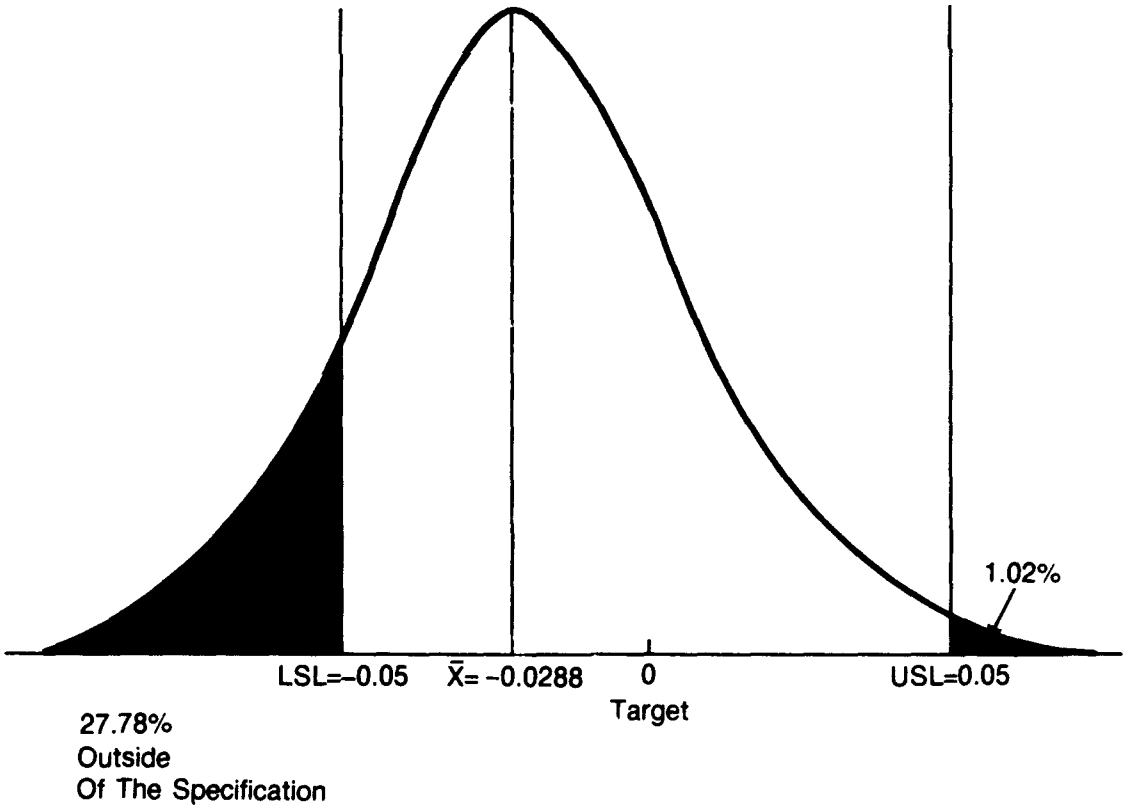


Figure 4A.

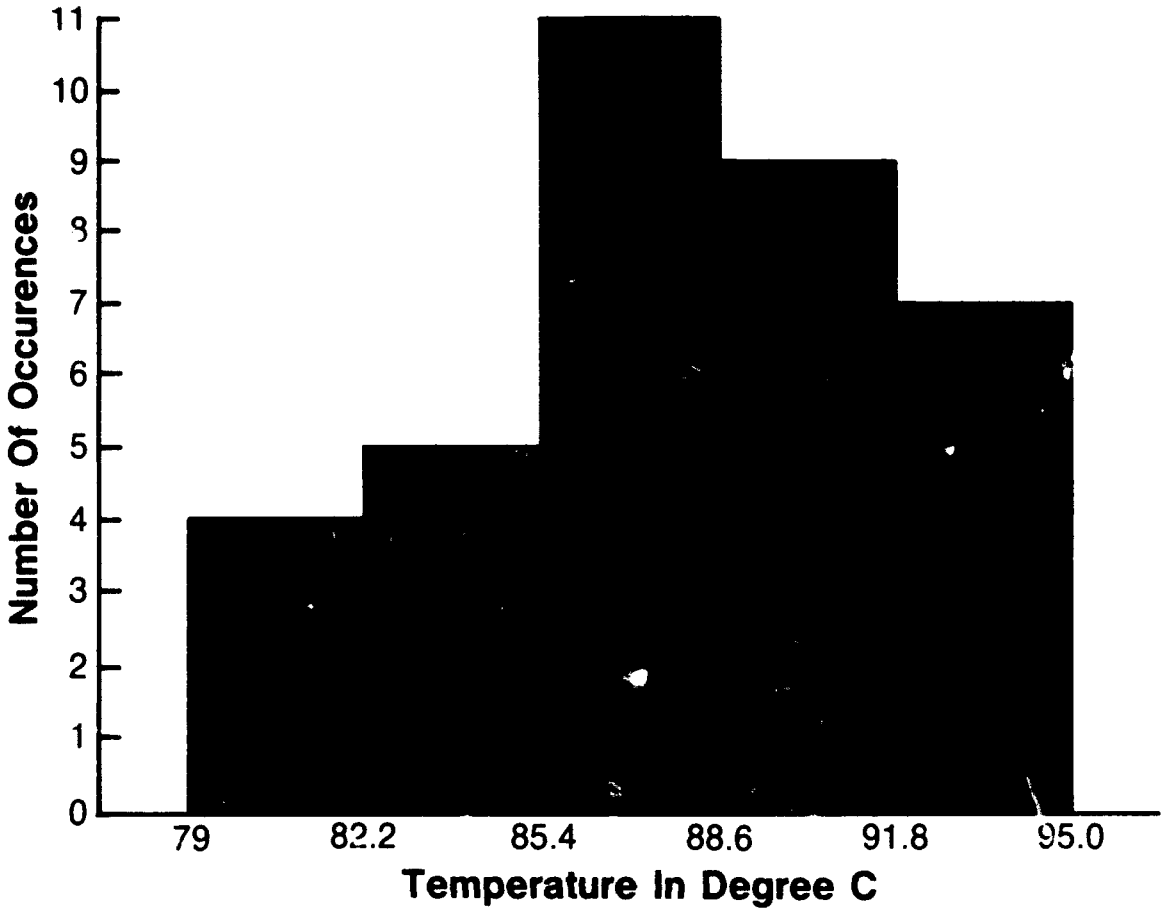


Figure 5. Turning Point Temperature Histogram.

very reliable. It is now understood that the variation is too large and the mean is shifted towards the lower specification limit. (Figure 4a) The target for the mean on this chart is 10 megahertz or 0. The same prediction could be made for the turning point data. Using these techniques it is much easier to make improvements in comparison to counting how many are in or out of specification with no regard to the location of the values.

1) variable data - measurement type data i.e. (length, weight, s/n ratio, modulation, frequency etc.) as apposed to attribute data (pass-fail, go-no go)

Scatter Plot/Regression Analysis

This tool can be used to try to understand the relationships between the causes and the effects. (Figure 6) This tool answers "what are the cause/effect relationships?." It is very helpful in understanding the major causes of variation of a particular effect. A simple regression can be plotted on an x-y grid or scatter plot while multiple regression will require more complex statistical methods. This tool is very effective when studying one or many causes and their relationship to the effect. The results can be used to monitor an important cause on a control chart or discover the best level of the cause which produces an optimum effect. [1]

Figure 6 represents a scatter plot or simple regression between the frequency stability (output/dependent variable/effect) and the distance of the thermistor from the heater (input/independent variable/cause). This is not actual data, but an example of what this relationship may look like.

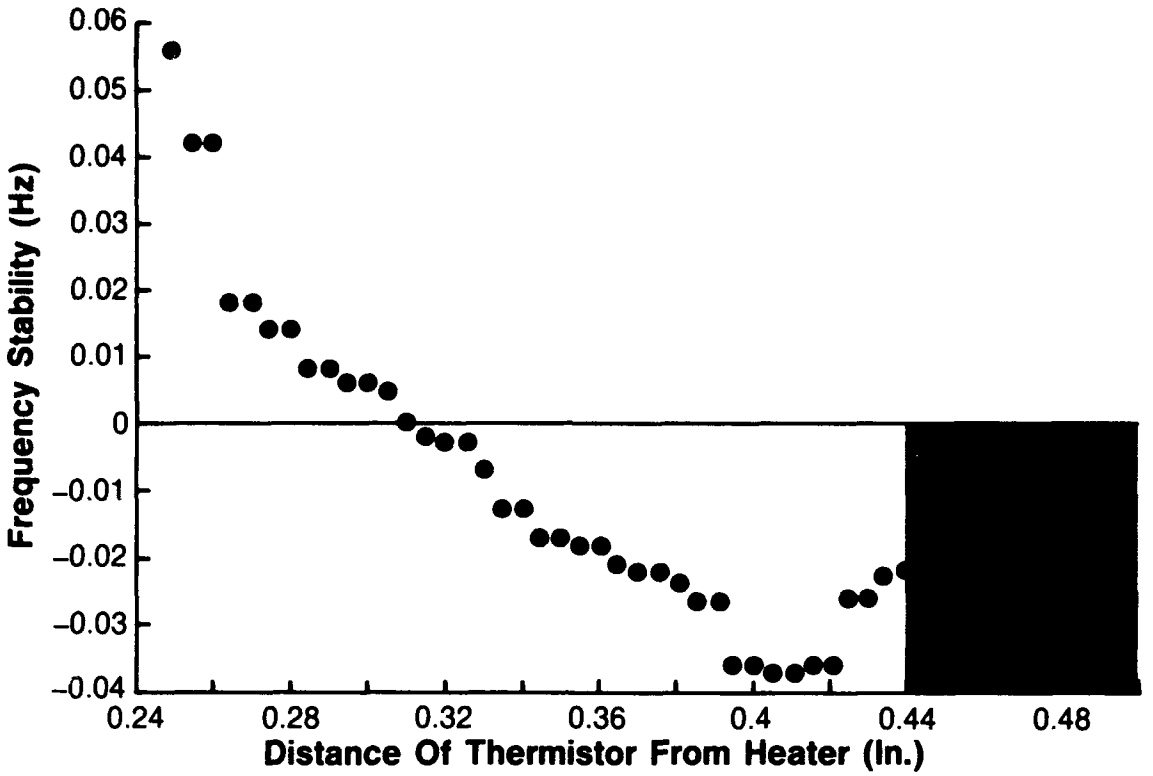


Figure 6. Thermistor Distance By Frequency Stability Scatter/Plot/Simple Regression Illustrative Data.

Design of Experiments

This tool is used often after a cause and effect chart to study many different causes and their relationship to the effect. This tool also answers "what are the cause/effect relationships?." It differs from regression in that the different levels of the causes are prudently picked to learn the maximum amount of information about the effect with the minimum amount of time, cost and effort. An example of a design layout is presented in Figure 7. [2,3]

This is a full factorial experiment with 2 levels and 3 factors (thermistor placement, turning point temperature, error in setting the turning point). This experiment would be able to answer the question of which factors or interaction of factors has the largest affect on frequency stability and possibly the optimum level of these factors to obtain the desired frequency stability mean and variation. It is much more effective than the normal shotgun or one factor at a time approach which frequently never solve the problems or answer the questions. These two approaches can never discover interactions. The designed experiment is a systematic and balanced matrix that ensures more understanding of the relationships. It helps understand why something works, not just if it works, and moves the experimenter towards the best solution instead of one that merely is acceptable. In addition designed experiments are often more economical and almost always generate more and better information.

Control Charts

This chart can be used to learn about a process, or after much of the learning is done to monitor a process for control and continually improve it. As a learning tool it can be used to learn the effects of many causes, one at a time, on the charted effect in order to continually improve the process. As a monitor it is statistical proof of quality and can always be used to improve a process. [4,5]

Figure 8 is the range chart for the frequency stability subgroups. It exhibits a state of statistical control. The

Run Number	Thermistor Placement	Turning Point Temperature	Error In Setting Oven To Turning Point
1	0.25 Inch	85 Deg	0 Deg
2	0.50 Inch	85 Deg	0 Deg
3	0.25 Inch	90 Deg	0 Deg
4	0.50 Inch	90 Deg	0 Deg
5	0.25 Inch	85 Deg	+2 Deg
6	0.50 Inch	85 Deg	+2 Deg
7	0.25 Inch	90 Deg	+2 Deg
8	0.50 Inch	90 Deg	+2 Deg

Figure 7. Full Factorial Experimental Design 2 Levels and 3 Factors.

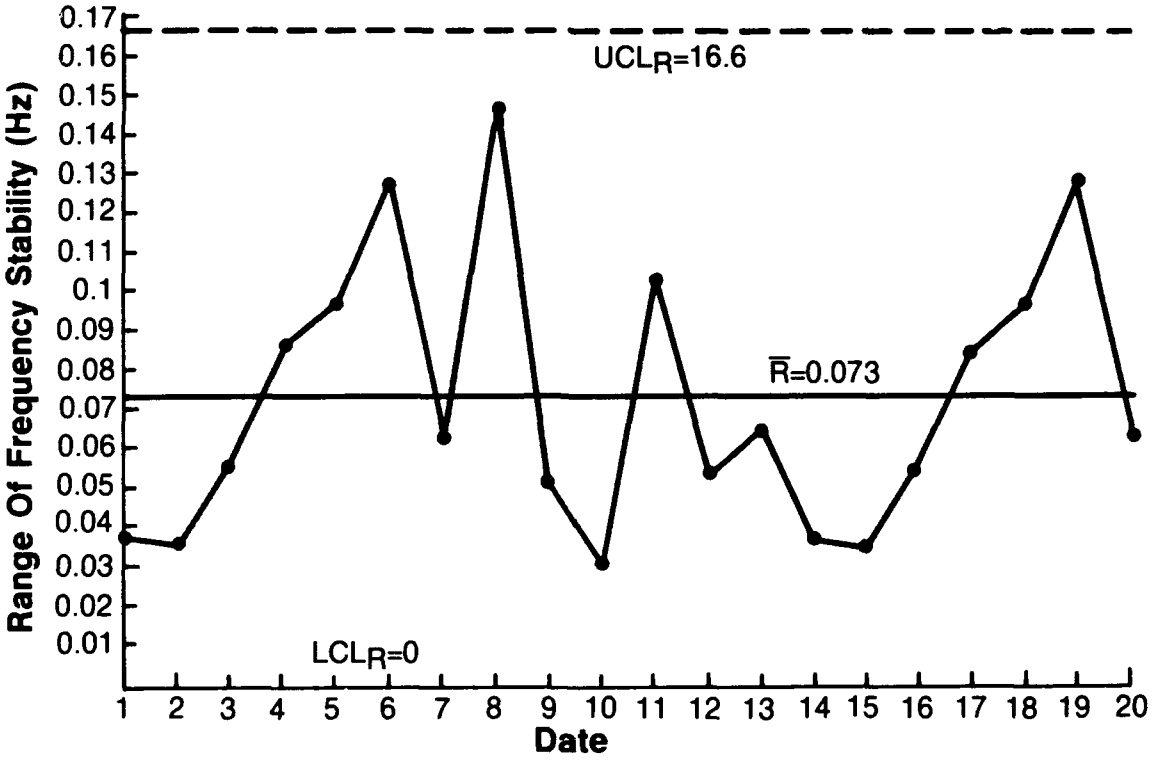


Figure 8. Frequency Stability 25 °C to 85 °C Range Chart for 10 MHz Frequency Std.

variation is larger than desired but the variation within the subgroups is consistent.

Figure 9 represents the plotted averages of subgroups of size 4 of frequency stability data from the histogram figure 4. This data is now in time order and more insight into the problem of too much variation and the shift of the mean towards the low side of the specification can be seen. When looking at the data a shift to a higher average is apparent from the 7th to 14th subgroup. Possible causes are dewar flask loss, thermistor lead length, thermistor heat sink compound, and crystal turning point temperature. This is just some possibilities of causes that can be discovered when using control charts.

Summary

The techniques discussed in this paper are the major techniques used in SPC. There are many additional methods and variations of the above methods that can be used. The statistical methods don't solve the problems , but they give people who can solve the problems the signal when and where to look for the problems. The SPC tools also give reliable feedback when changes are made and are a tremendous resource for quality and process history. They can help change peoples thinking of quality from merely being conformance to requirements towards thinking of quality as the reduction of variation around a target value. The last thought you should consider is that anyone in market competition whose competitor uses these techniques and does not use them also is at a distinct disadvantage.

[1] T. P. Ryan, "Statistical Methods For Quality Improvement." New York: Wiley, 1989

[2] G. E. P. Box, W. G. Hunter, J. S. Hunter. Statistics For Experimenters. New York: Wiley, 1978, chap 1, 6 & 7.

[3] D. C. Montgomery, Design And Analysis Of Experiments. New York: Wiley, 1984, chap 1, 7, 9 & 11.

[4] Western Electric, Statistical Quality Control Handbook. North Carolina: Western Electric, 1983, pp. 1-229.

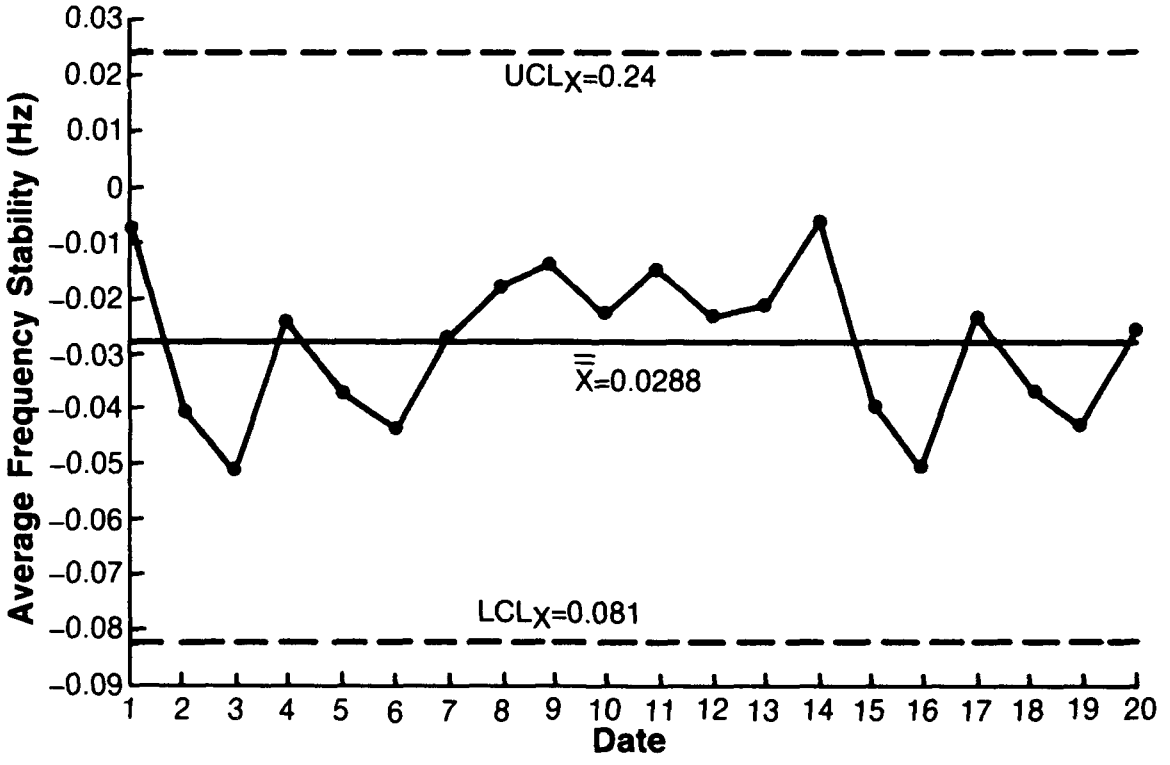


Figure 9. Frequency Stability 25 °C to 85 °C X-Bar Chart for 10 MHz Frequency Std.

[5] E. L. Grant, R. S. Leavenworth, Statistical Quality Control. New York: McGraw-Hill, 1974, pp. 1-347.

Biography of Bill Thomas

Fourteen years of industrial experience with the last six in the training and implementation of SPC, Experimental Design and other Statistical methods at Control Data Corp. and Rockwell International. Teaches numerous courses for the American Society for Quality Control and Kirkwood Community College. B.S. Degree from Western Illinois University, currently working on M.S. studies in statistics and engineering at the University of Iowa and is a Certified Quality Engineer.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A Synopsis of Quality Involvement/Improvement Programs and the Ramifications on our Industry

Charles Jensik, Ronald Zellers and Russell Lackey
Piezo Crystal Company
Carlisle Pa,
17013

Dr. Virgil Bottom, in his paper at the 35th FCS, "A History of the Quartz Crystal Industry in the USA", referred to the quartz crystal industry; "Like the North American Indian, a ward of the government, dependent upon it for financial and technical support, and struggling to stay alive to be in position to meet the challenge of the next military crisis."

Gentleman, our industry is dying. For many it is dead or will soon be. Referring to Fig. 1 which was supplied from the U.S. Department of Commerce, adjusted to reflect constant 1979 dollars, things don't look so good in our industry compared to inflated yearly dollars. In fact these adjusted numbers are most likely optimistic in that there is some offshore manufacturing included in these numbers. There is in my mind only one solution. It can be stated in somewhat different terms but they all say the same thing--Total Quality Mentality, Total Quality Management, Quality Involvement, etc.

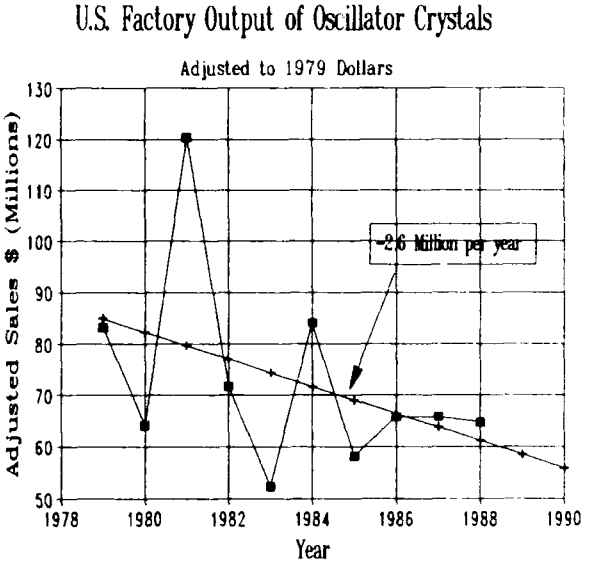


Figure 1

Identifying customers of our industry's products that really have a Quality Mentality, not "just for fun quality", was really rather easy. I would ask the purchasing person or contracts contact about his company's Quality systems or policy. If the contact just said "yes we have one" you could bet that it was just for fun. Even if the contact said that you must be preferred or approved or inspected to such and such a quality specification, it is still a good chance that it's just for fun. But if the customer says that they must work with you to assure themselves of your quality, you can be assured that they have a Quality mentality. Of the dozen or so companies that I tested only two really at this time demanded working assurances or proof of Total Quality Management. Those two companies were Motorola Inc. and Harris Corporation, specifically their Government Systems Sector. A brief description of each program is as follows:

MOTOROLA: Malcolm Baldrige National Quality Award participation and six sigma Quality are required to maintain standing as a supplier.

HARRIS : Supplier Partnership Program and subsequent preferred or best in class status to maintain standing as a supplier.

MALCOLM BALDRIGE NATIONAL QUALITY AWARD

The Malcolm Baldrige award was named after the former Secretary of Commerce in the Reagan Administration. Sigma is a statistical unit of measurement that describes the distribution about the mean of any process or procedure. A process or procedure that can achieve plus or minus SIX SIGMA capability can be expected to have a defect rate of no more than a few parts per million, even allowing for some shift in the mean.
The Malcolm Baldrige National Quality Award was

CH2818-3/90/0000-294 $1.00 © 1990 IEEE

established as a result of Public Law 100-107 "The Malcolm Baldrige National Quality Improvement Act of 1987" and signed by President Reagan on August 20, 1987. The purpose of the award is to promote quality awareness, recognize quality achievements of U.S. companies and publicize successful quality strategies. The Secretary of Commerce and the National Institute of Standards and Technology (NIST) is responsible for developing and administering the awards, with cooperation and financial support from the private sector.

The Baldrige award criteria are as follows:
1) Corporate quality leadership
2) Information and analysis
3) Planning
4) Human resource utilization
5) Quality assurance of products and services
6) Quality improvement results
7) Customer satisfaction

The examination process is as follows: Each written application is reviewed by three examiners. These examiners are quality experts, including retired quality professionals, selected from industry, professional and trade organizations and universities. High scoring applicants are selected as finalists, and must undergo a site-verification visit by one or more teams of examiners. A panel of judges reviews all data and information, and recommends award recipients. The Malcolm Baldrige National Quality Consortium, formed by the American Society for Quality Control and the American Productivity Center, administers the evaluation process.

SUPPLIER PARTNERSHIP PROCESS

The Partnership Program is a strategic initiative which allows HGSS (Harris Government Systems Sector) to focus its resources on the key suppliers. A stringent list of criteria is used to evaluate supplier performance, capability and commitment. It is as follows:

1. Quality rating for previous 4 quarters.
2. Scheduling rating for previous 4 quarters.
3. Internal program for continuous improvement.
4. Composite credit rating (D&B).
5. Consistently ships required technical data with hardware.
6. Consistently ships exact quantities as stated on Purchase Order.
7. Keeps Harris informed of production schedule, status.
8. Offers technical expertise in application of commodity.
9. Inspection system compliant to Mil-I-45208.
10. Monitors and reacts to GIDEP alert information.
11. Maintains FIFO inventory system.
12. Delivers valid SPC data with each delivery.
13. Internal standard parts program.

The objective of the program is to improve product quality, performance and decrease costs. In the case of the crystal/oscillator product area, 27 candidate suppliers were surveyed, with the "final four" identified either as "preferred" or "best in class". It is the intention of this program to then only use these suppliers. The initial reduction of supplier base allows for enhancement of HGSS communications with resulting committed suppliers. This leads to a reduction of inspection and data verification as a result of problem identification and resolution. HGSS then gets involved in new designs based on process capability. Once this happens the supplier gets involved with strategic planning and award of annual basic ordering agreements.

KEY CONCLUSIONS OF BOTH PROGRAMS

* Quality is a management leadership issue.
* Start with the basics.
* Quality is a management leadership issue.
* Implement systems and technical changes in concert with organization and cultural changes.
* Quality is a management leadership issue.
* Focus on implementing a few basic concepts.
* Quality is a management leadership issue.
* Shoot the gurus.
* Quality is a management leadership issue.
* Broaden the scope of quality improvement.
* Quality is a management leadership issue.
* Concentrate on a value-driven approach.
* Quality is a management leadership issue.
* Get training out of the classroom.
* Quality is a management leadership issue.
* Get personal involvement in the improvement process.

The basic components that are of a necessity to participate in the Baldrige or the Harris program are:

*Documentation of all processes.
*SPC utilized at every level at every work station.
*DFM applied to every aspect of design.
*MRPII utilized through-out the manufacturing area.
*CPM used on all development programs.

The results that are gotten are well worth the costs and emotional energy spent in making the cultural change to a organization. As an example, Figure 2 is a chart showing the rework, reorder costs over a period of 2 years at Piezo Crystal Company. In order to achieve this we had to get to three sigma process yields. It has taken us almost 6 years to get to this point in our development. We believe that six sigma is achievable but will take at least 6 more years. The return on investment calculations when viewed in retrospect yield a TEN fold return. The hand writing is on the wall, we must change our mentality; specifically our "just for fun" Quality mentality.

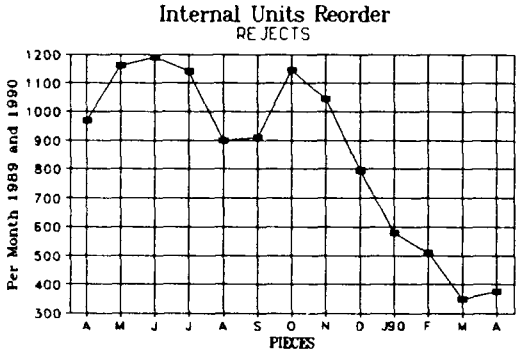


Figure 2

The universally accepted definition of Quality is compliance to customers specifications. So it is apparent that one way to change the Quality mentality of manufacturers is to change the specifications to reflect the yields of the processes to insure 3 or 6 sigma yields. The key is for the user or the manufacturer to develop specifications that will insure Quality. To build in Quality rather than inspect for Quality at the end of the production process. A typical specification would be based on the statistical capabilities of the processes that the device was manufactured to. For brevity only three characteristics are presented: resistance (Figure 3), Z-angle (Figure 4) and process lead time (Figure 5). By specifying only within these distributions 3 sigma process yields are assured.
In conclusion, from our experience at Piezo Crystal Company, the following benefits are gained by actively participating in a Quality program:

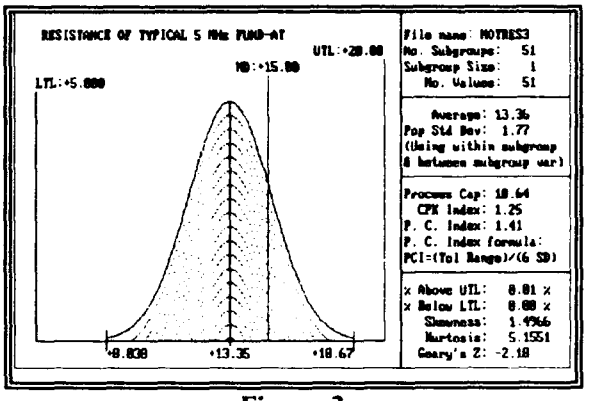


Figure 3

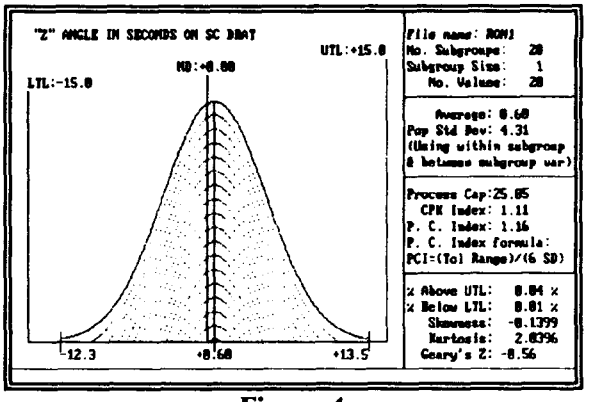


Figure 4

1. The manufacturer knows what the process capabilities are and can monitor changes to effect improvement in yields.
2. The customer gets a product that has Quality and is reproducible, not a "shop queen".
3. The manufacturer makes a profit (the ultimate goal of our society), is able to expense R&D and stay in business.

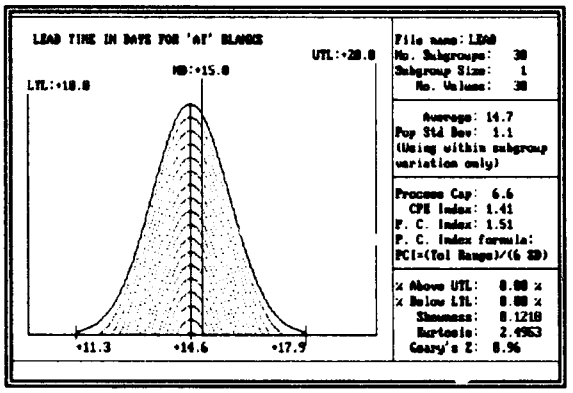


Figure 5

ACKNOWLEDGEMENTS

The author would like to thank Larry Barbary and numerous people at both Motorola and Harris Corporation.

REFERENCES

1. V.E. Bottom, "A History of the Quartz Crystal Industry in the USA", 35th Annual Symposium on Frequency Control, 3, (1981).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Raman-Nath Diffraction by Means of Thickness-Shear Oscillating AT-Cut Quartz Cristal Resonator

Sekiji YAMAGATA

Department of Physics, Asahikawa College, Hokkaido Univesity of Education Asahikawa, Hokkaido 070, Japan

SUMMARY

The Raman-Nath effect has been observed in quartz crystal resonator which is oscillating with an overtone mode. A plano-convex AT-cut quartz crystal resonator vibrating in 5th overtone thickness-shear mode, the resonant frequency is about 5 MHz, has been used as a Raman-Nath diffraction grating in this experiment.

When a laser beam is passing through the quartz resonator parallel to the thickness-shear-oscillation stress T_6', the stress creates an optical index variation inside the quartz crystal resonator. the variation functions as a diffraction grating, and the laser beam has been diffracted by the Raman-Nath effect. However, the diffraction angle of output light beams are very small, because the distance between two successive planes of maximun index is large, the diffraction phenemenon has been obseved only by using a special device. A single mode optical fiber has been used as an optical probe. It has been scanned in the output diffracted light beams for the measurement of diffraction patterns. The optical probe has been able to detect the diffraction light beams as modulated light beam with resonant frequency of the resonator.

This mechanism can be applied to the light beam spritter system and other optical wave guides, and a light-modulated quartz crystal oscillator.

INTRODUCTION

The princeples of acousticall devices for the light modulation are based on the scattering of light by sound waves. Refraction and diffraction effects can be observed when light beam passes through a transparent material at right angle to a high- frequency sound field propagation in the same medium.

The diffraction into a series of order of light incident normally on an acoustic diffraction grating in a liquid medium is termed the Debye-Sears effect.[1] The equivalent effect in a solid medium had been studied by Raman and Nath, energetically [2-6]. When a quartz crystal resonator which is vibrating in an overtone thickness-shear-oscillating frequency is used for a

CH2818-3/90/0000-298 $1.00 © 1990 IEEE

solid diffration grating, also Raman-Nath diffraction can be observed. Because th acoustic standing wave in the resonator creates a peiodic refractive index variation inside the quartz resonator.[7)]

A plano-convex AT-cut quartz crystal resonator (p.c.a. resonator) oscillating with the 5th overtone oscillation resonant frequency of the thickness-shear mode has been used in this experiment. Its resonant frequency is about 5MHz and the stress distributions inside the resonator have been studied in detail[8-10)].

SOME CONDITIONS FOR RAMAN-NATH EFFECT

There are some conditions if a quartz crystal resonator is used for the Raman-Nath diffraction grating. In generally, Sander experimentally confirmed the light intensities predicted by the Raman-Nath theary for both progressive and stationary wave[11)]. We can introduce to a quantitative condition from the Sander's confimation.

$$Q = \frac{2\pi}{n_0} \frac{L\lambda}{\Lambda^2} \qquad (1)$$

The Raman-Nath diffraction can be observed when $Q < 1$. In equ. (1), L is the pass length of a light beam in the medium, λ is the wave length of light, Λ is the wave length of the acoustic wave in the medium (it shows a distance between maximum refractive indices perpendicular to the light wave front), and n_0 is the refractive index of the medium. Fig. 1 shows the diffraction model with the acoustic wave in the p.c.a.resonator. If a 5th overtone of the thickness shear oscillation of it, the resonant frequency is about 5MHz. A He-Ne gaseous laser's wave length λ is 6.33×10^{-7}m, $L = 1 \times 10^{-2}$m, $\Lambda = 3.7 \times 10^{-4}$m, and $n_0 = 1.3$, then, Q-number becomes about 0.8, so the Raman-Nath effect can be achieved.

PERIODIC REFFACTIVE INDEX VALIATIONS IN THE QUARTZ CRYSTAL

The refractive index of the quartz crystal is specified by the indicatrix, which is an ellipsoid whose coefficients are the component of the relative dielectric impermeability, B_{ij}. The small change of refractive index produced by elecric field and mechanical stress is a small change in the shape, size and orientation of the indicatrix. This change is most conveniently specified by giving the small changes in the coefficients B_{ij}. When an electric field E_k and a stress T_{mn} are applied to the quartz crystal, small changes ΔB_{ij} in the B_{ij} are given by

$$\Delta B_{ij} = Z_{ijk} E_k + \pi_{ijmn} T_{mn} \qquad (2)$$

where Z_{ij} and π_{ijmn} are the electro-optical and the piezo-opticl coefficients, respectively. Note $B_{ij} = n_{ij}^2$, we can obtain

$$B_{ij} + \Delta B_{ij} = (n_{ij} + \Delta n_{ij})^{-2} \qquad (3)$$

where n_{ij} is a component of n_0 in the ij-

direction. Thus, in general, $\Delta n_{ij} \ll n_{ij}$, Δn_{ij} shows the small change of n_{ij}. Equ. (3) is reduced to

$$\Delta n_{ij} = -\frac{1}{2} n_0^3 \Delta B_{ij} \qquad (4)$$

and we rearrange (4) by substituting (2) into

$$\Delta n_{ij} = -\frac{1}{2} n_0^3 (z_{ijk} E_k + \pi_{ijmn} T_{mn}) \qquad (5)$$

The periodic refractive index variations are [12,13]

$$n_\perp = n_{ij} - \frac{1}{2} n_0^3 \left(z_{ijk} E_k + \pi_{ijmn} T_{mn} \right)$$

$$n_{\parallel} = n_{rs} - \frac{1}{2} n_0^3 \left(z_{rsk} E_k + \pi_{rsmn} T_{mn} \right) \qquad (6)$$

where $n_\perp$ and $n_{\parallel}$ are refractive indices at right angle each other inside the p.c.a. resonator. The stress T_{mn} can be genereted by the electric field E_k in the quartz crystal. When the field is alternating and its frequency ω_k coincides with the mechanical resonant frequency of the quartz crystal resonator, the photo-elastic effect becomes very large and the electro-optical effect is negligible. Then, ander the condition, Equ. (6) are rewritten as

$$n_\perp = n_{ij} - \frac{1}{2} n_0^3 \pi_{ijmn} \varepsilon_{kmn} E_k' \cos \omega_k t$$

$$n_{\parallel} = n_{rs} - \frac{1}{2} n_0^3 \pi_{rsmn} \varepsilon_{kmn} E_k' \cos \omega_k t \qquad (7)$$

In those Equ, the alternate stress T_{mn}' is replaced by the alternate electrical field $E_k = E_k' \cos \omega_k t$. When the p.c.a.resonator is oscillating in thichness-shear mode, the shear stress T_{12}' (T_6') is main stress. Here, we can introduce the stress-optical coefficients to the alternate thickness-shear stress T_6',

$$C_6(\Omega) = \pi n_0^3 \pi_{6mn} / \lambda \qquad (8)$$

where λ is the wave length of the light beam and Ω is an incident light beam angle from the z'-axis of AT-cut coordinate. The coefficients $C_6(\Omega)$ parallel to the (x', z') plane of AT-cut coordinate has been calculated[14-15]. It is shown in Fig. 2. In Fig. 2, the refractive index variation becomes maximum change when the incident angle of the light beam is chosen at $\Omega = 30°$. Fig. 3 shows the illustration of the relationship of between incident light beam and the AT-ct quartz crystal plate. Refractive indices $n_{\parallel}$ and $n_\perp$ are parallel and perpendicular components to the thickness-shear stress T_6'. These indices are most effectively changed by the stress T_6' when the incident angle of the light beam is chosed at 30° from z'-axis on the (x' ,z') plane. The alternate (vibrating) thickness-shear stress T_6' makes the periodic refractive index layers inside the p.c.a.resonator when the resonator is oscillating in 5th overtone oscillation. This aspect has been shown in Fig. 1.

PLANO-CONVEX AT-CUT QUARTZ CRYSTAL(p.c.a.) RESONATOR USED FOR RAMAN-NATH GRRATING

The p.c.a.resonator used for the Raman-

Nath modulator and its mounting aspect are illustrated in Fig. 4 (a),(b). Part of the resonator's circumference is cut at an angle of 30° against the z'-axis in order to pass the laser beam. The stress -optical coefficient $C_6(\Omega)$ is maximum value when the laser beam is passing through the p.c.a.resonator at the angle. The $C_6(\Omega)$ coefficient has been calculated and it is shown in Fig. 2. It was confirmed that cutting does not effect the thickness -shear oscillation mode in the resonant frequency of the p.c.a.resonator[14].

Vibration modes and thickness-shear stress distributions of the p.c.a.resonetor have been calculated and measured, too, when the p.c.a.resonator is oscillating in the fundamental, 3rd, and 5th overtone oscillation frequency, respectively[7,8,9,14,16]. Fig. 5 shows the relative calculated displacements of the p.c.a.resonator. And then, the thickness-shear stress T_6' distribution of the 5th overtone mode along the y'-direction is shown in Fig.6. Doted lines and solid lines in this Fig. are calculated values and measured values, respectively. The refractive index inside the p.c.a.resonator may be corrugated by the $|T_6'|$ stress distribution parallel to the (x', z') surface of the p.c.a.resonator, this phenomenon has been shown in Fig. 1. This effect can be applied to Raman-Nath difraction grating.

RAMAN-NATH DIFFRACTION THEORIES WITH p.c.a.RESONATOR

Consider a plane wavefront of laser beam of wave length λ and anglar frequency ω incident normally on the p.c.a.resonator, it has shown in Fig. 1. In this Fig., L is a light pass length in the resonator. A thickness-shear acoustic wave exists at right angle to the incident light beam. The stress $|T_6'|$ is generated inside it. Acording to Equ.(7), the $|T_6'|$ stress distribution makes up the refractive index corrugation shown as Fig. 1. If the distance between two successive planes of maximum density of the refractive index is Λ, Equ.(7) are rewritten as

$$n_\perp = n_{ij} - \frac{1}{2} n_0^3 \pi_{ij6} \varepsilon_{26} E_2' \cos\omega_k t \cdot \sin\frac{2\pi}{\Lambda} y'$$

$$n_{//} = n_{rs} - \frac{1}{2} n_0^3 \pi_{rs6} \varepsilon_{26} E_2' \cos\omega_k t \cdot \sin\frac{2\pi}{\Lambda} y' \qquad (9)$$

where y' is the position of the AT-cut coordinate shown in Fig. 3, and ω_k is the angler frequency of the alternate electric field applied to y' surface. Ander the condition k is equal to 2. We can conclude that the p.c.a.resonator oscillating in thickness -shear overtone frequencies is ableto use as a diffraction grating devices.

The refractive grating will thus give rise to diffraction order consisting of a zero order and first order at angles ϕ given

by

$$\sin \phi = \pm \frac{\lambda}{\Lambda} \qquad (10)$$

The real situation differs from the ideal in that other diffraction orders are presented, given by[17,18]

$$\phi = \frac{N\lambda}{\Lambda}, \quad N=0, 1, 2, 3, \qquad (11)$$

when the diffraction angle ϕ is very small.

An approximate relationship for the relative intensities in the various orders has been taken. Specifically, the intensity I_n in the Nth order is given by

$$I_n = J_n^2[\psi_m(y')] \qquad (12)$$

where J_n is the Nth-order Bessel function and $\psi_m(y')$ is the peak light deviation.

In this experiment, $\Lambda = 3.7\times10^{-4}$m and $\lambda = 6.33\times10^{-7}$m, then the first order diffraction angle ϕ_1 becomes 1.71×10^{-3} rad. (0.1 deg.).

EXPERIMENTS AND RESULTS

1. Experimental Apparatus

The diffraction angle ϕ of the output light beams will be very small becouse $\lambda \ll \Lambda$ in this experiment. The first-order diffraction angle has been estimated as 1.71×10^{-3}rad.(0.1deg.) when the p.c.a. resonator is oscillating in the 5th overtone resonant frequency. Since the diffraction angle is very small and diffraction patterns is flashing, like a stroboscope, we can not observe the diffration phenomenon by an ordinary lens system. The frashing frequency is equal to the resonant frequency of the p.c.a.resonator used for the diffration grating.

Therefore, a single mode optical fiber has been used for the detection probe of diffraction light beams. The construction of the experimental apparatus is shown in Fig.7 . The probe fiber may have only small scope from the output diffracted light beam becouse the single mode fiber's numerical aperture (NA) is very small[19]. The diffraction output light beams catched by the probe fiber are guided to a P.M.(photomutiplier) and changed to the electric signal which has same frequency of the p.c.a.resonator's resonant frequency. A tunig amplifier selects only the electrical signal which has the flashing modulated frequency from output signals. The amplitude of those signal from the amplifier are directly proportional to the intensity of the diffracted light beams given by Equ. (12). The diffraction patterns are recorded on the X-Y plotter when the top of the probe fiber is scanning along the y'-direction of AT-cut coordinate. The position of the probe fiber is detected by a differential transformer shown in Fig.7.

2. Experimental Results of Diffraction

Diffraction patterns of the otput light beam has been measured when the distance of between the p.c.a.resonator and the top of probe fiber, namely RP, was changed. Fig.8

shows diffraction patters of the output light beams. The diffraction patterns have been taken when the distance RPs was at 1, 5,10, 50, and 100 mm.

We can see that, from the measured diffraction patterns in Fig.8, the diffraction patters are noisy and the degree of light modulations is small when the distance RP is shorter than 10mm. For the reason, the condition of Franhofer diffraction should be not satisfied when the probe fiber is plased closely to the diffraction grating (p.c.a.resonator)[20]. The Diffraction phenomenons have been measured crialy when the distance RPs was chosen larger than 10mm.

APPLICATION TO THE LIGHT MODULATED QUARTZ CRYSTAL OSCILLATOR

This diffraction phenomenon is applicable to a new quartz crystal oscillator system, namely the Light Modulated Quartz Crystal Oscillator[14,21]. One of the diffracted output light beams recived by the probe fiber and the P.M. is transmitted to a ac amplifier, the output of the amplifier is lead to the electrodes of the p.c.a. resonator by a feedback loap and the optical-electric quartz crystal oscillator system is self-resonant. This oscillator system is shown in Fig.9.

Effectiveness of the light modulation can be estimated by measuring the threshold gain of the ac amplifier when the self-oscillation occurs. In Fig.9, μ and β are the gain of the amplifier and the light modulator, respectively. The product of $|\mu|$ and $|\beta|$ is equal to unity when self-oscillation occurs in this system. Since the value of $|\mu|$ is know, the gain $|\beta|$ may be obtained. The threshold gain $|\mu|$ of the ac amplifier is shown in Fig.10 when the distance RPs are changed. Fig.11 shows the output voutage of the light modulated oscillator when the amplitude of the ac amplifier is holded at 45 dB. From the experiment of self-oscillation, the gain of the light modulation was -36 dB in the 5th overtone frequency oscillation when RP=9 cm.

One of the important factors for the stability of a quartz crystal oscillator is the electro-mechanical energy coupling factor (e.m.e.c.factor)[14]. The e.m.e.c. factor can be made very small when the p.c.a.resonator is using as diffraction grating. If the e.m.e.c.factor is made very small by separating the excitation electrodes from the surfaces of the quartz plate, the oscillator system does not oscillate with the factor depending on other electric circuit constants of the oscillator system because the quartz plate used in this system works as a three-port network even though it is generally used as a two-port network in a standard quartz crystal osillator system. Therefore, the quartz crystal plate oscillates very near its purely mechanical resonant frequency. It may be concluded

from this reason that an oscillator of the present type will possibly have good frequency stability with time[14,21].

CONCLUSION

In this experiment, the overtone oscillating p.c.a.resonator can be used for the Raman-Nath diffraction grating. When the distance between the top of probe fiber and the p.c.a.resonator is lager than 10 mm, detected diffraction patterns become fine and clear. This phenomenon will be applicable to the high stability quartz crystal oscillator system. A trial oscillator has been oscillated in this experiment.

By the way, since the thickness of the quartz plate is thin, its resonant frequency becomes high and the acoustic wave length inside the resonator becomes very short, so the diffraction angle will be larger. It will be applied to the beam splitter of the optical communication system and other optical wave guide.

ACKNOWLEDGEMENT

Mr. A.Akazawa and Mr. R.Yamashita, students of our university, co-operate with the experiment.

REFERENCES

1. Debye,P., Sears,F.W., Proc. Natl. Acad. Sci., Washington, 18(1932),pp.409-414.
2. Raman,C.V., and Nath,N.S.N.,Proc. Indian Acad. Sci. 2A(1935), pp.406-412.
3. Raman,C.V., and Nath,N.S.N.,Proc. Indian Acad. Sci. 2A(1935), pp.412-420.
4. Raman,C.V., and Nath,N.S.N.,Proc. Indian Acad. Sci. 3A(1936), pp.75-84.
5. Raman,C.V., and Nath,N.S.N., Proc. Indian Acad. Sci. 3A(1936), pp.119-125.
6. Raman,C.V., and Nath,N.S.N.,Proc. Indian Acod. Sci. 3A(1936), pp.459-465.
7. Yamagata,S., Nara,S.,and Fukai,I., Microwave, Optics and Acoustics, IEE England, 3(1979), 6, pp.265-271.
8. Yamagata,S., Fukai,I., and Yasuda,I., Trans. IECE,Japan 60-A(1977),12, pp.1122-1129.
9. Yamagata,S., Yamamoto,K., Fukai,I., and Yasuda, I.,Elctro. Letter, IEE England, 14, 14(Jul.1979) pp.450-451.
10. Yamagata,S., The 10th Sympo. Ultrason. Electro., Japan, Nov.1989, pp.63-64.
11. Sander,H.F., Canad. J. Rev. A14(1963).pp.158-171
12. Mason,W.P., Crystal Physics of Interaction Process, Academic Press 1966, pp.157-200.
13. Nye,J.F., Physical Properties of Crystals, Oxford Press 1972, pp.235-258.
14. Yamagata,S., Fukai,I., and Yasuda,I., Trans. Sonic. & Ultrason.,SU-25(1978), pp.192-198.
15. Yamagata,S., Journal of Hokkaido Univ. of Educ. 2A, 29(1978), 1,pp.15-30.
16. Yamagata,S., Nara,S., Fukai,I., and Yasuda,I., Trans. IECE Japan,62-A(1979), 7, pp.436-443.
17. Narasimhamurty, Photoelastic and Electro-Optic, Plenum Press 1981, pp.160-186.
18. Walter,G.D., Hand Book of Optics, McGrow Hill

19. Forrest M. Mims III, A practical Introduction Light Wave Communications, Howard W. Sams & Co., Inc. 1982, pp103-108.

20. Ghatak,A., Optics, Mc Graw Hill 1977, pp325-330.

21. Yamagata,s., Electro. Letters, IEE England 19,19(Sept. 1983), pp.795-797.

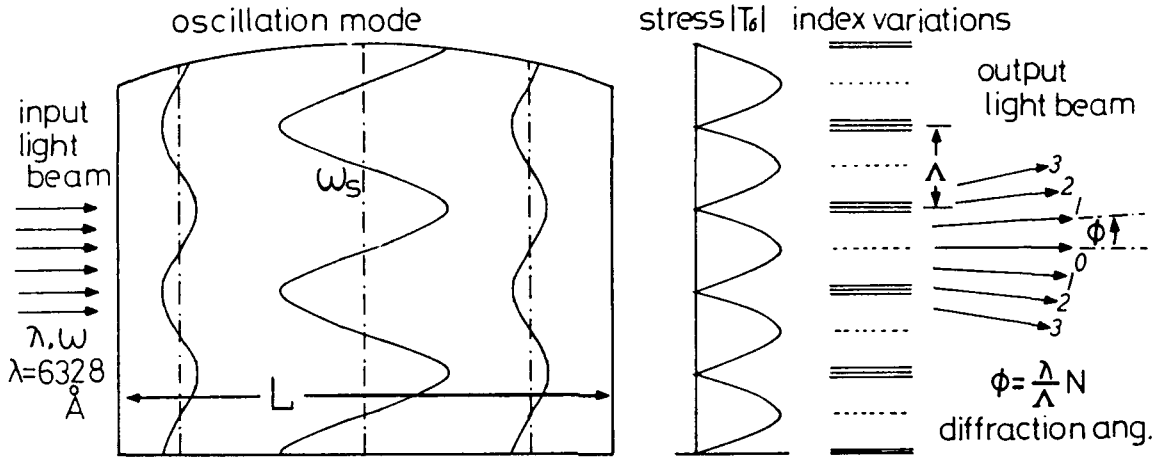


Fig.1 Diffraction grating model of plano-convex AT-cut quartz crystal resonator by thickness shear stress $|T_6'|$.

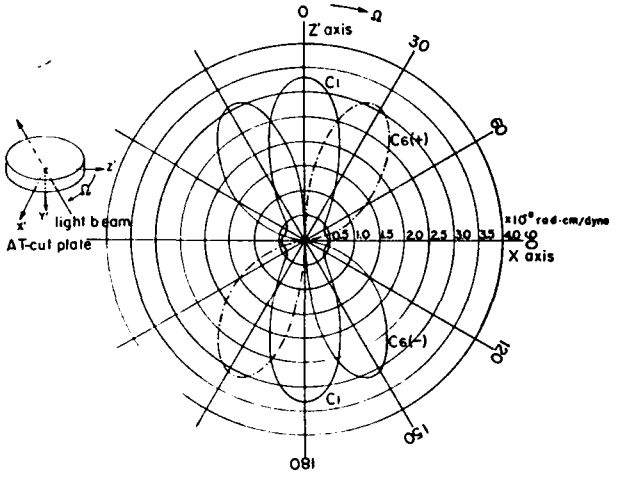


Fig.2 Stress-optical coefficient, C_1' and C_6' for T_1' and T_6', respectively. Angle Ω is measured from z' axis.

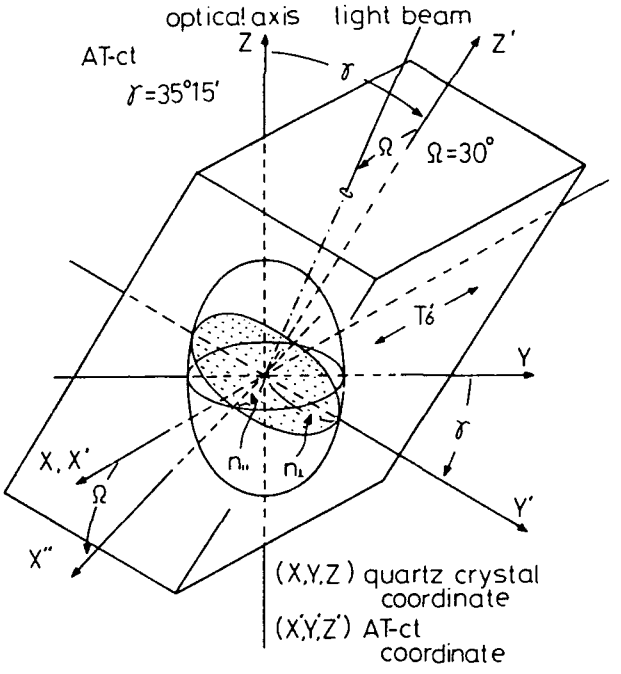


Fig.3 Indicatrix ellipsoid of AT-cut quartz crystal plate, $n_\perp$ and $n_\parallel$ are refractive indices perpendicula and parallel to stress T_6'.

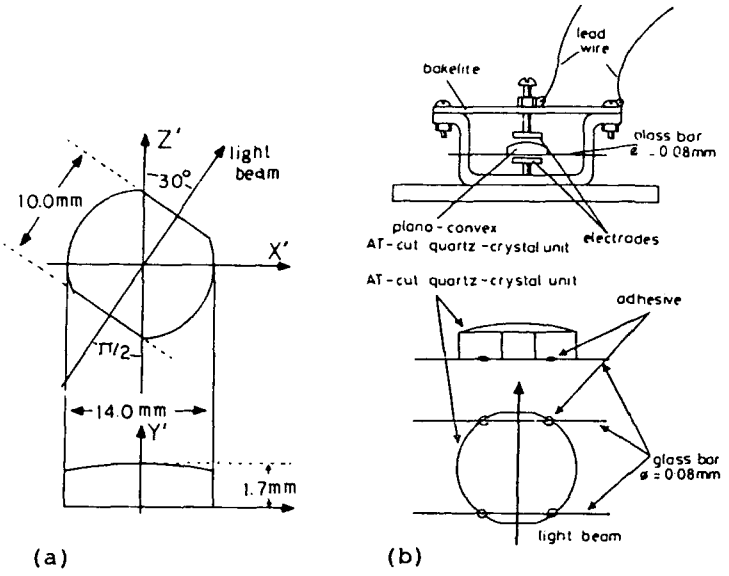


Fig. 4 (a) Plano-convex At-cut quartz crystal resonator using for diffraction grating in this experiment. (b) Glass-bar-mounted quartz crystal resonator using for diffraction experiment.

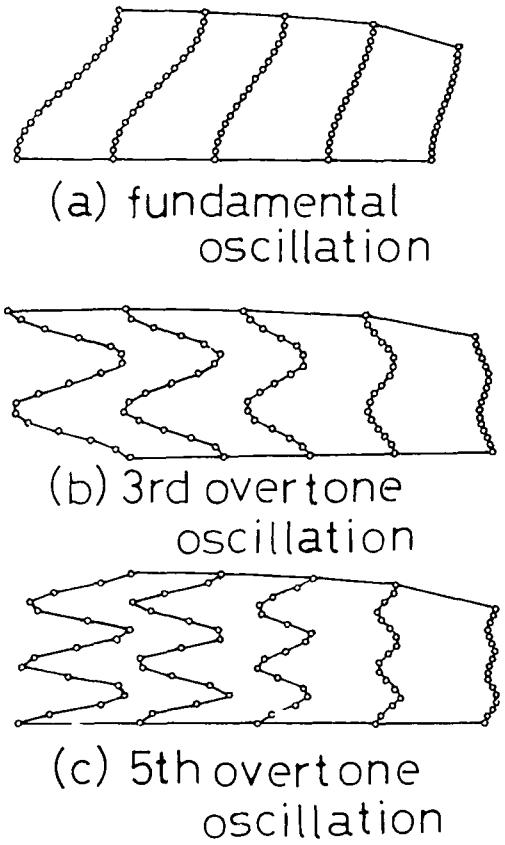


Fig.5 Displacement of thickness-shear vibration modes calculated for plano-convex AT-cut resonator in x'-y' surface.

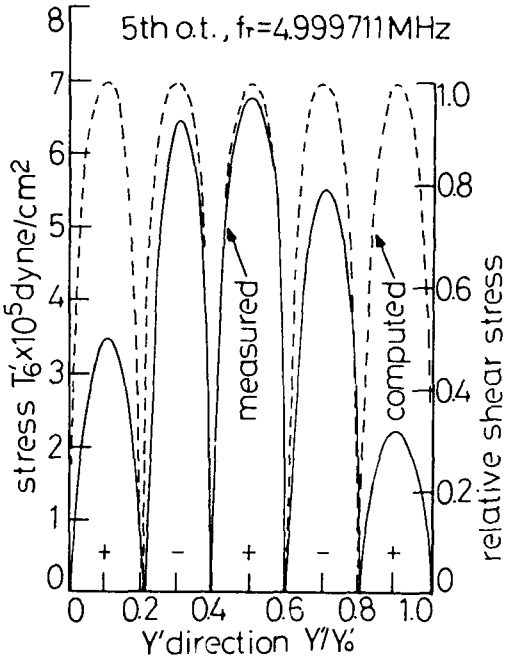


Fig.6 Thickness-shear stress distribution of plano-convex AT-cut quartz crystal resonator oscillating in 5th overtone resonant frequency. Solid lines and dotted lines show measurement and calculation values, respectively.

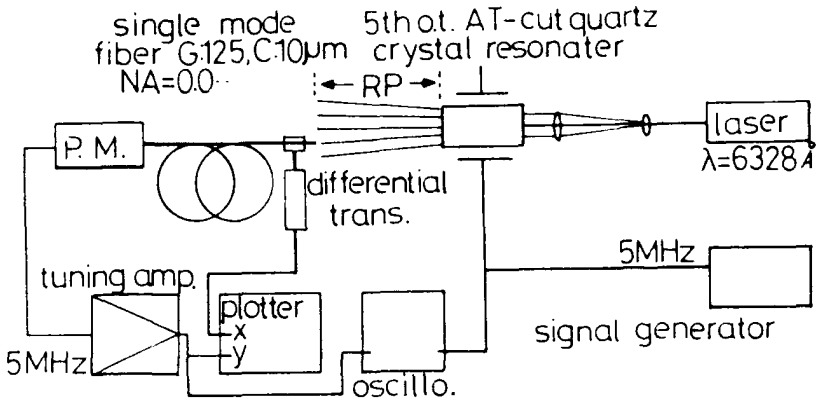


Fig.7 Experimental apparatus.for measurement of diffraction patterns.

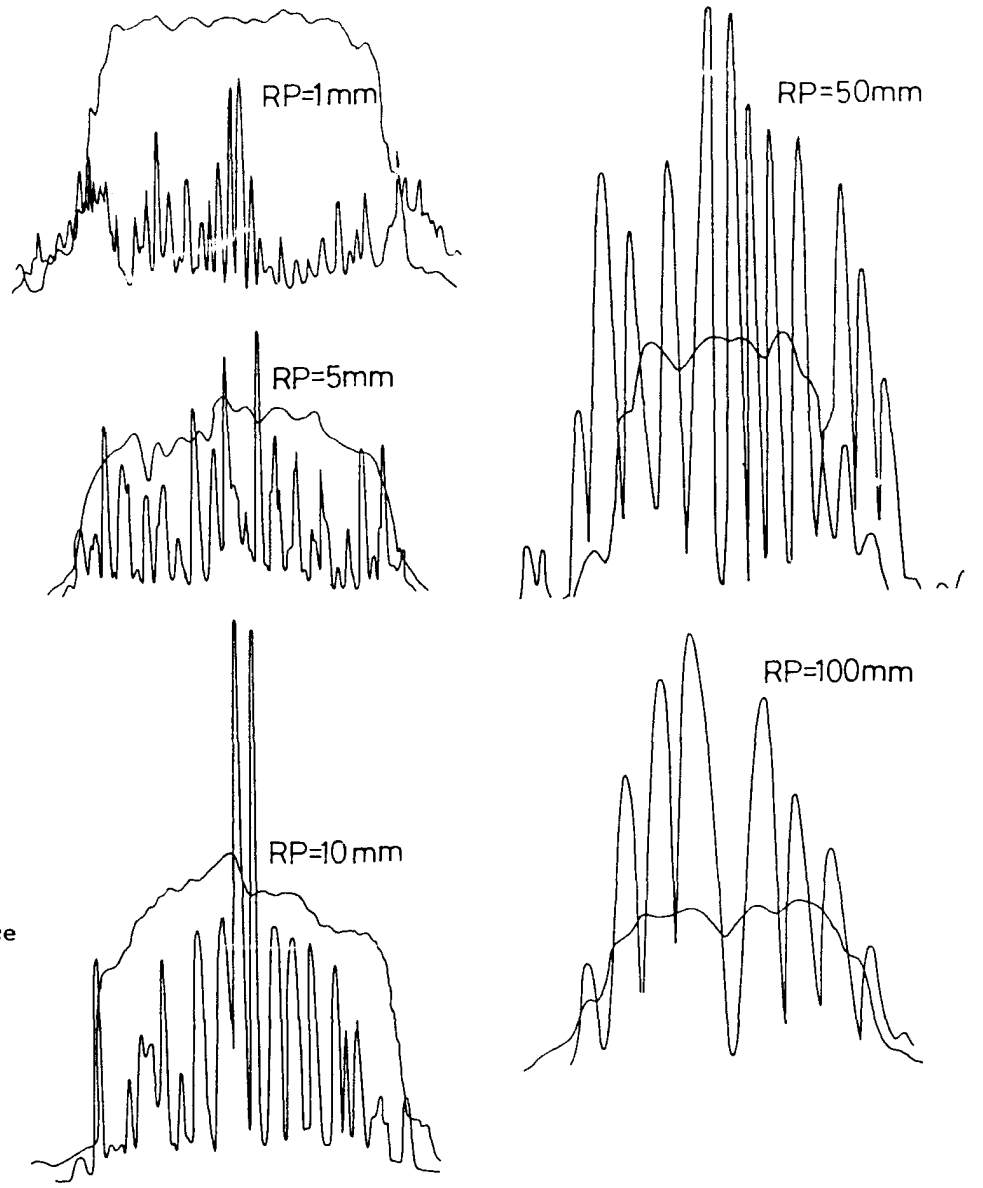


Fig.8
Measured diffractin patterns when distance between top of plobe fiber and resonator, RP, is changed.

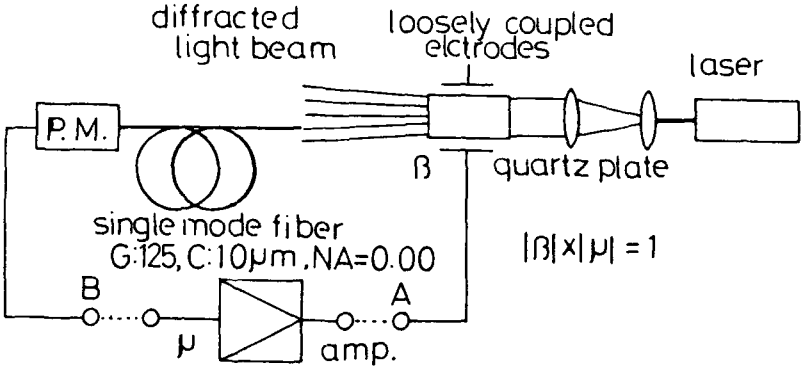


Fig.9
Diagram of quartz crystal oscillator by light modulaton system.

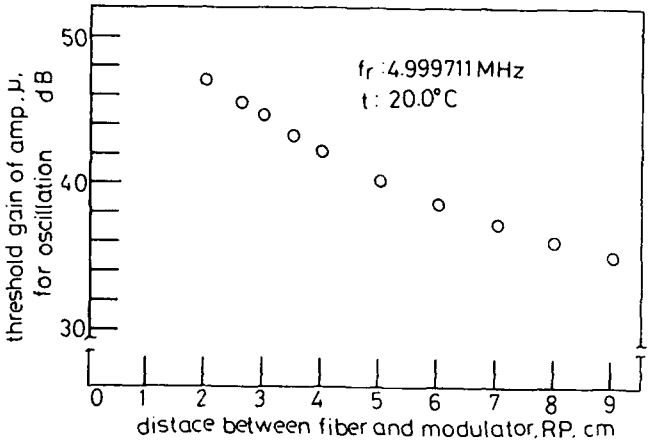


Fig.10 Threshold gain of ac amp. in this oscillator when RP distance was changed.

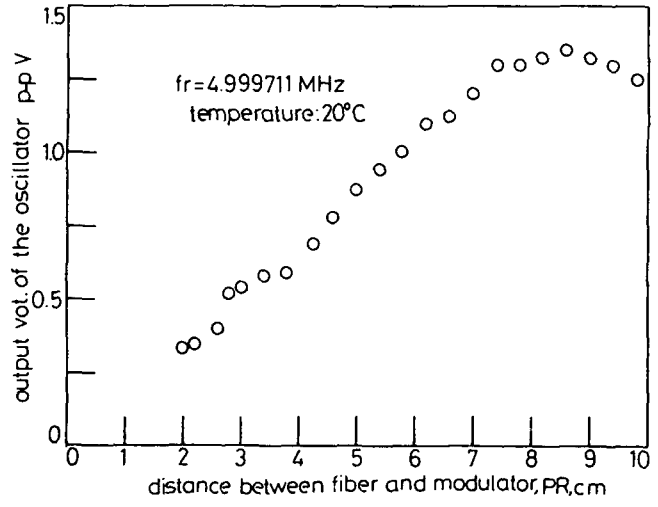


Fig.11 Output voltage from light light modulated oscillator when RP distance was changed.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SIMULATION NETWORKS FOR PIEZOELECTRIC BIMORPH ACTUATORS

Jan G. Smits[a] and Arthur Ballato[b],

a) Department of Electrical Engineering
Boston University
Boston, MA 02215
b) U.S. Army Laboratory Command
Electronics Technology and Devices Laboratory
Fort Monmouth, New Jersey 07703-5302

ABSTRACT— **The constituent equations for piezoelectric bimorphs are written in terms of redefined extensive and intensive variables which makes it possible to write the bimorph matrix in numerical form. This numerical matrix is diagonalised to give the eigenvalues of bimorphs. The modal matrix is used to find the corresponding eigenstates. A physical interpretation of bimorph eigenstates is proposed. Two realisations of network representation are given for the admittance matrix. The eigenvalue matrix is inverted to give the impedance matrix. The network representation of the impedance matrix is given.**

1 Introduction

Piezoelectric bimorphs have already a long history, they were invented by C. Baldwin Sawyer [1] in the late twenties. They were used in large quantities in grammophone recorders, [1], [2], as speaker element in telephone handsets, [3], [4], as optical chopper [5] or scanner [6], as pump element [7], [8], [9,] to deform mirrors to change the focal length [10], to tilt the angle of magnetic recorders, to reach an optimum reading and writing speed [11]

They have been used as motor elements [12], as element to extract wind energy from rotating Savonius rotors [13], as accelerometers [14], [15], [32]; as element in the control of vibration of mechanical constructions [16], [17], as gripper [18], [19], and as ultrasonic motor element [20], [21]. The latest class of applications is in Scanning Tunnelling Microscopy, where a bimorph serves as an element to move the pointed probe over the surface of the object under study [22], [23], [33].

The motion of the bimorph is small, but a clever design, and use of resonance may enhance the tip deflection considerably, as is clear from the commercial application of bimorph driven cooling fans [24], [31].

Network models of piezoelectric bimorphs have been proposed by Germano,[25], but these models do not take complete account of the effect of the two beams on each other, and are therefore not completely accurate. Also, if more than one kind of boundary condition exists, these models are not easily adapted.

In a previous paper [26] the constituent equations of piezoelectric bimorphs have been presented. As these devices are applied in industrial equipment as well as in consumer goods, it is useful to cast the equations in network representation, so that they are available to electrical engineers.

2 The bimorph matrix and its inverse

In the following text a bimorph is supposed to be connected to a rigid clamping support on its left side, while its right side is free. The bimorph matrix B relates the extensive variables M, F, p and V, respectively a moment at the end, a force at the end, a uniform load over the entire bimorph and the voltage over the electrodes, to the intensive variables $\alpha, \delta, \mathcal{V}$, and Q, the slope and the deflection at the end, the displaced volume and the charge on the electrode. The bimorph has a length L, and a width w, the thickness of the individual components is h. In [26] it has been derived that the relation between extensive and intensive parameters can be written as

$$\begin{pmatrix} \alpha \\ \delta \\ \mathcal{V} \\ Q \end{pmatrix} = \begin{pmatrix} \frac{3s_{11}^E L}{2wh^3} & \frac{3s_{11}^E L^2}{4wh^3} & \frac{s_{11}^E L^3}{4h^3} & \frac{-3d_{31}L}{4h^2} \\ \frac{3s_{11}^E L^2}{4wh^3} & \frac{s_{11}^E L^3}{2wh^3} & \frac{3s_{11}^E L^4}{16h^3} & \frac{-3d_{31}L^2}{8h^2} \\ \frac{s_{11}^E L^3}{4h^3} & \frac{3s_{11}^E L^4}{16h^3} & \frac{3ws_{11}^E L^4}{40h^3} & \frac{-d_{31}wL^2}{8h^2} \\ \frac{-3d_{31}L}{4h^2} & \frac{-3d_{31}L^2}{8h^2} & \frac{-d_{31}wL^2}{8h^2} & \frac{\epsilon_{33}^T Lw}{2h}(1-k_{31}^2/4) \end{pmatrix} \begin{pmatrix} M \\ F \\ p \\ V \end{pmatrix}$$

It should be noted that B has been written here as the matrix for an antiparallel or inward series bimorph. For a parallel bimorph some of the numerical coefficients change: the purely elastic terms remain identical, the piezoelectric terms are doubled, and the dielectric term is multiplied by four. So the bimorph matrix for that case can be found by multiplying cor-

US GOVERNMENT WORK IS NOT PROTECTED
BY US COPYRIGHT

responding coefficients by the following numbers:

$$\begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 1 & 1 & 1 & 2 \\ 2 & 2 & 2 & 4 \end{pmatrix} \tag{2}$$

In the bimorph matrix equation we define the vector which contains the external parameters M, F, p, V as the generalized force vector $\mathcal{F}$ and the vector which contains the device responses as the generalised displacement vector $\mathcal{D}$.

Equation (1) can now be written as :

$$\mathcal{D} = B \quad \mathcal{F} \tag{3}$$

The generalised force and deflection vectors can be redefined as the generalized current and voltage vector $\vec{\mathcal{I}}$ and $\vec{V}$, where

$$\vec{\mathcal{I}} = \begin{pmatrix} \alpha L \\ \delta \\ \mathcal{V}/wL \\ Qk \end{pmatrix} \quad \text{and} \quad \vec{V} = \begin{pmatrix} M/L \\ F \\ pwL \\ V/k \end{pmatrix} \tag{4}$$

in which $k = 4s_{11}^E L\hat{n}/2d_{31}wh$, and $\hat{n}$ is an arbitrary dimensionless number that can be set to unity, that is carried along to indicate that where it occurs, there is a piesoelectric term, and where it occurs squared, there is a dielectric term. With the further definition of $b = s_{11}^E L^3/w(2h)^3$, we obtain:

$$\vec{\mathcal{I}} = \begin{pmatrix} \alpha L \\ \delta \\ \mathcal{V}/wL \\ Qk \end{pmatrix} = b \cdot Y \begin{pmatrix} M/L \\ F \\ pwL \\ V/k \end{pmatrix} = b \cdot Y\vec{V} \tag{5}$$

in which the normalized admittance matrix Y is:

$$Y = \begin{pmatrix} 12 & 6 & 2 & -12\hat{n} \\ 6 & 4 & \frac{3}{2} & -6\hat{n} \\ 2 & \frac{3}{2} & \frac{3}{5} & -2\hat{n} \\ -12\hat{n} & -6\hat{n} & -2\hat{n} & 4a\hat{n}^2 \end{pmatrix} \tag{6}$$

In this matrix we have defined

$$a = \left(\frac{2}{k_{31}}\right)^2 - 1 \quad \text{with} \quad k_{31}^2 = \frac{d_{31}^2}{s_{11}^E \epsilon_{33}^T} \tag{7}$$

The determinant of Y is:

$$\det(Y) = \frac{4\hat{n}^2(a-3)}{5} = (\det Z)^{-1} \tag{8}$$

The determinant of Y is positive definite, because the total amount of energy stored in the bimorph is

$$U = \frac{1}{2}\mathcal{D}^t \cdot \mathcal{F} \tag{9}$$

or using (3) and the symmetry of B:

$$U = \frac{1}{2}\mathcal{F}^t \cdot B \cdot \mathcal{F}. \tag{10}$$

This amount cannot be negative, therefore the determinant is always positive (although it can be zero in limiting cases) [27].

This implies that $a \geq 3$ or $k_{31}^2 \leq 1$. The inverse of Y is Z:

$$Z = \begin{pmatrix} \frac{3a-8}{4(a-3)} & -3 & 5 & \frac{\hat{u}}{4(a-3)} \\ -3 & 16 & -30 & 0 \\ 5 & -30 & 60 & 0 \\ \frac{\hat{u}}{4(a-3)} & 0 & 0 & \frac{\hat{u}^2}{4(a-3)} \end{pmatrix} \tag{11}$$

where $Y \cdot Z = 1$ and $\hat{n}\hat{u} = 1$.

The occurence of zeros in the (2,4), (3,4), (4,2) and (4,3) positions deserves a comment. At first it seems unlikely that these zeros can occur in the inverse matrix, as they relate the deflection to a force and the displaced volume to a load on the bimorph. Let us consider the zero in the (4,2) position. In itself this zero relates a voltage V to the deflection δ that is its cause. Given that we have calculated the relation between the force and the charge in equation (1), and that this relation is nonzero, it seems impossible that the coefficient in the inverse relation would be zero. Here we have to remark that in equation (1) the elements in the right side vector are all independent. The elements in the left side vector are partially interrelated. The slope α is the derivative of the deflection δ and the displaced volume $\mathcal{V}$ is the integral of the deflection over the bimorph. We now realise that for instance the equation that is associated with the fourth row of Z is to be read as:

$$V/k = \left(\frac{\hat{u}}{4(a-3)}\alpha + 0\delta + 0\mathcal{V}/wL + \frac{\hat{u}^2}{4(a-3)}Qk\right)/b \tag{12}$$

Equation (12) tells us that if a certain deflection δ is applied and an associated volume displacement $\mathcal{V}$, but there is no slope α at the tip, that then the voltage at the electrodes is zero. But we *know* that: in a case as this the top element undergoes the same elongation as the bottom element and consequently there is no voltage difference between the top and bottom electrodes.

Alternatively the zero in the (2,4) position teaches us that there cannot exist a force on the device that

generates a voltage, while at the same time the slope, deflection and displaced volume are zero. We had not expected otherwise.

3 The Eigenvalues and Modal Matrix of the Bimorph Matrix

The matrix Y can be diagonalised to Y_d by a transformation matrix M_d. The transformation is as follows:

$$Y_d = M_d^t \cdot Y \cdot M_d \tag{12}$$

where the superscript t denotes transposition and in which

$$Y_d = \begin{pmatrix} 12 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/60 & 0 \\ 0 & 0 & 0 & 4(a-3)\hat{n}^2 \end{pmatrix} \tag{13}$$

and

$$M_d = \begin{pmatrix} 1 & -1/2 & +1/12 & +\hat{n} \\ 0 & 1 & -1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14a}$$

while

$$M_d^{-1} = \begin{pmatrix} 1 & 1/2 & 1/6 & -\hat{n} \\ 0 & 1 & +1/2 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{14b}$$

Alternatively the matrix Y can be diagonalised to a matrix which has eigenvalues along its main diagonal. The numerical values of the eigenvalues are the solutions of the equation:

$$(Y - \lambda I) = 0 \quad \text{or} \quad \det(Y - \lambda I) = 0. \tag{15}$$

These eigenvalues are difficult to compute in general, because they are the solutions of a fourth order equation, of which the coefficients depend on the piezoelectric term a. In the extreme case that the material is either not piezoelectric at all, or has a piezoelectric coupling factor $k_{31} = 1$, the resulting matrices Y and Z become somewhat simpler and it is possible to find the eigenvalues.

Once the eigenvalues have been found, the corresponding eigenstates X_λ are easy to find, namely by substituting these eigenvalues one by one in the matrix $Y - \lambda_i I$ and developing the determinants of the minors of the (arbitrarily chosen) first row [28]. The arrangement of the four eigenstates in a four by four matrix is the modal matrix M_λ, with which we can transform the generalised current and voltage vectors $\vec{I}$ and $\vec{V}$ so that their transforms are:

$$\vec{V}' = M_\lambda \cdot \vec{V} \qquad \text{and} \qquad \vec{I}' = M_\lambda \cdot \vec{I} \tag{18}$$

The reverse transformations are:

$$\vec{V} = M_\lambda^{-1} \cdot \vec{V}' \qquad \text{and} \qquad \vec{I} = M_\lambda^{-1} \cdot \vec{I}' \tag{19}$$

Using the defining equation (4) and (5) of $\vec{I}$, $\vec{V}$ and Y, we can write the energy contained in the bimorph using (9) and (10) as:

$$U = \frac{1}{2}\vec{I}^t \cdot \vec{V} \tag{20}$$

or with (5) and using the symmetry of Y:

$$U = \frac{1}{2}b\vec{V}^t \cdot Y \cdot \vec{V} \tag{21}$$

We now use (19) to rewrite $\vec{V}$ as $M_\lambda^{-1} \cdot \vec{V}'$:

$$U = \frac{1}{2}b\vec{V}'^t M_\lambda^{-1t} Y M_\lambda^t \vec{V}' \tag{22}$$

When the eigenvectors X_λ are normalised, the modal matrix becomes a unitary matrix, of which the inverse is the reciprocal, so that (22) reduces to :

$$U = \frac{1}{2}b\vec{V}'^t M_\lambda Y M_\lambda^t \vec{V}' \tag{23}$$

This can be rewritten as

$$U = \frac{1}{2}b\vec{V}'^t (M_\lambda^t Y^t M_\lambda)^t V' \tag{24}$$

$$U = \frac{1}{2}b\vec{V}'^t Y_\lambda^t \vec{V}' = \frac{1}{2}b\vec{V}'^t Y_\lambda \vec{V}' \tag{25}$$

In this expression Y_λ is a diagonal matrix with the eigenvalues at its diagonal and the matrix multiplication can be carried out quite easily to give:

$$U = \frac{1}{2}b(\lambda_1 V_1'^2 + \lambda_2 V_2'^2 + \lambda_3 V_3'^2 + \lambda_4 V_4'^2) \tag{26}$$

From this expression the meaning of the eigenvalues and the corresponding eigenstates becomes apparent. If we normalise the energy in the bimorph to be equal to 1 joule, or U=1, then (26) is the expression of an ellipsoid in a four dimensional space. The lengths of the semimajor axes are (except for the factor $b/2$) the reciprocals of the coefficients of the $V_i'^2$. This implies that a small value of λ corresponds with a large value of

the semimajor axis in the direction X_λ associated with that particular eigenvalue. So in this direction a large generalized voltage has to be exerted in order to get 1 joule into the bimorph. On the other hand, a large value of λ implies a short semimajor axis and a small generalized voltage required to store 1 joule into the bimorph. Apparently, in the four dimensional space spanned up by the modified moments, forces, uniform loads and voltages which make up the generalized voltage vector, there are "soft" and "hard" directions, in which it is respectively easy and difficult to introduce 1 joule.

3.1 Nonpiezoelectric Limit

In particular when the material is nonpiezoelectric, or $k_{31} = 0, a = \infty$ the matrix becomes simpler to manipulate because the entire fourth row and column of Z become zero, while $z_{11} = \frac{3}{4}$. We rewrite (11) as

$$Z(k_{31} = 0) = \begin{pmatrix} \frac{3}{4} & -3 & 5 & 0 \\ -3 & 16 & -30 & 0 \\ 5 & -30 & 60 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad (27)$$

As the product of the eigenvalues is equal to the determinant,(See e.g. [29]) and the determinant of (27) is zero because one entire row or column is zero, we see immediately that at least one of the eigenvalues λ_1 must be zero.

The other eigenvalues are found to be as: $\lambda_2 = 75.654, \lambda_3 = 1.032, \lambda_4 = 0.064$. The eigenstates corresponding to these eigenvalues are found as the columns of the adjoint matrix. [28,p63] We then find the modal matrix M_λ as follows:

$$M_\lambda = \begin{pmatrix} 0 & -1 & -1 & 1 \\ 0 & 5.82 & 1.54 & .54 \\ 0 & -11.48 & 0.87 & 0.187 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad (28)$$

3.2 Strong Piezoelectric Limit

The maximum value a piezoelectric coupling factor can have is unity. [27] In that case $k_{31} = 1$, or $a = 3$. When we substitute this value in Z the (1,1), (1,4), (4,1) and (4,4) elements become infinite, and therefore we use the Y matrix, which becomes:

$$Y = \begin{pmatrix} 12 & 6 & 2 & -12 \\ 6 & 4 & \frac{3}{2} & -6 \\ 2 & \frac{3}{2} & \frac{8}{5} & -2 \\ -12 & -6 & -2 & 12 \end{pmatrix} \qquad (29)$$

An observation of this matrix reveals that the first row and the fourth row only differ by a minus sign, so the derminant and therefore at least one of the eigenvalues must be zero. Call that λ_1. A further observation shows that the associated eigenstate must be $X_1 = (1, 0, 0, 1)$. The component values in this eigenstate imply that when total coupling exists between mechanical and electrical domains, energy introduced electrically into the bimorph, can be totally removed mechanically, and vice versa. If X_1 is applied to the bimorph, the moment M introduces rotation at the tip of the bender and a charge at the electrodes, while the voltage V completely cancels both the charge and the rotation. This is an interesting situation: the bimorph can be under tremendous mechanical and electrical stress, but it is still completely straight, and when the moment and voltage of X_1 are removed instantaneously and simultaneously, the beam will not snap or vibrate, but just remain still.

The other eigenvalues are found to be:$\lambda_2 = 27.5$, $\lambda_3 = 1.0975, \lambda_4 = 0.01319$

The modal matrix N_λ is constructed as:

$$N_\lambda = \begin{pmatrix} 1 & -8665 & 2.0262 & 2.69 \times 10^{-4} \\ 0 & -4521 & -6.5685 & 68.74 \times 10^{-4} \\ 0 & -1540 & -3.5065 & -135.57 \times 10^{-4} \\ 1 & 8657 & -2.0279 & 14.47 \times 10^{-4} \end{pmatrix} \qquad (30)$$

It can be observed from this modal matrix that the most effective way to get energy into and out of the bimorph is to apply a moment and a voltage together. The effects of forces and moments are less strong in conjunction with a voltage than that of the moment. The effectiveness of coupling through a moment, force or voltage is the subject of a later paper.

4 Simulation Networks

For the purpose of creating simulation networks, it is advantageous to use a diagonal matrix Y_d as in (13) where the diagonal values are not the eigenvalues, and the transformation matrix is not the modal matrix. Here it is useful to reduce the number of elements in the3 transformation matrix as much as possible. This is accomplished by using elementary operations [30], of a triangular form. The seros in this matrix do not show up in the networks as network elements, and hence make the networks simpler.

The voltage and current transformations leading to $Y_d = M_d^{-1} Y M_d$ are $\vec{I} = M_d^{-1}\vec{I}'$ and $\vec{V} = M_d\vec{V}'$.

There are actually two realisations leading to the same transformation; these are sketched in Figure 1.

A few remarks are in order here. The "current " variables of course should have a factor of $(j\omega)$ stuck

on to convert Q to $\dot{Q} \rightarrow I$ and displacements to velocity. We then remark that the diagonal Y_{ii} values are capacitors ($Y_{d11}, Y_{d22}, Y_{d33}$ are mechanical, Y_{d44} electrical capacitors). The value $m_{d14} = \hat{n}$ indicates that this is a piesoelectric transformer coil, or perhaps more properly this coil should be — at least for the conceptual purposes for the electrical engineer — broken into two transformers as in Figure 2.

The transformers represented here are multiwinding, common core type obeying visual rules —• volts per turn are general equal for each coil on a given core, and • ampere turns sum to sero for each core.

It may seem that the networks are a bit complex, but not overly so; if sensors and actuators are to be fully and rapidly incorporated into the world of electronics, they must be represented by networks. One cannot dispute that the equations of physics will give the same answer as the circuits; as Electrical Engineers are more comfortable with the latter, however, and the networks can moreover be directly embedded into simulation programs, it is useful to have a network representation of the bimorph. Furthermore, when the number of variables is reduced, as will often occur in practice, the effect is rapidly and easily deduced from the networks simply by realising what happens to a transformer when a coil is shorted or opened. For example, consider network of Figure 1a, when the electrical voltage with F and p set equal to sero; then $V_2 = V_3 = 0$ and ports (2) and (3) are short- circuited.

The network of Figure 3 is seen immediately to simplify to the network of Figure 4 because the short circuit at port (3) appears directly across the winding on the third core and produces short circuits on the other coils of the core. This uncouples port (3) from port (2) and, with the short circuit at port (2), produces a short circuit on one of the coils of the second core, which results in short circuits on its remaining coils, and leads to the simplified network of Figure 5. The same result would, of course, be obtained (with a different topology) by starting with the network on the right side.

Each of the four ports of a four port network as in Figure 3 can be

a an input-output port;

b a hidden short circuit port; or

c a hidden open circuit port, giving $3^4 = 81$ possible networks for each realization; obviously some of these are more useful than others!

If s_{11}^E and/or d_{31} are lossy (for example in the case we are dealing with acoustic viscosity), then Y_{dii} are complex elements represented by a capacitor and a resistor.

The above has considered the two realisations of Y. In like manner Z may also be realised in two forms. These are treated below. The Z matrix is given in (11) and (27). It is diagonalised to

$$(Z_d) = \begin{pmatrix} \frac{(3a-8)}{4(a-3)} & 0 & 0 & 0 \\ 0 & \frac{4(3a-5)}{(3a-8)} & 0 & 0 \\ 0 & 0 & \frac{5a}{(3a-5)} & 0 \\ 0 & 0 & 0 & \frac{a^2}{4a} \end{pmatrix} \tag{31}$$

by the congruence transformation

$$Z_d = C^t Z C \tag{32}$$

with

$$C = \begin{pmatrix} 1 & \frac{12(a-3)}{(3a-8)} & \frac{10(a-3)}{(3a-5)} & \frac{-3a}{a} \\ 0 & 1 & \frac{15(a-2)}{2(3a-5)} & \frac{-3a}{2a} \\ 0 & 0 & 1 & \frac{a}{2a} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{33}$$

and

$$C^{-1} = \begin{pmatrix} 1 & \frac{12(a-3)}{(3a-8)} & \frac{20(a-3)}{(3a-8)} & \frac{a}{(3a-8)} \\ 0 & 1 & \frac{-15(a-2)}{2(3a-5)} & \frac{3a}{4(3a-5)} \\ 0 & 0 & 1 & \frac{a}{2a} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{34}$$

The voltage and current transformations leading to $Z' = C^t Z C$ are again of two types, each with its distinct network. They are:

$$\vec{V}' = C^{-1}\vec{V}; \vec{I}' = C^{-1}\mathcal{I} \tag{35}$$

giving a realisation like the of Figure 1a with series connected coils on the diagonalised side.

For the Z cases, the networks are a bit more complicated because of the absence of off-diagonal seros on the upper-side of C, as occured with the Y_d matrix the networks take the form of Figure 6

As with the admittance forms, one has the simplifications that follow from imposition of short circuits and open circuits to the various ports.

CONCLUSION AND SUMMARY

The networks associated with the bimorph matrix have been presented. The bimorph matrix has been diagonalised, and eigenvalues and eigenstates have been found for the extremes of nonpiesoelectric and strong piesoelectric coupling. A physical interpretation of the eigenvalues and eigenstates has been proposed. The bimorph matrix has also been diagonalised by means of

elementary operation matrices, which gave the matrix in equivalent form. The transformation matrices are of upper triangular form. This choice of transformation yields networks of the simplest possible connection schemes. It has been shown that in case of less than all possible boundary conditions applied, the networks are easily simplified. The networks of the admittance and impedance matrix are presented.

REFERENCES

1. C.B.Sawyer, The Use of Rochelle Salt Crystals for Electrical Reproducers and Microphones. Proceedings Institute of Radio Engineers, Vol. 19, No. 11, Nov. 1931, pp.2020-2029.

2. B.B.Bauer, Sound recording method and apparatus 3,490,771, Jan. 20, 1970, United States Patent.

3. W.J.Denkmann, R.E.Nickell, and D.C.Stickler, Analysis of Structural Acoustic Interactions in Metal-Ceramic Transducers. IEEE Transactions on Audio and Electroacoustics, Vol. AU-21, No. 4, 1973, pp.317-324.

4. W.D.Cragg and N.W.Tester, Telephone Transducers using Piezoelectric Polymer Foil. Electrical Communication, Vol. 52, No. 4, 1977, pp.312-315.

5. M.R.Steel, F.Harrison and P.G.Harper, The piezoelectric bimorph: an experimental and theoretical study of its quasistatic response. J. Phys. D: Appl. Phys., Vol. 11, 1978, pp.979-989.

6. J.K.Lee. Piezoelectric bimorph optical scanners: analysis and construction. Applied Optics, Vol. 18, 1979, pp 454-459.

7. W.J.Spencer, W.T.Corbett, L.R.Domingues and B.D.Shafer, An Electronically Controlled Piezoelectric Insulin Pump and Valves. IEEE Transactions on Sonics and Ultrasonics, Vol. SU-25, No. 3, May 1978, pp.153-156.

8. J.G.Smits, Piezoelectric micropump with microvalves. Proceedings Eighth Biennial University / Government / Industry Microelectronics Symposium, Westborough, MA June 12-14, 1989, pp 92-94.

9. J.G.Smits, Piezoelectric Micropump with Three Microvalves Working Peristaltically. Conference Abstracts of the Transducers '89, Montreux, Switzerland, p 114.

10. S.A.Kokorowski, Analysis of adaptive optical elements made from piezoelectric bimorphs. J. Optical Soc. Am., Vol. 69, 1979, pp.181-187.

11. Y.Kubota, Rotating Magnetic head Piezoelectric assembly and drive circuitry for video tape recorder. 4,233,637, Nov. 11, 1980, United States Patent.

12. S.Osaka and M.Toda, Rotative motor using plural arrays of piezoelectric elements. 4,399,386, Aug. 16, 1983, United States Patent.

13. V.H.Schmidt, M.Klakken, and H.Darejeh, PVF2 bimorphs as active elements in wind generators. Ferroelectrics, Vol. 51, 1983, pp.105-110.

14. P.L.Chen, R.S.Muller, R.D.Jolly, G.L.Halac, R.M.White, A.P.Andrews, T.C.Lim, and M.E. Motamedi, Integrated Silicon Microbeam PI-FET Accelerometer. IEEE Transactions on Electron Devices, Vol. ED-29, No. 1, Jan. 1982, pp.27-33.

15. B.Bill, A.L.Wicks, Meausuring Simultaneously Translational and Angular Acceleration with the New Translational–Angular– Piezobeam (TAP) System. Abstract of Transducers '89. 5th International Conference on Solid-State Sensors and Actuators and Eurosensors III, June 25-30, 1989, Montreux, Switzerland, p.1229.

16. T.Bailey and J.E.Hubbard Jr., Distributes Piezoelectric - Polymer Active Vibration Control of a Cantilever Beam. J.Guidance, Vol. 8, Sept- Oct.1985, pp605-611.

17. E.F.Crawley and J.de Luis, Use of Piezoelectric Actuators as Elements of Intelligent Structures. AIAA Journal, Oct. 1987, pp1373-1385.

18. D.L.Heiserman, Piezoelectric polymer micromanipulator 4,610,475, Sept. 9, 1986, United States Patent.

19. M.F.Barsky, D.K.Lindner, and R.O.Claus, Robot Gripper Control System Using PVDF Piezoelectric Sensors. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. Vol. 36, No. 1, Jan. 1989, pp129-134.

20. A.Kumada, A Piezoelectric Ultrasonic Motor. Proceedings of the Sixth International Meeting on Ferroelectricity, Kobe 1985, Japanese J. Applied Physics, Vol. 24, 1985, Supplement 24-2, pp.739-741.

21. O.Kawasaki, T.Nishikura, Y.Imasaka, M.Sumihara, K.Takeda, and H.Yoneno, Ultrasonic Motors. IEEE Tokyo Section, Denshi Tokyo, No. 26, 1987, pp.158-161.

22. C.F. Quate, Imaging with the Tunneling and the Force Microscope, Proceedings of the IEEE conference on Micro Electro Mechanical Systems, Napa Valley, 11-14 February 1990, p 188.

23. T.W.Kenney, S.B.Waltman, J.K.Reynolds. W.J.Kaiser. A Micromachined Silicon Electron Tunneling Sensor.Proceedings of the IEEE conference on Micro Electro Mechanical Systems, Napa Valley, 11-14 February 1990, p 188.

24. Piezoelectric Fans, Brochure of the Piezo Systems Company, Cambridge MA.

25. C.P.Germano, Flexure Mode piezoelectric

Transducers, IEEE Transactions on Audio and Electroacoustics, Vol. AU-19, No. 1, 1971, pp.6-12.

26. Jan G.Smits and Susan I.Dalke, The Constituent Equations Of Piezoelectric Bimorph Actuators, IEEE 1989 Ultrasonics Symposium, Program and Abstracts, October 3-6, 1989, Montreal, Canada. p54.

27. J.G.Smits, Eigenstates of Coupling Factor and Loss Factor of Piezoelectric Ceramics, Ph.D. Thesis Twente University, 1978, the Netherlands.

28. Shahen A. Hovanessian and Louis A.Pipes, Digital Computer Methods in Engineering. McGraw – Hill Book Company, New York, 1969, Chapter 2.

29. R.Bronson, Matrix Methods, an Introduction, Academic Press, 1969, p96.

30. F.Hohn, Elementary Matrix Algebra, Macmillan Company, London, 1969, p113.

31. M.Toda, Theory of Air Flow Generation by a Resonant Type PVF2 Bimorph Cantilever Vibrator. Ferroelectrics, Vol. 22, 1979, pp.911-918.

32. P.L.Chen, R.S.Muller, and A.P.Andrews, Integrated Silicon PI-FET Accelerometer with Proof Mass. Sensors and Actuators, Vol. 5, 1984, pp.119-126.

33. P.Muralt, D.W.Pohl, and W.Denk, Wide-range, low-operating-voltage, bimorph STM: Application as potentiometer. IBM J. Res. Develop., Vol. 30, No. 5, Sept. 1986, pp.443-450.

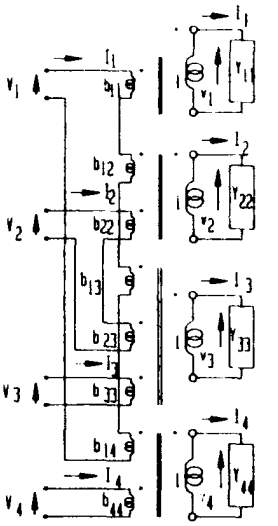


figure 1a

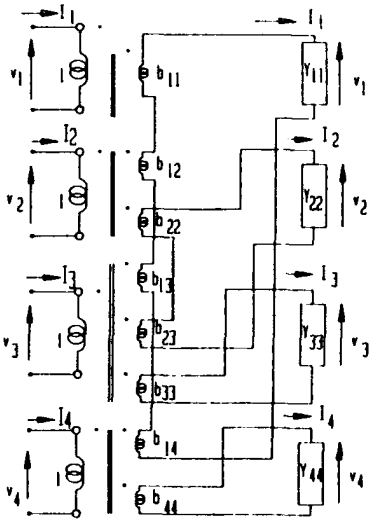


figure 1b

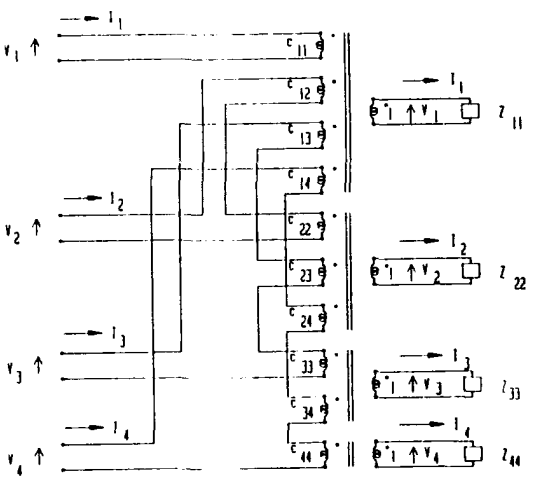


figure 2a

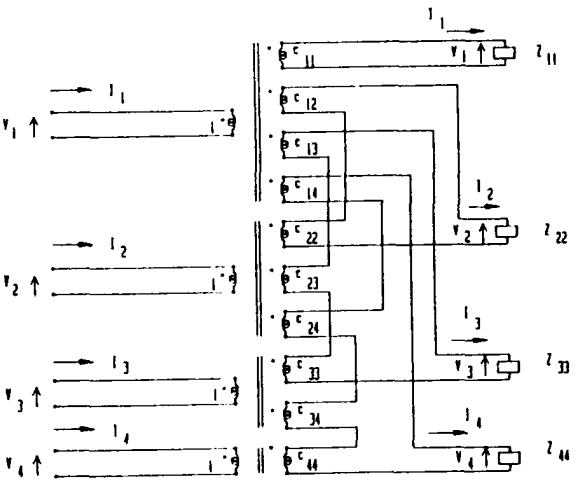


figure 2b

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NARROWBAND SAW FILTERS FOR IF APPLICATIONS

B. Horine

Sawtek, Inc. PO Box 609501, Orlando, FL 32860-9501

ABSTRACT

Surface acoustic wave (SAW) filters offer Q's comparable to crystal filters from 50 MHz to well over 1 GHz. Resonator filters using in-line coupling are the most familiar SAW approach to narrowband filtering. On the other hand, the newer waveguide coupled resonator filter offers superior close-in rejection. For the same filter size, the RFW achieves 30 to 40 dB of close-in rejection while the RFI achieves 10 to 20 dB. Since both approaches exhibit losses of only 3 to 6 dB, either may be easily cascaded to obtain greater rejection.

For narrowband applications where linear phase or phase tracking is important, resonant type filters may not meet performance requirements and a transversal design may be needed. In normal SAW transversal filter design, however, distortion due to overcoupling becomes a significant design challenge below 0.3% fractional bandwidth. Reflective array filters avoid this pitfall and result in a shorter device with little distortion for the same bandwidth and shapefactor.

Because of the wide selection of narrowband SAW techniques now available, the receiver design can flow smoothly from system requirements to choice of architecture and filter specifications. This paper will start with several different typical receiver requirements, point out the critical factors, and develop the SAW solution for each.

INTRODUCTION

Narrowband (less than 0.2%) IF filtering in the high VHF and UHF frequency band represents a special challenge although it offers many benefits. Only a few techniques produce high quality responses in a reasonable packaging configuration. Three main types of SAW filters have found common use in this area: the in-line coupled resonator filter (RFI), the waveguide coupled resonator filter (RFW), and the reflective array filter (RAF). Several tradeoffs are involved in designing and selecting a particular filter approach. This paper will review the general concerns of narrowband SAW filtering, the specific characteristics of each approach, and finally study the application of these filters in particular IF subsystems.

CH2818-3/90/0000-316 $1.00 © $1.00 IEEE

Frequency accuracy is critical in narrowband filtering applications. For SAW devices, the most important sources of frequency error are the manufacturing set accuracy and the temperature stability. Aging is also important but is nearly an order of magnitude smaller. The total required frequency tolerance is added to the information bandwidth in order to determine the passband width required of the filter. The tolerance requirement is also subtracted from the specified rejection bandwidths, thereby narrowing the actual filter rejection skirts. The tolerances required, therefore, for narrowband filters can have a dramatic effect on the shape factor of the filter. Also, the rejection floor tends to degrade at the wider bandwidths needed to cover the tolerances. For these reasons, it is important to maintain a high degree of control over the sources of error and minimize the frequency tolerance required.

The manufacturing set accuracy can range from +/- 30 ppm to +/- 200 ppm. The narrower range is used for low frequency applications where it is more effective to deal with a tight set accuracy rather than be driven to a lower shape factor and higher order filter. The wide end of the range is typically used for the narrowest band signals (BW3 < 300 ppm) where the rejection bandwidth is not critical. Most filters are designed assuming a +/- 75 ppm tolerance for manufacturing set accuracy.

ST cut quartz is the most commonly used substrate for narrowband filters, although other singly rotated cuts, particularly those with lower coupling coefficients, are of interest [1,2]. The complexities of beam steering, especially over a large temperature range has slowed the use of these other cuts. ST quartz exhibits a parabolic frequency versus temperature curve, and by varying the cut angle and metal thickness slightly, the turnover temperature can easily be set within +/- 10°C. Assuming a conservative second order coefficient, this leads to the required tolerances listed in Table I.

Range	Frequency Drift
0°C to 60°C	-55 ppm
-20°C to 75°C	-113 ppm
-54°C to 85°C	-220 ppm

Table I

In order to develop an appreciation for the significance of these frequency tolerances in SAW devices, Figure 1 details an example design. Note that the addition of the individual tolerances is done by adding the peak to peak value of each component.

Example:

Set accuracy =	± 75 ppm
Temperature stability =	- 113 ppm
Lifetime aging =	± 20 ppm
Total required tolerance =	303 ppm
Information BW =	400 ppm
Rejection BW =	2000 ppm
Shape Factor = (w/o tolerances)	5 : 1
Filter BW = 400 + 303 =	703 ppm
Rej. BW = 2000 - 303 =	1697 ppm
Shape Factor = (w/ tolerances)	2.4 : 1

Figure 1. Impact of Frequency Tolerances

As this example shows, the addition of the required tolerances has increased the number of poles needed to synthesize the filter function.

SAW Resonator Filters

Multipole resonator filters are the workhorse of narrowband filtering in SAW technology. Their center frequencies range from 50 MHz to over 1250 MHz. At 50 MHz the package required is 2 inches long. At 1250 MHz the line size is about 0.6 μm. Physical size and lithography capabilities therefore set the soft limits on the frequency range.

Bandwidth is limited on the low side to approximately 0.02% by the finite Q of the individual resonant cavities and by the required frequency tolerances. On the other end, 0.2% bandwidth is achievable. Wider bandwidths usually do not meet rejection requirements.

Insertion loss is sensitive to the center frequency and the number of poles. It ranges from 2 dB to 7 dB for most applications. Because this loss is usually easily handled in IF applications, cascading filters is an effective technique to increase the rejection. The filters are simply coupled resonant cavities and they exhibit the familiar Butterworth or Chebychev characteristics. The rejection floor is highly dependent upon the coupling mechanism between poles of the filter. These mechanisms are the basis for the following two resonator filter classifications.

In-Line Coupling

In-line coupling, diagramed in Figure 2, is the most established monolithic coupling technique for SAW resonator filters [3]. In this design, the outside acoustic reflector arrays trap the energy within the device ensuring a low loss and high Q response. A center reflector splits the device into two separate resonant cavities (or poles) and controls the degree of coupling. The interdigital transducers convert the energy from electrical to acoustical and vice-versa. This configuration results in a monolithic 2 pole building block. Higher order filters are normally implemented by cascading these 2 pole sections using the transducers and a matching inductor (often referred to as transducer coupling) [3].

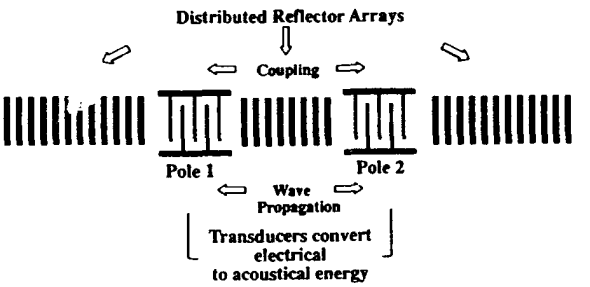


Figure 2. Inline Coupled Resonator Filter

The reflectors in the arrays are distributed uniformly on half-wavelength centers and are designed to trap the energy over a very narrow bandwidth. Beyond their finite bandwidth, the reflectors become essentially transparent so that the filter response is given simply by the transmission from one transducer to the other in a simple delay line configuration. This results in a

broad $(\sin x / x)^2$ response, as shown in Figure 3, with a resonance peak at the center where the reflectors trap the energy launched by the transducers. If one is interested in high selectivity, one would cascade the 2 pole sections to achieve better than the 18 - 20 dB near-in rejection shown. On the other hand, where absolute rejection more distant from the frequency of operation is most important as in some synthesizer applications, these filters provide low loss with outstanding rejection floors.

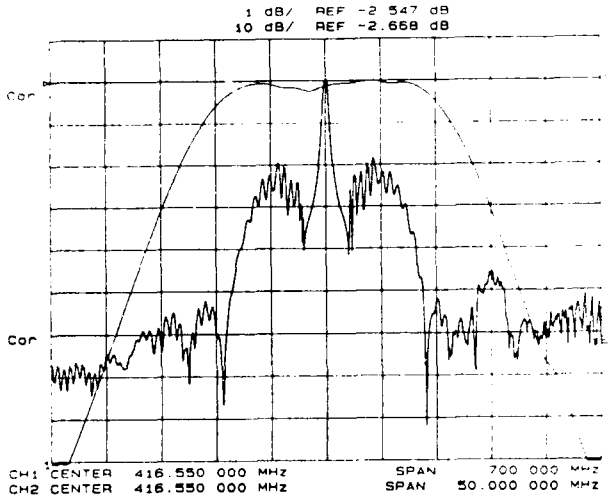


Figure 3. Inline Coupled Frequency Response

In general, the RFI approach achieves 45 dB to 55 dB of ultimate rejection with 10 dB to 18 dB of near-in rejection per 2 pole section. The near-in rejection degrades with increasing bandwidth resulting in a maximum practical bandwidth of 0.20% for most applications.

Waveguide Coupling

Waveguide coupling alleviates the near-in rejection limitation by coupling the energy transversely to the normal traveling wave direction [4,5]. The topology of the design is illustrated in Figure 4. In this case, a waveguide is formed by embedding a single cavity resonator in a slow wave structure. The fundamental mode profile across the beamwidth of the device is cosinusoidal within the guide with an exponential evanescent tail immediately outside the guide. If an identical guide is placed alongside the first one, energy may be transferred from pole to pole via the evanescent tail. The amount of intersection of each evanescent tail determines the coupling and therefore the bandwidth of the resulting filter.

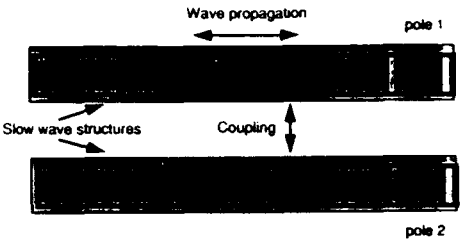


Figure 4. Waveguide Coupled Resonator Filter

When the reflectors become transparent in this design, outside the narrow stopband, the evanescent tail is no longer supported resulting in zero coupling, and the wave simply propagates to the end of the crystal and is absorbed or dispersed. Since there is no longer any path from input to output, the near-in rejection is excellent. In practice, this level is limited by spurious waveguide modes and electromagnetic feedthrough. The response of this type of filter is plotted in Figure 5. This is a 4 pole filter implemented monolithically using transducer coupling without a coupling inductor by taking advantage of a zero susceptance effect of a long transducer [6]. This direct cascading is typical of RFWs which are normally designed as 4 pole building blocks rather than the 2 pole blocks in the RFI case.

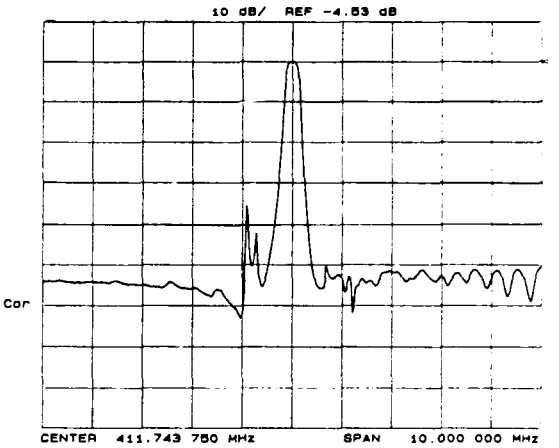


Figure 5. Waveguide Coupled Frequency Response

RFWs exhibit 40 dB to 50 dB of rejection with a few narrowband spurs as high as 30 dB to 40 dB. Because of crosstalk concerns and material property constraints, the bandwidth is currently limited to just over 0.1%. Extending the bandwidth range in order to cover broader applications is the subject of current research.

REFLECTIVE TRANSVERSAL FILTERS

In some applications, linear phase over the bandwidth and a reasonably steep shape factor are required. In this case, resonator type filters may not meet the objectives and a transversal design must be used. The problem with narrowband transversal filters is illustrated by consulting Figure 6a. The transducers are very long and are comprised of many electrodes, even in harmonic designs. The passband distortion largely attributable to overcoupling is worsened by the reflections and energy scattering from the large number of electrodes.

One way to avoid the overcoupling is by separating the frequency selection and energy transduction process as indicated in Figure 6b [7]. In this configuration, denoted as a reflective array filter (RAF), the launching and receiving transducers have relatively few fingers eliminating the overcoupling in the transduction. The reflective array performs the principle frequency selection.

Figure 6a. Traditional SAW Transversal Filter Design

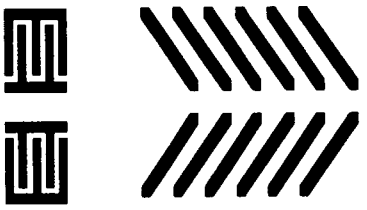


Figure 6b. Reflective Array SAW Transversal Filter Design

The transversal design, using reflector arrays for frequency selection, exhibits linear phase characteristics. Shape factors similar to traditional SAW transversal filters are readily achieved, although it is important to realize that both the bandwidth and the shape factor are a function of the length of the device, and a steep and narrow filter can be unrealistically long.

Because of the second order tradeoffs associated with designing a pair of low loss 90° reflector arrays, the RAF is normally designed with an insertion loss of 18 dB to 26 dB.

APPLICATIONS

IF applications may require a steep shape factor for channel selectivity, high rejection, minimum flyback, or a flat passband. In some applications linear phase and low VSWR may be required. The choice of filter element will be dependent upon these specific requirements. A few examples will clarify this.

ANTI-JAM MODEM

An anti-jam modem is intended to accept a modulated signal from a radio and demodulate and decode the information in the presence of jammers. An example is a Qualcomm modem that accepts a frequency hopped signal centered at 70 MHz. The information bandwidth is hopped over a 40 MHz bandwidth. In this application, good power linearity is important in order to reduce jamming effects. The noise figure is dominated by the radio so the insertion loss of the filter is not critical. Since the signal is M'ary FSK modulated, a flat frequency response is desired.

Figure 7 is a simplified block diagram of the IF section. Up conversion is the obvious choice in view of the large bandwidth. Since IMD is a concern, it is important to eliminate any extraneous tones as early in the chain as possible. A high resolution synthesizer is used as the first LO and a narrowband filter is used in the first IF. By setting the first IF at 289.368 MHz, a convenient 300 MHz fixed second LO can be used to convert the signal to approximately 10.6 MHz where very inexpensive filters and other components are available. At this point the signal is demodulated.

Clearly, the high frequency of operation, narrow bandwidth, and good IMD rejection suggest a SAW filter. An RFW was designed with ±77 kHz 1 dB bandwidth. The specified 20 dB bandwidth of ±215 kHz sets the noise bandwidth of the system while the 45 dB bandwidth of 1 MHz sets the jamming bandwidth of the modem.

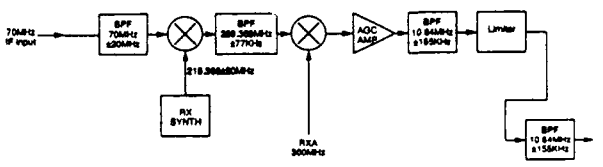


Figure 7. AntiJam Modem IF Section (Qualcomm)

Figure 8 is a response plot of the filter showing both the narrowband and wideband response. Note that the narrowband spurious responses on the low side are located just inside the jamming bandwidth. The insertion loss and ripple are 4.5 dB and 0.25 dB respectively.

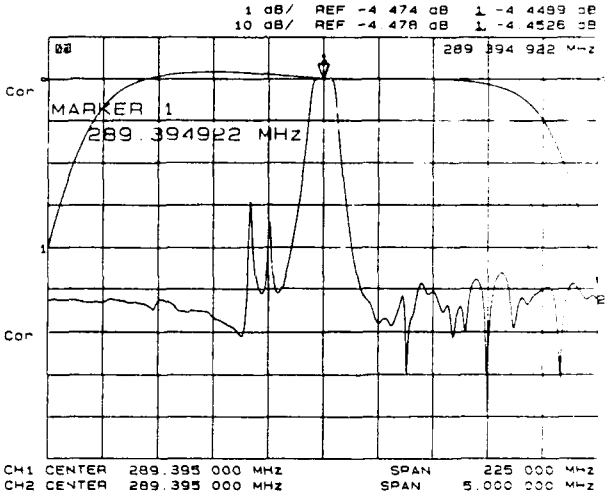


Figure 8. RFW 4 Pole for A-J Modem IF

SATELLITE TRANSPONDER

Satellite transponders have special requirements that are not always as critical in other applications. For example, ruggedness, reliability, small size, and low weight often rank at the top of the requirements list. In addition, loss is important to the extent of avoiding being downlink limited. Finally, the interference needing rejection is dominated by the many other signals used in the satellite.

The Data Collection Platform (DCP) of the GOES satellite built by Ford Aerospace Company for NASA includes both interrogate and reply transponders that communicate between a control station and hundreds of remote stations scattered throughout the footprint of the satellite. In order to transmit the data efficiently, it is frequency and time multiplexed.

Figures 9 and 10 illustrate the simplified block diagrams of the interrogate (DCPI) and reply (DCPR) channel respectively. In both channels, two filters are cascaded in order to meet the ultimate rejection requirements. The DCPI filter is used primarily to reject other incoming signals that are used elsewhere in the communications subsystem. On the other hand, the DCPR filter is more concerned with eliminating unwanted signals coming from the ground and especially reducing the out-of-band noise. Both filters operate in the 400 MHz range with narrow bandwidth, must survive a 1500 G mechanical shock, and be small in size and weight. These requirements suggest a SAW resonator filter.

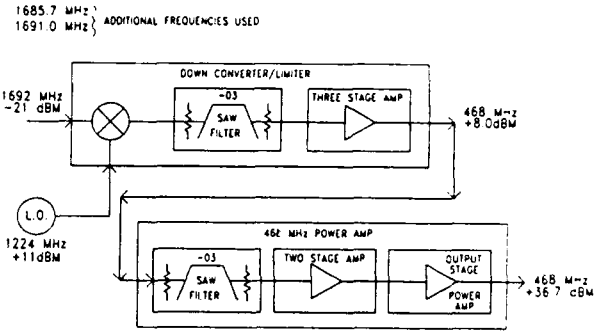


Figure 9. GOES DCPI Transponder (Ford Aerospace)

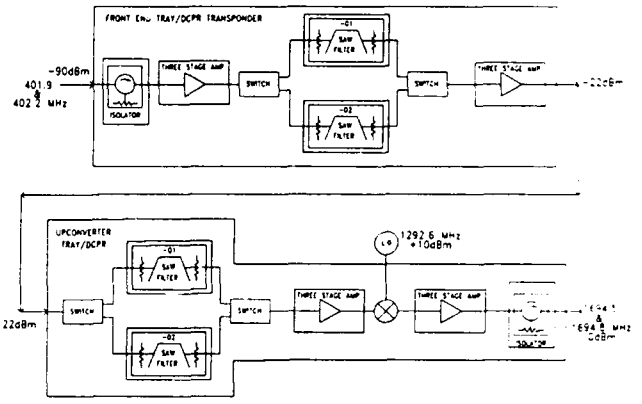


Figure 10. GOES DCPR Transponder (Ford Aerospace)

DCPR

The DCPR channel is specified with a 400 kHz bandwidth at 1.4 dB. Once the appropriate frequency tolerances have been added, the passband width is too large to be implemented using RFW technology. Instead, a 12 pole RFI is used to achieve 60 dB of near-in rejection. These are delivered as a pair of 6 pole devices so that amplifiers may be placed in between them. The noise bandwidth is specified at 500 kHz. The VSWR over the 3 dB bandwidth is important so

that the integration of the satellite can proceed smoothly and at a low risk. This is specified at 1.92:1 which will require some resistive padding inside the device. Finally, because the signal is actually a composite, frequency multiplexed signal, no sharp transitions can be tolerated since they can affect a single channel significantly. This requirement leads to a passband slope specification of 0.1 dB / kHz. The response of a 6 pole unit is plotted in Figure 11. The insertion loss of nearly 17 dB resulted in a VSWR of 1.3 : 1 maximum across the 3 dB bandwidth. This bandwidth is 495 kHz wide.

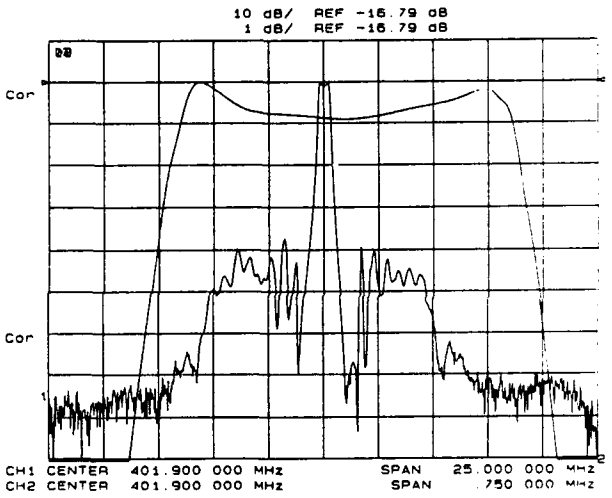


Figure 11. RFI 6 pole for DCPR Channel

DCPI

The DCPI channel is specified at only 180 kHz bandwidth at 1.2 dB. This is narrow enough for an RFW approach. In this case 8 poles are required to achieve the 60 dB near-in rejection specification. The filters are delivered as separate 4 pole devices so that amplifiers may be used in between. The specifications are similar to those for the DCPR filter, scaled accordingly for the narrower bandwidth. The response of a single 4 pole device is plotted in Figure 12.

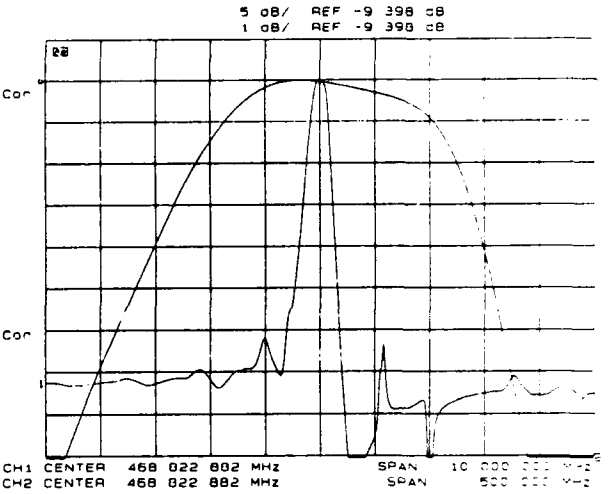


Figure 12. RFW 4 pole for DCPI Channel

DIRECTION FINDING

Direction finding is required in electronic warfare applications and navigation and landing system applications. The principle is to measure relative phase from multiple sources (antennas). It is obviously important to have linear phase, and in *some applications*, phase tracking over temperature is required. Normally a calibration signal is used to effectively zero any overall phase offset at any given temperature.

Since linear phase is a major issue, a RAF type filter is normally used. The IF processing should be done below 600 MHz (about 1/2 the frequency limit of resonators) because of fabrication limitations.

Although the specific program with its particular system design cannot be given, a filter has been designed and modeled that is representative of typical designs for this application. Figures 13 and 14 illustrate the amplitude and phase response of a 400 MHz filter with a 3 dB bandwidth of 500 kHz. The 35 dB bandwidth is 1.0 MHz and the near in rejection is approximately 38 dB with the ultimate rejection limited by crosstalk at 50 dB to 55 dB. It exhibits less than 1 dB of amplitude ripple at 19 dB of loss. The delay has been subtracted from the response in Figure 14 in order to display the phase linearity. Very little non-linearity is displayed. (The phase ripple in an actual device would probably be around 6 degrees due to miscellaneous wave reflections.) The design details and experimental results will be the subject of a future paper [8].

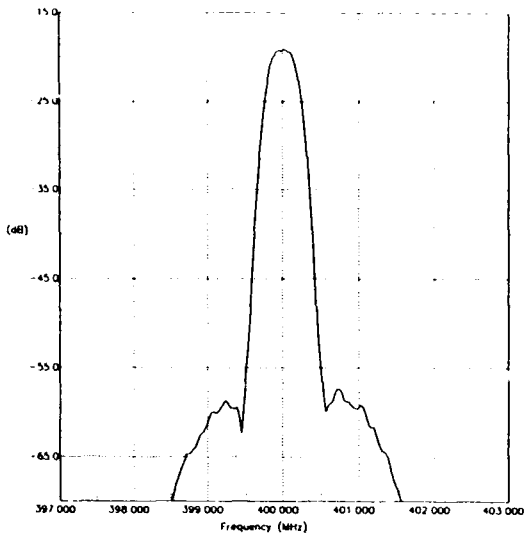


Figure 13. RAF for Direction Finding Application

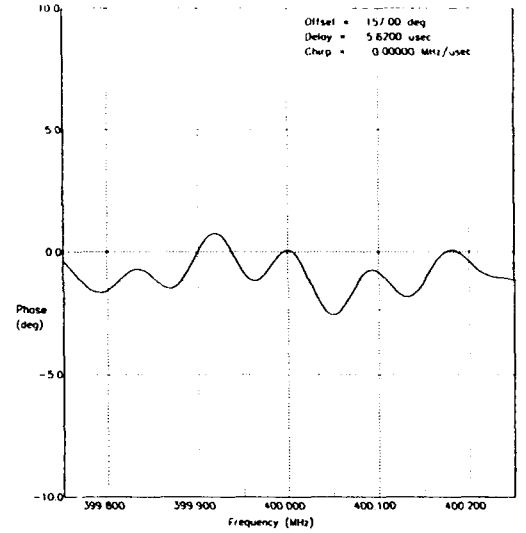


Figure 14. RAF Phase Linearity

CONCLUSION

From the above discussion, it is clear that SAW filters can and are being used in a wide variety of narrowband IF applications. RFW filters offer the best near-in rejection performance for a given size and insertion loss but are limited to 0.1% bandwidth. When greater bandwidths are called for but low loss is important or near-in rejection is not a concern, RFI filters are effective. Because of their low loss, resonator type filters are often cascaded when greater rejection or steeper shape factors are required. Finally, when linear phase is required, RAF filters can satisfy the specifications provided a larger size and higher insertion loss can be tolerated.

ACKNOWLEDGMENTS

The author gratefully acknowledges the contributions of R. Kornfeld of Qualcomm and J. J. MacFarland of Ford Aerospace Company who provided the details of their subsystem designs for this paper. The author also acknowledges the contributions of S. Gopani, T. O'Shea, R. Brown, and J. Sunkin who contributed to the design of these filters and R. Hays who contributed much helpful guidance in the preparation of this paper.

REFERENCES

1. R. Webster, "X-cut Quartz for Improved Surface Acoustic Wave Temperature Stability," in Journal of Applied Physics, vol 56, 1984, pp 1540-1542.

2. B. K. Sinha, "A Stress and Temperature Compensated Orientation and Propagation Direction for Surface Acoustic Wave Devices," in "IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. UFFC-34, NO1, 1987, pp 64-74.

3. P. S. Cross, R. V. Schmidt, and H. A. Haus, "Acoustically Cascaded ASW Resonator Filters," in 1976 Ultrasonics Symposium Proceedings, pp. 277-280.

4. H. F. Tiersten and R. C. Smythe, "Guided Acoustic Surface Wave Filters," in 1975 Ultrasonics Symposium Proceedings, pp. 293-294.

5. M. Tanaka, T. Morita, K. Ono, and Y. Nakazawa, "Narrow Bandpass Filter Using Double-Mode SAW Resonators on Quartz," in Proceeding of the 38TH Annual Symposium on Frequency Control, 1984, pp 286-293.

6. M. Hikita, Y. Kinoshita, H. Kojima, and T. Using New Phase Weighting," in IEEE MTT International Microwave Symp. Digest, 1982, pp. 46-48.

7. J. Melngailis, J. M. Smith, and J. H. Cafarella, "Bandpass Surface Wave Filters," in 1972 Ultrasonics Symposium Proceedings, pp 221-225.

8. S. Gopani, R. Brown, J. Hines, and B. Horine, "Reflective Array SAW Narrowband Filters," to be published.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NOVEL TYPES OF GUIDED S.A.W. REFLECTORS FOR MICRO-RESONATORS

G.Golan, G.Griffel, A.Seidman, and N.Croitoru.
Department of Electron Devices and Materials
Faculty of Engineering
Tel-Aviv University
Ramat-Aviv 69978
Israel

Surface acoustic waves for micrograting reflectors have been characterized. Based on the perturbation theory, eight different types of structures on an acoustic waveguide were analyzed. Results of simulations of all eight types of corrugation structures were evaluated in order to find the least leaky waveguide, the most efficient reflector (with minimum necessary perturbations), and the optimal mode shape for improved performances. General design curves are presented in order to illustrate the behavior of the incident and reflected waves under a variety of structural conditions. Analytic expressions for the calculations of the mode amplitude and mode shape, and for general acoustic corrugations are derived and then the simulations results are presented.

I. INTRODUCTION

The analysis of perturbed channel waveguides plays a significant role in the design of integrated microresonators or Surface Acoustic Wave (SAW) couplers, since channeled waveguides form the basic building block of many devices. The perturbed waveguide mode dispersion is crucial to device design in order to have prior knowledge of waveguide behavior in a practical device. In the field of electrooptics a number of methods for analyzing optical perturbed waveguides exist in the literature[1]. Marcatili[2] has analyzed a step index optical channel by ignoring the fields in the corner regions. For such an approximation, the results are accurate only far from cut off since, in this case, the fields at the corner can be ignored.

The integration of waveguides and distributed reflectors in 'integrated acoustic' device applications have recently been studied by our group[3].

Using coupled mode theory, it was shown that a periodically perturbed conventional $\Delta V/V$ is (the relative velocity reduction due to a deposited metallic strip over a piezoelectric substrate) acoustical waveguide in a microresonator structure, can serve as a Fabry-Pérot cavity. A general method, which was given in that work, could be applied to any perturbed waveguide structure in the sense that it could analyze and obtain the number of transverse modes, their cutoff lines, and their reflecting response without having to solve the corresponding eigenvalue equations. An analytical expression for the incident and reflected fields in a periodically perturbed acoustic waveguide was presented. The main disadvantage in all the above methods is that neither the analytical nor the numerical methods can provide normalized solutions for periodically perturbed waveguides of arbitrary corrugation shapes and aspect ratios. Rounded, diamond-shape, or star-type corrugations are not suitable for analytical treatment. Numerical methods, on the other hand, are applicable only to the specific case under investigation, thus losing generality. This makes it necessary to search if the associated wave equation (or in a periodic case, the Helmholtz equation) can be normalized prior to a numerical solution, giving results that would become general and applicable to any kind of corrugation shape on a certain waveguide. The demand for new kinds of periodic corrugations in an acoustical waveguide configuration increases, since conventional perturbations excite leaky modes that deteriorate the isolating properties of a typical confining waveguide.

CH2818-3/90/0000-323 $1.00 © 1990 IEEE

In this paper several kinds of complex periodic corrugations will be discussed and numerically analyzed in acoustical micromirror configurations. The final goal of this work is to gain better isolation and crosstalk reduction along with improvement of reflection properties. This work is novel in the sense that it enables one to evaluate different waveguide structures, accepting structures of any kind, and its results are applicable to all kinds of waveguides with aspects ratios, not necessarily equal to one.

In Sec. II, the main principles of the numerical approach of the normalized Helmholtz solution is presented and discussed. In Sec. III, various corrugation shapes are described. These shapes are mainly distinguished in the sense that some of them are externally mounted while the rest are internally mounted. Sec. IV describes evaluation and results of the simulations, pointing to the most preferable acoustical structure.

II. THE NUMERICAL APPROACH OF AN ARBITRARY PERTURBATION STRUCTURE

The numerical approach, which was basically developed by Matsuhara[4], used the variational method. Forsythe and Wasow[5] on the other hand, used the finite difference method. In this paper a combination of both variational and finite difference method were used. The finite difference approach for solving the normalized Helmholtz equation requires a large number of subintervals since the field is discontinuous at the subinterval boundaries. In the variational approach with bilinear interpolation, the field is continuous at the subinterval boundaries, independent of the number of subintervals, enabling the use of a smaller interval for an accurate solution.

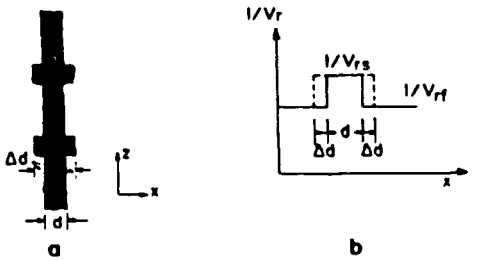


Fig. 1: Basic corrugation structure (a) and its corresponding index profile (b).

A basic waveguide perturbation is of the form shown in Fig.1. Obviously different types of corrugations can fit the basic structure as long as their dimensions remain smaller than the waveguide width.

In order to find the perturbation response parameters, we express the perturbed Rayleigh channel phase velocity as[3]

$$\left[\frac{1}{V_r}\right]_{(x,z)} = \frac{1}{V_r(x)} + F_{(x)}\Delta\ \frac{1}{V_r(z)}\ , \qquad (1)$$

where F(x) is the normalized window function given by

$$F(x) = \begin{cases} 1,\ d < |x| < d + \Delta d \\ 0,\ \text{else,} \end{cases}$$

and

$$\Delta\,[1/V_r(z)] \ll [1/V_r(x)] \qquad (3)$$

is the variation of the small perturbations. Here $V_{r(x)}$ is the Rayleigh velocity variation due to the small transversal perturbation. (Fig.1.)

The wave velocity gets its minimal value when F(x) = 1 in the perturbed region and seems to remain constant at this new value ($\Delta V_{r(x)}$) along the perturbation width. The overall field in the perturbed waveguide can be approximated by a linear combination of back and forth propagating modes, which are coupled to each other by the applied z-perturbation. The approximated modal field expression in such a structure is[3]:

$$w(x) = W(x)\,[R(z)\exp(-i\beta z) + S(z)\exp(i\beta z)]\ , \qquad (4)$$

where β is the wave number of the propagating wave.

The expressions for R(z) and S(z) (the reflected and incident waves) include the decay factor along the z direction of the waveguide and the periodic perturbations upon it. These expressions were already derived in our previous work[3].

W(x) is the local normalized modal field along the x axis and is approximately independent of z. Numerical representation of this modal field using the variational method eases the solution of a partial differential equation since the variational approach typically converges more rapidly than other conventional solving methods. Matsuhara[4] and Taylor[6] have employed this approach to solve the Helmholtz wave equation.

Application of their method to the normalized Helmholtz equation yields the normalized solution directly. From physical considerations, the fields outside an acoustical waveguide vanish at infinity. This provides a boundary condition for the variational problem. Since the fields decay exponentially in the substrate underneath, it is possible to assume some finite points in which the field vanishes. This assumption is necessary to facilitate the numerical calculation.

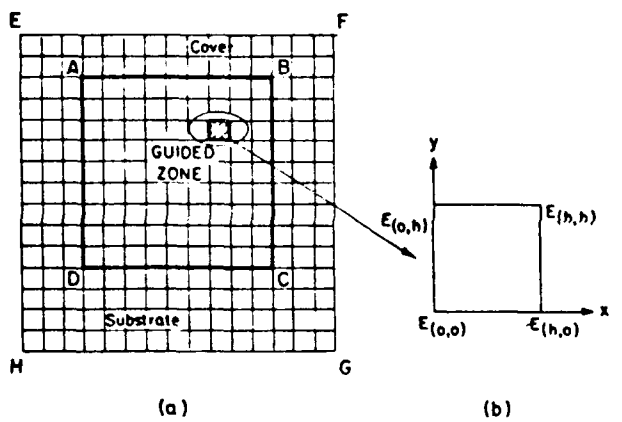


Fig. 2: Schematic presentation of the ABCD guided zone surrounded by the boundaries EFGH where the field vanishes. (a) Each grid point represents a local field (b). An expanded grid cell B3 (shaded in (a), B4 is shown in (b)).

Figure 2 shows an ABCD waveguide, surrounded by a boundary EFGH in which the modal field is assumed to vanish for modes far from cutoff. It should be mentioned that this assumption is correct only for a skin guiding zone underneath a $\Delta V/V$ waveguide. The solution method can be illustrated by a square grid 14x14 (Fig.2). The inner part of the grid is the guided zone whereas the outer part is the boundary. The field in the grid is specified by 14x14 vector components $(E_1, E_2 \ldots E_{196})$. Inside ABCD the field is expressed as a bilinear interpolation of the four grid point values[5]. Expansion of a typical cell in the grid is shown in Fig.2(b). The linear field distribution inside this cell is given as[7]

$$E(x',y') = \frac{1}{h^2}[E_{oo}\, x'y' + E_{\infty o}(h - x')(h - y') + E_{ho}\, y'(h - x') + E_{oh}(h - y')x']. \quad (5)$$

$E(x',y')$ is the field amplitude that satisfies the normalized, two dimensional scalar Helmholtz wave equation. E_{hh}, E_{oo}, E_{oh}, and E_{ho} are shown in Fig. 2.

Linearity of this interpolation even on the cell edges, necessitates continuity throughout the whole region (except for the EFGH boundary nodes) in which the field vanishes. x' and y' are the normalized width and depth space variables, defined as

$$x' = x/w \qquad y' = y/\delta\,. \quad (6)$$

where w is the waveguide width and δ is the skin depth of the guiding zone.

A numerical solution technique, similar to the one given in Ref.7 involves solution of a (196x196) matrix and therefore has 196 eigenvalues. We are interested in finding only those eigenvalues that correspond to the mode propagation constants. The parameters β and ϕ, that are defined later, are used to convert the two-dimensional scalar Helmholtz wave equation into a normalized standard partial differential equation of the form[7]:

$$\frac{\partial^2 E}{\partial x'^2} + \frac{1}{r^2}\frac{\partial^2 E}{\partial y'^2} + \phi^2[f(x'y') - \beta]E = 0\,. \quad (7)$$

The parameters β and ϕ of a d width waveguide are defined as:

$$\beta = \frac{\frac{1}{(V_{eff})^2} - \left(\frac{1}{V_f}\right)^2}{\left(\frac{1}{V_s}\right)^2 - \left(\frac{1}{V_f}\right)^2} \tag{8}$$

$$\phi = \frac{2\pi}{\lambda_0} d \left[\left(\frac{1}{V_s}\right)^2 - \left(\frac{1}{V_f}\right)^2 \right]^{1/2} \tag{9}$$

where $1/V_{eff}$ is the effective mode index. $1/V_f$ and $1/V_s$ are the indices in the fast and slow regions, respectively, and λ_0 is the center wavelength of operation (under the waveguide the wave propagates in a slower velocity ΔV than outside the waveguide V). The resultant eigenvalues λ are given as $\lambda = \phi^2\beta$, thus for a given ϕ, the normalized propagation constant β can be found.

Since $0 \leq \beta \leq 1$, only the eigenvalues that lie between zero and ϕ^2 correspond to the modes of the waveguide[8]. A numerical computational technique that enables one to find the largest eigenvalues of a matrix between 0 and ϕ^2 was used. In our simulations ϕ was taken as a parameter and yielded the mode propagation constant eigenvalue β as a function of the various corrugation structures.

Using a 3D computational technique the relevant modes were then illustrated. The x,y axes are the real topographical directions and the z axis describes the mode amplitude. Figs.3(a)-(d). This 3D illustration describes the eigenfunction of the corresponding β of each of the perturbation structures. As β increases, its shape becomes more confined with steeper walls. Smaller values of β yield a smeared mode across the waveguide section with lower magnitude.

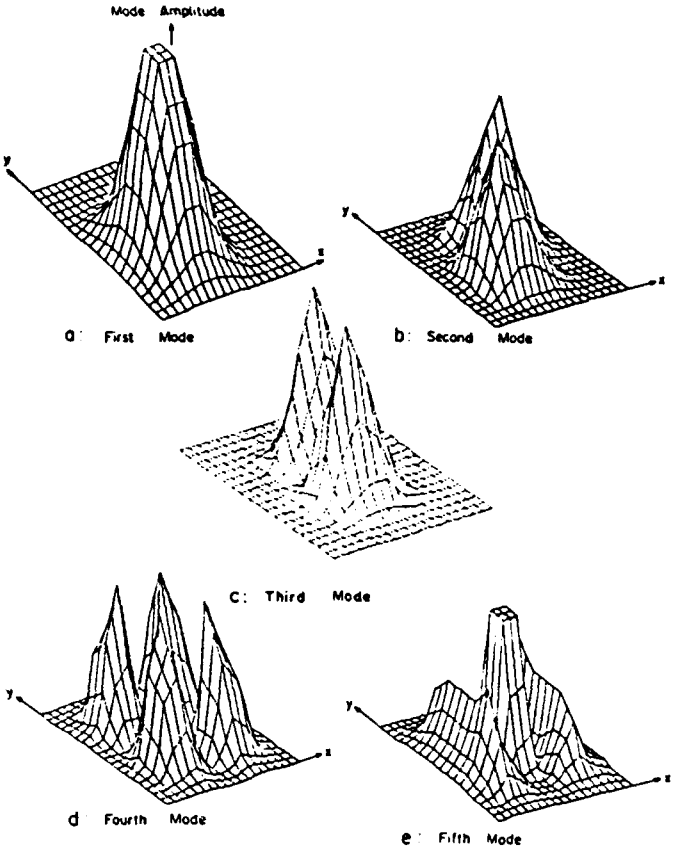


Fig. 3: Basic mode description of first five fundamental modes in the guiding zone underneath a $\Delta V/V$ waveguide. The z axis represents the mode amplitude along the waveguide. The waveguide parameter ϕ was 5 and yielded β for each simulation.

III. CORRUGATION STRUCTURES

Eight types of corrugation structures were devised, investigated, and simulated (Fig.4). A first category consisted of three types of waveguides: cross type, pallet structure and arrow type with externally mounted corrugations. The second category consisted of five internally corrugated waveguides: H type, V groove, normal well, diamond well and tunnel type.

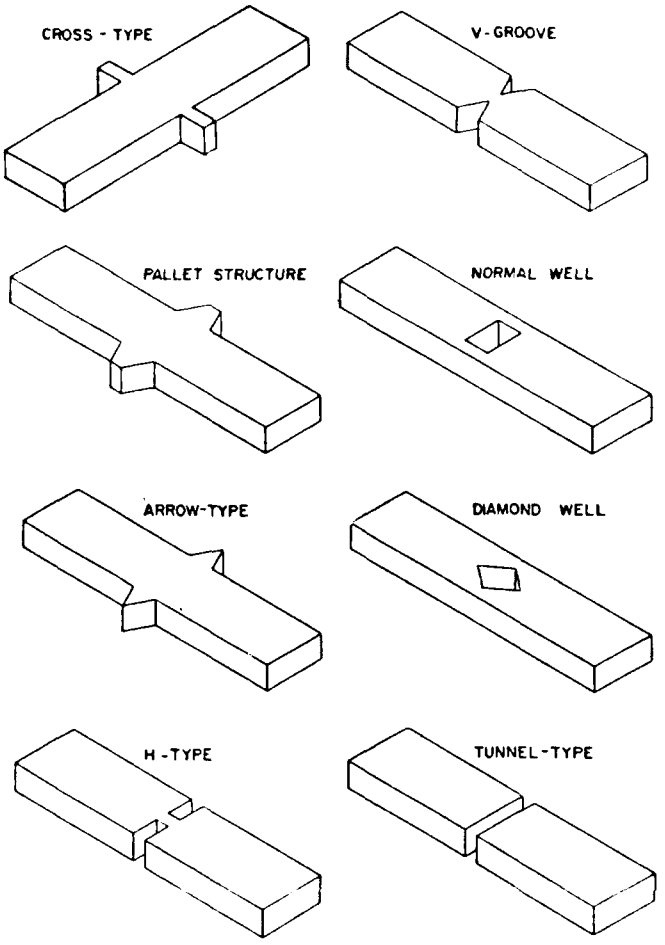


Fig. 4: Eight types of internal and external corrugation structures.

A. Simulation procedure

The main interest in perturbed waveguides is their transmission properties as a function of various physical factors. Among these factors are the number of perturbations, the perturbation geometry, and the ability to guide and control the reflected waves with maximum confinement to prevent interferences with adjacent acoustical elements. In each of the studied structures, the normalized modal fields contours and their 1/V profiles as a function of x and y are presented. This was done for the first five fundamental modes. The importance of these modal field distributions is in the design of micromirrors for microresonators, acoustic directional couples, splitters, and filters. Figure 3 shows an example of normalized modal shapes of the five fundamental modes. These modal illustrations emphasize the exact propagating energy locations of each mode and its profile.

The parameter ϕ in most cases has a typical acoustic value of 5 and yields β as a result. The modal field distribution is presented in a 3D figure where the x,y axes indicate the geometrical location and the x axis illustrates the mode amplitude distribution in that specific location. This procedure is repeated for up to 200 perturbations. The transition from one perturbation to another is obtained using the overlapping matrix multiplication between a perturbed and an unperturbed waveguide section.

In such a multiplication of the modal fields matrices, part of the field energy is lost in a radiative leakage mechanism. There could be three major possible reasons for this energy loss in such a transition: (1) energy transfer to a reflected mode; (2) leakage to the substrate by bulk modes; and (3) surface scattering beyond the acoustic waveguide section. In the following simulation the perturbed waveguide was examined as a complete unit of reflecting mirror without distinguishing between the three mechanism of energy loss.

The matrix multiplication for one transition from the main guide to the perturbed region and then back to the main guide requires a complicated procedure since many matrices are involved in it. The cross-section shape of the polygonal structures creates slices of series planes that require many multiplications, one by the other, in order to achieve a proper guide geometry description. The accuracy of this matrix multiplication technique is a function of the number of slices that are taken in this computation. The percentage of energy loss in such a mode transition from one cross section to another is then calculated for the whole transition, which is equal to one perturbation. The total energy loss after a certain number of perturbation is calculated as 1/aN where a is the characteristic loss of one perturbation and N is the number of perturbations. This function decays asymptotically to zero with the increase of N. The total number of perturbations that causes maximum energy reflection in a waveguide section is the required parameter in a micromirror fabrication. This parameter, however, is not precise, since it is influenced by the three additional energy loss mechanisms, causing supplementary losses in the wave propagation. The physical interpretation of these overlapping matrices resembles that of the optical case in which a stack of slides are positioned in series representing a periodic perturbation structure.

A number of 100 perturbations upon the waveguide was taken as a maximum in the incident wave simulations, since all the modes lost most of their energy at that point. In the reflected wave simulations, however, the maximum number of perturbations was taken as 230 for a total reflection, but with a different initial condition of the input signal.

The simulations were repeated twice, once for the transmission waves and once for the reflected waves. In each simulation 16 dependencies were examined and analyzed in the following four sets:

(1) a^{-N} (mode amplitude) versus the number of perturbations (p) in the investigated devices for modes one to five.

(2) a^{-N} versus the number of perturbations for K=1 :- 5 in devices 1 :- 4.

(3) a^{-N} versus the mode number (K) for 10,20,50, and 100 perturbations for the investigated devices.

(4) a^{-N} versus the mode number (K) for 10,20,50, and 100 perturbations in devices 1 :- 4.

B. Incident waves simulations

Mode amplitude (a^{-N}) versus the number of perturbations for constant K

Fig. 5 shows $a^{-N}(p)$, in the first set, for an effective fundamental mode (K=1 -: 5) in the investigated devices. A basic fact that is revelead by this simulation is the nonlinear decay curve of the propagating wave in relation to the number of perturbations. Device No.4, (the diamond well) has a stronger decay and a corresponding reflectivity than the normal well, the V groove and the star type.

The explicit expression for R(z) and S(z) (the reflected and incident waves) including the decay factor (α) along z direction of the waveguide and the periodic perturbations, are[3] the set of coupled mode equations:

$$\frac{\partial}{\partial z} R(z) = \alpha R(z) + iK_r e^{i(2\delta)z} S(z) , \qquad (10)$$

$$\frac{\partial}{\partial z} S(z) = -\alpha S(z) - iK_r e^{(-i(2\delta)z)} R(z) \qquad (11)$$

where δ is defined as the grating perturbation mismatch, or[3]

$$\delta \underline{\Delta} \beta - \frac{\pi}{\lambda_r} = \beta - \beta_r . \qquad (12)$$

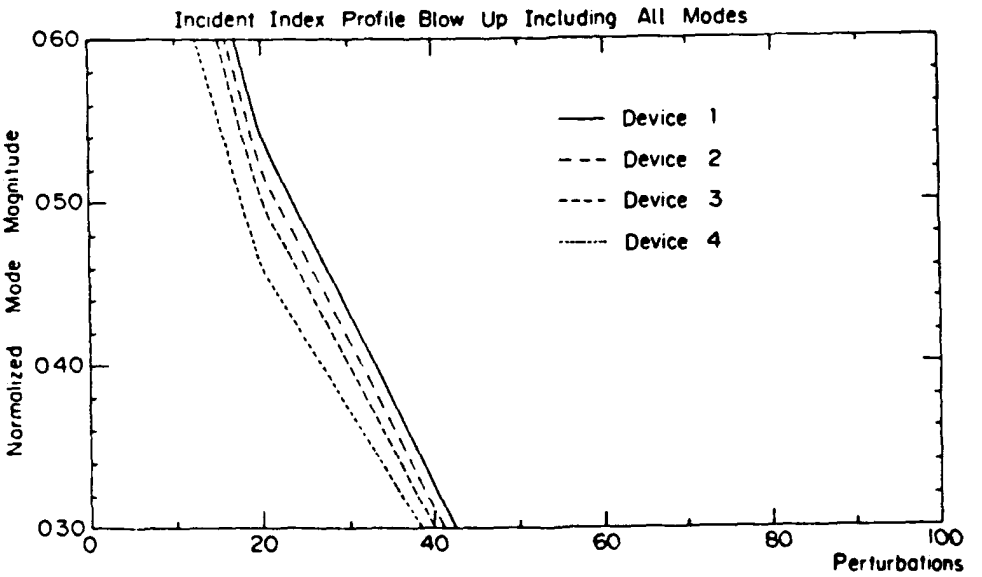


Fig. 5: A blowup of the incident wave modal decay in four sample devices: diamond well (device No.4), normal well (device No.3), V groove (device No.2), and star type (device No.1).

K_r is the coupling coefficient[3] and a is the longitudinal decay factor due to dissipative losses of the waveguide material. λ_r is the grating period of a z dependent periodical perturbation.

Final expressions for R(z) and S(z) are also given in Ref.3 as well as their graphical representation. For frequencies around the center frequency f_0, the mode power function of the transmitted signals decreases exponentially along the corrugated waveguide section whereas the reflected signal rises in the same manner in the opposite direction due to coupling between the forward and backward propagating modes. The two external perturbed devices decay faster than the internally perturbed devices, but almost in the same way.

Figure 6 in the second set, is an example of a^{-N} dependence on the number of perturbations (p) for one of the examined devices (diamond well) with the five fundamental modes as parameters. it is obvious that the first mode is the strongest and fifth one is the weakest. Around 50 perturbations, a^{-N} decreases to less than 10% of its magnitude. The behavior of the diamond well was found to be typical for the other examined devices.

In the third set of simulations an attempt to compare the four examined devices performance as a function of their mode numbers in 10,20,50 and 100 perturbations was made.

In the reflected wave simulations the same four basic comparisons were done: (1) $a^{-N}(p)$ with the device type as a parameter for modes one to five; (2) $a^{-N}(p)$ with the mode number as a parameter for devices one to four; (2) $a^{-N}(K)$ with the device number as a parameter for 10,20,50, and 100 perturbations; and (4) $a^{-N}(K)$ with the perturbation number as a parameter for devices 1 -: 4.

Figure 7 shows the reflected wave profile for all the examined devices. The star type device which is externally perturbed has the best reflections properties while the normal and diamond wells are less efficient. However, these internal perturbed devices still remain advantageous since most of the the propagating wave remains confined underneath them.

Investigation of the reflectance of a well type device is illustrated as an example in Figure 8. Here K is a parameter and K=1 is the strongest. The index, n line in a typical index value, representing this device.

IV. DISCUSSION AND CONCLUSIONS

The goal of the investigation of microreflectors was to achieve equivalent performances as in conventional grid reflectors, in order to obtain low loss integrated SAW devices. Based on coupled mode theory, a microresonator prototype was developed and analyzed in a previous publication without micromirror optimization. This device had high insertion loss, due to its leaky perturbed waveguide and inefficient electromechanical transduction into and out of the substrate crystal.

In order to a void the high insertion loss of the conventional micromirrors, investigation of optimal corrugation structures was performed to prevent or reduce, at least, the leakage from the waveguide reflectors. Three possible mechanisms are responsible for this leakage in a perturbed waveguide: (1) wave diffraction, (2) decay into the substrate, and (3) energy transfer between modes.

In order to minimize the wave diffraction eight different types of corrugation structures were studied. The star type and pallet structure devices were suggested. Their smooth contour lines decrease somewhat the modal energy loss in the waveguide.

The ingrowing perturbations devices; the V groove and the H type structures improved the outgrowing perturbations performances of the star type and pallet structure since the diffraction took place inside the system. The internal perturbations devices: the normal well and the diamond well had a preliminary advantage of reduced internal diffraction. This fact was proven by the simulations since in reflection simulations less perturbations were required for better transmission. As a result the internal perturbed waveguide, particularly the diamond well is the optimal corrugation structure.

ACKNOWLEDGMENT

The authors wish to express their appreciation to R. Dahan for his assistance in this work.

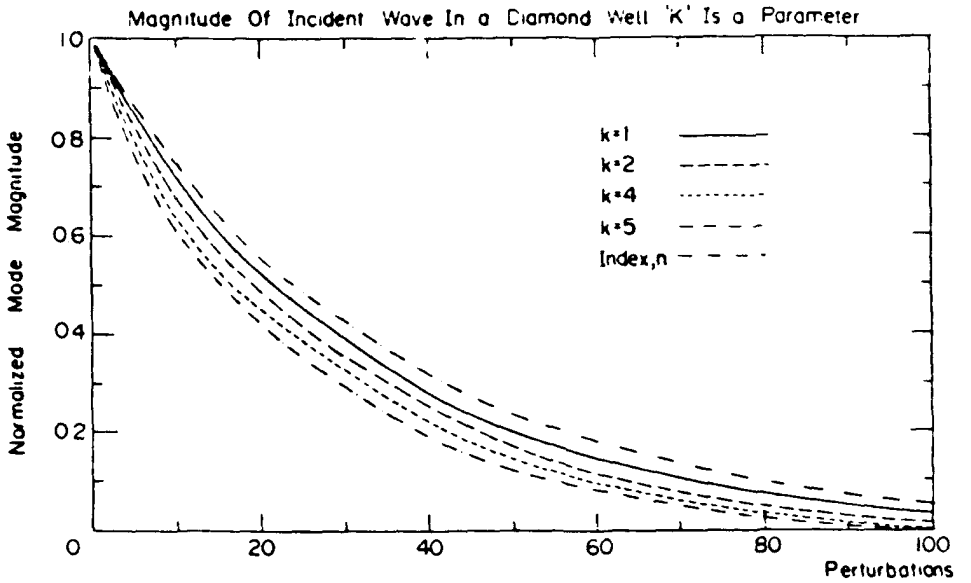


Fig. 6: Five fundamental modes in a diamond well device. The index, n line represents an effective index value, typical to a specific device.

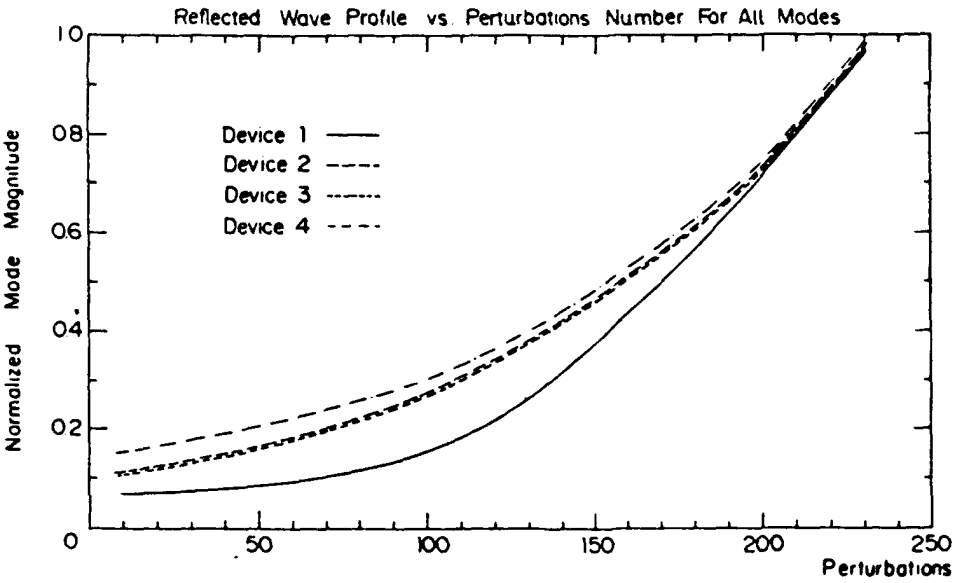


Fig. 7: Reflected wave profile for four sample devices. Each device was represented by its effective index. Devices 1-4 are the star type, V groove, the normal well and the diamond well, respectively.

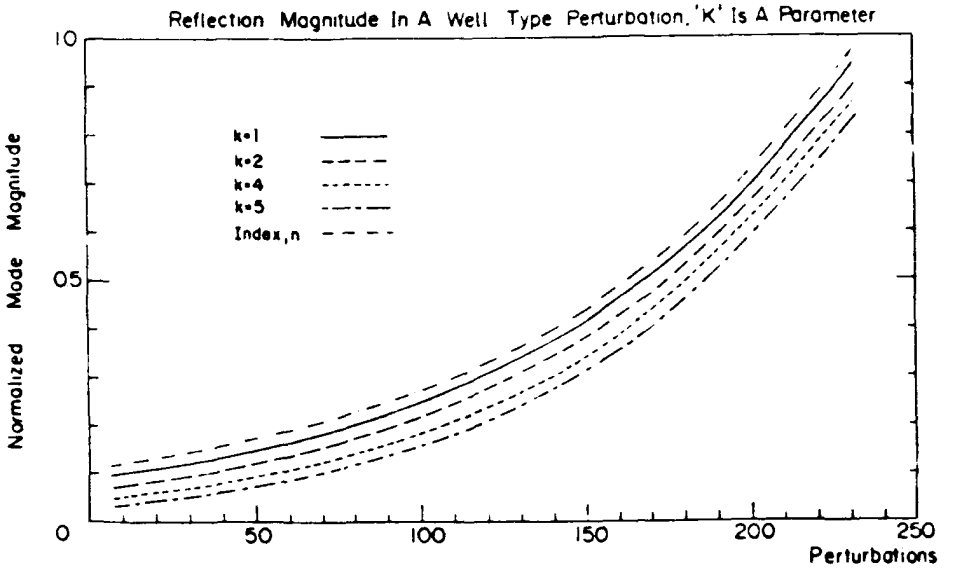


Fig. 8: Reflections in an internally perturbed device (well type) with the mode number K as a parameter.

1. H. Kogelnik and C.V. Shank, J. Appl. Phys. **43**, 2327 (1972).

2. E.A.J. Marcatili, Bell Syst. Tech J. **48**, 2071 (1969).

3. G. Golan, G. Griffer, A. Seidman, and N. Croitoru, J. Appl. Phys. **66**, 1428 (1989).

4. M. Matsuhara, J. Opt. Soc. Amer., **63**, 1514 (1973).

5. G.E. Forsythe and W.R. Wasow, Finite Difference Methods for Partial Differential Equations (Wiley, New York, 1960, p.150).

6. H.F. Taylor, IEEE J. Quantum Electron, **QE-12**, 748 (1976).

7. R.K. Lagu and R.V. Ramaswamy, IEEE J. Quantum Electron. **QE-22**, 968 (1986).

8. G. Grifferl, G. Golan, S. Ruschin, A. Seidman, and N. Croitoru, IEEE U.F.F.C. **35**, 503 (1988).

9. G.B. Hocker and W.K. Burns, IEEE J. Quantum Electron -**11**, 270 (1975).

10. H.F. Tiersten, J. Appl. Phys. **40**, 720 (1969).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FBAR FILTERS AT GHz FREQUENCIES

C. VALE, J. ROSENBAUM, S. HORWITZ, S. KRISHNASWAMY AND R. MOORE

Westinghouse Electric Corp.
Electronic Systems Group
Baltimore, Md. 21203

ABSTRACT

The Film Bulk Acoustic Resonator, or FBAR, is a new technology which has been used in the design and fabrication of microwave filters. These filters exhibit a combination of low loss, small size, frequency and bandwidth not attainable by any other type of device. Using a piezoelectric film of zinc oxide as a high Q crystal resonator along with passive components it is possible to form an integrable, high performance bandpass filter. The range of center frequencies achievable is roughly 1 to 5 GHz using current resonator designs. The technique features potentially low cost (since the filters are fabricated by batch processing), and, perhaps most intriguing, the capability of being combined with more complex functions (such as a complete radar receiver) as part of a larger mask set. Other advantages are moderate power handling capability, good temperature stability, and high dynamic range.

INTRODUCTION

The performance requirements of front end filters above 1GHz are increasingly difficult to meet with traditional approaches of lumped element, dielectric or surface acoustic wave filters. In particular size and weight must be increased if low insertion loss is desired and vice versa. Bulk acoustic wave filters offer unique advantages because they are much smaller than dielectric or lumped element devices and possess much lower insertion loss than surface wave devices. Furthermore, since they are fabricated directly on semiconductor wafers they can be integrated with active circuitry to form a component in a complete receiver on a chip.

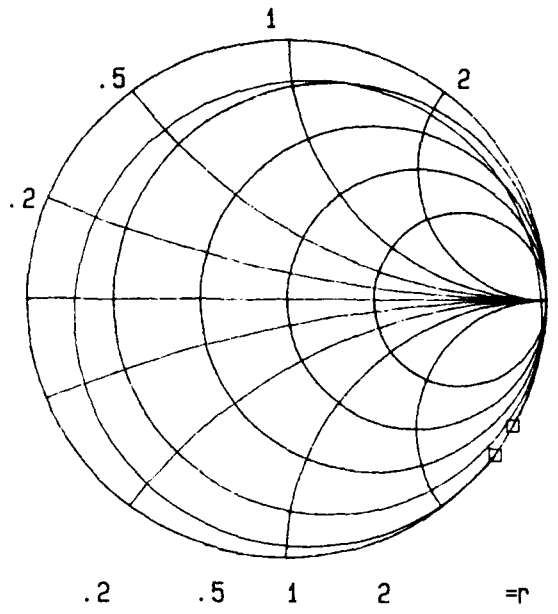
FILM BULK ACOUSTIC RESONATOR DESCRIPTION

The most important components in a multipole FBAR filter are the acoustic resonators. They consist of a four layer composite structure in which sputtered SiO_2 is used as a support layer both to provide mechanical rigidity and for temperature compensation[1]. A thin layer (usually less than .25 microns) of crystallographically ordered gold is deposited over the amorphous SiO_2 which acts as the growth bed for the piezoelectrically active ZnO layer which is deposited by r.f. magnetron sputtering. The top electrode is the fourth layer and consists of an evaporated aluminum film of thickness between .2 and .5 microns.

Since the resonant frequencies and piezoelectric coupling of the FBAR depend on the material properties and thicknesses of all the layers it is essential to have a circuit model which accurately describes this relatively complex configuration. Since the metal layers are quite thick relative to the piezoelectric layer at frequencies above 1GHz they cannot be neglected in the analysis and thus closed form expressions (which can be written only for one or two layer structures) cannot be used.

The one dimensional Mason model in which an arbitrary number of acoustic layers (piezoelectric and non piezoelectric) can be cascaded to form a complex structure has been shown to give excellent correlation with experimental results if the resonators operate in a trapped mode[2]. This model yields the input impedance as a function of frequency; it is relatively straightforward to extract a Butterworth-VanDyke equivalent circuit by inspecting the impedance characteristic near the series and parallel resonances of the structure[3]. Figure 1 shows the complex impedance using the Mason model of a typical resonator near 1 GHz and the experimentally measured responses. Removing the electrical resistance (approximately 2 ohms) which contributes to the motional resistance at series resonance the agreement is seen to be quite good. One important factor which is not predicted by the one dimensional model is the presence of in-band spurious responses as seen in the experimental curves. While it is not possible to entirely eliminate the spurs their size can be reduced by careful attention to cleanliness during the resonator fabrication. Their effect on the the pass band can be further reduced by coupling techniques as explained in the next section.

CH2818-3/90/0000-332 $1.00 © 1990 IEEE



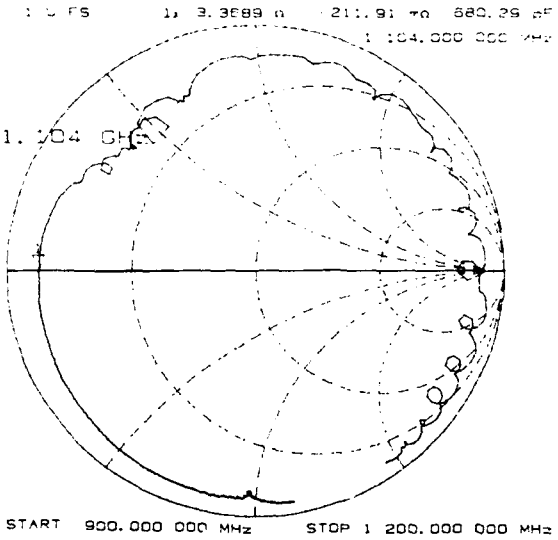
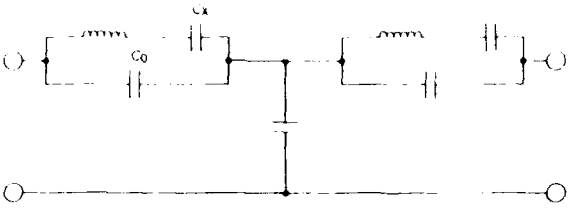


Fig. 1. Complex impedance of computer model and measured FBAR data

FILTER DESIGN

The ideal configuration for a filter using FBAR resonators is the resonator-and-capacitor-only ladder of figure 2. It is a configuration which can make a symmetrical passband if the fractional bandwidth is less than about:

$$\text{i.e.:}\quad \frac{\Delta f_3}{f_o} < \frac{\frac{1}{5}(C_x/C_o)}{5} = \frac{1}{5}\left[\frac{C_x}{C_o}\right]$$



BULK ACOUSTIC RESONATOR NARROW BAND LADDER

$$BW_3 < \frac{f_o(C_x/C_o)}{5}$$

Fig. 2. Two-pole narrow band ladder filter schematic

Non-symmetrical passbands can be made wider than this but they go hand-in-hand with severe group delay distortion.

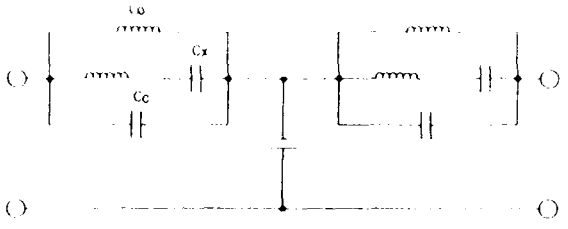
Since the capacity ratio limits bandwidth, one simple approach to wider bandwidth symmetrical filters is resonating the C_o with an inductor at f_o thus effectively canceling it out over a limited bandwidth. By computer modeling and practical demonstration this is known to be an effective technique. However, it has the disadvantage of introducing a spurious passband extending from D.C. to $f_o/\sqrt{2}$. This must be suppressed by cascading with a high pass or band-pass "guard" filter. Alternatively, other sources of selectivity may already reject that band from the system.

Full-lattice and half-lattice balanced networks would offer symmetrical passbands; but they are not practical for the very high frequency bands of FBAR. The chosen filter circuit for this resonator is the ladder configuration of figure 3 with deposited spiral inductors. The network topology meets these desirable goals:

1. Cascaded resonator ladder configuration
2. All resonators identical
3. No variable elements
4. Temperature stable
5. Completely monolithic

The simple two-pole filter can be designed by presuming that the filter consists of two series resonant circuits in series, coupled together by a capacitor connected to ground between them. Thus, we make the approximation that the C_o-L_o combination is negligible in the vicinity of the center frequency. This is a valid assumption over the band of frequencies $f_o(L_o/L_x)^{1/2}$ centered at f_o.

Desired low pass prototype characteristics (Butterworth or Chebychev) can be selected from Zverev[4] and converted to the band-pass form described above using the Humpherys[5]



BULK ACOUSTIC RESONATOR WIDE BAND LADDER RESONATING C_o WITH L_o

Fig. 3. Two-pole wide band ladder filter schematic

procedure. The latter is an extremely flexible technique leading to many variations of circuit configurations through the use of constant reactance inverters; the importance of this will be made evident. Figure 4 shows the ideal computed gain response of a Butterworth two pole filter, and the resulting response when finite Q is included in the inductors, capacitors and resonators. The most significant contributor to insertion loss is the Q_x of the resonators (.8 db); finite Q in the coupling capacitor adds about 0.15 db and the loss in the coil L_o adds about .1 db. Thus improvement in the resonator Q offers the best return; but there is a point of diminishing returns because doubling Q_x merely drops the insertion loss to .65 db.

If two identical units are cascaded without any regard to the exactitude of the design the resulting filter would not necessarily result in a recognized four-pole response. However, performing a low-pass to band-pass transformation with constant reactance inverters can solve two problems: exact design of multi-pole Butterworth and Chebychev filters can be performed; and some relief from the effects of unwanted spurious responses can be achieved. It is our intention to use simple two-pole filters as building blocks to achieve well known tabulated filter characteristics of even order. A four pole network can be designed having the resonant frequency located near the center of the filter instead of at the lower band-edge as in figure 2. This requires a coupling inductor connected to ground between the building-block pairs. It

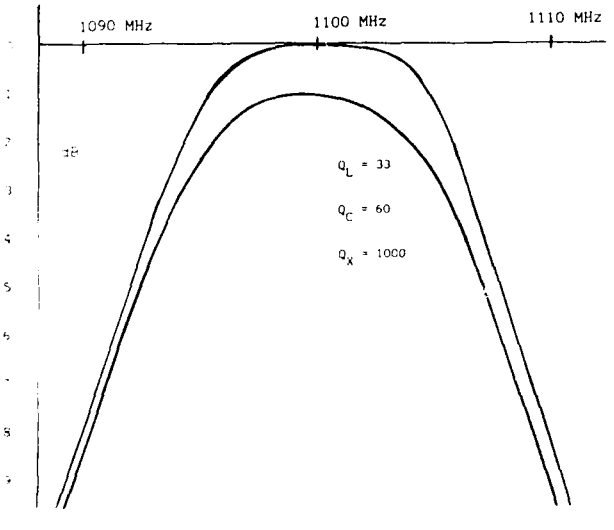


Fig. 4. Ideal and realizable filter response of two-pole wide band ladder filter

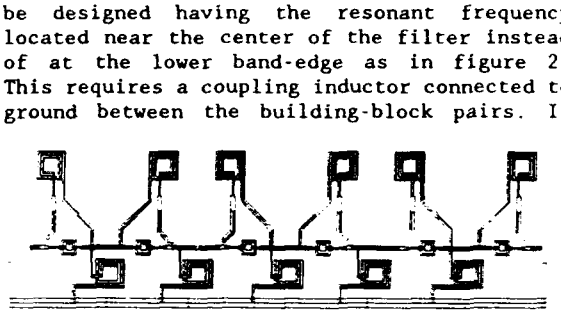


Fig. 5. Six-pole wide band ladder filter layout with all inductive coupling

thus moves all spurious (which are always located above the resonator series resonance) one-half a bandwidth upwards assuring a spurious free passband and moving those in the stopband down to lower level on the skirt. The circuit is achieved by using two negative constant reactance inverters and one positive inverter. All four resonators are identical in frequency and motional inductance. A six-pole network as shown in figure 5 would require two different building blocks: two with resonator pairs located slightly below center frequency (Pair I) and one pair almost exactly at center frequency (Pair II). Figure 6 shows the computed response for both the lossless case and for the case where the elements have realizable Q. The design features only two different resonator pairs; the frequency of each is located near the middle of the passband; the coupling is accomplished with alter-

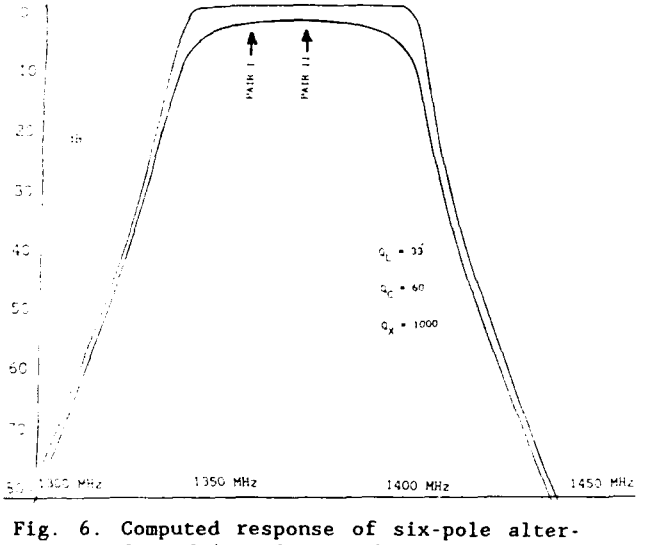


Fig. 6. Computed response of six-pole alternate coupling filter for lossless (infinite Q) case and for realizable filter finite Q

nating reactance signs; and two extra inductors are needed on the ends to make the configuration an exact cascade of FBAR two-resonator chips.

FABRICATION

A two-pole FBAR filter operating in the 1GHz range and using a ladder configuration has been fabricated and tested. The design incorporates techniques which emphasize monolithic low cost fabrication which is compatible with MMIC circuitry. The device layout for the two pole filter is shown in figure 7. The FBARs are the most critical components and their fabrication is the most difficult step. The ordered gold deposition is performed at critical temperature, rate and vacuum requirements. The zinc oxide will generally be strongly active if the gold is well ordered. The sputter deposition must be performed so that the ZnO film has the lowest possible mechanical stress consistent with piezoelectric activity. Low stress films are essential both for mechanical integrity and electrical performance. Compressive films bow upward and have been observed to shatter even with the slightest applied pressure. Stress also seriously degrades the resonator Q and increases insertion loss.

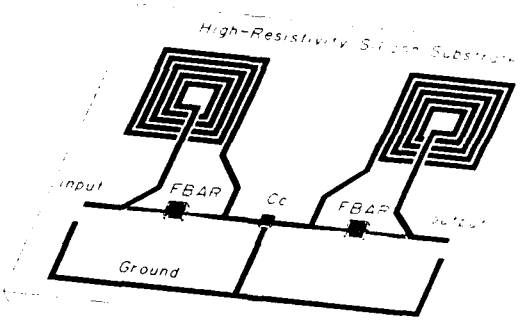


Fig. 7. Device layout for fabricated two-pole

The compensating inductors and coupling capacitor are fabricated using standard microelectronic techniques. The capacitor dielectric consists of 2500Å of silicon nitride deposited using a plasma enhanced chemical vapor deposition process. The inductor is a standard spiral configuration whose value and geometry are determined using computer aided design techniques[6]. A thick gold metalization of at least 4 microns is used to increase the Q. This metalization is performed using electron-beam evaporation and the delineation of the inductors is done by either etching or lift-off. The inductor lines are broken and air bridges are fabricated over the center conductor. Air bridges are also fabricated at the coupling capacitor connections. These bridges are quite sturdy and hold up well in subsequent processing.

Fabrication of the acoustic resonators requires that the substrate crystal be entirely removed under the resonating area. This is accomplished for both silicon and gallium arsenide by reactive ion etching the substrate the same way that via holes are formed in analog MMIC's. Using RIE allows the etching to be performed in a batch process (with up to 25 wafers at a time) without requiring protection of the front side of the wafer as in wet chemical etching. For gallium arsenide a chlorine based chemistry is used while for silicon sulfur hexafluoride gives excellent results. Etch rates of up to 8 mils per hour for silicon and 6 mils per hour for gallium arsenide have been achieved.

DEVICE PERFORMANCE

Figure 8 shows the measured gain response of a two pole monolithic FBAR filter designed on these principles. It is remarkable for its low insertion loss and deep stop bands. For this value of insertion loss the filter performance is primarily limited by the quality of the FBAR's. There are, however, spurious responses occurring in the passband and on the upper stopband skirt. These spurs cause the small ripple in the pass band. The effect of the spurs can be minimized however, by improving the processing. The out of band

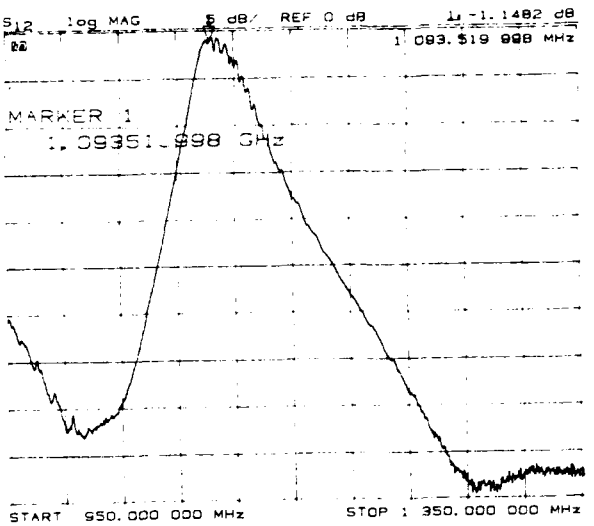
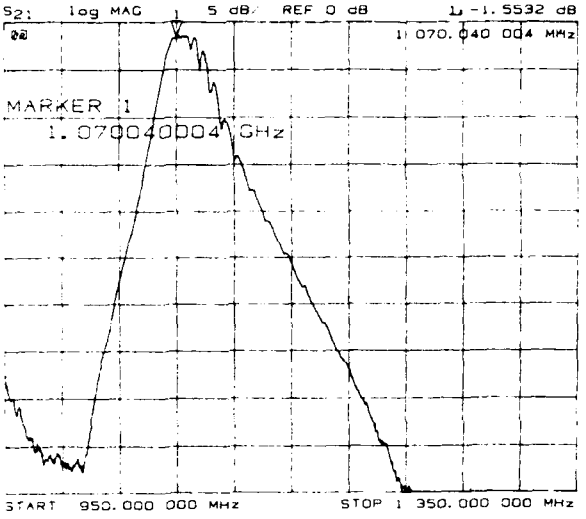


Fig. 8. Typical performance curve of low loss two pole filter.

rejection also depends on the quality of the FBARs as well as their dimensions and is typically greater than 50 dB for a two pole device. Figure 9a show a typical filter with relatively large spurious ripple; in figure 9b the coupling capacitor has been replaced with a coupling inductor whose value (approximately 8 nH) is determined by Humpherys' design procedure. Note that in this case the spurs have been shifted to the high end of the response and do not appear in the pass band. Figure 10 shows the response of a cascade connection of two two- pole filters and offers phenomenal

100 dB stopbands and less than 4 dB insertion loss. It is desirable to achieve the sharper selectivity of multiple resonator filters; and the Humpherys technique is a viable approach to achieving it.

For the prototype filter the chip size was chosen for convenience of handling to be 300 mils per side. Using this oversized design there are 69 devices on a 3 inch wafer. The size of the individual chips can be reduced by more than 50% without changing the basic design. Thus it should be possible to pack at least 150 devices into a 3 inch or nearly 300 devices into a 4 inch wafer. Increasing the FBAR dimensions reduces the inductor size allowing even more devices on a wafer.



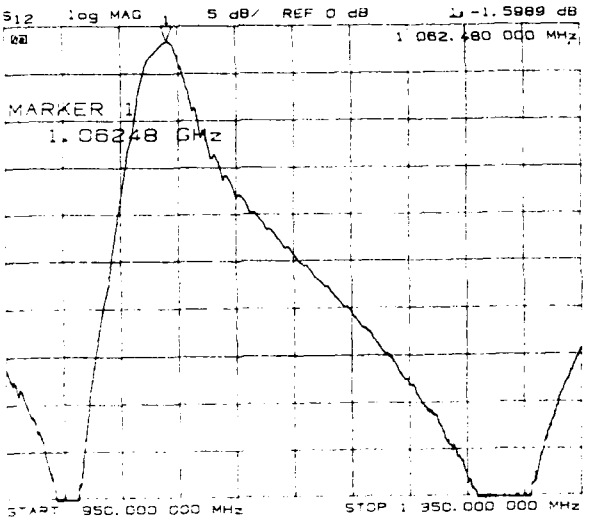


Fig. 9. Effect of coupling inductor on the response of two-pole filter

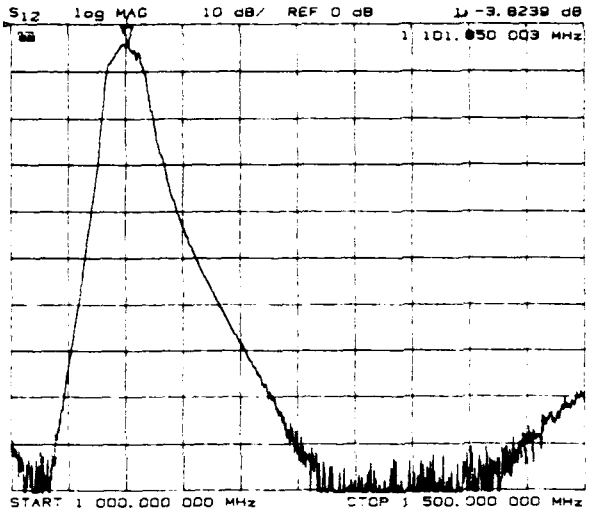


Fig. 10. Four pole response obtained by cascading two two-pole filters

REFERENCES:

1. M. Driscoll, R. Moore, J. Rosenbaum, S. Krishnaswamy, and J. Szedon, "Recent Advances in Monolithic Film Resonator Technology" IEEE Ultrasonics Symposium Proceedings 1986 pp.365-369

2. J. Rosenbaum, H. Salvo and S. Krishnaswamy, "Overtone Response of Composite Bulk Acoustic Resonators" Proc. 40th Annual Frequency Control Symposium 1986 pp. 206-209

3. J. Rosenbaum, "Bulk Acoustic Waves: Theory and Devices", Artech 1988 Chapter 11

4. A. Zverev, "Handbook of Filter Synthesis", John Wiley, 1967.

5. D. Humpherys, "The Analysis, Design and Synthesis of Electrical Filters", Prentice Hall 1970 p466

6. D. Krafcsik and D. Dawson, "A Closed Form Expression for Representing the Distributed Nature of the Spiral Inductor" IEEE Microwave and Millimeter Wave Monolithic Symposium 1986 pp. 87-92

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NEW DESIGNS FOR RESONATORS AND FILTERS USING LITHIUM TANTALATE

J. Détaint, J. Schwartzel, H.Carru, R. Lefèvre, C. Joly
CNET dept PAB/BAG/MCT/CMM 196, Av. Henri Ravera Bagneux France
And
B. Capelle, Y. Zheng, A. Zarka
L.M.C. Université P. & M. Curie Place Jussieu Paris France

ABSTRACT: Lithium tantalate is a material which has very large electromechanical coupling coefficients together with compensated orientations for the shear modes. In this paper we report the results of investigations made in view to improve the characteristics of fundamental and overtone mode resonators. Using the stroboscopic topography technique, we have shown that, the rather poor Q factors of the fundamental thickness shear resonators are due to non-stationnary plate modes coupled to the fast shear. The use of bar shapped thickness shear resonators permits to obtain higher Q factors. To obtain overtone resonators with a zero first order temperature coefficient, we have considered doubly rotated cuts, the compensation of the T.C. of the overtones of the X cut by a layer of silica, and, the use of the lateral field excitation. This latter type of resonator is very well adapted to lithium tantalate and permits to obtain, using the plano-convexe geometry, very large Q factors for overtones with very elevated ranks .

I - INTRODUCTION

Lithium tantalate is a material which has for the thickness modes a very large electromechanical coupling factor (k reaching 45 % for the fast shear mode) together with compensated cuts (1) (2). These properties are very attractive to obtain filters with large bandwidth and oscillators with important shifts. Moreover , despite a rather poor crystalline quality (non stoechiometry, dislocations etc...) this material presents of one the smallest acoustic dissipation known (3).

However, excepted for some doubly rotated cuts (4) (5), only the fundamental fast thickness shear mode of cuts near the X one is thermally compensated, and, with a conventional design this mode has a rather modest Q factor. During the course of a study presently made in view to research an optimal solution for the I.F. Filters of the new pan-european numerical radiotelephone system several conventional and new designs were investigated to enhance the Q factors of devices using the fundamental mode and to obtain compensated overtones with a sufficient capacitances ratio to have the required bandwith (150 to 250 kHz at center frequencies in the range 45-75 MHz).

CH2818-3/90/0000-337 $1.00 © 1990 IEEE

Also, using the sophisticated technique of X ray strosboscopic topography, a new insight on the properties of the vibration modes in this material was obtained which explains the low Q factor observed with the fundamental mode and several features of the different new designs for overtone resonators.

II - THICKNESS SHEAR MODES IN LITHIUM TANTALATE

Several loci of zero first order temperature coefficients exist for the two basic families of shear modes in lithium tantalate (2). They are represented as a function of the plate orientation , for the fast and the slow shear modes, in Figure 1 and 2. In these figures the plate orientation is located by the two polar angles α and β defining the orientation of the normal to the plate.

For the fast shear modes (B mode) (Figure 1) the most important zero F.T.C. locus is that situated around the X cut. These orientations provide a very high coupling coefficient. In this case the thermal compensation for the resonance frequency is obtained due to a large negative value of the first order T.C. of the coupling coefficient (6). This compensation disappears for the overtones (Figure 1). The second locus of zero F.T.C. is situated around the Y rotated cuts, it exists for the fundamental mode and the overtones, but, requires the use of doubly rotated plates. For the slow shear mode (C mode) two loci of zero first order temperature exist, one is around the Z cut, the second is close to the Y-10° cut (Figure 2). For a conventional excitation by a field in the direction of the normal to the plate, doubly rotated cuts are required to have a non zero coupling coefficient.

Using, the same representation, level curves for the electromechanical coupling coefficient are displayed for the fast shear (B mode) and the slow shear (C mode) in Figures 3 and 4 (2).

III - DEVICES USING THE FUNDAMENTAL MODE

The most interesting cuts are those near the X one. For these cuts the coupling coefficient is about 45 % and the frequency constants, without mass loading, are near 2000kHz.mm.

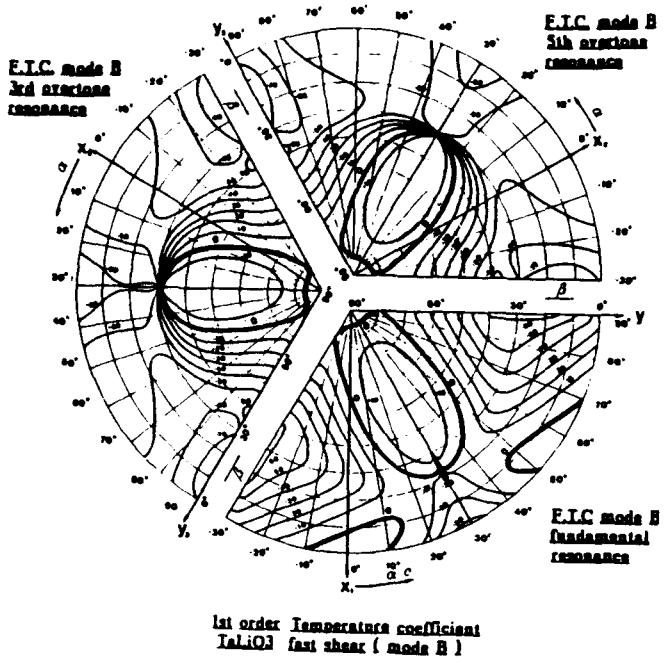


Figure 1: 1st Order F.T.C. (B mode).

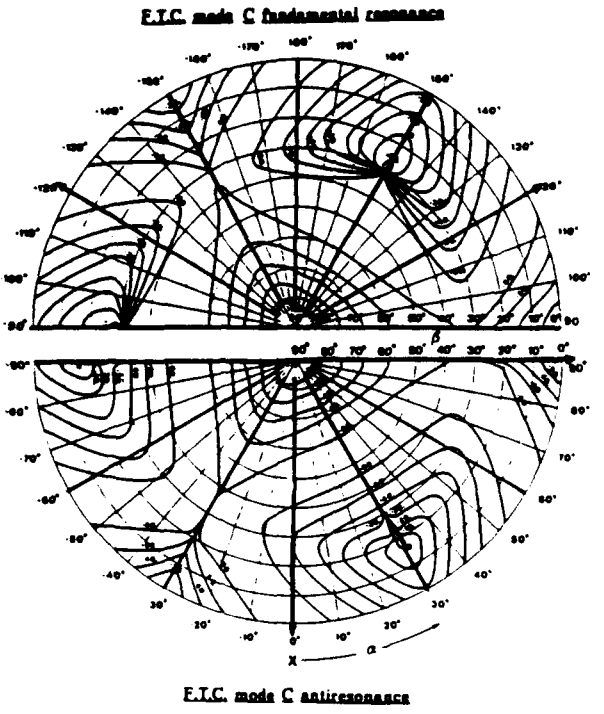


Figure 2: 1st Order F.T.C. (C mode).

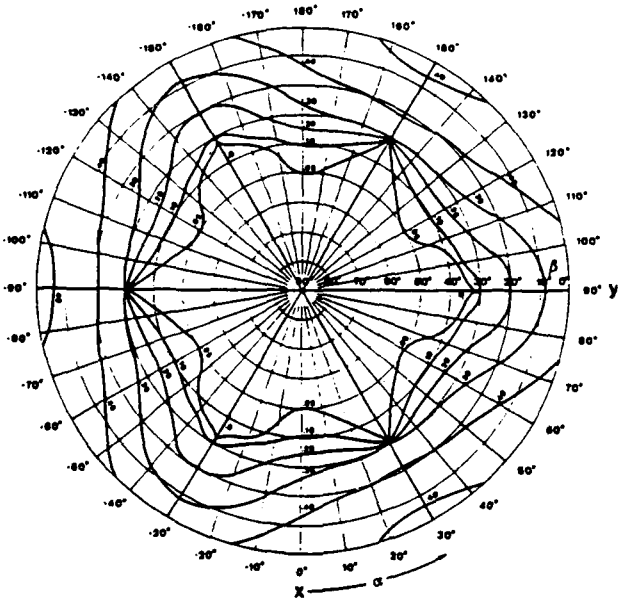


Figure 3: Coupling coefficient (B mode).

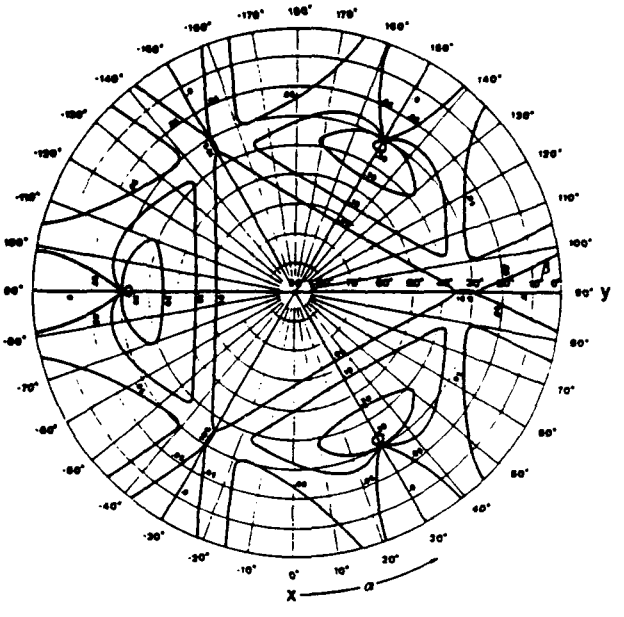


Figure 4: Coupling coefficient (C mode).

TABLE I: PROPERTIES OF THE X CUT

MODE	$\vec{u}$	VELOCITY	$\bar{C}$(MKSA)	k	T.C1Fr [°C^{-1}]	TC1Fa [°C^{-1}]	TC2Fr.1 [°C^{-2}]	TC2Fa [°C^{-2}]	Fr_1*2h
B	u_1=0.0 u_2=0.5749 u_3=-0.8181	4212 m/s	1.323 10^{11}	45,1 %	+4,0 10^{-6}	-44.10^{-6}	114.10^{-9}	-53 10^{-9}	1916 (kHz.mm)
C	u_1=0.0 u_2=0.8181 u_3=0.5749	3368 m/s	0,845 10^{11}	3,5 %	-63.10^{-6}	-63.10^{-6}	-30.10^{-9}	-30 10^{-9}	1683 (kHz.mm)

On Figure 5 are represented the most important properties of the B mode of X rotated cuts (around the y axis). A typical frequency spectrum of an X cut resonator is given on Figure 6. It can be noticed that the C mode is also excited but has a rather low coupling coefficient. The properties of the B and C modes of fundamental X cut resonators are given in table I.

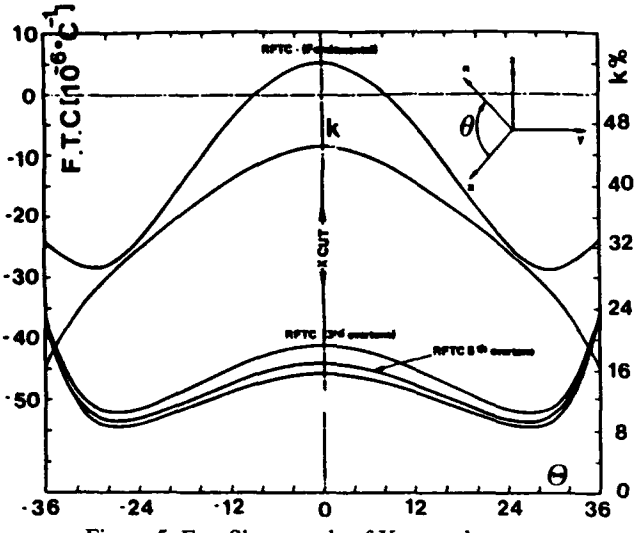


Figure 5: Fast Shear mode of X rotated cuts

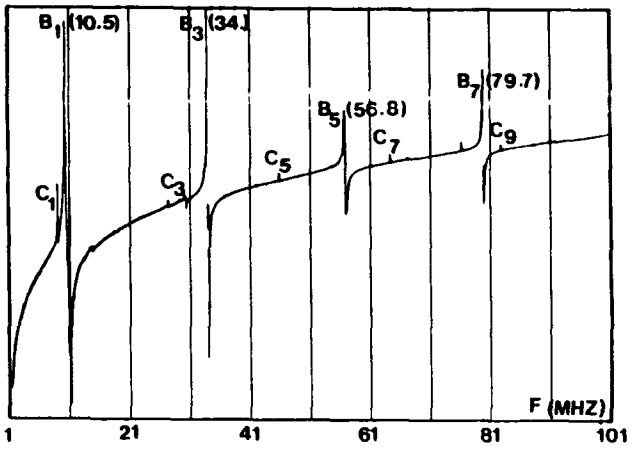


Figure 6: Frequency spectrum of an X cut plate.

III.1 - Conventional X cut resonators

The Q factors of X cut resonators using the fundamental mode are very low, in most cases in the range of a very few thousands. This is one of the main concern with this type of resonator (particularly for applications to filters with bandwith lower than 1%); one other concern being the possibility to have unwanted responses near the resonance or the antiresonance not due to anharmonic modes (as for other materials, the anharmonic modes can be suppressed by the choice of proper dimensions for the electrodes , a proper mass loading and the fulfilment of certain conditions of symmetry).

An examination of the vibration mode of such resonators was made by stroboscopic X ray. topography using synchrotron radiation (7) (8). This method allows to determine the components of the displacement u_k (x_1, x_2, x_3, t) provided they have an $e^{j\omega t}$ time dependence where ω is the angular recurrence frequency of the synchrotron radiation. These experiments were made with X cut fundamental resonators having a resonance frequency equal or close to 12667120 Hz or 15846400 Hz, using the white radiation of D.C.I at the L.U.R.E (ORSAY-FRANCE). The time resolution obtained is 1 ns and the spatial resolution in the order of some ten microns depending mostly, in this case, on the absorbtion and the perfection of the material. Some of the topographs obtained are displayed in Figure 7 and 8.

On figure 7a is displayed mostly the main component of the thickness shear displacement (component u_1 in the direction of the $\vec{u}$ vector given in table 1 for the B unidimensional mode). This topograph was obtained using the stroboscopic technique ; it can be observed that the conventional trapped shear mode observable at the center of the plate is coupled to a plate mode extending up to the edge of the plate. On Figure 7b is represented a conventional topograph obtained in the same electrical and diffraction conditions (with integration over the time of $\vec{u}$ (x_1, x_2, x_3, t). The coupled plate mode can no longer be observed in these conditions. This means that this plate mode is not a standing wave but a progressive one. It was directly verified in similar cases, by a video imaging technique using the stroboscopic effect provided by a slow shift of the relative phase

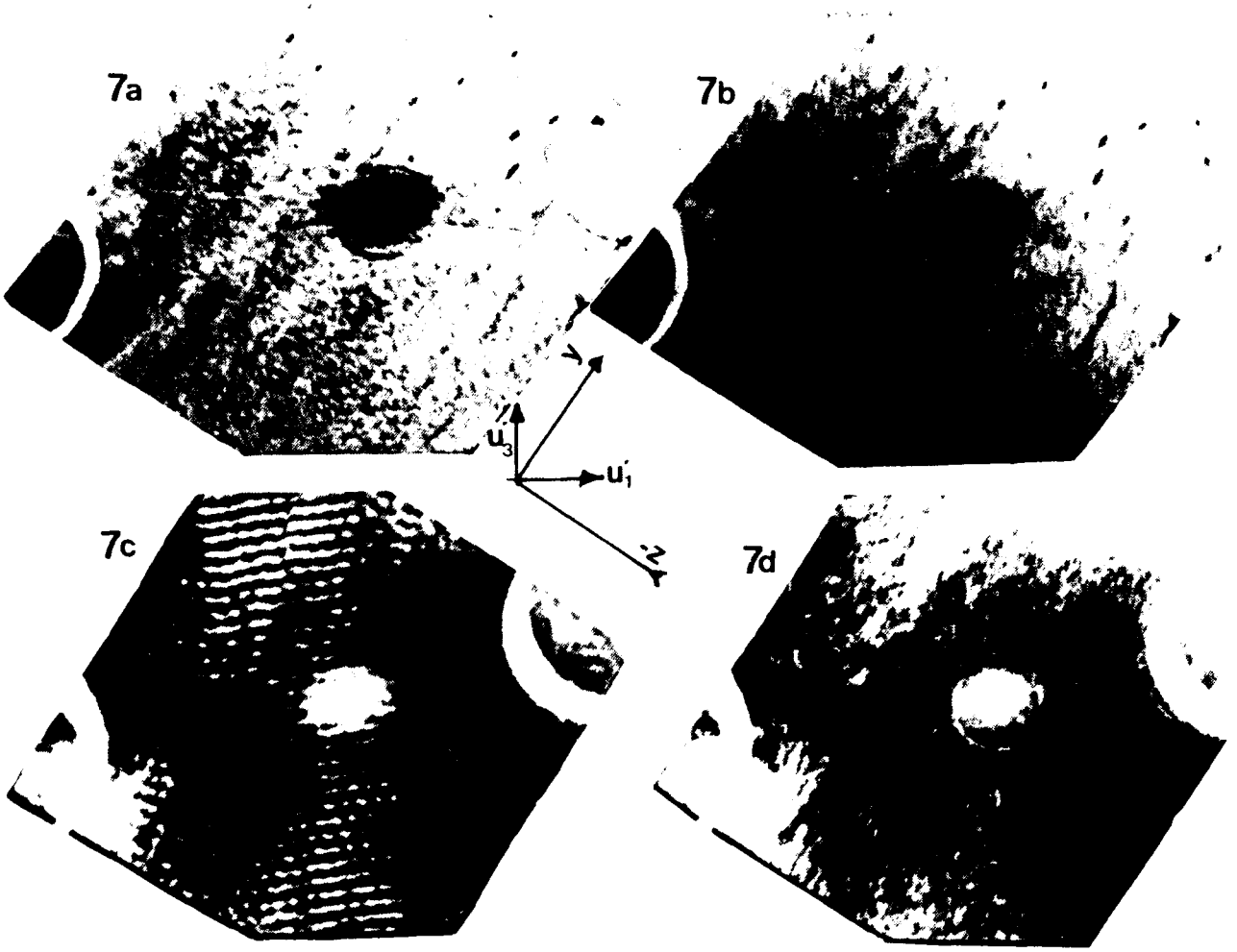


Figure 7: Stroboscopic topographs.

between the X ray pulses and the excitation current, that the plate waves were travelling from the electroded region towards the mounting clips of the resonator. On figures 7c and 7d are displayed respectively a stroboscopic and a time integrated topograph representing a linear combination of the two shear components. Again it appears that the plate coupled mode is a progressive one. In this case (corresponding to the same resonator as in figures 7a and 7b), it appears that the second progressive component has a very large amplitude.

On figure 8 are represented 3 stroboscopic topographs that display the following components of the vibration mode
- On figure 8a is displayed a linear combination of the u_2' and u_3' components of the mode (mostly u_3').
- On figure 8b is displayed a linear combination of the u_1' and u_2' (mostly u_1').
- On figure 8c we have a linear combination of the three components of $\vec{u}$.

[Where the u_j' are the projections of the observed mode on the three (orthogonal) eigen vectors representing the displacement (polarization) of the three one dimensional thickness modes of an X plate (the axis x_1', x_2', x_3' corresponding respectively to the displacements of the B, A, and C modes].

From these observations it can be inferred that the coupled plate mode(s) have in this case 3 non negligible components. Other observations made on about 10 resonators with different designs, have indicated that the u_2' and u_3' components are, in all the observed cases, progressive waves. Also it is possible to observe that the lateral anisotropy of the stationary part of u_1' (which corresponds to the usually considered vibration mode of a trapped energy resonator using the B mode) is non negligible.

The trapped part of the mode extends more on the direction x_1' of its polarization than on x_3'. As indicated by the theories (9) (10) (11), this component displays an elliptical shape.

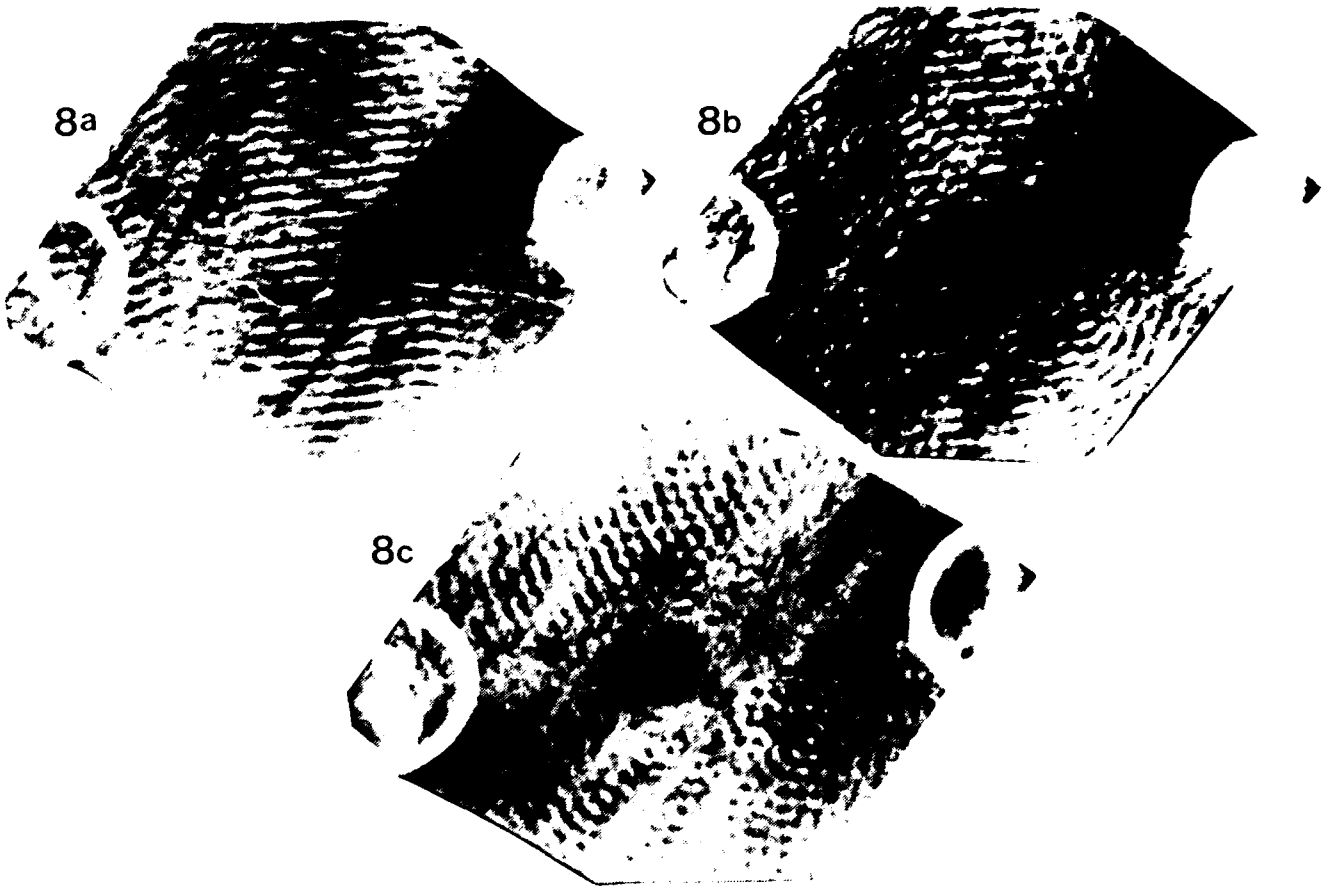


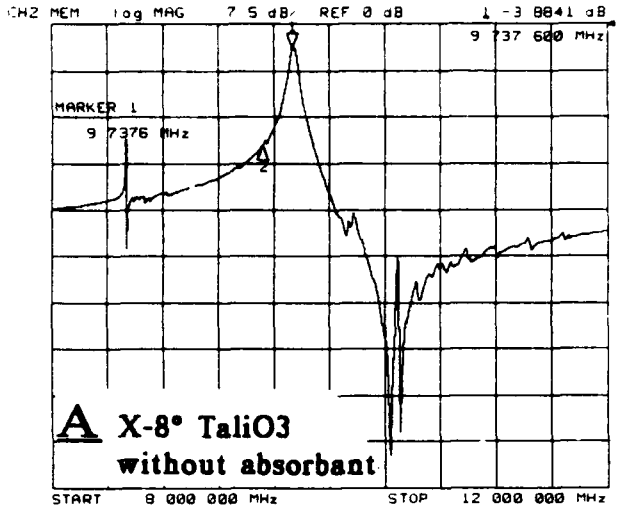
Figure 8: Stroboscopic topographs.

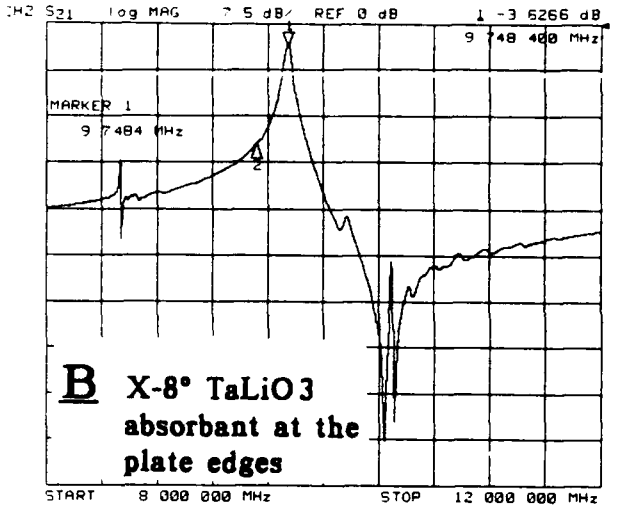
One important thing to notice in the very large amount of elastic energy relative to the progressive (but with a large standing wave ratio) components. From the plate theories established by Professors Mindlin, Tiersten and Lee, it is believed that these progressive components can become resonant (standing waves) at some frequencies but probably in a system including more that the plate (plate + mounting clips ?+ holder ? + ...?). We believe also that, when they are resonant, they are the cause of the activity dips and of the parasitic responses observed in many cases near the resonance of the B mode. To get further insight on this question, an absorbing compound was put on the periphery of an X cut resonator to absorb at least in part the plate waves. On figures 9a and 9b are represented the electrical responses of the resonator without absorbant and with absorbant. On figure 9c, the difference between these two responses is displayed. The obtained curve, which is a representation of the absorbed energy as a function of frequency (to be corrected proportionally to S_{21} of figure 9a) has a discrete structure. In this Figure each peak of energy absorption corresponds much probably to a resonant coupled plate mode.

III.2 - Miniature bar shaped thickness shear resonators

In the present state of the theory it seems very difficult to find a mean to avoid any coupling to plate modes for the X cut resonators of conventional design; principally because the very large coupling coefficient leads to have a large amplitude of vibration over a very large frequency range.

If one of the lateral dimensions of the plate is greatly reduced, the spacing of many of the plate modes can be enlarged so that the probability of coupling can be reduced. An experimental investigation of such bar resonators was made. Our experiments have confirmed (12) that it is possible by a choice of the lateral dimensions of the bars to achieve greater Q factors than with conventional X cut resonators. A gain of a factor about four was achieved. The shape of such resonators is indicated in figure 10. We have chosen to put the length along the x_1' axis previously defined to have rectangular electrodes with a lateral anisotropy similar to that observed experimentally. This choice was also dictated by the observation in plates of coupled modes looking like flexure modes with their nodal lines in the x_1' direction.





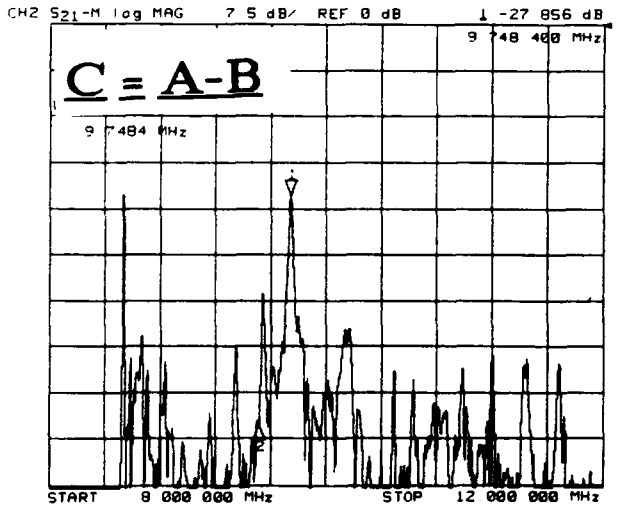


Figure 9: Effect of an acoustic absorbant.

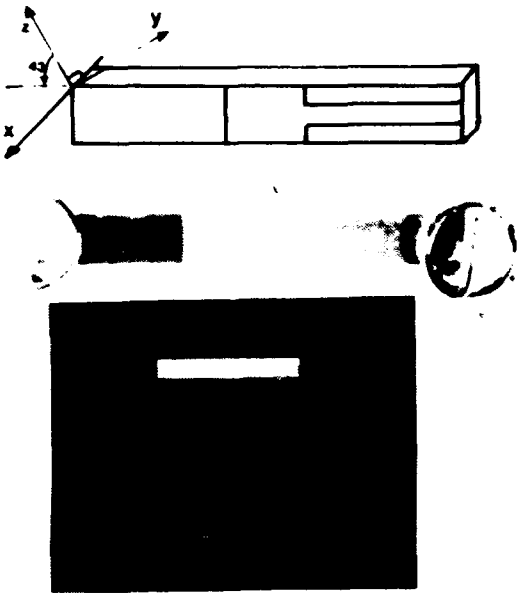


Figure 10: Bar shaped resonator.

During this investigation it was observed that mode coupling can occur near the resonance of the main mode for improper width (Figures 11a and 11b), that, above a certain value, the length is not an important parameter, and also that the anharmonic spectrum of the resonator remains mainly dependant of the electrode dimensions (Figure 11c). The maximum observed Q factors were in the range of 6000. The experimental observations have suggested that much better values of the Q factor could be obtained using some modification of the geometry. A theoretical analysis of the vibration modes of such devices is currently being made using the latest developments of the theories of piezoelectric plates. On the whole, this type of resonators have the interest to give enhanced Q factors and also to use much less material by resonator function than conventional resonators.

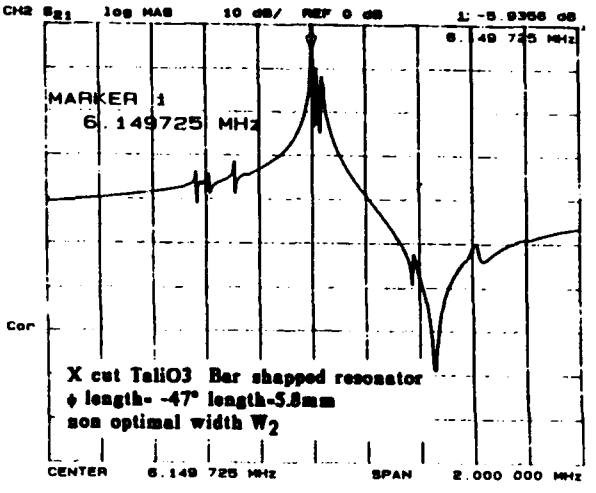


Figure 11a

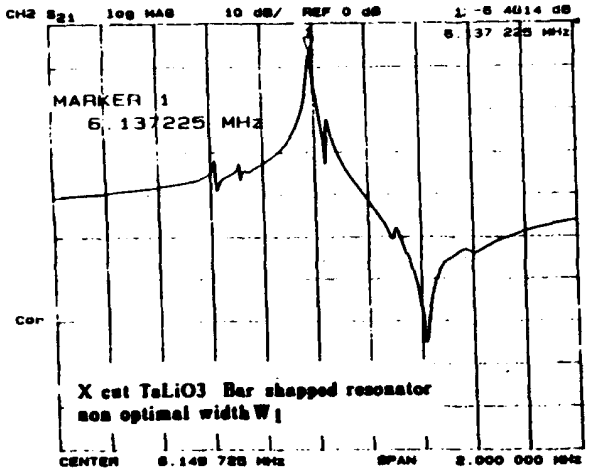


Figure 11b

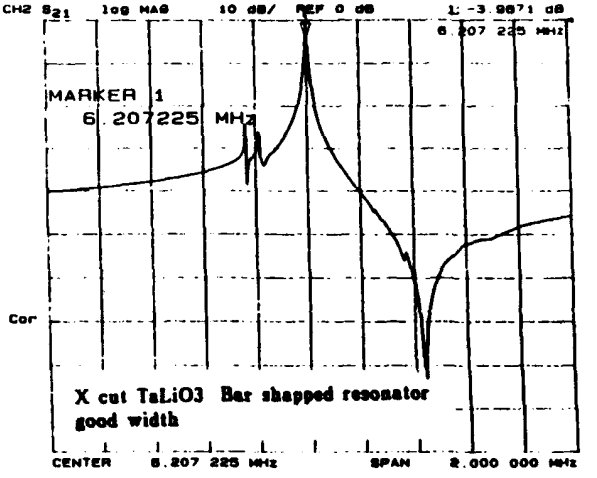


Figure 11c

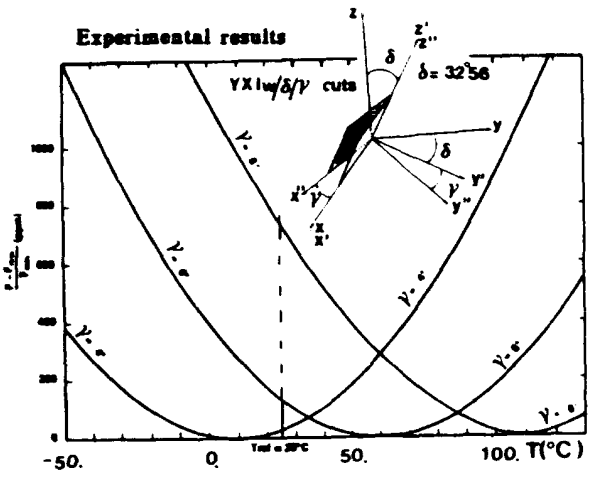


Figure 12: Thermal behaviour of doubly rotated cuts (Experimental).

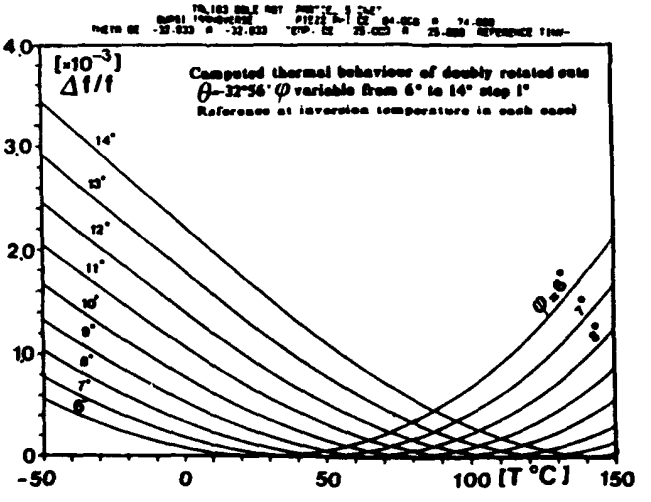


Figure 13: Thermal behaviour of doubly rotated cuts (computed).

IV - DEVICE USING OVERTONE MODES

The overtone modes of cuts with a high coupling coefficient display a much higher Q factor than the fundamental mode. They have an "effective coupling coefficient" that can reach 15 %. We have previously shown that for conventionally designed resonators, a thermal compensation can be obtained using doubly rotated cuts (4). Other designs can be found to obtain overtone resonators with a zero F.T.C., several will be discussed in the following.

IV.1 - Doubly rotated cuts with zero T.C. for the overtone

The properties of cuts near the second locus of zero T.C. of the B mode were examined. It was noticed that quite large coupling coefficients can be obtained. But since the three thickness modes are excited in these cuts they are not very favourable for filter applications, unless a method to eliminate certain of these modes is applied. In figures 12 and 13 are represented experimental and computed thermal behaviours for several cuts belonging or near this locus.

IV.2 - Overtone resonators using a T.C. compensating layer

This type of compensation is mostly used for surface wave devices (13) for which it is difficult to find materials having simultaneously a zero T.C. and a large coupling coefficient. In the case of high frequency bulk wave resonators this method is also interesting since the thickness of the compensating layer is compatible with the existing techniques of film deposition. The most appropriate material for a compensating layer is the

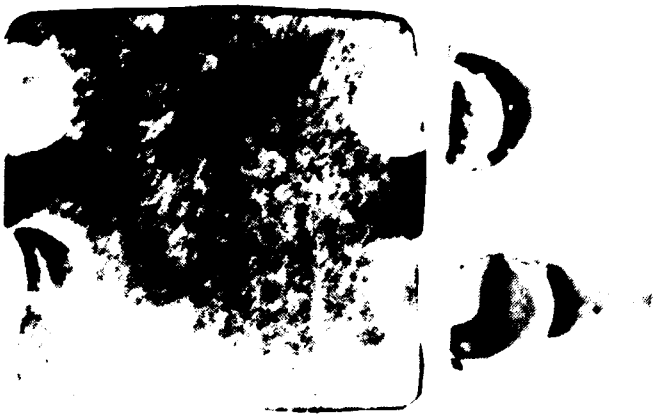


Figure 14:One resonator excited.

Figure 15:Integrated filter.

Figure 16:Filter excited.

amorphous silica (and also probably the amorphous aluminium phosphate). The largely positive T.C. of silica (70 to 80 . $10^{-6}°C^{-1}$ for the shear waves) is sufficient to compensate the moderately negative T.C. of the third or of the filth overtone of cuts of lithium tantalate with high coupling coefficients. Again in this case the most appropriate cuts are those around the X cut which have a first order T.C. around -40 $10^{-6}°C^{-1}$ for the third overtone.

Properties of the third overtone of the X cut.

The Q factor of the third overtone mode of X cut resonators is much higher than that of the fundamental mode (in the range 10^4 to 510^4) and the response presents much less coupling to plate modes. It was found using stroboscopic X ray topography that the lateral anisotropy of this mode is much greater than for the fundamental mode (Figure 14). The very small lateral dimensions of the electrodes required to trap no anharmonics make possible to integrate several resonators on the same plate (Figure 15). In this case to avoid any coupling betwen the resonators by the plate mode (Figure 16) it is necessary to use some form of acoustic isolation between the resonators.

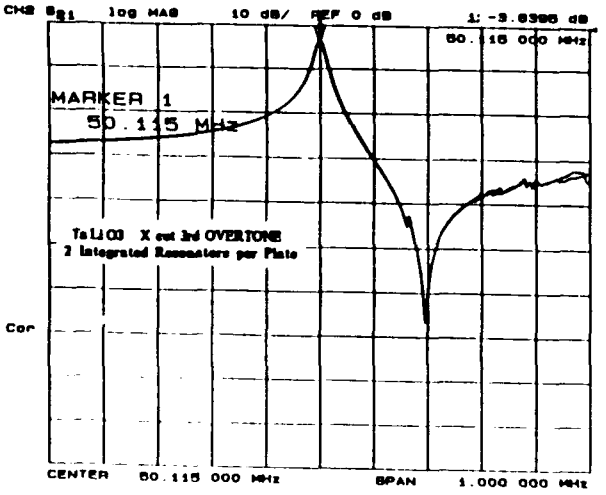


Figure 17 (a) : Response of 2 integrated resonators

An example of electrical response of resonators integrated on one plate is given in Figure 17a the response of a 2 Poles Jaumann filter using similar integrated resonators is given on Figure 17b. The 3 dB bandwith of this filter is 150kHz.

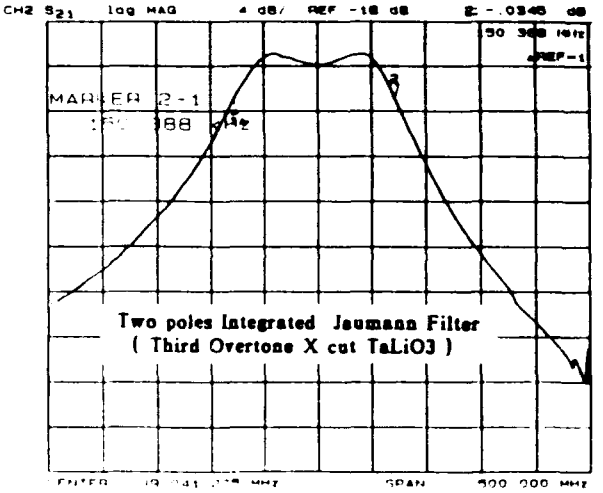


Figure 17 (b): Response of an integrated filter

IV.3 - Modelisation of composite compensated resonators

As it was observed (Figure 14 and 16) that, when no coupling to plate modes are present, the vibration mode of overtone resonators is extremely confined, the analysis was conducted assuming a one dimensional propagation in the different layers (compensating material, electrodes, lithium tantalate) constituting the composite resonator (Figure 18). This assumption is sufficient for the purposes of finding a compensated design and also to have approximate values of the equivalent scheme.

The model can take into account all the layers constituting the resonator and the effect of gas loading (by viscosity for the shear modes) on the external surfaces. It uses exacts solutions of one dimensional piezoelectricity for the piezoelectric

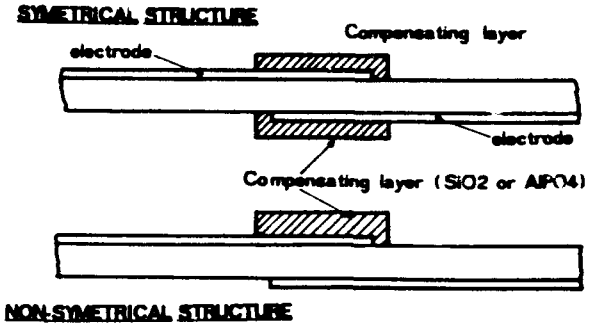


Figure 18: Compensation by a Silica layer.

materials and of elasticity for the electrode layers and the compensating material. The acoustic and dielectric losses in any material are taken into account using imaginary parts for the constants (C or ε) of the material. The model uses a chain matrix formalism (Figure 19) based upon the electromechanical chain matrix (4x4) proposed by Sittig (14). The obtained result is the

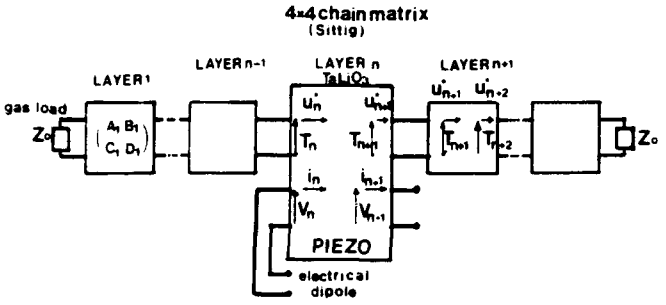


Figure 19: 1 D. Modelization of composites.

electrical impedance of the resonator, from the impedance computed as a function of frequency, the zero phase resonance frequencies are searched for by the program. Since the temperature coefficients of all the quantities are considered, it is possible to compute the temperature coefficients of the resonance frequencies (This is done calculating the resonance frequencies for 5 to 10 temperatures and then by extraction of the temperature coefficients by a polynomial least square fitting). The corresponding computer program was used to analyse the most interesting structures for our application. It was found that it is possible to consider either the symmetrical or the non symmetrical structures. The non symmetrical structures have the property to have a strong resonance (if the non symmetry is sufficient) on the second overtone (and more generally on even overtones). In this case, it is appeared that the T.C. of the lithium tantalate layer is compensated by a small layer of silica. The symmetrical structures operating on the third overtone have the interest to permit the obtention of very large bandwith (large resonance-antiresonance spacing) two examples of calculated response for composite resonator are given on figures 20 and 21.

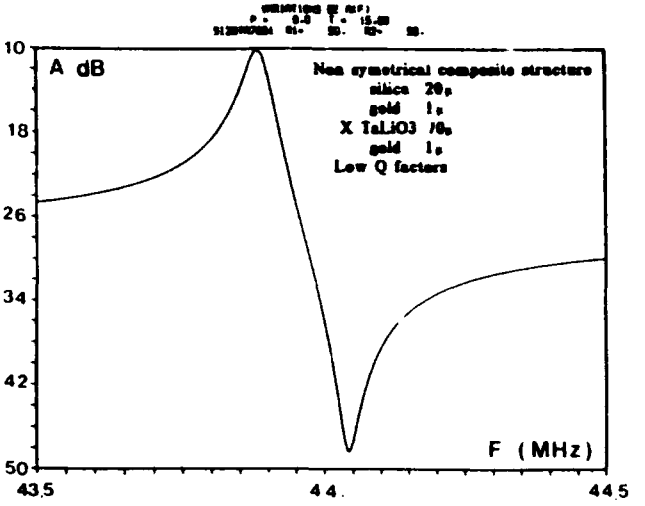


Figure 20 : Example of a computed response

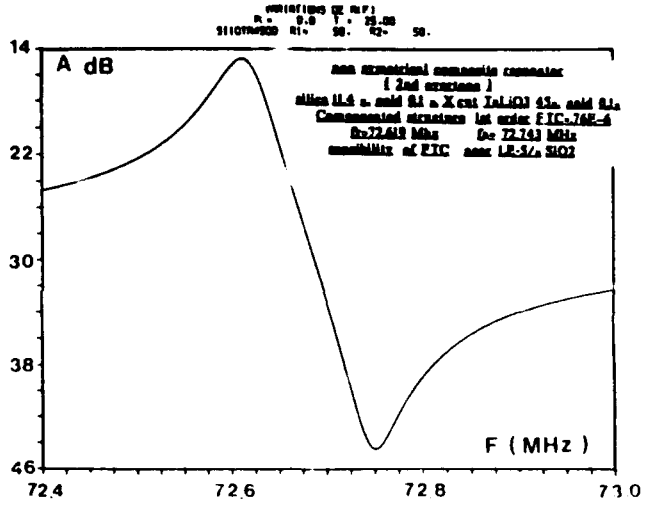


Figure 21 : Other example of a computed response

IV.4 - Overtone resonators using the lateral excitation

A. Ballato (15) (16) (17) has demonstrated the interest of using the lateral excitation to obtain severall interesting properties with the thickness mode resonators (excitation of modes which are non-piezoelectric using the conventional design, selection of only one thickness mode in a cut etc...). Due to the high values of the dielectric constant and of the coupling coefficients (the lateral coupling coefficients are also very high), the use of lateral field excitation is more interesting in lithium tantalate than in any other known compensated material. The principle of lateral field excitation is recalled in figure 22. Different types of resonators were made using different design to obtain an energy trapping. Very interesting results have been obtain using plano-convex resonators. In table II are given the

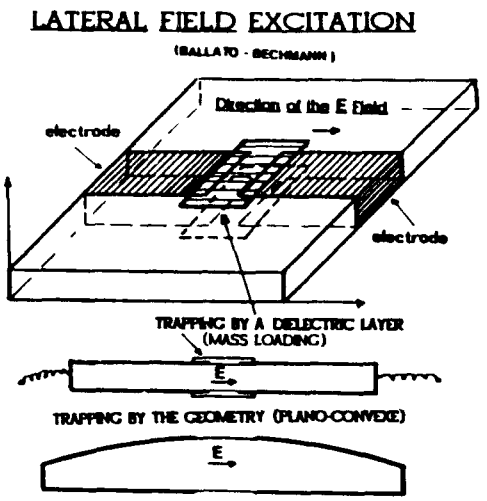


Figure 22: Lateral Field excitation.

measured characteristics of a resonator using the B mode of the Y-32°56° cut. This cut is very close to the second locus of zero T.C. for the B mode. As can be observed in figure 3 this mode is not excited with an electric field normal to the plate.

TABLE II
-32°56' PLANO-CONVEX RESONATOR (L.E.)
2h=1,5mm R = 150 mm

N	Frequency (Hz)	Q factor	Q at 5 MHz	R (Ohm)	L (Henry)
1	1.321.625	9130	-	144.	-
3	3.614.263		(untrapped ?)	-	-
5	6.375.086	149.000	190.000	38.35	0.142
7	8.936.515	191.000	342.000	27.01	0.092
9	11.447.612	107.000	245.000	87.32	0.129
11	14.012.703	471.000	1.310.000	16.5	0.088
13	16.562.111	576.000	1.908.000	18.4	0.101
15	19.089.865	262.000	1.003.000	38.35	0.084
17	21.639.792	489.000	2.120.000	30.3	0.108
19	24.190.249	343.000	1.663.000	52.40	.118
21	26.719.321	347.000	1.856.000	73.98	0.133
23	29.266.075	457.000	2.675.000	51.70	.128
AT QUARTZ 5TH OVERTONE (T.E.)					
5	5.000.000	2.400.000	2.400.000	70	5.6

In table II we observe that overtones with very high ranks have exceptionally large Q.f products and have also very favourable characteristics for use in oscillator circuits. Their impedance level, even for the 26th overtone are much more favourable than the 5th overtone of an AT quartz plano-convex resonator. These measurements are a confirmation of the observations made in the GHz frequency range which have indicated extremely low acoustic propagation losses for this material and an intrinsic Q.f product higher than 30 10^{12} (3).

Several response curves obtained for the overtones of one Y-32°56 resonator are shown on figure 23. It was also observed that these resonators were able to withstand extremely high excitation levels before that the electrical response shows a slight form of non linear behaviour. All these properties make them very attractive to obtain, by direct generation in the range from

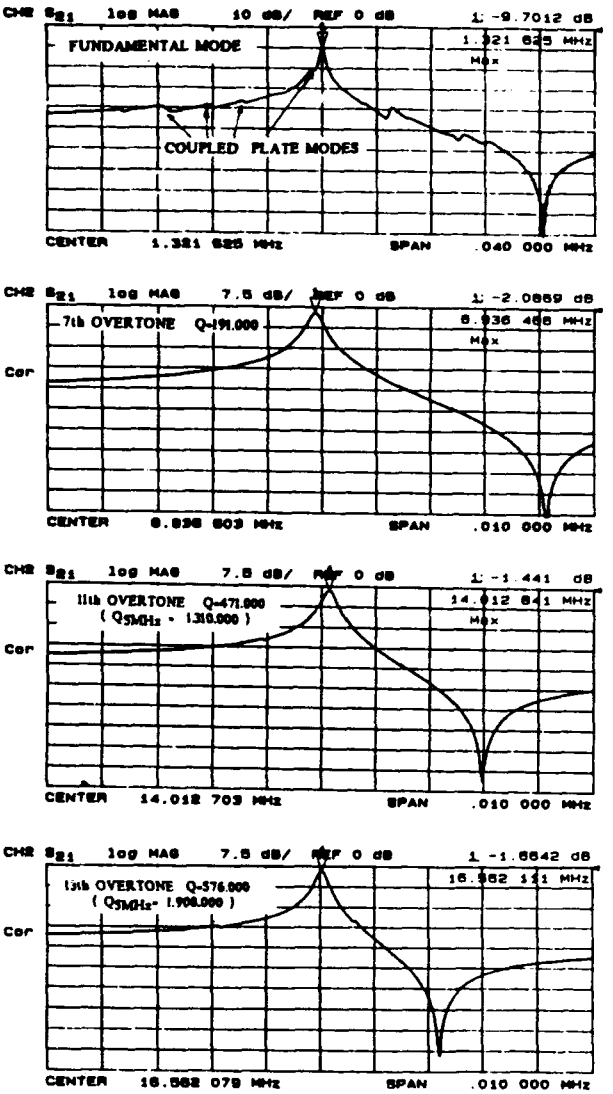


Figure 23 (a): Responses of a plano-convex resonator [LE].

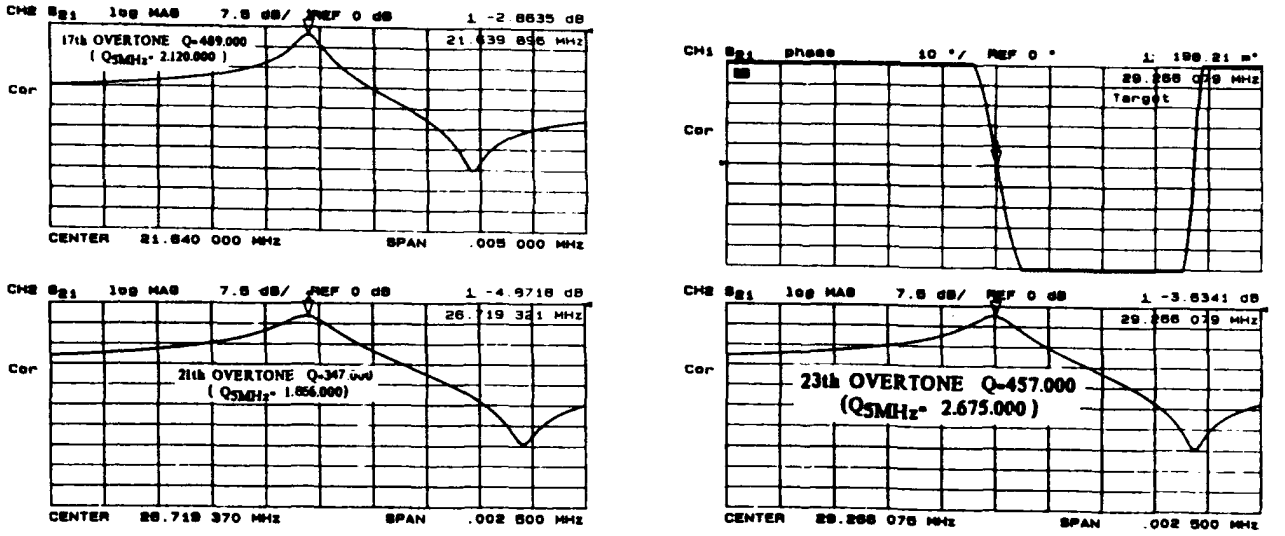


Figure 23 (b): Responses of a plano-convex resonator [LE].

a few MHz to over 150 MHz, oscillators with outstanding phase noise properties (18). The response curves for different excitation levels of the overtone of a Y-32°56' resonator are displayed in figure 24. Other preliminary experiments with L.E. plane resonators, using the z cut, have indicated the possibility to have a sufficient "effective coupling coefficient" to obtain the required bandwidth for the I.F. filter.

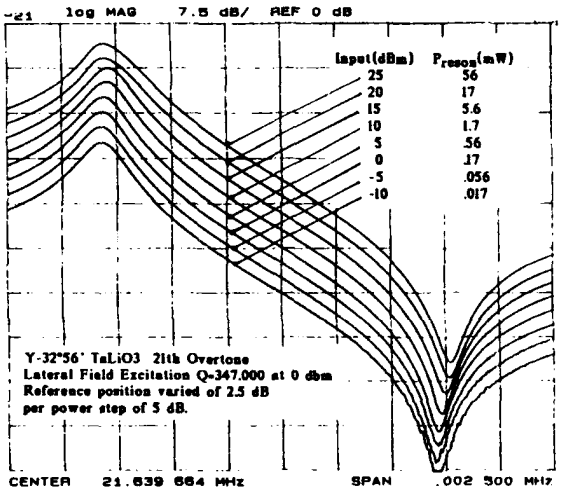


Figure 24: Influence of the excitation level.

V - CONCLUSION

Several conventional and new types of resonators using lithium tantalate where investigated in view to define the most interesting solutions, using this material, to make I.F. Filters having the characteristics required by the new numerical pan European radiotelephone system (G.S.M.). Beside several conclusions relative to this specific question, these investigations have revealed many new features of the considered devices.

Among them, the most important points are :

- The Q factor limitation by couplings to dissipating plate modes of the devices using a conventional design and the fundamental mode
- The possibility to obtain high frequency overtone resonators and integrated filters with a zero T.C. using composite structures.
- The major interest of lateral field excitation in devices using this material and the outstanding characteristics obtained with overtones of very high ranks.

This study has confirmed that due to its energy trapping characteristics, lithium tantalate is particularly adapted to obtain extremely miniaturized integrated devices and also that several new interesting solutions to realize the I.F. Filters for the G.S.M. radiotelephone system can be implemented using this material.

ACKNOWLEDGMENTS

The authors are indebted to A. Ballato for the revelation of the great interest of the lateral field excitation in lithium tantalate. They acknowledge the skilful assistance of J. Douis for cutting and polishing the many samples used in this study and thanks particularly M.P. Louis for the preparation of this manuscript.

REFERENCES

(1) A. Ballato Doubly rotated thickness mode plate vibrators in Physical Acoustics Vol XII Academic Press New York 1977

(2) J. Détaint, R. Lançon Proc. 30th. Annual Freq. Control Symp. 1976

(3) R.A.Moore, J.T.Haynes, B.R.McAvoy Proc. 1981 Ultrasonics Symposium p 414 (1981)

(4) J. Détaint Electron Letters vol 13 n° 1 p 20 1977

(5) J.W. Burgess M.C. Hales Proc. IEE Vol 123 n° 6 June 1976

(6) M. Onoe Proc. IEEE Vol 57 p 702 (1969)

(7) A. Zarka, B. Capelle, J. Détaint, J. Schwartzel Proc. 41st Annual Frequency Control Symp. 1987 p. 236-240

(8) B. Capelle, J. Détaint, Y. Zheng, J. Schwartzel, A. Zarka 4 th European Frequency and Time Forum (these proceedings)

(9) D.S. Stevens, H.F. Tiersten J. Acoust. Soc. Am Vol 79 n° 6 p 1811 (1986)

(10) K. Nakamura, M. Shimizu Proc. IEEE Ultrasonics Symposium p 606 (1976)

(11) J. Détaint, J. Schwartzel, C. Joly, E. Philippot Proc. 41th. Annual Frequency Control Symp. p 314 1987

(12) Y. Fujiwara and N. Wakatsuki Strip. IEEE Trans U.F.C.C vol UFCC34 n° 1 p 39-44 (1987)

(13) G. Cambon, J. Lassale, J.M. Saurel, J. Attal Proc. 1981 Ultrasonics Symposium p 364

(14) E.K. Sittig Physical Acoustics vol IX Academic Press 1972

(15) A. Ballato, J. Kosinski, M. Mizan, T. Lukaszek Proc. 43rd Annual Frequency Control Symp. p 557 (1989)

(16) E.R. Hatch, A. Ballato Lateral Field excitation of quartz plate Proc. 1983 IEEE Ultrasonics Symposium p 512 (1983)

(17) A. Ballato, E. Hatch, M. Mizan, B.H.T. Chai, R.S. Tilton, T.J. Lukaszek Proc. 38th Annual Freq. Control Symposium (1984)

(18) MM. Driscoll, C.R. Vale, R.W. Weinert Proc. 1987 IEEE Ultrasonics Symposium p 347 (1987)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

LINEAR PHASE FILTERING

B. D'ALBARET - L. BIDART - G. SILLIOC

Q K PIEZOELECTRONIQUE
30 Bld Galliéni, F 92392 VILLENEUVE LA GARENNE CEDEX

SUMMARY

Digital Communication processing is placing the importance of the linearity information in a prominent position.

Today reliability as well as the rapidity of transmissions widely depend on the phase linearity or on the constant group delay in filtering systems.

Concerning classical filters, it appears that linear phase is incompatible with selectivity. Most of the time, it is necessary to find the best compromise between these requirements.

Whatever might be the system, the purity of the output signal and selectivity remain two very important parameters. It is important to keep them in all environmental conditions and especially during vibrations.

INTRODUCTION

This presentation draws up the inventory of the different types of linear phase Xtal filters and the ways in which they can improve, control, correct the phase linearity so that they obtain a group delay adapted to needs of up to date electronic.

Theoritical possibilities and practical limitations afforded by different polynomial or elliptical filters will be discussed as well as the correction which can be made by equalizers.

Finally phase noise measurements of different types of filters under vibrations will be presented.

COMPARISON BETWEEN DIFFERENT TYPES OF FILTERS

With classical approximations, it appears that linear phase is incompatible with selectivity. In order to compare the different approximations, we have chosen a 6 poles Xtal filter with 5 types of transfer function :

1. Butterworth
2. Tchebycheff with 0.1 dB of ripple
3. Bessel
4. Linear phase with equiripple 0.5°
5. Semi linear

Fig. 1 shows the 5 types of attenuation

Fig. 2 shows the 5 types of BW and G.Delay

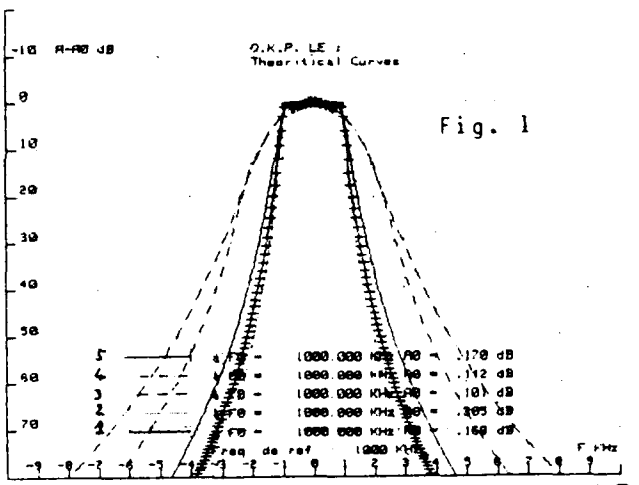


Fig. 1

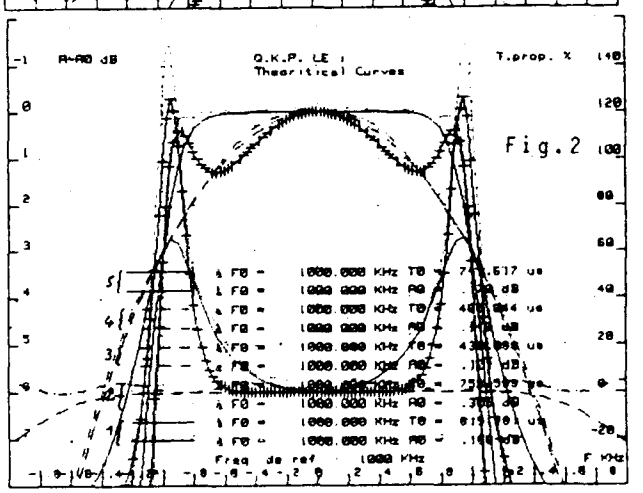


Fig.2

CH2818-3/90/0000-349 $1.00 © 1990 IEEE

The following Table 1. compares the different attenuations and group delay variations

Table 1.

FUNCTION TYPE	K 60/3	dτ/τ 60 %	dτ/τ 100 %
Butterworth	3.2	20	66
Tchebycheff 0.1 dB	2.3	25	150
Bessel	5.4	0	0
Linear equiripple 0.5°	4.4	4	4
Semi linear	2.6	2	125

K is the shape factor, i.e. the ratio : 60 dB BW/3 dB BW.

$d\tau/\tau$ is the relative variation of the group delay for 60 % or 100 % of the 3 dB bandwidth.

If we observe the 4 first lines, we can see that the shape factor is bad when the group delay variation is good, so it was necessary to find a solution between both.

COMPROMISE

In order to preserve a constant group delay in part of the bandwidth, while preserving relatively high selectivity, special transfer function have been developed : these filters called semi-linear filters present a constant group delay in part of the 3 dB bandwidth, this is called the usefull bandwidth. It is like a Bessel filter for this part of band, but with an out of band attenuation like a Butterworth filter.

As we can see on the table, the "semi-linear" filter is sharper that the other 2 linear filters. Its group delay is completely constant on the usefull bandwidth (60 % of the total bandwidth).

Fig. 3 shows the real curves of a 6 poles 90 MHz Bessel filter using 6 quartz resonators. The group delay is constant over all the 3 dB BW $\pm$ 14 KHz, we observe a variation of 3μs for an absolute group delay of 32 μs, i.e. 10 %.

Fig. 4 presents the practical realiza - tion.

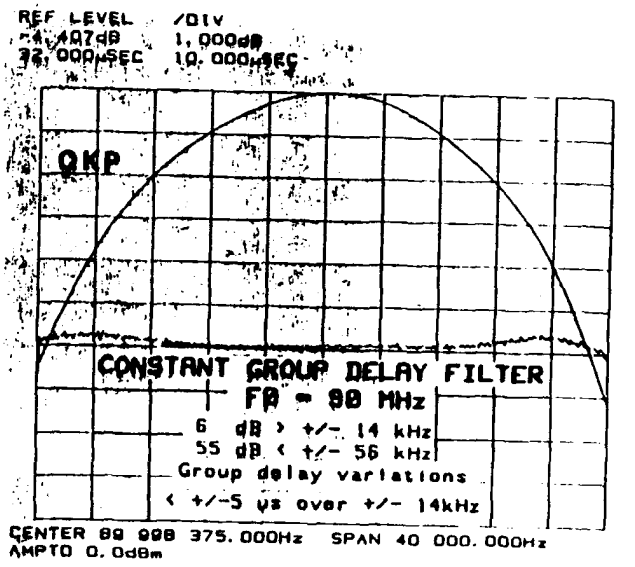


FIG 3

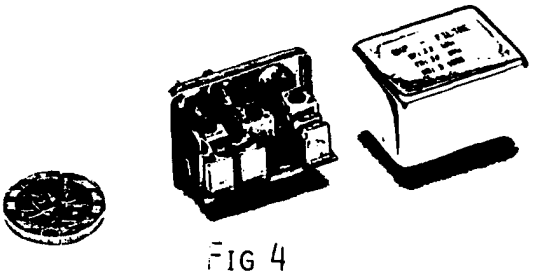


FIG 4

Fig. 5 shows a " semi-linear " 6 poles filter at 21.380 MHz (6 dB BW $\geq \pm$ 13 KHz)using 3 monolithic resonators. The group delay is constant over 60 % of the bandwidth : $d\tau/\tau \leq 3\mu s$ over $\pm$ 8.5 KHz, i.e. a variation $\leq$ 5 %. Fig. 5 and 6 show a shape factor 60/6 dB of 2.3.

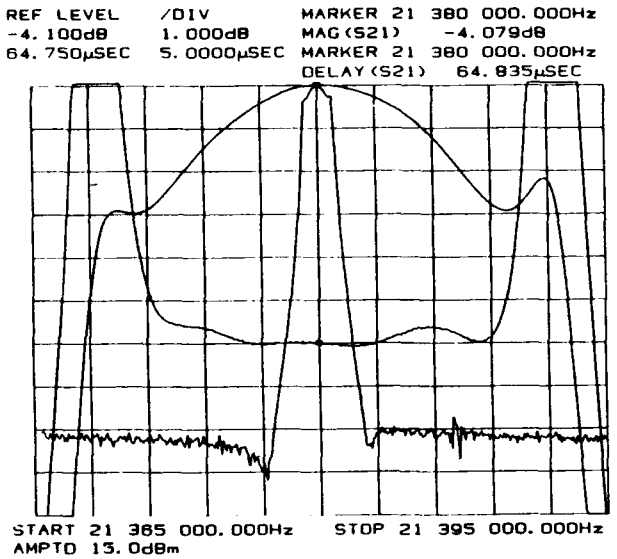


FIG 5

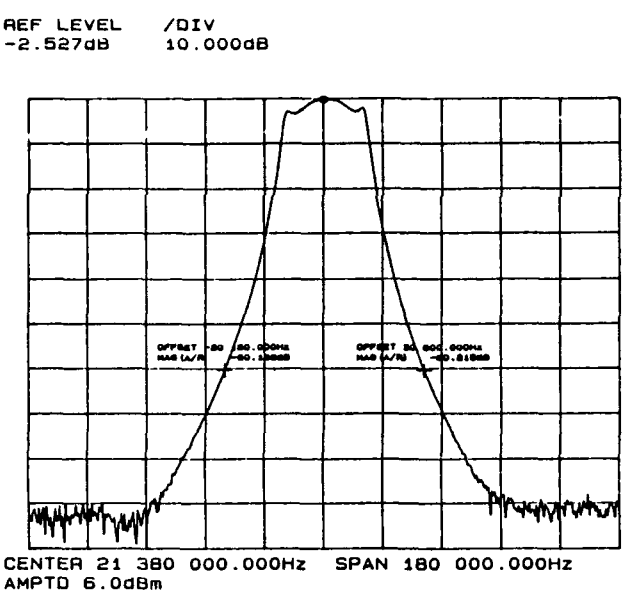


FIG 6

FEASIBILITY

For a Xtal filter, the maximum realizable relative bandwidth is approximately for quartz : 0.3 %. But this value is only available for non linear filters.

For a linear filter, it is necessary to take into account twice the real bandwidth (in fact the resonators frequencies are twice as far apart as usual).

Therefore, filters which were realizable in quartz with a Butterworth or Tchebycheff function are no longer feasable with linear phase function if the relative bandwidth is greater than 0.15 %.

In this case, the Lithium Tantalate can be used. It allows us to go up to 5 % of relative bandwidth (2.5 % for linear filters).

EQUALIZERS

Sometimes, the attenuation conditions are so hard that it is impossible to use a traditional function.

A typical application is the Single side Band filter where 2 schemes are employed :

- a scheme in ladder with resonators in series and in parallel (see Fig.7)

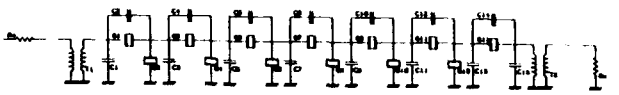


FIG 7

- a scheme in bridge with several resonators per arm (see Fig. 8).

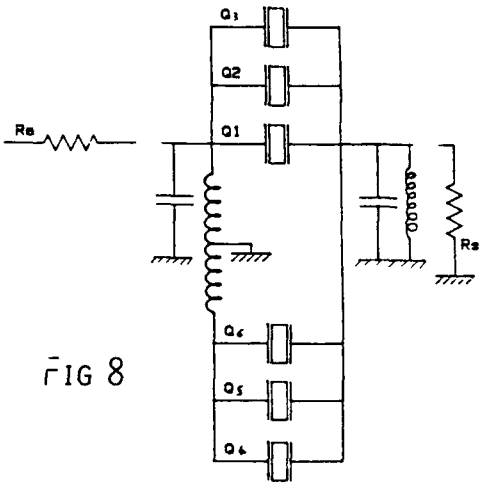


FIG 8

All these filters present several poles of infinite attenuation and are very sharp, i.e. a shape factor K 60/3 dB less than 1.5. Of course, for such filters, the group delay is basically distorded.

For certain application, it is in fact necessary to obtain a relatively constant group delay.

In this case, we use a second filter called equalizer which follows the first. It presents a very flat attenuation in the bandwidth in order to preserve the attenuation curve of the first filter.

Furthermore, it has a group delay opposite to the first in order to obtain by the sum the flattest possible variation of group delay.

The typical curves of attenuation and group delay of a network "pass all", i.e. a network which changes the phase but not the attenuation, with the use of Xtal resonators is given Fig. 9.

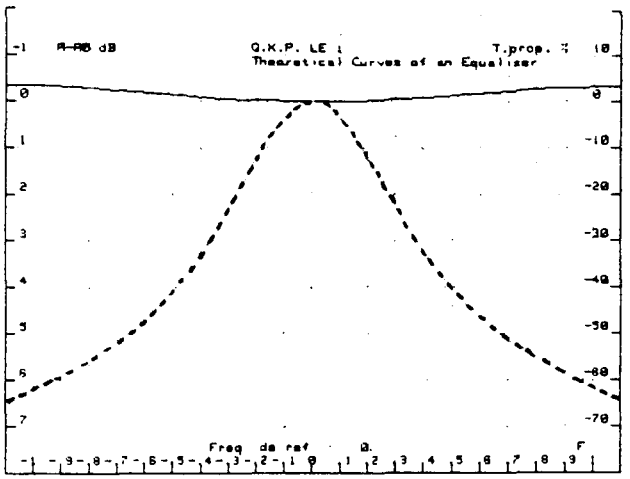


FIG 9

To obtain such network, it is necessary to make the synthesis of a non minimum phase function, so called, if it has zeros in the right half part of the complex plan.

A synthesis program gives us all the values of the elements for a bridge network (the Jaumann scheme) with 2 Xtals in parallel in one arm (see Fig. 10).

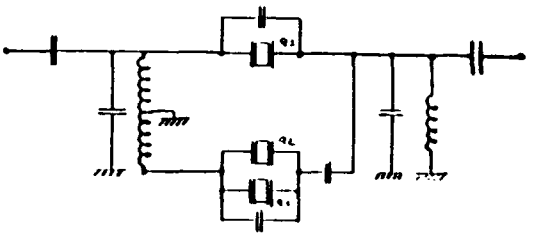


FIG 10

If the correction of the group delay is not enough sufficient, of course 2 cells in series can be used.

Ex 1 : Picture 11 shows a 10 MHz filter S.S.B. made with a ladder scheme using 13 quartz resonators.

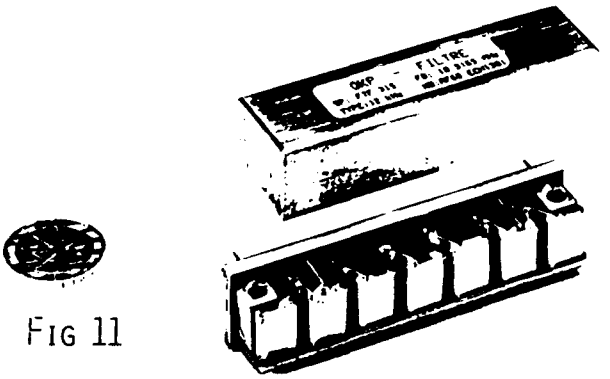
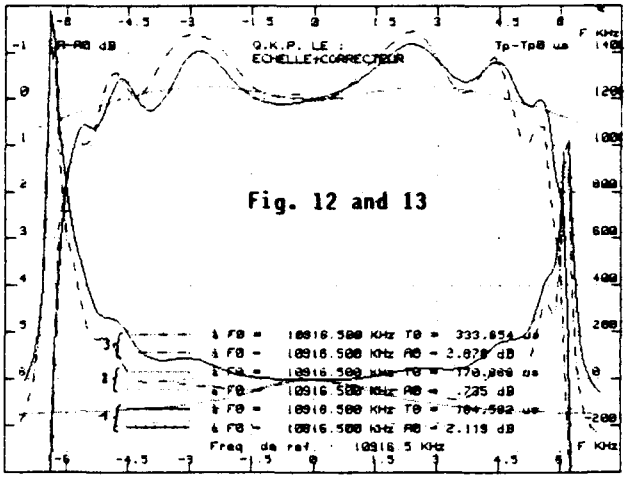


FIG 11

The specification was 3 dB BW $\geq \pm$ 6 KHz
Group delay variation in the BW $\pm$ 5 KHz less than 200 μs.

We can observe on the curve 1 of Fig. 12 and 13 the theoritical simulation. The variation of the group delay makes 280 s, therefore it is necessary to correct a minimum of 100 μs with an equalizer.

One cell has been calculated and the theoritical curve can be seen on curve 2 of Fig. 13. Curve 3 shows the results of the two filters in series after simulation on computer.



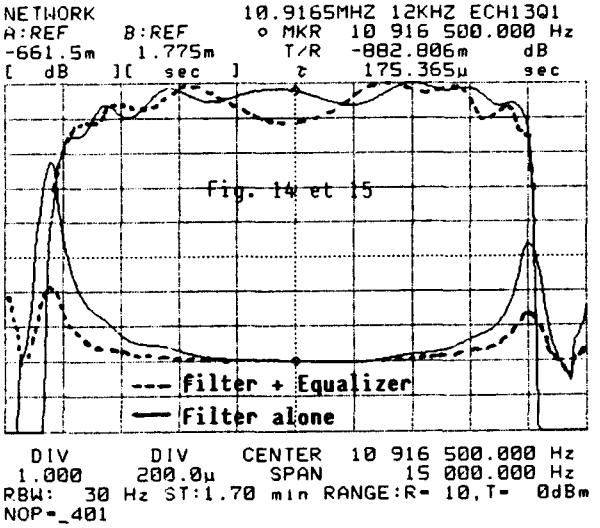


Fig. 14 and 15 show the practical results without and with equalizer that the specification achieves.

Ex. 2 : Fig. 16 presents the scheme of a 10 MHz filter calculated with a Cauer synthesis Jaumann type using 6 LiTa03 resonators.

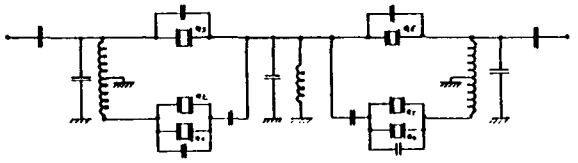


FIG 16

In fact the 3 dB bandwidth of ± 50 KHz was too wide for quartz.

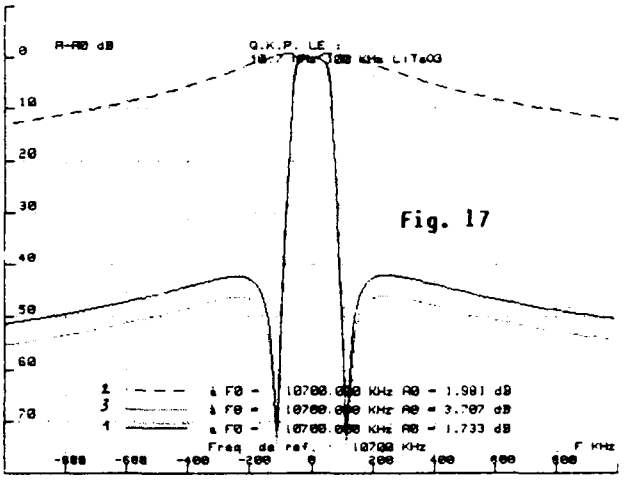
The specification for the group delay was :

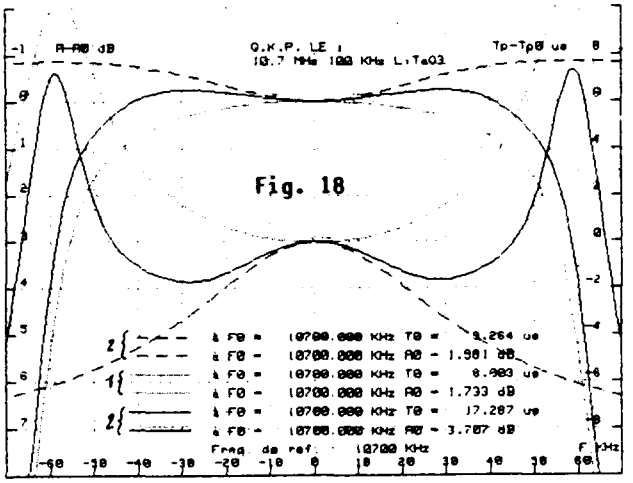
Maximum variation over ± 45 KHz : 2 µs

Curve 1 of Fig. 17 and 18 shows the theoritical simulation of the filter which makes 5 µs of variation at extremities of the bandwidth.

An equalizer has been calculated with a correction of 5 µs (curve 2 of Fig. 18).

Curve 3 shows the result of the two filters in series, the group delay variation is less than 1.5 µs.





INTRODUCTION TO "PHASE NOISE SENSITIVITY" IN CRYSTALS FILTERS UNDER VIBRATIONS

Apart from problems regarding the mechanical filter structure, the main factor which disturbs the filter output signal is the " g sensitivity" of the resonator itself.

In fact, resonators are more or less sensitive to mechanical vibrations, and this sensitivity is characterized by a Fm Modulation of the filter output signal.

Currently, we know that this phenomenon is mainly in relation with the following parameters :

- raw material type (SiO_2 , $LiTaO_3$..)
- cut (AT, BT, SC...)
- adhesion of metallisation to the crystal wafer
- number and geometrical position of the mounting points
- length and rigidity of these points.

The following part of this paper aims to describe method, tests and results currently achieved in QKP.

MEASUREMENTS OF "ACCELEROMETRIC SENSITIVITY" ON RESONATORS

a) SYSTEM DESIGN (See Fig. 19)

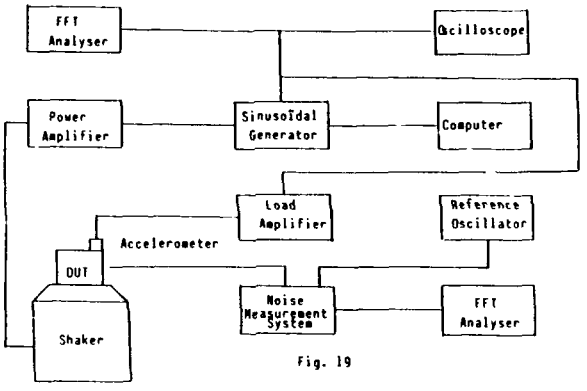


Fig. 19

b) THEORY

Associated to an oscillator, the resonator is submitted to a given frequency and acceleration in sinusoïdal mode.

The "g sensitivity" of the resonator , when submitted to vibrational input, generates sidebands on the output spectrum of the oscillator.

Knowing the level of these sidebands we can calculate the modulation index (m) and the phase variation (Fc). By definition the "g sensitivity" is defined by :

$$\mathcal{L}(f)_{dBc} = 20 \log \left[\frac{\Delta F}{Fo} \cdot g \cdot \frac{Fo}{2 fm} \right]$$

and

$$S = \frac{2}{Fo} \cdot \frac{fm}{g} \cdot 10^{\frac{\mathcal{L}(f)}{20}}$$

where :

S = $\Delta F/Fo$ " g sensitivity" for the resonator
g = mechanical acceleration level
Fo= oscillator frequency output
fm= vibrational frequency

This formula is used for a modulation index (m) $\leq 0,01$.
For a modulation index (m) $\geq 0,01$ it is necessary to use " Bessel functions ".
To calculate " g sensitivity " in random mode the following formula is used :

$$\mathcal{L}(f)\ dB/Hz = 20 \log \left[\frac{\sqrt{2}}{2} \cdot g_{RMS} \cdot S \cdot \frac{Fo}{fm} \right]$$

and

$$S = 10^{\frac{\mathcal{L}(f)}{20}} \cdot \left[\frac{2}{\sqrt{2}} \cdot \frac{fm}{g_{RMS} \cdot Fo} \right]$$

The acceleration is defined as g^2/Hz, but it has to be redefined as g RMS by Hz.

c) RESULTS

Table 2 summarizes some experimental figures realized by QKP :

TABLE 2
Resonator Accelerometric Sensitivity ($\Delta F/F$ by g)

			5 MHz	10 MHz	20 MHz	50 MHz	100 MHz
QUARTZ	AT_{CUT}	P 1	3.10^{-9}	3.10^{-9}	1.10^{-9}	1.10^{-8}	1.10^{-8}
		P 3	-	-	5.10^{-10}	5.10^{-10}	3.10^{-10}
		P 5	1.10^{-9}	1.10^{-9}	-	5.10^{-10}	3.10^{-10}
	BT_{CUT}	P 3	-	1.10^{-10}	1.10^{-10}	1.10^{-10}	1.10^{-10}
		P 5	-	-	-	3.10^{-10}	1.10^{-10}
	SC_{CUT}	P 3	7.10^{-10}	1.10^{-9}	1.10^{-10}	-	1.10^{-10}
		P 5	-	-	-	-	-
$LiTaO_3$	X CUT	P 1	-	-	2.10^{-9}	1.10^{-8}	5.10^{-8}

This table shows that performance degradation occurs as resonators get thinner in quartz as well as in $LiTaO_3$.

FILTER PHASE NOISE MEASUREMENTS UNDER MECHANICAL VIBRATIONS

a) By definition, we suppose that mechanical design (can, printed-board, fixations...) as well as test fixtures (mechanical interface, cables, ...) have been optimized in such a manner that mechanical Q is the nearest possible to 1.

All measurements listed below have been defined for a random acceleration level of 0.02 g2/Hz from 20 Hz to 2000 Hz (see Fig 20) and given for the worst vibration axis.

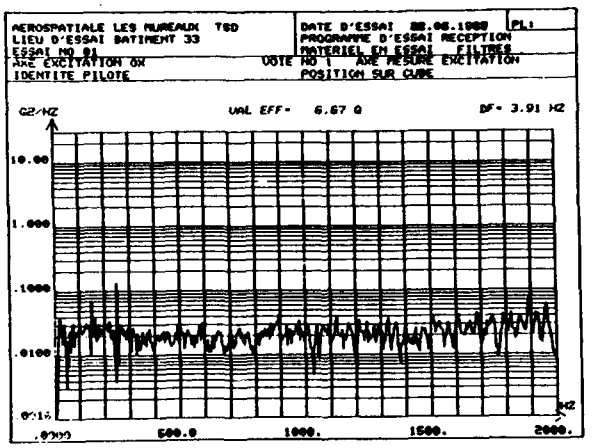


Fig. 20

b) SYSTEM DESIGN (see Fig. 21)

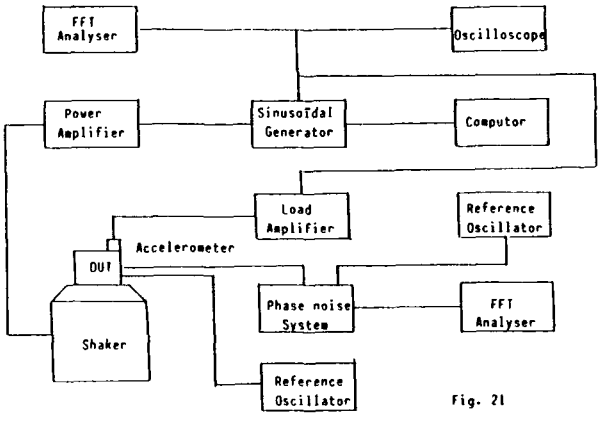


Fig. 21

c) DETAIL OF PHASE NOISE MEASUREMENT SYSTEM

. Description

The output signal of the filter is locked on the signal of a QKP reference oscillator. The signal error V(F) at the mixer output is proportional to the phase variations (Δ Fc) For "Fourrier frequency" outside of the "lockingbandwidth" V(F) is amplified by G and measured with a BF analyser,

where : G = Go + Gl

and Go= 40 dB

Gl= tunable from 1 to 40 dB

If B is the analysis Bandwidth and Vcc is the peak to peak level of the "mixing voltage" between the two signals, the system calibration is defined as follows :

$$\mathcal{L}(f)\,dB/Hz = 20 \log \frac{\sqrt{2} \cdot V(F)}{G\sqrt{B} \cdot Vcc}$$

Fig. 22 gives the skeleton diagram of this measurement system.

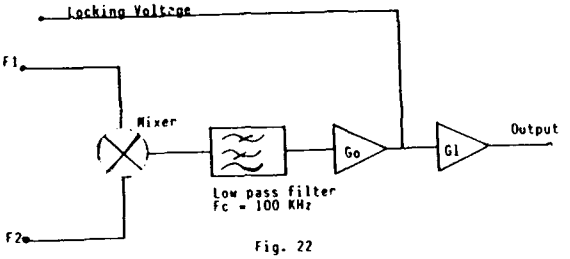


Fig. 22

RESULTS

1. 5 MHZ QUARTZ FILTER(BW3 = 3 KHZ

Number of poles : 6
Raw material : Quartz
Mode : AT cut Fundamental
Technology : Standard Xtal HC 18 U
$\Delta F/F$ by g : 1.10-9(see table 2)

For results see Fig. 23.

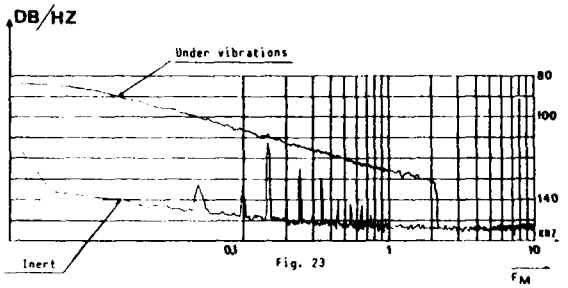


Fig. 23

2. 20 MHZ QUARTZ FILTER (BW6 = 26 KHz

Number of poles : 8
Raw material : Quartz
Mode : AT cut Fundamental
Technology : Monolithic HC 45 U
$\Delta F/F$ by g : 3.10-9 (see table 2)

For results see Fig. 24.

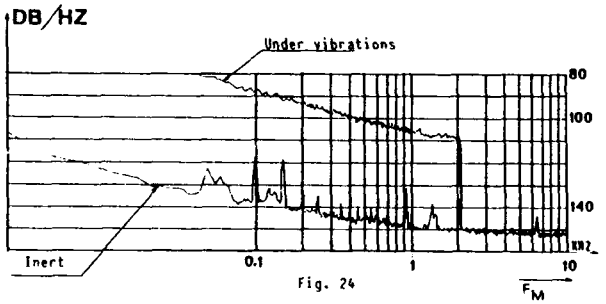


Fig. 24

3. 100 MHZ QUARTZ FILTER (BW3 = 10 KHz)

Number of poles : 4
Raw material : Quartz
Mode : Third overtone SC cut
Technology : Standard TO 5
$\Delta F/F$ by g : 1.10-10 (see table 2)

For results see Fig. 25

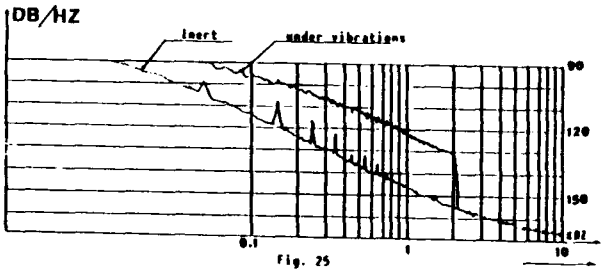


Fig. 25

Table 3 gives a summary of the results.

	100 Hz	1000 Hz	2000 Hz
1	106 dB/Hz	126 dB/Hz	132 dB/Hz
2	85 dB/Hz	105 dB/Hz	111 dB/Hz
3	100 dB/Hz	120 dB/Hz	126 dB/Hz

TABLE 3 - SUMMARY OF RESULTS

CONCLUSION

In this paper we have exposed the possibilities which exist to obtain filters with a good phase linearity and a good phase noise sensitivity under vibrations.

If the phase linearity is given by the right choice of the theoritical transfer function :

- linear phase
- semi-linear phase
- equalizer,

the phase noise sensitivity under vibrations depends on the right technological choice :

- filter structure
- resonator "g sensitivity"

A filter with a good phase linearity and/or a good phase noise sensitivity under vibrations, leads to an increase in volume and price and also to a bandwidth restriction.

BIBLIOGRAPHY

1. V.R. Rosati and R.L. Filler, "Reduction of the effects of Vibartion on SC cut Quartz crystal oscillators". Proc. 35th Annual Frequency Control Symposium (1981)
2. R.L. FILLER and J.R. VIG " The acceleration and Warm-up Characteristics of Four-point mounted SC and AT cut Resonators" Proc. 35th Annual Frequency Control Symposium
3. R.L. FILLER " The effect of vibration on Frequency standards and clocks" Proc. 35th Annual Frequency control Symposium (1981
4. RL. FILLER, J.A. Kosinski and J.R. VIG "Further studies on the acceleration sensitivity of Quartz resonators" Proc. 37th Annual Frequency control Symposium (1983).
5. Anatol I. Zverev "Handbook of Filter synthesis.
6. Humpherys " The Analysis and Synthesis of Electrical filters" Prentice Hall Inc. 1970
7. B.D. Rakovich & B.M. Djuric " Proc. IEEE 59 n° 11 Nov. 1971.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

TWO-DIMENSIONAL ANALYSIS OF THICKNESS-SHEAR AND FLEXURAL VIBRATIONS IN RECTANGULAR AT-CUT QUARTZ PLATES USING A ONE-DIMENSIONAL FINITE ELEMENT METHOD

Hitoshi SEKIMOTO and Yasuaki WATANABE
Tokyo Metropolitan University, Tokyo, 158 Japan
Mitsuo NAKAZAWA
Shinsyu University, Nagano, 380 Japan

Summary

In this paper, the two-dimensional coupled vibrations of thickness-shear and flexure in miniature rectangular AT-cut quartz resonators are theoretically studied. The analysis is based on Mindlin's plate equations. To calculate the mode chart, a one-dimensional finite element technique is employed. It is shown that the theoretical curves match the experimental data well.

I. Introduction

Strip AT-cut quartz resonators have been recently developed and widely used in frequency applications because of their small size [1]. However, the spurious characteristics of such miniature rectangular resonators have not been fully understood. The reason for this seems to be that there have been no analytical and numerical approaches to circumvent the difficulties in solving the two-dimensional boundary-value problem of rectangular AT-cut plates. This paper presents a method to analyze the two-dimensional coupled vibrations of thickness-shear and flexure. This method is based on Mindlin's two-dimensional plate equations [2] and effectively utilizes a one-dimensional finite element technique.

It is well known that Mindlin's plate equations have formed a basis for analyzing the thickness-shear and flexural vibrations in a rectangular AT-cut quartz plate. The equations have been successfully applied to the investigation of the strong spurious resonances with phase reversals along the X axis [3]. However, two-dimensional vibrations, in particular unwanted flexural vibrations, depending on both X-length and Z'-width have not been fully solved. In the famous work of Mindlin and Spencer [4] they have omitted the contribution of the second thickness-shear wave with the predominant strain component S_{23}. This S_{23} however plays an important role in studying the spurious characteristics of miniature resonators. In fact, Milsom et al. [5] have already showed that the second thickness-shear wave affects the main resonance characteristics in the strip resonators elongated along the Z' axis. Lee et al. [6] have included the influence of S_{23} in the analysis of trapped energy resonators, but the use of two-dimensional finite element technique limits the practical application of their method to the analysis of the thickness-shear family. To the authors' knowledge, the spurious characteristics in the miniature rectangular resonators still remains to be solved theoretically.

In this paper, an efficient method is proposed for analyzing the thickness-shear and flexural vibrations in strip resonators elongated along the X axis. The three plate equations which include two thickness-shear waves and one flexural wave are solved using a one-dimensional finite element method (FEM). Our approach here follows that of Tiersten's variational formalism [7]; two-dimensional eigenmodes in a semi-infinite plate are determined, and then a frequency equation is derived from a variational expression equivalent to the remaining boundary conditions. A one-dimensional FEM is effectively employed in these two processes. The ratios of Z'-width to plate thickness to fulfill the resonant condition are calculated as a function of frequency. It is shown that the calculations are in good agreement with the experimental results, and most of spurious resonances depending on Z'-width belong to the two-dimensional flexural family.

II. Plate Equations of Flexural Motion

Consider a rectangular AT-cut quartz plate of length 2l, thickness 2b and width 2w as shown in Fig.1, where the x_1, x_2 and x_3 coordinates are chosen as the X, Y' and Z' axes, respectively. Mindlin's two-dimensional stress equations of flexural motion, which take into account the c_{66} and c_{44} thickness-shear deformations and the flexural deformation, are (Eq.(2) in [4])

$$Q_{1,1}+Q_{3,3}+2b\,\rho\,\omega^2 u_2=0 \tag{1a}$$
$$M_{1,1}+M_{6,3}-Q_1+(2/3)b^3\,\rho\,\omega^2\psi_1=0 \tag{1b}$$
$$M_{6,1}+M_{3,3}-Q_3+(2/3)b^3\,\rho\,\omega^2\psi_3=0 \tag{1c}$$

where ω is the radian frequency, ρ is the plate mass density, u_2 is the component of the displacement and ψ_i are the components of the rotation in the x_i direction of a plate element. Since we are concerned with only the flexural motion, the shearing stress resultants Q_i normal to the plane of the plate and the bending and twisting couples M_i may be written as

$$\begin{aligned} Q_1&=2bk_1^2c_{66}(u_{2,1}+\psi_1)\\ Q_3&=2bk_3^2c_{44}'(u_{2,3}+\psi_3)\\ M_1&=(2/3)b^3(\gamma_{11}\psi_{1,1}+\gamma_{13}\psi_{3,3})\\ M_3&=(2/3)b^3(\gamma_{13}\psi_{1,1}+\gamma_{33}\psi_{3,3})\\ M_6&=(2/3)b^3\gamma_{66}(\psi_{1,3}+\psi_{3,1}) \end{aligned} \tag{2}$$

where

$$\begin{aligned} &\gamma_{pq}=c_{pq}'-c_{4p}'c_{q4}'/c_{44}',\\ &c_{pq}'=c_{pq}-c_{2p}c_{q2}/c_{22},\\ &\gamma_{66}=c_{66}-c_{56}^2/c_{66},\\ &k_1^2=\pi^2/12,\\ &k_3^2=\pi^2[c_{22}+c_{44}-\sqrt{(c_{22}-c_{44})^2+4c_{24}^2}]/24c_{44}'. \end{aligned} \tag{3}$$

CH2818-3/90/0000-358 $1.00 © 1990 IEEE

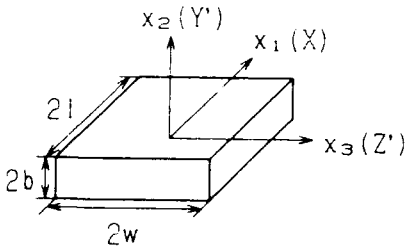


Fig.1 A retangular AT-cut quartz plate.

and c_{ij} are the elastic stiffnesses of the AT-cut quartz plate. Accordingly, the proper boundary conditions of a rectangular plate with all four edges free are

$$Q_1=M_1=M_6=0 \quad \text{on } x_1=\pm l \tag{4}$$
$$Q_3=M_3=M_6=0 \quad \text{on } x_3=\pm w \tag{5}$$

III. Guided Modes

The solutions of (1)-(5) can be expressed in terms of a sum of eigenmodes guided by the plate edges parallel to the x_3 axis [7]. In this paper, both the transverse fields along the x_1 direction and the longitudinal propagation wavenumbers in the x_3 direction for the guided modes are determined by using a one-dimensional FEM. To utilize the FEM, we will derive a variational expression for the guided modes.

When the impressed electric field is uniform over the plate, the vibrations excited will have u_2 odd in x_1 and even in x_3, ψ_1 even in x_1 and x_3, and ψ_3 odd in x_1 and x_3. Hence the appropriate form of a guided mode is given as

$$\begin{aligned} u_2 &= b\,U_2(x_1)\cos\beta' x_3 \\ \psi_1 &= \Psi_1(x_1)\cos\beta' x_3 \\ \psi_3 &= \Psi_3(x_1)\sin\beta' x_3 \end{aligned} \tag{6}$$

where β' is the wavenumber. The field components U_2 and Ψ_3 are odd functions of x_1, and Ψ_1 is the even function. The substitution of (6) into (1) and (2) yields

$$bk_1^2c_{66}D_{6,1}+\beta'k_3^2c_{44}'D_4+\rho\omega^2b^2U_2=0 \tag{7a}$$
$$b(\gamma_{11}E_1+\gamma_{13}E_3)_{,1}/3+\beta'\gamma_{65}E_6/3 - k_1^2c_{66}D_6+\rho\omega^2b^2\Psi_1/3=0 \tag{7b}$$
$$b\gamma_{65}E_{6,1}/3-\beta'(\gamma_{13}E_1+\gamma_{33}E_3)/3 - k_3^2c_{44}'D_4+\rho\omega^2b^2\Psi_3/3=0 \tag{7c}$$

where $\beta'=\beta b$ and

$$D_4=-\beta'U_2+\Psi_3,\ D_6=bU_{2,1}+\Psi_1,\ E_1=b\Psi_{1,1},$$
$$E_3=\beta'\Psi_3,\ E_6=-\beta'\Psi_1+b\Psi_{3,1} \tag{8}$$

Multiplying (7a),(7b) and (7c) by U_2, Ψ_1 and Ψ_3, respectively, and integrating through the half x_1-length and adding, we obtain

$$\int_0^1 \{k_3^2c_{44}'D_4^2+k_1^2c_{66}D_6^2+(\gamma_{11}E_1^2+2\gamma_{13}E_1E_3 + \gamma_{33}E_3^2+\gamma_{65}E_6^2)/3\}dx_1 = \rho\omega^2b^2\int_0^1\{U_2^2+(\Psi_1^2+\Psi_3^2)/3\}dx_1 \tag{9}$$

Here the boundary conditions of (4) have been applied. It is noted that (9) is a variational expression for the guided modes, and therefore can be solved by applying a one-dimensional FEM.

According to the FEM [8], we discretize the transversal structure of the plate into line-elements consisting of four nodes. Then, using a third-degree interpolation polynomial $S_j(x_1)$ which is with a value of 1 at the jth node and zero at others, the transversal fields are approximated as

$$[U_2,\Psi_1,\Psi_3]=\sum_{j=0}^{N_p-1}S_j(x_1)[U_{2j},\Psi_{1j},\Psi_{3j}],\quad 0\leqq x_1\leqq 1 \tag{10}$$

where N_p is the number of nodes, and U_{2j}, Ψ_{1j} and Ψ_{3j} denote the nodal values at the jth node. Since U_2 and Ψ_3 are odd in x_1, U_{20} and Ψ_{30} at $x_1=0$ are set to zero. When (10) is substituted into (9), the result can be stated as the following matrix equations

$$\begin{aligned} (\beta'^2A_{11}+C_{11})\{U_2\}+C_{12}\{\Psi_1\}+\beta'B_{13}\{\Psi_3\} &= 0 \\ C_{12}\{U_2\}+(\beta'^2A_{22}+C_{22})\{\Psi_1\}+\beta'B_{23}\{\Psi_3\} &= 0 \\ \beta'B_{13}^T\{U_2\}+\beta'B_{23}^T\{\Psi_1\}+(\beta'^2A_{33}+C_{33})\{\Psi_3\} &= 0 \end{aligned} \tag{11}$$

where $\beta'=\beta b$, $\{\cdot\}$ represents a column vector and the matrices C_{ij} are functions of ω^2. Defining a total column vector $\{y\}$ by

$$\{y\}^T=[\{U_2\}^T,\{\Psi_1\}^T,\beta'\{\Psi_3\}^T] \tag{12}$$

(11) is reduced to a standard eigenvalue equation

$$F\cdot\{y\}=\beta'^2\{y\} \tag{13}$$

where F is the N × N matrix (N=3•N_p-2)

$$F=\begin{bmatrix} -A_{11}^{-1}C_{11} & -A_{11}^{-1}C_{12} & -A_{11}^{-1}B_{13} \\ -A_{22}^{-1}C_{12} & -A_{22}^{-1}C_{22} & -A_{22}^{-1}B_{23} \\ -A_{33}^{-1}B_{13}^T & -A_{33}^{-1}B_{23}^T & -A_{33}^{-1}C_{33} \end{bmatrix} \tag{14}$$

From (13) we can compute the dispersion relation between a given frequency ω and the wavenumbers β'. An example of dispersion curves is shown in Fig.2 for a plate with l/b=8.039, of which the half x_1-length was discretized into 15 line-elements (N_p=15, N=136). The ordinate is the frequency Ω normalized to the c_{66} thickness-shear frequency $\omega_0(=\sqrt{3k_1^2c_{66}/\rho b^2})$ of an infinite plate and the abscissa is the dimensionless wavenumber $\phi=2\beta b/\pi$. We note that there are many complex branches shown as the thick lines, as well as the real or imaginary branches shown as the thin lines. The occurrence of these complex branches is a consequence of including the c_{44} thickness-shear deformation in the analysis [5],[9]. At $\phi=0$, ψ_3 is uncoupled to u_2 and ψ_1, and the guided modes may be separated into three groups of an uncoupled c_{44} thickness-shear mode, a predominantly flexural mode and a predominantly c_{66} thickness-shear mode which are designated as TS'-i, F-j and TS-k in Fig.2, respectively. The number in these mode labels refers to the number of antinodes across the length 2l of the plate. In case of $\phi\neq 0$, ψ_3 is coupled to u_2 and ψ_1. This coupling produces the complex branches, and also greatly changes the real and imaginary branches obtained by neglecting ψ_3 [9] which Mindlin and Spencer have used in [4].

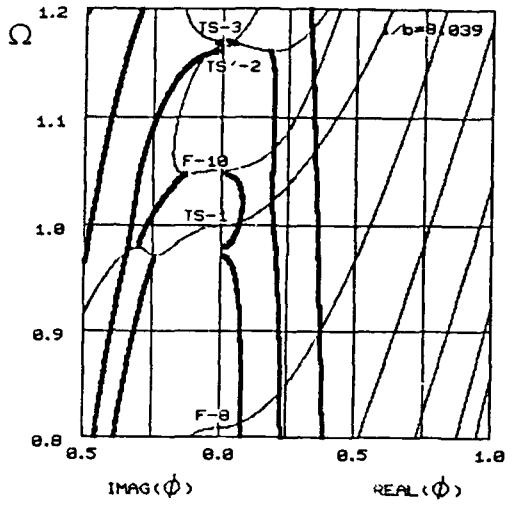


Fig.2 Calculated dispersion curves for thickness-shear and flexural waves propagating in the x_3 direction in a semi-infinite AT-cut quartz plate(l/b=8.039).

IV. Frequency Equation

The guided modes satisfy the boundary conditions (4) at $x_1=\pm 1$ automatically. Therefore the remaining boundary conditions (5) at $x_3=\pm w$ are written by approximating with a linear combination of N guided modes as follows.

$$\begin{aligned} Q_3 &= \sum_{n=1}^{N} a_n Q_3^{(n)}(x_1)\sin \beta^{(n)} w/\beta^{(n)'}=0 \\ M_6 &= \sum_{n=1}^{N} a_n M_6^{(n)}(x_1)\sin \beta^{(n)} w/\beta^{(n)'}=0 \\ M_3 &= \sum_{n=1}^{N} a_n M_3^{(n)}(x_1)\cos \beta^{(n)} w=0 \end{aligned} \tag{15}$$

where a_n and $\beta^{(n)}$ are the amplitude and wavenumber of the nth guided mode, and

$$\begin{aligned} Q_3^{(n)} &= 2bk_3^2 c_{44}'(-\beta^{(n)'2} U_2^{(n)} + \beta^{(n)'} \Psi_3^{(n)}) \\ M_6^{(n)} &= (2/3)b^2 \gamma_{66}(-\beta^{(n)'2} \Psi_1^{(n)} + \beta^{(n)'} b \Psi_{3,1}^{(n)}) \\ M_3^{(n)} &= (2/3)b^2(\gamma_{13} b \Psi_{1,1}^{(n)} + \gamma_{33}\beta^{(n)'} \Psi_3^{(n)}) \end{aligned} \tag{16}$$

Here $U_2^{(n)}$, $\Psi_1^{(n)}$ and $\Psi_3^{(n)}$ represent the nth eigen functions determined from (10) and (12).

Since we take into account only N terms of the guided modes, (15) cannot be fulfilled for all x_1. Consequently we attempt to utilize both Tiersten's variational technique [7] and one-dimensional FEM employed above. Now we introduce arbitrarily adjustable field components $\underline{U_2}$, $\underline{\Psi_1}$ and $\underline{\Psi_3}$ at the edge $x_3=w$. Then (15) may be transformed, according to the variational procedure, into

$$\int_0^1 (Q_3 b\underline{U_2} + M_6 \underline{\Psi_1} + M_3 \underline{\Psi_3})dx_1=0, \quad x_3=w \tag{17}$$

The arbitrarily adjustable field components are expanded in terms of the same third-degree interpolation polynomials S_j which were used in (10), as follows.

$$[\underline{U_2}, \underline{\Psi_1}, \underline{\Psi_3}] = \sum_{j=0}^{N_p-1} S_j(x_1)[\underline{U_{2j}}, \underline{\Psi_{1j}}, \underline{\Psi_{3j}}], \quad 0 \leqq x_1 \leqq 1 \tag{18}$$

where N_p is the number of nodes, $\underline{U_{2j}}$, $\underline{\Psi_{1j}}$ and $\underline{\Psi_{3j}}$ denote the nodal values at the jth node, and $\underline{U_{20}}$ and $\underline{\Psi_{30}}$ at $x_1=0$ are set to zero. Substituing (15) and (18) into (17) yields the following approximation for the boundary conditions:

$$\begin{aligned} &\sum_{i=1}^{N_p-1} \underline{U_{2i}} \Big[\sum_{n=1}^{N} f_{in} a_n\Big] + \sum_{j=0}^{N_p-1} \underline{\Psi_{1j}} \Big[\sum_{n=1}^{N} g_{jn} a_n\Big] \\ &\quad + \sum_{k=1}^{N_p-1} \underline{\Psi_{3k}} \Big[\sum_{n=1}^{N} h_{kn} a_n\Big] = 0 \end{aligned} \tag{19}$$

where $N=3*N_p-2$ and

$$\begin{aligned} f_{in} &= \Big(\int_0^1 S_i b Q_3^{(n)} dx_1\Big)\sin \beta^{(n)} w/\beta^{(n)'} \\ g_{jn} &= \Big(\int_0^1 S_j M_6^{(n)} dx_1\Big)\sin \beta^{(n)} w/\beta^{(n)'} \\ h_{kn} &= \Big(\int_0^1 S_k M_3^{(n)} dx_1\Big)\cos \beta^{(n)} w \end{aligned} \tag{20}$$

The N nodal values of $\underline{U_{2i}}$, $\underline{\Psi_{1j}}$ and $\underline{\Psi_{3k}}$ are arbitrary. Hence we obtain a set of N homogeneous algebraic equations which governs the amplitudes a_n. For these equations to have nontrivial solutions, the determinant of the coefficients must vanish, that is

$$\det [\, f_{ni}, g_{nj}, h_{nk} \,] = 0 \tag{21}$$

for n=1,...N, i,k=1,...N_p-1 and j=0,...N_p-1.
Equation (21) represents a functional relation between the resonant frequency ω and the width-to-thickness ratio w/b, and is the frequency equation for the present problem.

V. Numerical Results

For two fixed length-to-thickness ratios l/b of 8.039 and 15.7, the width-to-thickness ratios w/b to satisfy (21) were calculated as a function of frequency and compared with the experimental results. We here omitted the piezoelectric effect and employed Bechmann's elastic constants [10] which yields a frequency constant of 1.6546 (MHz·mm) for the AT-cut. The results are illustrated in Figs.3 and 5 where the solid curves show the calculated results. In the computation of the curves, the 10^{-4} or better accuracy for w/b could be attained by taking 15 (N=136) and 20 (N=181) line-elements for l/b=8.039 and 15.7, respectively.

Fig.3(a) shows the results for l/b=8.039. The experimental data represented by small dots were obtained by one of the authors, Nakazawa, for a plate with the thickness 2b and length 2l of 1.737 mm and 13.964 mm, respectively, making the fundamental main thickness-shear frequency about 0.95 MHz. The plate width 2w was carefully reduced from 15.3 mm (w/b=8.81) to 3.2 mm (w b=1.84) in steps of about 0.1 mm. The plate was completely covered with a pair of metallic electrodes, and the frequency and

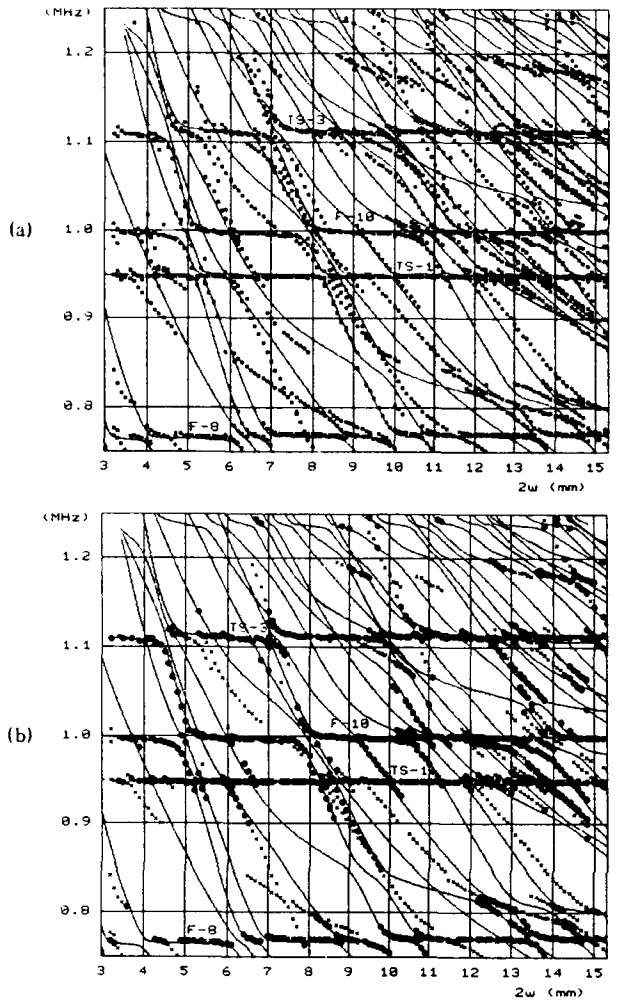


Fig.3 Comparison of calculated and measured resonant frequencies of thickness-shear and flexure in rectangular AT-cut quartz plates (thickness 2b=1.737mm, X-length 2l=13.964mm), as a function of Z'-width 2w. In (a) all the measured data, of which each strength havs been originally classified into six ranks, are plotted with square dots (□) of the same size, and in (b) only the resonances corresponding to the higher four ranks are plotted with marks (●, ×).

strength of each resonance were recorded. Although the ralative amplitudes of resonances have been originally classified into six ranks, all the data in Fig.3(a) are plotted with dots of the same size. For clarity, a comparison of the calculated results with only the resonances corresponding to the higher four ranks is also shown in Fig.3(b). Here the experimental data in good agreement with the present theory are emphasized by putting the large dots on them, and the data in disagreement are denoted by the crosses. The discrepancies between theory and experiment are most probably due to the face-shear and extensional modes not included in the analysis.

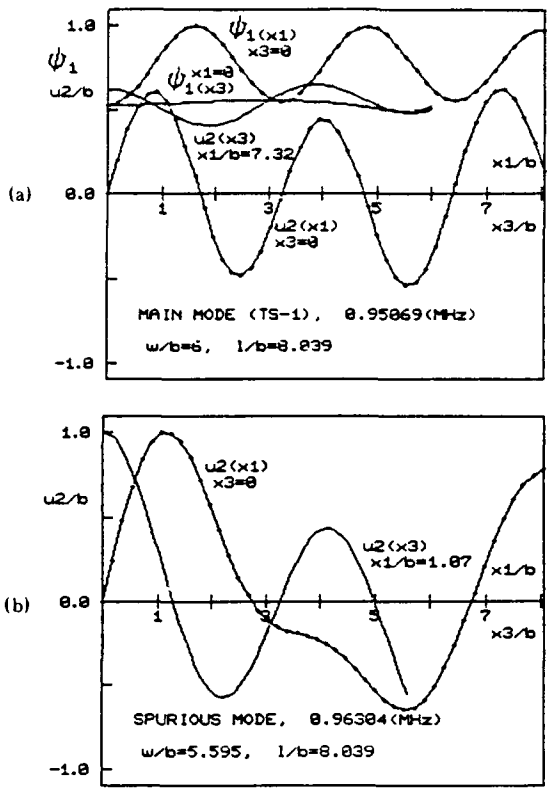


Fig.4 Calculated vibrational patterns for (a) a main thickness-shear mode and (b) a flexural spurious mode.

In Fig.3(b), there are four horizontal lines designated as the 8th and 10th modes of flexure (F) and the 1st and 3rd modes of thickness-shear (TS). These are nearly straight crested in the Z' direction as depicted in Fig.4(a). The other steep curves belong to the flexural family strongly depending on the Z' dimension. Fig.4(b) shows a calculated example of its predominan. component u_2. The phase reversals in both the X and Z' directions are visible. The theory predicts quite well the intersections of these two-dimensional flexural modes and the main thickness-shear mode TS-1.

The results for l/b=15.7 and w/b=2～5 are shown in the form of frequency constant (KHz·mm) vs. w/b, in Fig.5. The dots in the figure were plotted from the data obtained by Yamashita et al.(Fig.6 in [11]). Here we adjusted the data in order that the measured frequencies of main resonances nearly agree in the calculated values, because the measured values were about 13KHz·mm high compared with the data plotted in Figs.9-11 in [11]. The good agreement is observed again between the calculated results and the experimental data marked by the large dots.

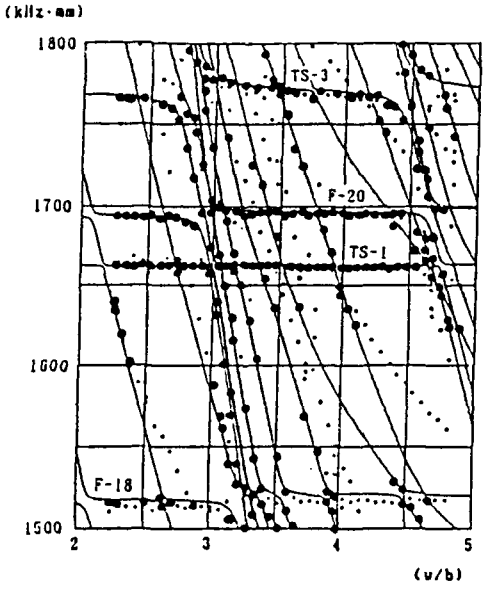


Fig.5 Comparison of calculated and measured resonant frequencies for 4.19 MHz rectangular AT-cut quartz plates (X-length-to-thickness ratio l/b=15.7), as a function of Z'-width-to-thickness ratio w/b. The dots (●, •) are the measured data by Yamashita et al.[11].

VI. Conclusions

On the basis of the comparison between theory and experiment given in Figs.3 and 5, we conclude that most of spurious resonances in strip type of miniature resonators belong to the two-dimensional flexural family. The numerical method presented here is well suited to determine the points at which these flexural modes will coincide with the fundamental main thickness-shear mode in the miniature resonators elongated along the X axis. It is not so difficult to include the face-shear and extensional modes in the analysis. If the coupling with the thickness-shear and flexural modes is omitted, the spurious characteristics associated with the face-shear and extension can be easily solved by using the present technique. In fact, we have found that such an analysis may remarkably improve the discrepancies between theory and experiment. We will report these results in the near future.

Acknowledgements

The authors would like to thank M.Okazaki of Nihon Dempa Kogyo Co. Ltd. for help in performing the experiment.

References

[1]M.Onoe and M.Okazaki:"Miniature AT-cut strip resonators with tilted edges", Proc.Ann.Freq.Control Symp., 29,pp.42-48(1975)

[2]R.D.Mindlin:"High frequency vibrations of crystal plates", Quart.Appl.Math., 19,pp.51-61(1961)

[3]R.D.Mindlin and D.C.Gazis:"Strong resonances of rectangular AT-cut quartz plates", Proc.U.S.Natl.Congr.Appl.Mech.,4th,pp.305-310(1962)

[4]R.D.Mindlin and W.J.Spencer:"Anharmonic,thickness-twist overtones of thickness-shear and flexural vibrations of rectangular,AT-cut quartz plates", J. Acoust. Soc. Amer., 42,6, pp.1268-1277(1967)

[5]R.F.Milsom, D.T.Elliott, S.Terry-Wood and M.Redwood, "Analysis and design of coupled-mode miniature bar resonators and monolithic filter",IEEE Trans. Sonics & Ultrason. , 30,3,pp.140-155 (1983)

[6]P.C.Lee, C.Zee and C.A.Brebbia:"Thickness-shear, thickness -twist, and flexural vibrations of rectangular AT-cut quartz plates with patch electrodes", Proc.Ann.Freq.Control Symp., 32, pp.108-119 (1978)

[7]H.F.Tiersten:"Linear piezoelectric plate vibrations", Plenum Press, pp.137-138(1969)

[8]O.C.Zienkiewicz:"The finite element method",McGraw Hill (1977)

[9]H.Sekimoto and Y.Watanabe:"On the thickness-twist approximation for rectangular AT-cut quartz resonators proposed by Mindlin and Spencer", Trans. Inst. Electron. Inform. Commun. Eng. Japan, J72-A,11, pp.1934-1936(1989),in Japanese

[10]R.Bechmann:"Elastic and piezoelectric constants of alpha -quartz", Phy. Rev., 110,p.1060(1958)

[11]S.Yamashita et al.:"A 4.19MHz beveled rectangular AT-cut quartz resonators", Proc.Ann.Freq.Control Symp.,32, pp.267 -276(1978)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PRECISE DETERMINATION OF VHF QUARTZ CRYSTAL RESONATOR EQUIVALENT PARAMETERS

Yoshimasa OOMURA and Yasuaki WATANABE

Tokyo Metropolitan University
Tokyo, 158 Japan

Summary

A novel method has been developed for accurate determination of VHF quartz crystal resonator equivalent parameters and resonant frequencies. This method employs an electrical equivalent circuit which includes the spurious responses close to the main mode and the resistance originated from the evaporated electrodes. The equivalent parameters and the resonant frequencies are determined by a non-linear least squares method from the impedance characteristics of the resonators under test.

It shows that an excellent agreement is obtained between the impedance characteristics calculated by the equivalent parameters and experimental data. Moreover, the good accordance is obtained between the calculated non-reactive frequency(fr) and the experimental value. Although the strong spurious responses exist in the vicinity of main resonance, the exact determination of equivalent parameters can be carried by this method.

I. Introduction

Recently the standard measurement of a piezoelectric resonator, which includes a quartz crystal unit, a saw resonator and a ceramic resonator, has been examined by an automatic network analyzer in I.E.C. , TC49 WG6 [1]. Therefore, Japanese National Committee has investigated the same measurement method according with the national round robin and so on. In this paper, a new method, which is suited for an automatic network analyzer system, has been developed for precise determination of VHF quartz resonator equivalent parameters.

Generally speaking, the equivalent circuits of the quartz crystal unit is constructed by four elements, that is, the motional arm contained by L1, C1, R1, and the parallel capacitance Co. This circuit has been almost adopted in measuring the electrical characteristics of crystal unit. However, in the range of VHF, the four elements equivalent circuit is not useful for obtaining the exact impedance-frequency characteristics. It is reason that there are an energy loss by the evaporated electrodes and the support[2]. The five elements equivalent circuit, therefore, is considered instead of the mentioned above circuit, the fifth element corresponds to the energy loss and is denoted by R_o. The electrical behavior is more exactly described by this five elements circuit than the traditional circuit. On the other hand, the higher harmonic overtone is generally used in VHF crystal resonator units, and the spurious responses exist nearby the main resonance frequency. It is, therefore, necessary to consider the influence of these nearby responses as well as the R_o for determination of the exact equivalent parameters of the main resonance.

CH2818-3/90/0000-363 $1.00 © 1990 IEEE

In this paper, an effective technique is proposed for the purpose of the more precisely determination of the equivalent parameters in VHF quartz resonator units. In the 2nd chapter, it is shown that how to determine the equivalent circuit which contains both R_o and nearby responses, and the procedure of the determination of the equivalent parameters. In the 3rd chapter, the experimental results are shown by measurements with the automatic network analyzer concerning the crystal unit in the range of frequency from 97MHz to 200MHz. The estimation of this method is shown in the 4th chapter.

II. Determination method of equivalent parameters

II.1. Equivalent circuit of VHF crystal units

Figure 1 shows a typical impedance-frequency characteristics of the VHF crystal unit. In this figure it is understood that spurious responses exist in the vicinity of the main resonance frequency. In this paper, therefore, both R_o and nearby resonances are added to the traditional four elements equivalent circuit. This circuit is shown in Figure 2.

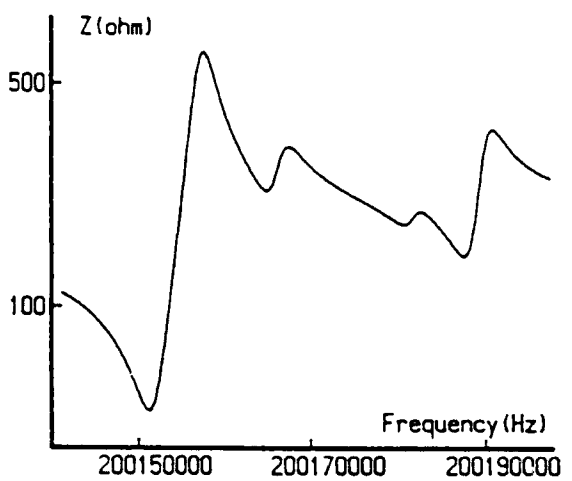


Fig.1 Typical impedance - frequency characteristics of a VHF quartz unit.

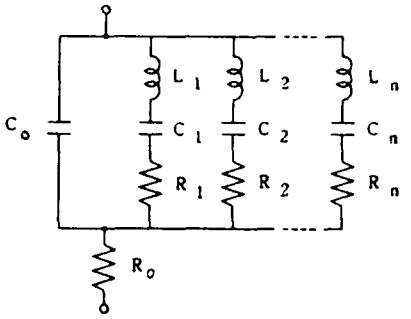


Fig.2 Equivalent circuit of a VHF quartz crystal resonator unit.

II.2. Determination of equivalent parameters

In case of obtaining the equivalent parameters by using the circuit shown in Figure 2, the determination method based on a circular diagram may not be adopted. Therefore a non-linear least squares method is used by expanding to the circuit containing Ro and nearby responses. According to the method, at first step the approximate values of equivalent parameters are necessary as the initial values, and their values are obtained from the impedance behavior of a crystal unit under test. Next, the obtained values are utilized as the initial values for progressive procedure. Determination of equivalent parameters is carried out by iterative least squares method. This method brings so least deviation that results agree with values between the measuring and the calculating of impedance characteristics. In this method the equivalent parameters, which are determined as experimental subjects, are as follows:

(1) Reactance $X_o(1 / 2 \pi f_{s1} C_o)$ in parallel arm.
(2) Energy loss R_o.
(3) Series resonance frequencies ($f_{sn} = 1 / 2 \pi \sqrt{L_n C_n}$).
(4) Parallel resonance frequencies ($f_{pn}=1 / 2 \pi \sqrt{L_n C_n // C_o}$).
(5) Series resistances (R_n).

where n is the number of motional arms containing main resonance frequency, therefore the number of the equivalent parameters becomes to the sum of 2+3n.

At the beginning, Z_{m1} correspond to the absolute impedance value at measuring frequencies of a crystal unit under test, and Z_{o1} is the calculated absolute values of the impedance at the same frequencies mentioned above.

After this, the estimation function S is defined as follows,

$$S=\sum_{i=1}^{k}(Z_{o1}-Z_{m1})^2 \qquad (1)$$

where k is the total number of measuring points. And also, the minimum values condition of the estimation function is given the following formula by the definition of least squares method,

$$\frac{\partial S}{\partial x_o}=\frac{\partial S}{\partial R_o}=\frac{\partial S}{\partial f_{s1}}=\frac{\partial S}{\partial f_{p1}}=\frac{\partial S}{\partial R_1}=\cdots=\frac{\partial S}{\partial R_n}=0 \qquad (2)$$

However, the analytical solution of formula (2) may not be obtained, for the reason why each function, which is decided by subject, does not contain the formula of the linear function Z_{o1}. Therefore, the estimation function S is expressed as follows adopting in case of the 1st term considered by the Taylor series expansion of Z_{o1}.

$$S=\sum_{i=1}^{k}\left(\frac{\partial Z_{o1}}{\partial X_o}\Delta X_o+\frac{\partial Z_{o1}}{\partial R_o}\Delta R_o+\frac{\partial Z_{o1}}{\partial f_{s1}}\Delta f_{s1}+\cdots+\frac{\partial Z_{o1}}{\partial R_n}\Delta R_n+Z_{c1}-Z_{m1}\right)^2 \qquad (3)$$

where sign Δ shows a correction quantity to the initial condition value of each decided subject.

The solution which gives minimum condition for (3) is shown by next equation.

$$\frac{\partial S}{\partial \Delta x_o}=\frac{\partial S}{\partial \Delta R_o}=\frac{\partial S}{\partial \Delta f_{s1}}=\frac{\partial S}{\partial \Delta fp_1}=\frac{\partial S}{\partial \Delta R_1}=\cdots=\frac{\partial S}{\partial \Delta Rn}=0 \qquad (4)$$

From (4), correct values according to initial values are obtained. On the other hand, equation (3) is the first order approximation, therefore, the calculation is repeated in practice until each value brings to convergence. In this calculation, the initial values are determined by treating that each spurious resonance is independent respectively.

III. Measurement results

III.1. Measurement apparatus and Crystal units under test

The automatic network analyzer (Anritsu MS620J & SWR Bridge) was selected for actual measurement of resonator impedance. Figure 3 shows a block diagram of measurement setup. The initial calibration of the system is carried out by one-port method using 50 Ω termination, open and short[3]. The Rb atomic oscillator is applied as an external frequency standard, and the temperature in oven is kept at 25 ± 0.01 ℃ in order that the equivalent parameters of crystal unit under test may not change during measurement.

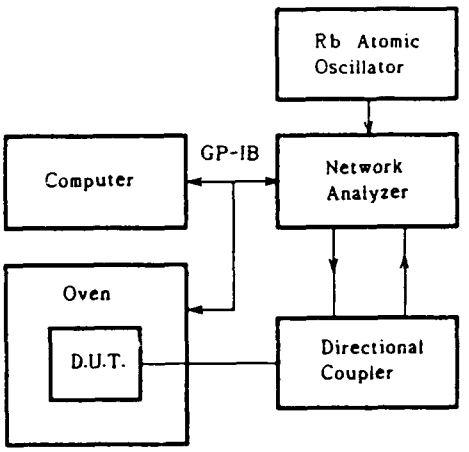


Fig.3 Block diagram of the measuring system.

The rating resonance frequencies of crystal units are (a)200MHz, (b)160MHz, (c)125MHz, (d)97MHz respectively and four samples are all AT-cut. In these resonators, four or five spurious responses are adopted according to the near order from the main response, to evaluate the influence of them together with the main resonance in this experiment.

III.2. Experimental results of equivalent parameters

Table 1 (a) to (d) show the decided results of the sampled crystal units. Table 2 shows the mean value and the standard deviation after series of 10 times measurements are repeated. It is understood from Table 2 that the standard deviation of resonance frequencies is 9.8×10^{-9} and each equivalent parameter is within 0.63%, and also that the experimental repeatability is excellent.

Table 1. Analysis Results for quartz units under test. (25℃, 0.5mA)

(a) 200MHz (7th overtone, $Q=5.8\times10^{4}$)

Motional Arm	f_s(Hz)	R_s(Ω)	L_s(mH)	X_o(Ω)	R_o(Ω)
1	200151840	42.5	2.29	177	7.53
2	200165941	593	29.3		
3	200181635	1.09k	62.4		
4	200188506	222	11.7		

(b) 160MHz (9th overtone, $Q=7.2\times10^{4}$)

Motional Arm	f_s(Hz)	R_s(Ω)	L_s(mH)	X_o(Ω)	R_o(Ω)
1	160781944	50.3	3.86	193	3.01
2	160789786	1.98k	151		
3	160794632	2.18k	163		
4	160799897	349	26.9		

(c) 125MHz (7th overtone, $Q=9.5\times10^{4}$)

Motional Arm	f_s(Hz)	R_s(Ω)	L_s(mH)	X_o(Ω)	R_o(Ω)
1	125112527	26.4	3.95	253	6.50
2	125119200	1.08k	153		
3	125121492	1.92k	266		
4	125128334	826	118		
5	125134107	196	28.1		

(d) 97MHz (5th overtone, $Q=5.9\times10^{4}$)

Motional Arm	f_s(Hz)	R_s(Ω)	L_s(mH)	X_o(Ω)	R_o(Ω)
1	97105772	30.0	3.52	275	6.60
2	97120680	250	29.3		
3	97133316	780	79.2		
4	97141014	145	16.8		

Table 2. Standard deviation of equivalent parameters. (200MHz quartz unit, N=10, 25.0℃, 0.5mA)

	fs1(Hz)	Rs1(Ω)	Ls(mH)	Xo(Ω)	Ro(Ω)
Average χ	200151840	42.5	2.29	177	7.53
S.D. σ	2.0	0.06	0.0014	0.61	0.04
σ / χ	9.8×10^{-9}	0.14%	0.062%	0.34%	0.63%

IV. Evaluation of equivalent parameters

IV.1. Impedance characteristics

The evaluation method in this paper is the mean square deviation between experimental values and impedance behavior calculated from equivalent parameters obtained by this method. The considering frequency range is restricted in the neighborhood of the main resonance(from f_{s1} to f_{p1}) under this condition, the decision method by the traditional four elements equivalent circuit is compared to the same method by the five elements equivalent circuit. In Table 3, the results are shown. In the same table, the least deviation is shown by the method presented in this paper whichever sample resonators' results are selected. That is to say, it is understood that the main resonance impedance characteristics can be the most exactly explained by the equivalent parameters according to this method. Figure 4 shows the deviation between the measured value and the calculated results of impedance characteristics at main resonance of the sample crystal unit (a). From both Table 3 and Figure 4, it is cleared that the exact impedance characteristics is at the main resonance is obtained by the equivalent circuit including R_o and nearby resonances.

IV.2. Phase characteristics

The evaluation of phase characteristics is carried by the comparison of the non-reactive (zero phase) frequency calculated by the equivalent parameters and the measured one. The experimental results of both four elements and five elements equivalent circuits are also shown in Table 4. It is understood from the same table that the deviation between the experimental and calculating results by this method is within 3.6×10^{-8}, and the best accordance is obtained in each resonator unit. Finally the main resonance phase behavior is well explained by the equivalent circuit used in this paper.

Table 3. Mean square deviations between experimental impedance and calculated one.

	(a)200MHz	(b)160MHz	(c)125MHz	(d)97MHz
4 elements	6.75	5.88	18.11	11.86
5 elements	2.20	0.99	4.13	2.29
Nearby response	2.38	0.97	2.62	3.22
This method	0.20	0.55	0.37	0.43

V. Conclusions

The equivalent parameters of VHF crystal units, which is contained by R_o and the spurious responses in the vicinity of main resonance, is precisely decided by this method mentioned in this paper. Moreover, this method is compared to the traditional measuring method. In fact, the results of measuring VHF crystal unit explained that the main resonance impedance characteristics more exactly explained than the well-known equivalent circuit

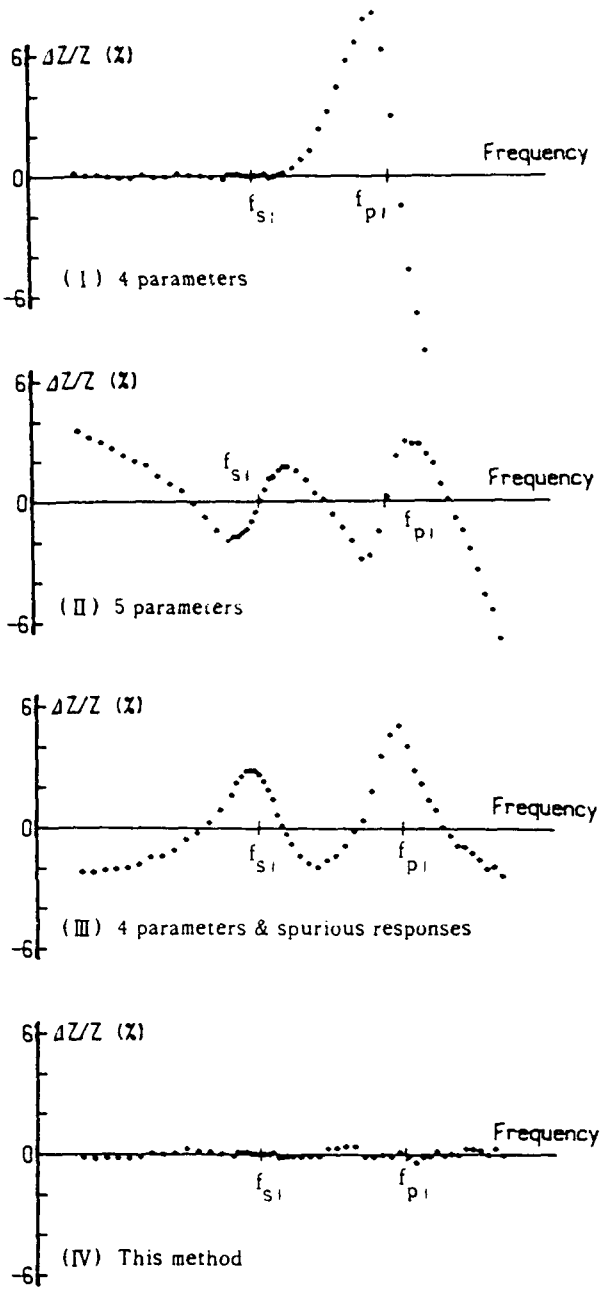


Fig.4 Deviations between measured impedance values and calculated ones.

Table 4. Comparison of measured non-reactive frequencies and Calculated ones.

	(a) 200MHz	(b) 160MHz	(c) 125MHz	(d) 97MHz
Measured Value	200152259	160782251	125112589	97105856
4 Parameters	200152413 ($+7.7\times10^{-7}$)	160782302 ($+3.2\times10^{-7}$)	125112620 ($+2.4\times10^{-7}$)	97105894 ($+3.9\times10^{-7}$)
5 Parameters	200152389 ($+6.5\times10^{-7}$)	160782321 ($+4.4\times10^{-7}$)	125112662 ($+5.8\times10^{-7}$)	97105899 ($+4.4\times10^{-7}$)
Nearby Responses	200152299 ($+2.0\times10^{-7}$)	160782268 ($+1.1\times10^{-7}$)	125112600 ($+8.8\times10^{-8}$)	97105859 ($+3.0\times10^{-8}$)
This Method	200152252 (-3.6×10^{-8})	160782254 ($+1.7\times10^{-8}$)	125112588 (-1.5×10^{-8})	97105854 (-2.3×10^{-8})

of the traditional four parameters. And also from the experimental results of the non-reactive frequency, the equivalent circuit in this paper is very good for the phase characteristics.

In conclusion it is understood that this method is very good for deciding the equivalent parameters of VHF crystal unit. Moreover, this method is going to be investigated in Japanese National Round Robin for the Network Analyzer transmission method.

Acknowledgement

The authors would like to thank Mr. M.Okazaki and Mr. M.Koyama, who are with NDK Co. Ltd., for equipment of the sample crystal units.

References

[1] IEC Document 49(U.S.A.)166: "A Reference Method of Measurement for Quartz Resonator Units. Using Automatic Network Analyzer Techniques for the Determination of Equivalent Electrical Parameters."(1986)

[2] I. Koga : "Equivalent Circuit of VHF Crystal Resonator equipped with Evaporated Electrodes" Published by Institute of Electrical Communication, written by Japanese (April 1983) (This was translated into English by Y.Oomura for the class of NIU and was distributed by EIA.)

[3] Watanabe.Y, Y.Oomura, Y.Iizuka: "Determination of Parameters of VHF Quartz Crystal Units Equipped with Evaporated Electrodes by the use of an Fr-meter Fixture" IEICE Trans. of Japan A. Vol.J70-A No.11 pp.1674-1680 (Nov.1987)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A NEW LOW FREQUENCY THERMALLY COMPENSATED CONTOUR MODE RESONATOR

C. BOURGEOIS

Swiss Center for Electronics and Microthechnology Inc.
Research and development
Neuchâtel, Switzerland

ABSTRACT

A new thermally compensated quartz resonator for wrist-watch applications is described. It oscillates at a low frequency (524 kHz) in order to minimize oscillator power comsumption. It could be mass-produced by using photolithography and chemical etching. The structure has been analysed and optimized with a contour mode computer program based on variational techniques.

The thermal compensation is achieved by the mechanical coupling of two low frequencies contour modes of the resonator, namely a flexure-type mode and an extensional-type mode. The structure is such that the elastic part of the mechanical coupling can be balanced with an inertial one for any crystallographic orientation of the substrate. The coupling factor vanishes, but not its thermal derivatives. Thus it can be shown that its influence is limited to the second and higher orders of the temperature coefficients of the frequencies.

INTRODUCTION

In piezoelectric resonators, electromechanical coupling makes it possible to excite elastic modes. Most of them are characterized by shear (thickness or face), flexural or extentional motion. Due to the finite dimensions and reflexion on the bounderies of the resonator, these different modes can be elastically-coupled. One mode is usually desired and all the others are unwanted.

In some cases two modes are deliberately elastic-coupled in order to achieve a particular property, most often the frequency-temperature dependence of the resonator. The flexural mode of the tuning fork can be coupled with a torsional mode of the arms[5]. In the GT or ZT resonators[1,2,3,4] two extensional contour modes of a rectangular plate are coupled together through an elastic cross-constant (Poisson's ratio σ_{12}). The resulting frequency is very near constant over a wide temperature range.

CH2818-3/90/0000-367 $1.00 © 1990 IEEE

This paper concerns a thermally compensated resonator for wrist-watch applications which can be clamped and which oscillates at a low frequency in order to minimize the oscillator power consumption. This compensation is achieved by the coupling of an extensional-type mass-loaded mode with a flexure-type mode of the structure, as shown in Fig. 1. These two modes belong to the same family of dilatation contour modes[7,8].

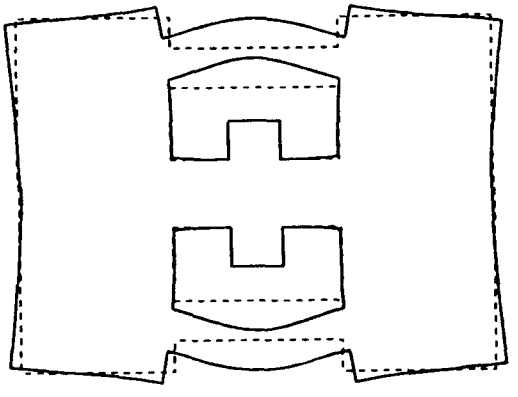


Figure 1a : Flexure-type mode

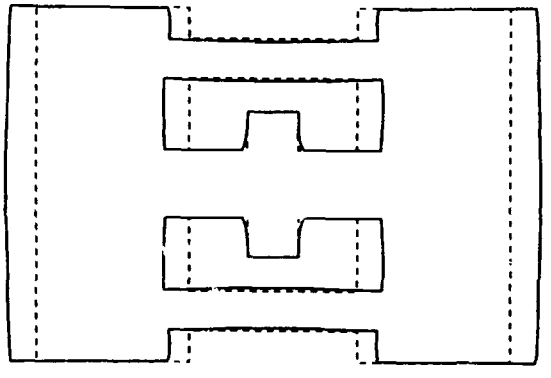


Figure 1b : Extensional-type mode

The structure is such that the elastic part of the mechanical coupling can be balanced with an inertial one for any crystallographic orientation of the substrate. The global coupling vanishes, but not its thermal derivatives. Thus its influence is limited to the temperature derivatives of the two frequencies.

It is suggested that this resonator be designated **FXT**, to stress out the fact that it uses a **F**le**X**ural-e**XT**entional coupling.

The frequency can be related to the outline dimensions by the **resonator extent** defined as

$$E = F\ L\ 2V_L^{-1}$$

where F = frequency
L = maximum dimension
V_L = longitudinal wave velocity (~ 5500 m/s)

Thus a low frequency resonator is characterized by a low extent.

In a GT or ZT rectangular plate E = 1;

ELASTIC COUPLING

The f-T curve of a resonator (frequency temperature dependence) can be expressed as

$$F(T) = F_o\ (1 + \alpha(T-T_o) + \beta(T-T_o)^2 + \gamma(T-T_o)^3\ )$$

where T = temperature
T_o = reference temperature
F_o = frequency at T_o
α = first-order temperature coefficient
β = second-order temperature coefficient
γ = third-order temperature coefficient

Elastic cross-constant (Poisson's ratio or mutual constants σ_{ij}) or inertial forces allow weak energy exchanges between two coupled modes[6p84]. The frequency of one mode as well as its thermal coefficients depend on the proximity of the other.

In a first approximation the coupling effect on one mode can be expressed through a coupling factor as

$$F = F_o\ [1 + K^2/4\cdot\Delta_f]$$
$$\alpha = \alpha_o + KK'/2\cdot\Delta_f$$
$$\beta = \beta_o + [KK'' + K'^2\ ]/2\cdot\Delta_f$$
$$\gamma = \gamma_o + [KK''' + K'K'']/2\cdot\Delta_f$$

where F = frequency of the compensated mode with coupling
F_o = frequency of the compensated mode without coupling
F_c = frequency of the compensating mode without coupling

$$\Delta_f = (\ F_o - F_c\ )\ /F_o$$

$\alpha_o\ \beta_o\ \gamma_o$ = frequency-temperature coefficient without coupling
$K = K_E - K_I$ = coupling factor
K_E = elastic coupling factor
K_I = inertial coupling factor or coupling with inertial forces

$K'\ K''\ K'''$ = thermal derivatives of coupling factor K

The elastic coupling K_E depends mainly on the cut angles of the resonator. The inertial coupling K_I basically depends on the shape and the boundary conditions. Thus the thermal derivatives of K_I depend on the thermal expansion and are much smaller than those of K_E, which depend on the elastic constants. On the other hand, one notices that K becomes negative if $K_I > K_E$.

K will be chosen as small as possible in order to minimize the sensitivity of the first thermal coefficient α to the dimensions (in the ZT-cut $K = \sigma_{12} = 0$).

On the other hand, for quartz resonators the second thermal coefficient β of an uncoupled (or free) mode is generally negative when the first coefficient α vanishes[9] (excepted for some thickness modes, such as the thickness shear of the AT-cut or SC-cuts). That means that the frequency of the compensated mode must be greater than that of the compensating mode in order to minimize the second thermal coefficient β.

RESONATOR STRUCTURE

FLEXURAL-EXTENSIONAL COUPLING

The structure of the FXT resonator has been analysed and optimized with a contour mode computer program based on variational techniques. Holland and Eernisse[7] have shown that Rayleigh-Ritz techniques can be used to approximate the normal modes of the mechanically free, rectangular isotropic plates. In order to apply this method to anisotropic plates of any shape, special techniques have been developed. The kinetic and elastic matrices as well as their thermal derivatives up to the third order are reduced to a tridiagonal form. This procedure precludes iterative methods to compute eigenmodes and their temperature coefficients. Furthermore the convergence of the process depends on the choice of trial functions. In most cases, it has been found that simple power series are more advantageous than circular expansions for describing displacements. Meanwhile a set of circular, hyperbolic and power series expansions are often suitable to describe different kinds of modes (flexure, shear, torsion, dilatation type modes).

The FXT resonator is composed of two rectangular plates connected by three parallel arms. The central arm includes a clamping area (Fig. 5) where stresses and movements are evanescent, so that it can be glued with epoxy onto a tiny pedestal. The two plates are used as mass loading in order to decrease the frequency of the desired contour mode. The thermal compensation is achieved by the mechanical coupling of the first two contour modes of the structure (Fig. 1), namely a flexure-type mode and an extensional-type mass-loaded mode.

The extensional mode is chosen as the thermally compensated mode, specially for its high quality factor and a simple electrode configuration.

The mechanical coupling factor K can be adjusted by the distance between the two external arms, the width ratio of the external arms to the central arm and by the width ratio of the mass-loaded plates to the central arm. K becomes negative if the first two parameters decrease or if the last one increases.

These three parameters can also, at the same time, minimize the coupling factor K over a wide range of the arms length (Fig. 2a), and the sensitivity due to the under-cutting (X faces) during the chemical etching.

The coupling factor K is little sensitive to the ratio of arms length to total length, as well as the frequency of the extensional mode, if this ratio is about 1/3 (Fig. 2a and Fig. 2b). This last parameter is used to adjust the frequency of the compensating mode (flexure-type Fig. 1a) in order to minimize the thermal dependence of the extentional mode (Fig. 1b).

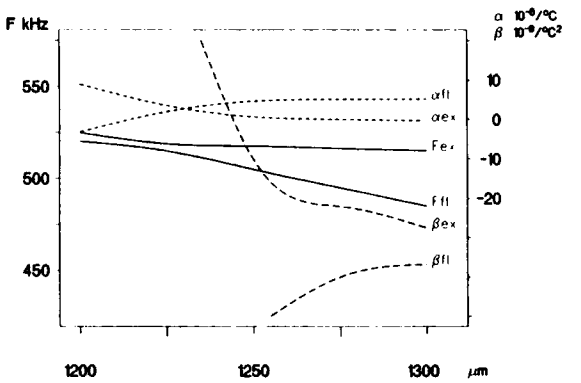


Figure 2a : Arms length dependance with K=0 KK'=0

The cut of a FXT resonator is chosen such that the first temperature coefficient α is close enough to zero without coupling in order to minimize its sensitivity to the dimensions.

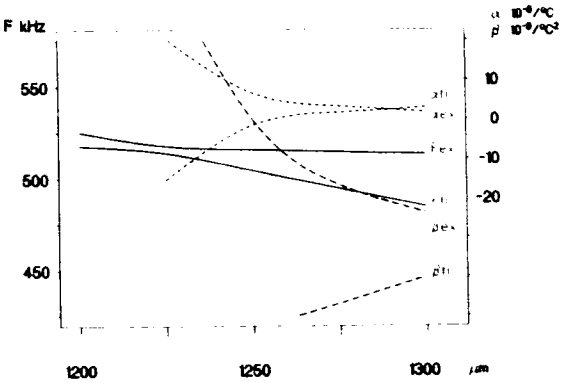


Figure 2b : Arms length dependance with K>0 KK'<0

CUT CHOICE

As far as almost any quartz crystal cut could be used, we realied the first FXT resonators with ZT-cut substrates. This cut allows a piezoelectric coupling with a perpendicular field (Fig. 3). The Poisson's ratio σ_{12} is almost zero. On the other hand the lateral faces are tilted through the chemical etching. Thus the contour modes are strongly coupled mechanically with unwanted out of plane modes.

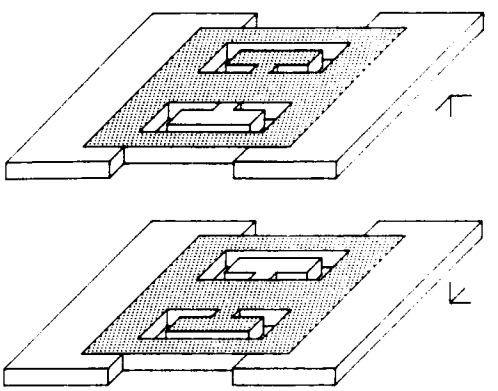


Figure 3 : Electrode configuration with perpendicular field

Afterwards we chose a simple cut, namely the +2° X-cut. This cut is well adapted to chemical etching and the lateral faces are not tilted. Nevertheless the two rectangular plate must be shifted by 10 μm in the X^+ direction in order to compensate the effect of the Poisson's ratio σ_{12} on the movement of the clamping zone. The piezoelectric coupling is obtained with an in-plane field (Fig 4).

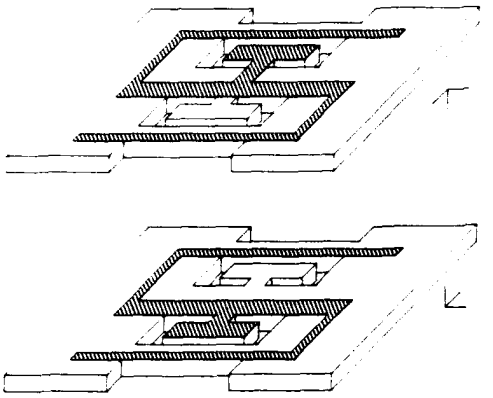


Figure 4 : Electrode configuration with in-plane field

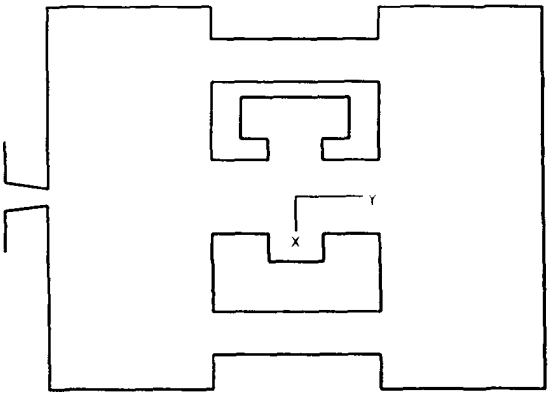


Figure 5 : Outline of the +2°X-cut

SEM OBSERVATIONS

Scanning electron microscopy has been used to observe the vibration patterns of the FXT structure. These patterns characterize the electric potential distribution onto the resonator surface due to the piezoelectric effect. They are dependent on the substrate cut, which governs the anisotropy and the distribution of the piezoelectric coefficients.

Figures 6 and 7 show the SEM micrographs of the flexural-type mode (compensating mode) and of the extensional mode (compensating mode) in a +2°X-cut FXT resonator. For this cut and on the two external arms the electric potential distribution is governed mainly by the piezoelectric coefficient d_{12}. This is pointed out by two white lines in the pattern of the extensional-type mode and the three white lines in the flexural-type pattern. The different brightnesses of the +Y and -Y regions are due to a symetry about the X axe of the piezoelectic matrix.

Figure 6 : SEM pattern of the flexure-type mode

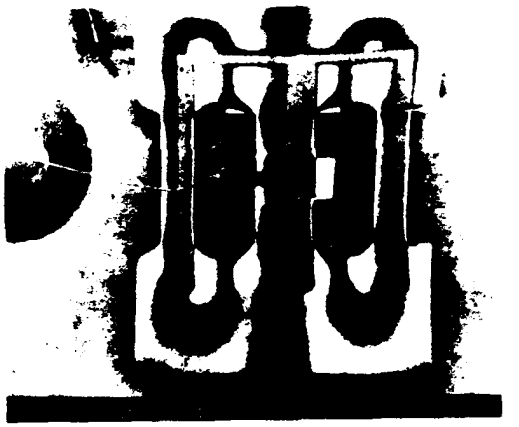


Figure 7 : SEM pattern of the dilatation-type mode

EXPERIMENTAL RESUTS

Cut	+2° X
Frequency	524 kHz
Length	3.3 mm
Width	2.75 mm
Thickness	0.1 to 0.2 mm
Quality factor	> 500'000
Extent	< 0.7
β	< 10 $10^{-9}/°C^2$
Aging at 25°C	< 0.2 10^{-6}/first year (without preaging)

CONCLUSIONS

The thermal properties of the miniature quartz resonator described in this paper are at least 4 times better than those of the flexural tuning-fork.

The structure with 3 arms minimizes the sensitivity to the chemical etching process, yields a high quality factor and a good aging.

ACKNOWLEDGMENTS

The author wishes to acknowledge many helpful discussions with J. Hermann. Many thanks also to C. Frossard and F. Crevoisier for their assistance in resonators fabrication.

This work has been partially funded by the Swiss watchmaking industry (CEH) and by the Commision pour l'encouragement des recherches scientifiques (CERS).

REFERENCES

[1] W.P. MASON, "A New Quartz-Crystal Plate, Designated the GT", Proc. IRE 1940, 220-223.

[2] H. KAWASHIMA et al, "New Frequency Characteristics of Miniaturized GT-Cut Quartz Resonators", Proc. 34th Annual Frequency Control Symp. 1980, 131-135.

[3] J. HERMANN and C. BOURGEOIS, "A New Quartz Crystal Cut for Contour Mode Resonators", Proc. 33th Annual Frequency Control Symp. 1979, 255-262.

[4] J. HERMANN, "A Novel Miniature ZT-Cut Resonator", Proc. 39th Annual Frequency Control Symp. 1985, 375-380.

[5] S. KOGURE et al, "New Type Twin Mode Resonator". Proc. 34th Annual Frequency Control Symp. 1980, 160-166.

[6] W.G. CADY, "Piezoelectricity", Dover Publications, 1962.

[7] R. HOLLAND and E.P. EERNISSE, "Contour Extensional Resonant Properties of Rectangular Piezoelectric Plates", IEEE Trans. Sonics and Ultrasonics, SU-15, 1968, 97-105.

[8] C. BOURGEOIS, "Decoupled Families of Contour Modes of Planar Thin Plates", Proc. 34th Annual Frequency Control Symp. 1980, 419-425.

[9] E.A. GERBER and A. BALLATO, "Precision Frequency Control", Academic Press, Inc. 1985.

TRAPPED-ENERGY PIEZOELECTRIC RESONATORS WITH ELLIPTICAL RING ELECTRODES

KIYOSHI NAKAMURA, RYOICHI YASUIKE*, KOICHI HIRAMA* and HIROSHI SHIMIZU**

Tohoku University, Sendai, Japan
*Toyo Communication Equipment Co.,Ltd., Kanagawa Pref., Japan
**The University of Electro-Communications, Chofu, Japan

Summary

This paper explores and investigates the trapped-energy piezoelectric resonators with elliptical ring electrodes conforming to the anisotropy in the piezoelectric plate. In this piezoelectric resonator, the electrode size can be arbitrarily large without causing inharmonic overtone resonances, and therefore the resonator has a low series resistance. Some resonators with elliptical ring electrodes have been fabricated using ceramic and AT-cut quartz plates. It has been shown that these resonators have series resistances lower than those of the conventional resonators.

1. Introduction

It is known that, if a trapped-energy thickness mode piezoelectric resonator is designed with the product of the electrode dimension and the square root of the plateback set below a certain value, then a single resonance characteristic with no inharmonic overtone response is obtained[1],[2].

Hence when the magnitude of frequency-lowering due to piezoelectric effect is large, or when overtone modes are to be used, it is necessary to obtain a single resonance characteristic. This requires either making the electrodes thin, or to reduce the lateral electrode dimensions. However, if the electrodes are made too thin, the resulting electrical resistance lowers the Q of the resonator; thus the lateral dimensions of the electrodes are limited to fairly small values. As a consequence, the shunt capacitance of the resonator cannot be made large, and so the impedance level and series resistance are both high. In applications for oscillators, often a low impedance resonator with a large electrode area is needed.

In this paper, we propose a resonator with electrodes in either circular or elliptical ring shapes, as an trapped-energy resonator of which electrode area can be increased arbitrarily while maintaining a single resonance characteristic. In order to confirm the feasibility of the proposed device, we have performed related analyses and experiments, and here report the results of these studies.

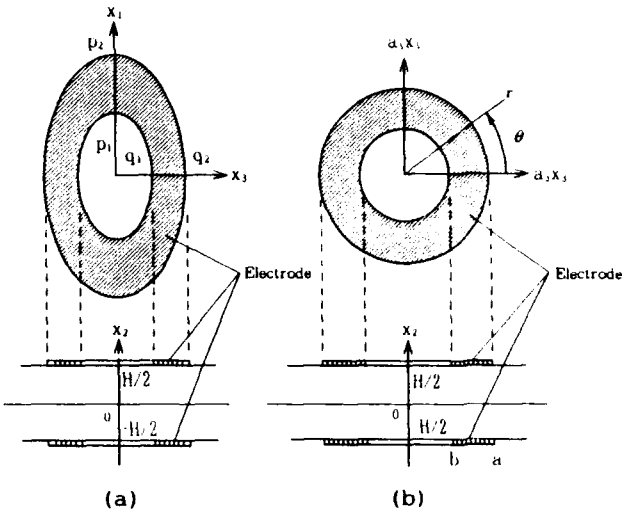


Figure 1. Electrode configuration and coordinate transformation.

2. Analyses

2.1 The wave equation and its solutions

We consider a trapped-energy thickness mode resonator with elliptical ring electrodes, such as that shown in Fig.1(a). The resonances of trapped-energy modes occur very close to the cut-off frequencies of thickness-mode waves propagating along the plate. The elestic property concerning wave propagation is generally different in the x_1 and x_3 directions in the plane of the plate. By using a potential-wave approximation the wave equation for thickness vibrations near this cut-off frequency may be written as follows[3]:

$$\frac{\partial^2 u}{\partial (a_1 x_1)^2} + \frac{\partial^2 u}{\partial (a_3 x_3)^2} + \frac{1}{v_s^2} \cdot (\omega^2 - \omega_c^2)\, u = 0 \qquad (1)$$

where u is the displacement, v_s is the velocity of the plane wave propagating in the thickness direction, and ω_c is the angular cut-off frequency which is equal to ω_0 at the electroded region and to $\omega_{0'}$ elesewhere. In eq.(1), the constants a_1 and a_3 vary depending on the materials of the piezoelectric plate, the cut of the crystal, and the mode of propagation. If the wave propagation

CH2818-3/90/0000-372 $1.00 © 1990 IEEE

is isotropic in the plane of the plate, then $a_1=a_3$.

As is clear from eq.(1), when the electrode dimensions p_1,p_2,q_1 and q_2 satisfy the following relation:

$$\left.\begin{aligned} a_1p_1&=a_3q_1 \\ a_1p_2&=a_3q_2 \end{aligned}\right\} \qquad (2)$$

then the extent of energy trapping in the x_1 direction may be considered to be the same as in the x_3 direction. Such an electrode shape enables us not only to analyze vibration easily but also to obtain analysis by transforming into polar coordinates[3]. There is also the added advantage that with such an electrode shape, a resonance characteristic free from spurious non-axisymmetric mode response is obtained.

For the resonator with the electrode shape given by eq.(2), we transform the (x_1,x_3) coordinate plane in Fig.1(a) into the plane (a_1x_1,a_3x_3) shown in Fig.1(b), and in turn transform this into the polar coordinate plane (r,θ). At these polar coordinates, the electrode shape is reduced to a circular ring, whose outer and inner radii, a and b, are given by

$$\left.\begin{aligned} a&=a_1p_1=a_3q_1 \\ b&=a_1p_2=a_3q_2 \end{aligned}\right\} \qquad (3)$$

As the result of these coordinate transformations, the wave equation (1) can be rewritten:

$$\frac{\partial^2 u}{\partial r^2}+\frac{1}{r}\frac{\partial u}{\partial r}+\frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2}+k_{r0}^2u=0 \qquad (4)$$

where k_{r0} is given by

$$\left.\begin{aligned} k_{r0}=k_r&=\frac{1}{v_s}(\omega^2-\omega_0^2)^{1/2} && (b<r<a) \\ k_{r0}=jk_r'&=j\frac{1}{v_s}(\omega_0'^2-\omega^2)^{1/2} && (r<b,\ r>a) \end{aligned}\right\} \qquad (5)$$

The solutions to eq.(3) are expressed as

$$\left.\begin{aligned} u_n&=(A_nJ_n(k_rr)+B_nY_n(k_rr))\cos n\theta\cdot\sin(x_2\omega_0/v_s)\cdot e^{j\omega t} && (b<r<a) \\ u_n&=C_nI_n(k_r'r)\cos n\theta\cdot\sin(x_2\omega_0'/v_s)\cdot e^{j\omega t} && (0\leqq r\leqq b) \\ u_n&=D_nK_n(k_r'r)\cos n\theta\cdot\sin(x_2\omega_0'/v_s)\cdot e^{j\omega t} && (a\leqq r) \end{aligned}\right\} \qquad (6)$$

where n is an integer expressing the circumferential order; J_n, Y_n are Bessel functions; I_n, K_n are the modified Bessel functions.

As $\sin(x_2\omega_0/v_s)\simeq\sin(x_2\omega_0'/v_s)$, continuity of u and $\partial u/\partial r$ at the boundary between the electroded and unelectroded regions gives the following equation.

$$\left.\begin{aligned} &\begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{21} & M_{22} & M_{23} & M_{24} \\ M_{31} & M_{32} & M_{33} & M_{34} \\ M_{41} & M_{42} & M_{43} & M_{44} \end{pmatrix}\begin{pmatrix} A_n \\ B_n \\ C_n \\ D_n \end{pmatrix}=0 \quad (i,j=1\sim4) \\ &M_{13}=M_{24}=M_{33}=M_{44}=0 \\ &M_{11}=J_n(k_ra) \\ &M_{12}=Y_n(k_ra) \\ &M_{14}=-K_n(k_r'a) \\ &M_{21}=J_n(k_rb) \\ &M_{22}=Y_n(k_rb) \\ &M_{23}=-I_n(k_r'b) \\ &M_{31}=k_r\{J_{n-1}(k_ra)-J_{n+1}(k_ra)\} \\ &M_{32}=k_r\{Y_{n-1}(k_ra)-Y_{n+1}(k_ra)\} \\ &M_{34}=k_r'\{K_{n-1}(k_r'a)+K_{n+1}(k_r'a)\} \\ &M_{41}=k_r\{J_{n-1}(k_rb)-J_{n+1}(k_rb)\} \\ &M_{42}=k_r\{Y_{n-1}(k_rb)-Y_{n+1}(k_rb)\} \\ &M_{43}=-k_r'\{I_{n-1}(k_r'b)+I_{n+1}(k_r'b)\} \end{aligned}\right\} \qquad (7)$$

The frequency equation for determining the resonance frequencies is obtained by setting the determinant $|M_{ij}|$ equal to zero. The vibration modes are found by determining the ratios among A_n,B_n,C_n and D_n from eq.(7), and substituting into eq.(6).

2.2 The resonance frequency

As explained below, vibration modes with a circumferential order equal to unity will not be excited piezoelectrically. Hence we here concider only the modes of vibration with axial symmetry in the polar coordinates (r,θ), that is, the modes with n=0. The resonance frequency spectrum obtained by numerical calculations is shown in Fig.2. In this figure, N is the order of the overtone, ω_r is the angular resonance frequency, and Δ is the plateback given by

$$\Delta=(\omega_0'-\omega_0)/\omega_0' \qquad (8)$$

This spectrum was calculated by varying a/H for N=1, b/H=7.5 and Δ =1%, but there is almost no change in the curve for other values of b/H and Δ. In the figure, "0th", "1st" and "2nd" indicate the principal, 1st inharmonic, and 2nd inharmonic modes, respectively.

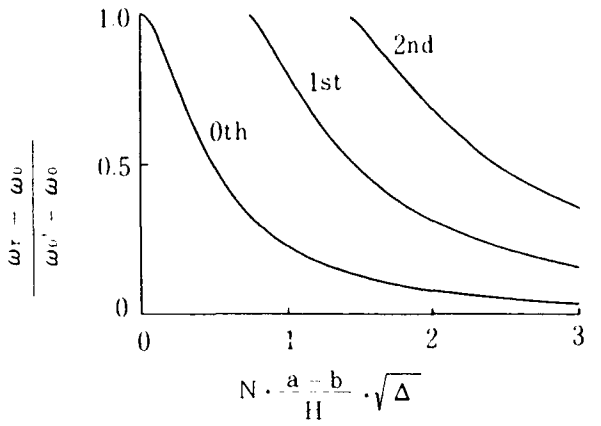


Figure 2. Resonance frequency spectrum (n=0).

Figure 3 shows the critical electrode ring widths at which first and second inharmonic modes begin to appear. We see that the advent of inharmonic modes is almost independent of the inner radius, and is determined by the ring width (a-b). Hence by choosing a ring width within the range where inharmonic modes do not appear, the inner radius of the electrode can be increased, resulting in a resonator with a single resonance characteristic and a low series resistance.

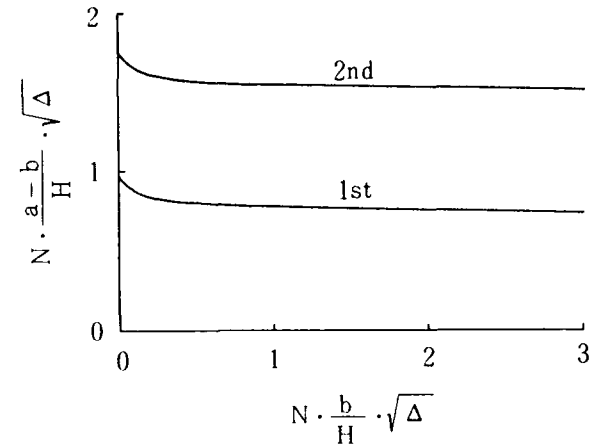


Figure 3. Critical ring width at which 1st and 2nd inharmonic modes begin to appear.

2.3 Vibration modes and capacitance ratios

Figure 4 shows the displacements of the 0th, 1st and 2nd modes, for Δ =0.1, a/H=13.2, b/H=7.5, N=1 and $(N\cdot((a-b)/H)\cdot\sqrt{\Delta}=1.8)$.

We next determine another important resonator parameter, the capacitance ratio. From the condition that the electrical energy stored in the motional series inductance L of the resonator is equal to the kinetic energy, we determine the inductance. This enables us to calculate the capacitance ratio, or the ratio of the shunt capacitance C_d to the motional series capacitance C in the equivalent circuit, as follows:

$$\frac{C_d}{C} = \frac{N^2}{k^2}\left(\frac{\pi\omega_r}{2\omega_0}\right)^2 \frac{\frac{2}{S}\int_0^{2\pi} d\theta \int_{-\infty}^{\infty} u_n^2 \cdot r\, dr}{\left(\frac{2}{S}\int_0^{2\pi} d\theta \int_b^a u_n \cdot r\, dr\right)^2} \qquad (9)$$

where k is the electromechanical coupling coefficient, and S is the electrode area. Figure 5 shows the capacitance ratios for the 0th and 1st modes as functions of $N\cdot((a-b)/H)\cdot\sqrt{\Delta}$. The

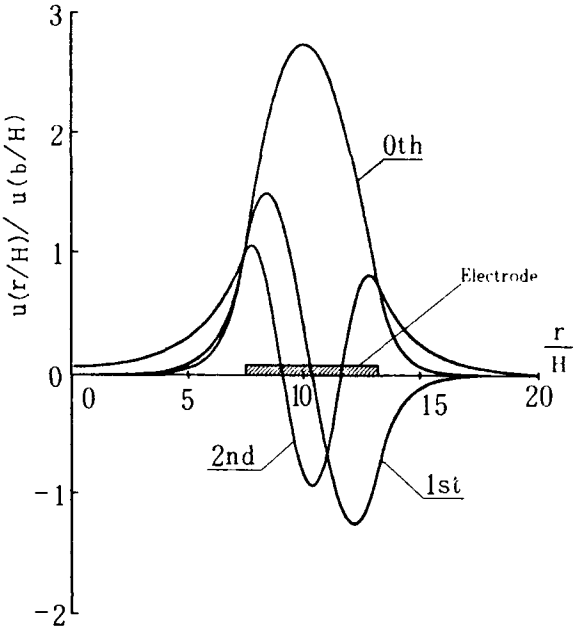


Figure 4. Vibrational modes.

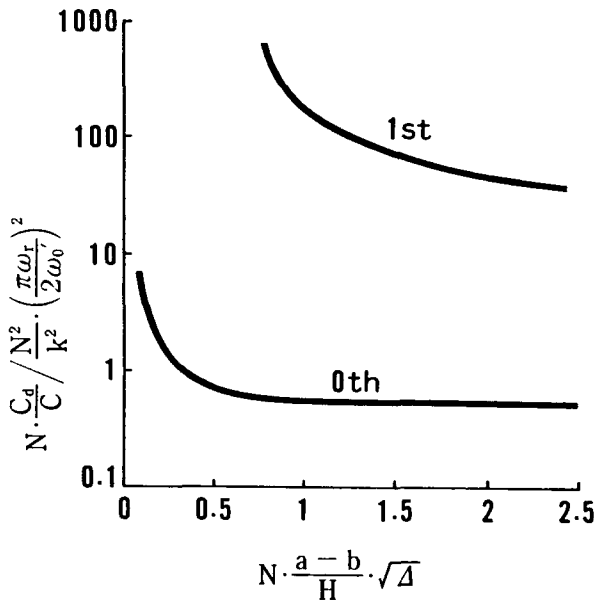


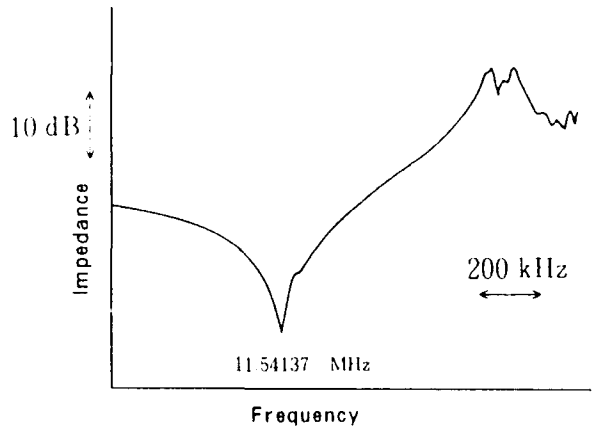
Figure 5. Normalized capacitance ratio as a function of $N\cdot((a-b)/H)\cdot\sqrt{\Delta}$

capacitance ratio for the 0th mode is sufficiently small if $N\cdot((a-b)/H)\cdot\sqrt{\Delta} > 0.5$. In contrast to this, that for the 1st inharmonic mode is roughly two orders of magnitude greater than for the 0th mode. This is due to the fact that the 1st mode is almost anti-symmetric with respect to the center of the width of the ring.

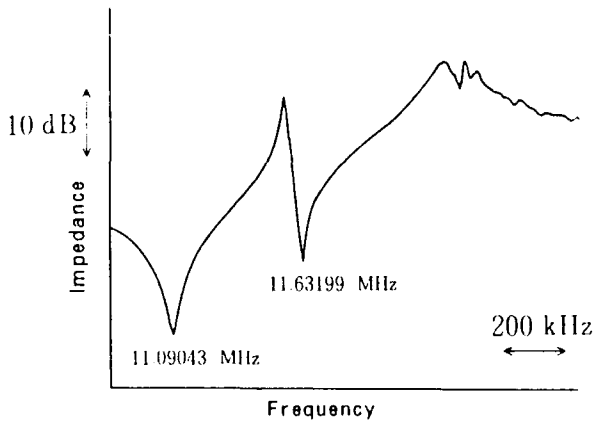
For all non-axisymmetric vibrations having $n \neq 0$, the integration in the denominator on the right-hand side of eq. (9) becomes zero, and the capacitance ratio becomes infinite. This means that such vibrations are not piezoelectrically excited. This is because the electrode shape is that of an elliptical ring corresponding to the anisotropy within the plane of the plate. If the electrode shape deviates from this elliptical shape, then non-axisymmetric vibrations will also be excited, as spurious modes.

3. Experiment

3.1 Thickness-extensional mode ceramic resonators



(a) Resonator with ring electrodes.



(b) Resonator with circular electrodes.

Figure 6. Impedance characteristics of fundamental thickness-extensional mode ceramic resonators.

In the case of thickness-extensional in a piezoelectric ceramic plate poled in the thickness direction, the elliptical ring shape of the electrodes reduce to a circular ring, because the plate is isotropic ($a_1=a_3$) in the plane of the plate. A fundamental thickness-extensional mode resonator with circular ring electrodes was fabricated. The impedance characteristic is shown in Fig.6(a). For comparison, the characteristic of a resonator with circular electrodes of the same area is shown in Fig.6(b). Design parameters for both resonators are given in Table 1.

An inharmonic mode appeared in the resonator with conventional circular electrodes, while no inharmonic mode appeared in the resonator with ring electrodes.

Table 1. Design parameters for ceramic resonators.

	Inner Radius (mm)	Outer Radius (mm)	Area (mm^2)
Resonator with Ring Electrodes	1.20	1.50	2.54
Resonator with Circular Electrodes	0	0.90	2.54

Plate Thickness 0.2 mm

Electrode Thickness Ag 3000 Å (Each Side)

3.2 Thickness-shear mode AT-cut quartz resonators

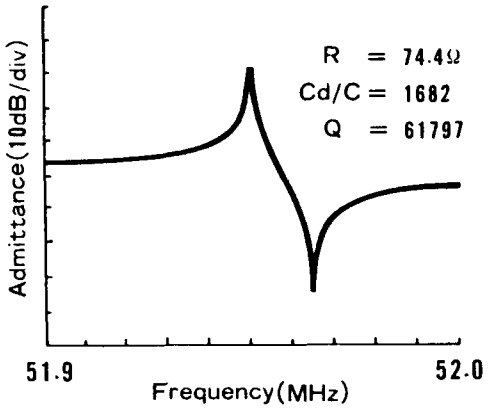
Third thickness-shear overtone resonators were fabricated using AT-cut quartz plates. For the third overtones in AT-cut quartz plates, $a_1=0.629$ and $a_3=0.649$. Hence, the elliptical ring electrodes were designed so that the ratio of major to minor axes might be equal to 1.031. For comparison, a resonator with circular electrodes was also fabricated. Both resonator electrodes were designed as large as possible, provided no inharmonic mode would appear. The design parameters are given in Table 2.

Table 2. Design parameters for AT-cut quartz resonators.

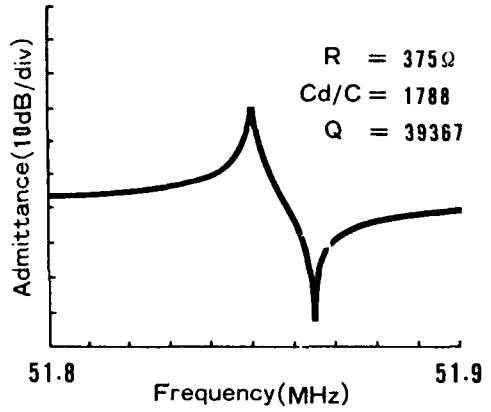
	Inner Radius (mm)		Outer Radius (mm)	
	x_1 Direction	x_3 Direction	x_1 Direction	x_3 Direction
Resonator with Elliptical Ring Electrodes	0.794	0.771	1.185	1.150
Resonator with Circular Electrodes	0	0	0.495	0.495

Plate Thickness 0.094 mm Plateback 1%

Figure 7 shows the admittance characteristics of both resonators. The resonator with elliptical ring electrodes had a fairly lower series resistance. This is attributed mainly to the larger electrode area of the resonator with elliptical ring electrodes.



(a) Resonator with elliptical ring electrodes.



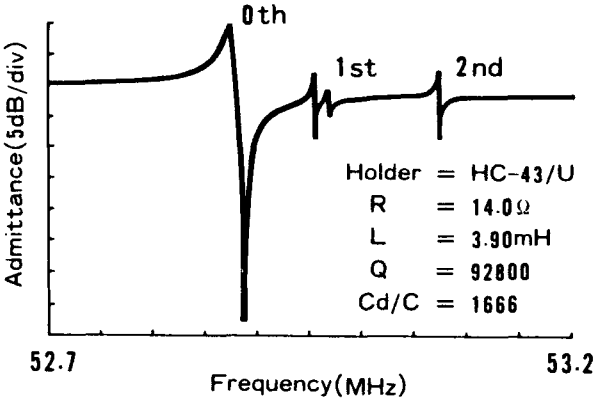
(b) Resonator with circular electrodes.

Figure 7. Admittance characteristics of 3rd-overtone AT-cut quartz resonators.

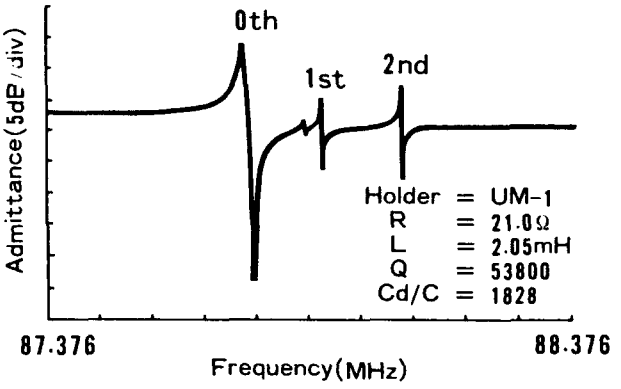
In order to obtain resonators with even smaller series resistences, we tried designing and fabricating third-overtone AT-cut quartz resonators, allowing the occurrence of inharmonic overtones to some extent. Three types of resonator were used in experiments, operating in the 52 MHz band and in the 87 MHz band and in the 150 MHz band. The electrode sizes were designed as large as possible, with the constraint that mounting losses and the influence of the plate edge on the characteristics be negligible. Table 3 shows the design parameters employed.

Table 3. Design parameters for resonators with low series resistances.

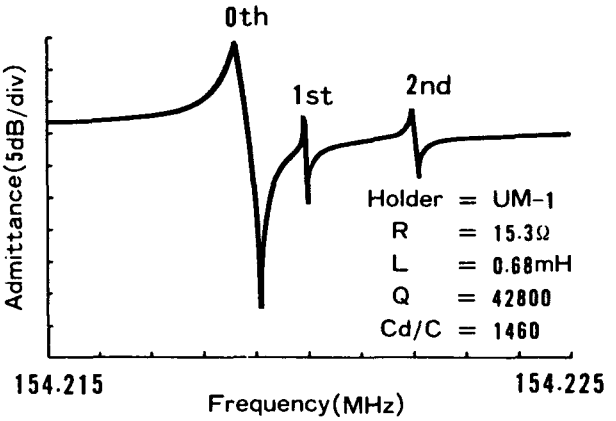
	Inner Radius (mm)		Outer Radius (mm)		Plate Dimension (mm)	Plateback (%)
	x_1 Direction	x_3 Direction	x_1 Direction	x_3 Direction		
52MHz Band Resonator	0.460	0.450	1.750	1.690	6.5φ×0.094	1.0
87MHz Band Resonator	0.300	0.291	1.200	1.163	5.0φ×0.054	1.0
150MHz Band Resonator	0.200	0.194	0.800	0.775	5.0φ×0.030	0.4



(a) 52MHz band resonator.



(b) 87MHz band resonator.



(c) 150MHz band resonator.

Figure 8. Admittance characteristics of 3rd-overtone AT-cut quartz resonators.

Figure 8 shows the admittance characteristics near the third overtone resonance frequency. The series resistances are fairly low for the main resonance, ranging from 14 to 21 ohm. Spurious 1st and 2nd inharmonic modes appeared in the characteristics of those resonators. The 150 MHz band resonator had fairly high resistances for the inharmonic modes, about 10 times as large as that of the main resonance.

The temperature dependence of frequency for the three types of quartz resonator described above was measured. It was confirmed that the temperature characteristic of frequency for the resonator with elliptical ring electrodes is a third-order curve with a point of inflection near room temperature, as in Fig.9. The third-order coefficient is of roughly the same magnitude as for ordinary AT-cut quartz plate resonators with circular electrodes.

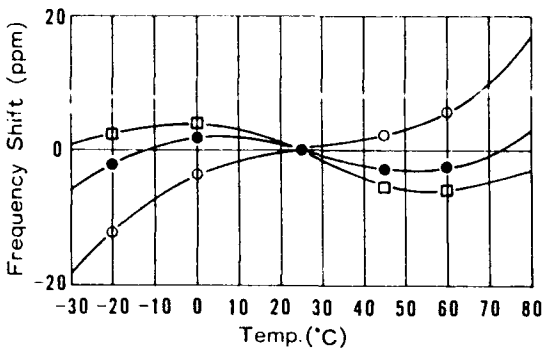


Figure 9. Frequency-temperature characteristics of 3rd-overtone AT-cut quartz resonators.

4. Conclusion

We have proposed the use of elliptical ring electrodes as one means of obtaining resonators with a low series resistance and a single resonance characteristic, and have analyzed the characteristics of such resonators.

The efficacy of the proposed elliptical ring electrodes has been verified experimentally for thickness extensional mode ceramic resonators, and for third-order thickness-shear mode AT-cut quartz resonators.

In the case of a third overtone resonator using a 5mmØ AT-cut quartz plate, whose electrode size has been set so as to allow the appearance of small inharmonic modes, the series resistance has been very low (15 ohm) for the main resonance and fairly high (> 150 ohm) for inharmonic modes.

Acknowledgements

The authors are grateful to Dr. H. Sekimoto of Tokyo Metropolitan Univ. for providing data of the anisotropy constants a_1, a_3 for the third-order thickness-shear mode in AT-cut quartz plates.

References

[1]W. Shockley, D. R. Curran and D. J. Konevals: "Energy trapping and related studies of multiple electrode crystals", Proc. 17th Ann. Freq. Control Symp. , p.88(1963).

[2]M. Onoe, H. Jumonji:"Analysis of Piezoelectric Resonators Vibrating in Trapped Energy Modes", J. Inst. Electron. Commun. Eng. Japan, vol. 48, no. 9, p.1574-1581, Sept. 1965.

[3]K. Nakamura and H. Shimizu:"Analyses of two-dimensional energy trapping in piezoelectric plates with rectangular electrodes", Proc. IEEE Ultrasonics Symposium, p.606(1976).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

VARIATIONAL ANALYSIS OF NEW SHAPE LENGTH EXTENSIONAL MODE QUARTZ CRYSTAL RESONATOR TAKING ACCOUNT OF LATERAL MOTION

Hirofumi KAWASHIMA and Mitsuhiro NAKAZATO

Seiko Electronic Components Ltd.
Hiraicho Tochigi shi 328, Japan

Abstract

This paper describes a new shape length extensional mode quartz crystal resonator with the resonant frequency of 500kHz to 1.536MHz, formed by an etching method. An object of this paper is, theoretically and experimentally, to clarify frequency characteristics, frequency temperature behavior and electrical characteristics of the new shape length extensional mode quartz crystal resonator. In analysis procedure, a vibration analysis including a vibrational portion and supporting portions of the new shape resonator is performed by an energy method, taking account of lateral motion, and its frequency equation is derived. From the obtained frequency equation, frequency characteristics and frequency temperature behavior versus a width-length ratio are, first, analyzed and then electrical characteristics are done, so that it is shown that the calculated results agrees well with the measured data. Finally, it is essentially understood that the present resonator is miniaturized, strong against shock and has small series resistance R_1 in the frequency range of 500kHz to 1.536MHz, because it is incorporated with the vibrational portion and the supporting portions by the etching method.

§ 1 Introduction

According to miniaturization and lightweight of various products such as portable products, communication equipment and consumer products, especially, in development of IC cards and pagers, the miniaturization of quartz crystal resonators which are used in the products is keenly required. However, the miniaturization by the conventional shape quartz crystal resonator [7] brings about large energy losses caused by vibration and weakness against shock, as things are, a quartz crystal resonator fully satisfying this requisite is not introduced. In addition, in manufacturing processes, automatic operation by introduction of a robot is being positively planned. Under these circumstances, a miniaturized quartz crystal resonator which is strong against shock and simultaneously crystal units suitable for surface mounting are, therefore, firmly required. In order to achieve these requirements, a new design method of a quartz crystal resonator including a supporting method is needed, because, by miniaturization of a resonator consisting of a vibrational portion and supporting portions, it can be, sufficiently, forecast that the supporting portions greatly influence the electrical characteristics, and also, a dimension omitted until now greatly influences the resonant frequency and the frequency temperature behavior.

This paper proposes a new shape length extensional mode quartz crystal resonator with resonant frequency of 500kHz to 1.536MHz and especially suitable for a surface mounting type. In addition, the resonant frequency, the frequecy temperature behavior and the electrical characteristics of the new shape resonator, taking into account lateral motion, are analyzed and examined. In analysis procedure, first, an equation of motion of the new shape length extensional mode quartz crystal resonator consisting of a vibrational portion and supporting portions is derived from an energy method. From the obtained equation of motion, the frequency equation is easily calculated. Next, the frequency characteristics, the frequency temperature behavior and the electrical characteristics of the present resonator are analyzed and compared with the measured values, so that it is shown that both results agree well. Let us describe concreatly below.

§ 2 Analysis Procedure

2.1 Equation of motion

Figure 1 shows a new shape length extensional mode quartz crystal resonator and its coordinate system. The present resonator consists of a vibrational portion and supporting portions connected at the center of the vibrational portion, furthermore, it is rotated with a rotation angle θ around x axis.

Figure 2 shows a modified shape of the new shape quartz crystal resonator illustrated in Fig. 1. As shown in Fig. 1, this resonator performs a vibration of extensional and compressional modes in the y axis direction by an electric field of the x axis direction, so called, length extensional mode vibra-

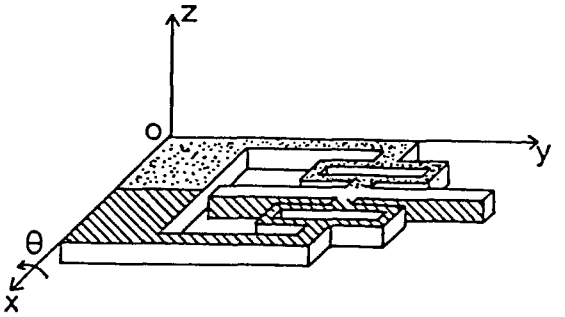


Fig. 1 New shape length extensional mode quartz crystal resonator and its coordinate system.

CH2818-3/90/0000-378 $1.00 © 1990 IEEE

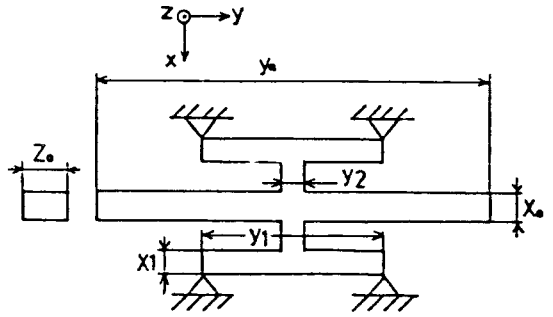


Fig. 2 Modified shape of the new shape quartz crystal resonator.

tion. According to this vibration, it also expands and contracts in the x and z axes direction. Therefore, it is conceivable that the resonator, as shown in Fig. 2, has the supporting portions vibrating in a flexural mode with each mass m at both ends of the x axis direction of the vibrational portion and also with the boundary conditions of "both hinged ends" [1][2]. That is, the supporting portions each have mass m and resistance to a vibration of the vibrational portion (work done by external force). Let us take width x_0, length y_0 and thickness z_0 in a dimension of the vibrational portion for this modified shape, which correspond to crystal axes x, y and z of quartz, respectively. Furthermore, take width x_1, length y_1 and thickness z_0 in a dimension of the supporting portions and the center of the vibrational portion to be the original point o. Now, when taking kinetic energy K_1, K_2 *of the vibrational portion* and one supporting portion, potential energy U_1 of the vibrational portion, and work W_2 done by external force of one supporting portion, the total kinetic energy $\hat{K}$ and the total potential energy $\hat{U}$ are as follows:

$$\hat{K}=K_1+K_2\big|_{+x_0/2}+K_2\big|_{-x_0/2} \tag{1}$$

$$\hat{U}=U_1 \tag{2}$$

Therefore, the Lagrangian L is obtained:

$$L=\hat{K}-\hat{U} \tag{3}$$

In addition, since resistance caused by the supporting portions can be considered as external force operated on the vibrational portion, the total work $\hat{W}$ done by the external force is expressed as follows:

$$\hat{W}=W_2\big|_{+x_0/2}+W_2\big|_{-x_0/2} \tag{4}$$

Therefore, when taking time t, from the variational theorem, the following relation is given:

$$\delta\int_{t_0}^{t_1}L\,dt+\int_{t_0}^{t_1}\delta\hat{W}\,dt=0 \tag{5}$$

Accordingly, an equation of motion is derived from equation (5). Next, let us describe its derivatives procedure in detail. Since the resonator in Fig. 1 vibrates in a length extensional mode by an electric field of the x axis direction and actually the electric field is only applied in the x axis direction, from the piezoelectric nature, stresses T_3, T_5, T_6 are satisfied with

$$T_3=T_5=T_6=0 \tag{6}$$

In addition, when thickness z_0 of the resonator is much smaller than length y_0 ($z_0 \ll y_0$), stress T_4 is given as

$$T_4=0 \tag{7}$$

Quite similarly, when width x_0 is sufficiently small compared with length y_0, stress T_1 becomes even smaller than stress T_2, and besides elastic compliance constant s_{11} is much larger than another elastic compliance constants s_{21}, s_{31}, s_{41}, as a result, the following relation is obtained:

$$s_{11}T_1\neq 0,\quad s_{21}T_1=s_{31}T_1=s_{41}T_1=0 \tag{8}$$

Therefore, strains S_i ($i=1\sim4$) which are given as a function of stresses T_1, T_2 and elastic compliance constants s_{11}, s_{12}, s_{22}, s_{32}, s_{12}, are expressed by

$$\left.\begin{aligned} S_1&=s_{11}T_1+s_{12}T_2\\ S_2&=s_{22}T_2\\ S_3&=s_{32}T_2\\ S_4&=s_{42}T_2 \end{aligned}\right\} \tag{9}$$

Now, when displacements u, v, w are taken in accordance with the x, y and z axes direction, first, derive stress T_2 from the second term of equation (9), next, calculate each displacement u, v, w by substituting this T_2 into the first, third and fourth terms of equation (9), moreover, taking into *consideration a* summation of displacements by length extensional mode and shear force to be $\bar{v}$, the displacements u, $\bar{v}$, w are given as

$$\left.\begin{aligned} u&=-(1+c_x)\sigma_x x\frac{\partial v}{\partial y}\\ \bar{v}&=v-\sigma_{yz}z\frac{\partial v}{\partial y}\\ w&=-\sigma_z z\frac{\partial v}{\partial y} \end{aligned}\right\} \tag{10}$$

Where c_x is a correction factor showing displacement quantity increased in the x axis direction, here, which is determined by an analysis of the finite element method. Namely, it is conceivable that this makes a Poisson's ratio large outwardly. However, it goes without saying that c_x becomes zero, when $T_1=0$ and there is nothing of supporting portion. In addition, the signs of σ_x, σ_{yz} and σ_z represent Poisson's ratioes, respectively, and each sign is given as $\sigma_x=-s_{12}/s_{22}$, $\sigma_{yz}=-s_{12}/s_{22}$ and $\sigma_z=-s_{32}/s_{22}$.

Next, by employing the displacements in equation (10) the total kinetic energy $\hat{K}$ is calculated as follows:

$$\hat{K}=\frac{1}{2}\rho A\int_{-y_0/2}^{y_0/2}\{(\frac{\partial v}{\partial t})^2+k^2(\frac{\partial^2 v}{\partial t\partial y})^2\}\,dy$$
$$+m(1+c_x)^2\sigma_x^2\frac{x_0^2}{4}(\frac{\partial^2 v}{\partial t\partial y})^2\Big|_{y=\pm y_2/2\fallingdotseq 0} \tag{11}$$

While taking approximately $T_1=0$ ($y_0 \gg x_0$) into account, the total potential energy $\hat{U}$ is given as

$$\hat{U}=\frac{1}{2} A\int_{-y_0/2}^{y_0/2}\gamma_{22}\left(\frac{\partial v}{\partial y}\right)^2 dy \qquad (\because c_{23}=c_{32}) \quad (12)$$

Where each sign shows the sectional area A ($=x_0z_0$), the density ρ and width y_2 of a connecting portion, k and γ_{22} are, respectively, expressed by

$$k^2 = \{(1+c_x)^2\sigma_x^2 x_0^2+(\sigma_y^2+\sigma_z^2)z_0^2\}/12$$

$$\gamma_{22}=(c_{22}-c_{23}^2/c_{33})=1/s_{22}$$

Next, when taking a displacement of the supporting portion vibrating in a flexural mode to be u', the work W_2 done by external force of the supporting portion is given by the following relationship:

$$W_2=\frac{1}{2} E I\int_{-y_1/2}^{y_1/2}\left(\frac{\partial^2 u'}{\partial y^2}\right)^2 dy \qquad (13)$$

Where E is young's modulus, and I is the moment of inertia. Furthermore, since the operating portion in the flexural mode can be considered with the boundary conditions of " both hinged ends " at both ends, the following relationship is obtained:

$$u'=0 \quad ; y=\pm y_1/2,\ x=\pm x_0/2$$

$$\frac{\partial^2 u'}{\partial y^2}=0 \qquad (14)$$

Therefore, from equation (4), the displacement u' is calculated as follows:

$$u'=\pm u_0\cos\frac{\pi}{y_1} y\cos\omega' t\Big|_{x=\pm x_0/2} \qquad (15)$$

Where u_0 is the amplitude, and ω' is the angular frequency. Accordingly, the external force $F|_{\pm x_0/2}$ is calculated from the below equation using the displacement u':

$$F|_{\pm x_0/2}=-\int_{-y_1/2}^{y_1/2} E I\left(\frac{\partial^4 u'}{\partial y^4}\right) dy$$

$$=\mp 2u_0 E I\left(\frac{\pi}{y_1}\right)^3\cos\omega' t \qquad (16)$$

The force given in equation (16) is the force at $x=\pm x_0/2$ and $|y|\leqq y_2/2\fallingdotseq 0$, and it operates on the vibrational portion in the y axis direction as resistance. Furthermore, since the force operating at $\pm x_0/2$ is the same, this force can be considered as the force per unit volume $\overline{F}\pm$ operating on the vibrational portion. Accordingly, the following relationship is obtained:

$$\overline{F}\pm=\frac{2F|_{\pm x_0/2}}{V_1} \qquad (17)$$

Where $V_1=x_0y_0z_0$.

Substituting equations (11), (12) and (13) into equation (5), since δv is arbitrary under the condition that the vibrational portion and the supporting portions are connected at $x=\pm x_0/2$ and $y\fallingdotseq 0$, taking into account that variation $\delta u'$ of the displacement u' is approximately expressed by $\delta u'=\nu\,\delta v$ {$\because \nu=(1+c_x)\sigma_x x_0/2y_0$}, the equation of motion is given as follows:

$$\rho\left(\frac{\partial^2 v}{\partial t^2}-k^2\frac{\partial^4 v}{\partial y^2\partial t^2}\right)-\gamma_{22}\frac{\partial^2 v}{\partial y^2}\pm\hat{F}\pm=0 \qquad (18)$$

Where $\hat{F}\pm-\nu\overline{F}\pm$

In addition, the boundary conditions are obtained quite similarly from equation (5):

$$m(1+c_x)^2\sigma_x^2\frac{x_0^2}{4}\frac{\partial^4 v}{\partial t^2\partial y^2}\Big|_{y=\pm y_2/2}\fallingdotseq 0$$

$$=\mp\gamma_{22}A\frac{\partial v}{\partial y}\Big|_{y=\pm y_0/2} \qquad (19)$$

$$v=0 \quad ; y=0$$

The equation of motion in equation (18) can be solved under the boundary conditions of equation(19). For this resonator, it is so designed that the resistance $\hat{F}\pm$ which is brought about by the supporting portions, hardly suppresses the vibration of the length extensional mode, namely, becomes very small.

2.2 Frequency equation

Next, let us solve the equation of motion for equation (18). When amplitude v_0, a constant ξ which is determined by the boundary conditions and angular frequency ω are taken, the displacement v from the second term in equation (19) can be written as follows, including a time expression:

$$v=v_0\sin\xi y\cos\omega t \qquad (20)$$

In addition, since the vibrational portion and the supporting portions are connected at $x=\pm x_0/2$ and $|y|\leqq y_2/2\fallingdotseq 0$, $\omega=\omega'$ from eqations (15) and (20). Therefore, $\hat{F}\pm$ in equation (18) is given as the following form, namely,

$$\hat{F}\pm=\pm F_0\cos\omega t \qquad (21)$$

Where F_0 has the following relation:

$$F_0=-4\frac{\nu u_0}{V_1} E I\left(\frac{\pi}{y_1}\right)^3$$

The resistance of equation (21) shows elongation of the vibrational portion produced by force $\hat{F}\pm$. When its quantity is taken as $v_1\cos\omega t$, including a time expression, v_1 is given as $F_0/\rho\omega^2$. Hence, when the total displacement v including the supporting portions is newly taken, the displacement v is given as

$$v=(v_1+v_0\sin\xi y)\cos\omega t \qquad (22)$$

Substituting equation (22) into equation (18) and after the arrangement, the resonant frequency f taken the lateral motion into account is derived:

$$f=\frac{n}{2y_0}\sqrt{\frac{\gamma_{22}}{\rho}}\frac{1}{\sqrt{1+k^2(n\pi/y_0)^2}} \qquad (23)$$

Where the sign n ($\xi=n\pi/y_0$) is determined by the boundary conditions and it is given as a root of equation (24) obtained from equation (19).

$$\cos\frac{n\pi}{2} = -\mu n^3 \sin\frac{n\pi}{2y_0} y_2 \qquad (24)$$

Where $$\mu = \frac{\pi^3}{4}\frac{m}{M}(1+c_x)^2\sigma_x^2\frac{\omega^2}{\omega_1^2}\left(\frac{x_0}{y_0}\right)^2$$

M :Mass of vibrational portion
ω, ω' :Angular frequencies when taking into account and omitting lateral motion

Inaddition of the supporting portions gives $\mu = 0$, while addition of the supporting portions gives rise to resonant frequency somewhat higher than that of its inaddition, because μ has a positive value by adding the supporting portions. For example, however, μ has a very small value of 2.8×10^{-4} under the condition of resonant frequency f=1.0MHz (y_2=0.1mm), so that frequency deviation is approximately several ten ppm and so very small as to neglect it. Therefore, the constant n can be regarded as a positive value of 1, 3, 5..... approximately, even if the supporting portions are added.

§ 3 Results and Discussion

First of all, a relationship of frequency constant ($f \cdot y_0$) versus a ratio Rxy($=x_0/y_0$) of width x_0 and length y_0 is shown from equation (23). Next, calculating the first and second order temperature coefficients α, β versus Rxy and Rzy (thickness z_0 to length y_0 ratio), from this, a relationship of a turn over temperature point T.p versus Rxy and Rzy is shown in comparison with the measured data. Furthermore, a relationship of a turn over temperature point T.p versus a cut angle θ is shown and simultaneously, an example of frequency temperature characteristics is illustrated. Next, a relationship of motional inductance L_1 versus width x_0 and length y_0 is exhibited. In addition, a capacitance ratio r which is defined as the ratio of shunt capacitance C_0 and motional capacitance C_1, is shown in a relation of a cut angle θ. Finally, a relationship of series resistance R_1 and a quality factor Q versus resonant frequency f is shown. Besides, elastic compliance constants [4] transformed from elastic stiffness constants of reference [3] are employed in an analysis of frequency temperature characteristics for this paper. Where the first and second order temperature coefficients $s_{11}'/s_{11}=7.39\times10^{-6}/℃$, $s_{11}''/s_{11}=78.8\times10^{-9}/℃^2$ ($' \equiv \partial/\partial T$, T: Temperature) of s_{11} are employed in this calculation, because the temperature coefficients of c_{ij} have basically a physical error and taking this fact into account.

3.1 Frequency constant

Figure 3 shows a relationship between frequency constant ($f \cdot y_0$) and Rxy ($=x_0/y_0$), which is given as the ratio of width x_0 and length y_0, of a length extensional mode quartz crystal resonator with the cut angle $\theta=0°00'$, length $y_0=2.50$mm (f=1.1MHz) of the vibrational portion and thickness $z_0=160\mu$m. That is, versus a change of width x_0 of the vibrational portion. As is apparent from equation (23), according to increase of the dimensional ratio Rxy, the frequency constant ($f \cdot y_0$) becomes somewhat small and this tendency in this calculation agree well with the measured values. However, this quantity of change, as shown

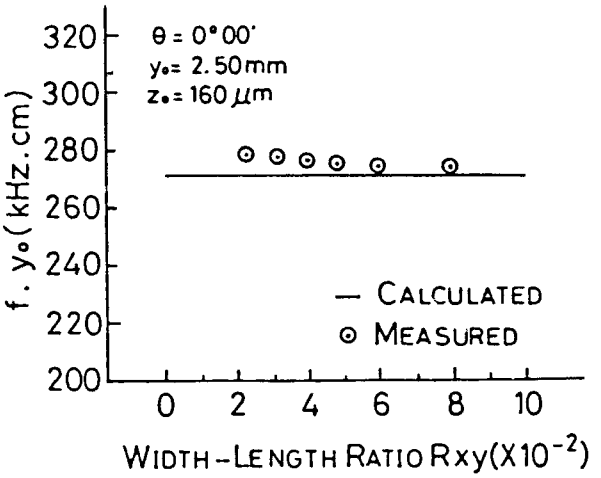


Fig. 3 Relationship between frequency constant ($f \cdot y_0$) and width-length ratio Rxy.

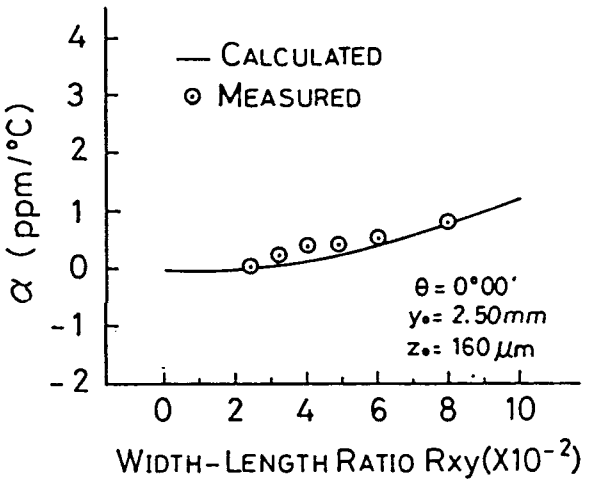


Fig. 4 Relationship between first order temperature coefficient α and width-length ratio Rxy.

in Fig. 3, is so extremely small as to neglect it. In addition, with respect to an absolute value of the frequency constant ($f \cdot y_0$), for example, when Rxy=3.2×10^{-2} ($x_0=80\mu$m), the frequency constant ($f \cdot y_0$) has a value of 272kHz・cm in the calculation, while it has a value of 278kHz・cm in the experiments. Both results comparatively agree well, even if there occurs an error about 2.2 per cent between them. In particular, the error between in the experiments and in the calculation shows a tendency to increase a little bit in the area of small Rxy, it is conceivable for the reason that etching-remainder remained in the length direction by chemical etching greatly affects it.

3.2 Frequency temperature coefficients and turn over temperature point

Figure 4 shows a relationship between the first order temperature coefficient α and the dimensional ratio Rxy of the present quartz crystal reso-

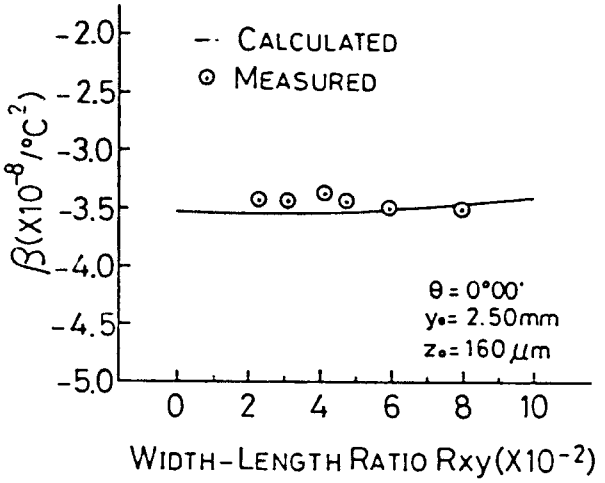


Fig. 5 Relationship between second order temperature coefficient β and width-length ratio Rxy.

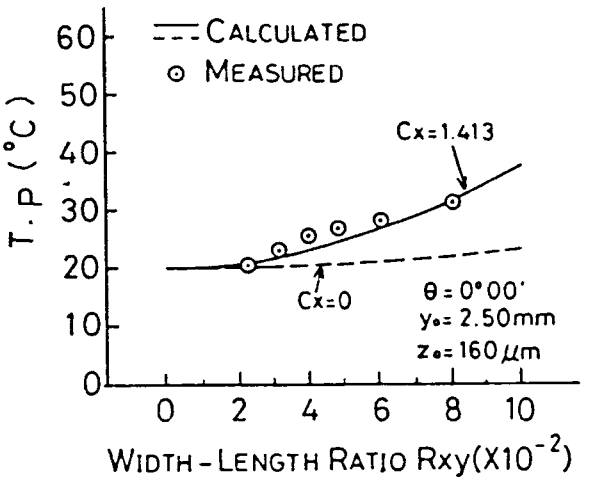


Fig. 6 Relationship between turn over temperature point T.p and width-length ratio Rxy.

nator with the cut angle $\theta = 0°00'$, length $y_0 = 2.50$mm ($f = 1.1$MHz) of the vibrational portion, thickness $z_0 = 160\mu$m and $c_x = 1.413$. In this case, also, versus a change of width x_0. As the dimensional ratio Rxy is increased, α also becomes large. For this result, it is suggested that the turn over temperature point T.p shifts by a change of width x_0. The solid line represents the calculated values, while the sign ⊙ represents the experimental data, both results are in good agreement.

Figure 5 shows a relationship between the second order temperature coefficient β and the dimensional ratio Rxy of a resonator with the same cut angle and dimensions as the resonator of fig. 4. As is obvious from Fig. 5, β versus Rxy almost shows the constant value in both the calculated values and the measured ones, which value gives approximately $-3.5\times10^{-8}/°C^2$. Both results agree very well.

Figure 6 shows a relationship between the turn over temperature point T.p and the dimensional ratio Rxy of a resonator with the same cut angle and dimensions as the resonator of Fig. 4. Namely, T.p is calculated from α in Fig. 4 and β in Fig. 5 computed under the condition of $c_x = 1.413$. As mentioned above, according to increase of Rxy, α also increases, then, since β almost shows the constant value, T.p becomes high according to increase of Rxy. That is, according to the change of Rxy $= 2.4\times10^{-2}$ ($x_0 = 60\mu$m) to 8.0×10^{-2} ($x_0 = 200\mu$m), T.p raises 9.9℃ in the calculation and 10.8℃ in the experiments. The calculated values shown by the solid line and the experimental ones by the signs ⊙, as shown in Fig. 6, are in good agreement, while the broken line is calculated in the case of $c_x = 0$, in other word, under the conditions of $T_1 = 0$ and no existing supporting portions. In this case, the variational quantity of T.p versus a change of Rxy is extremely small, a big difference between the calculated results and the measured values of the present resonator takes place. Judging from this result, it is easily understood that the displacement of the width direction caused by stress T_1 and the supporting portions added, is not so small as to neglect it.

Figure 7 shows a relationship between the turn over temperature point T.p and a dimensional ratio

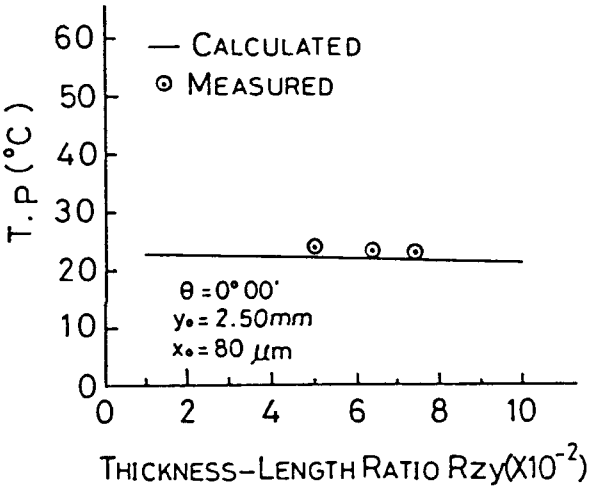


Fig. 7 Relationship between turn over temperature point T.p and thickness-length ratio Rzy.

Rzy, which is expressed by the ratio of thickness z_0 and length y_0, of a resonator with the cut angle $\theta = 0°00'$, length $y_0 = 2.50$mm ($f = 1.1$MHz) of the vibrational portion, width $x_0 = 80\mu$m and $c_x = 1.413$. As the dimensional ratio Rzy is increased, T.p exhibits a tendency to lower. For example, as the dimensional ratio Rzy has an increasing value such as 5.0×10^{-2} ($z_0 = 125\mu$m), 6.4×10^{-2} ($z_0 = 160\mu$m) and 7.4×10^{-2} ($z_0 = 185\mu$m) in the experiments, the T.p has conversely a decreasing value such as 24.3, 23.2 and 23.1℃. This tendency agrees well with the calculated values. Furthermore, the variational quantity of T.p versus the dimensional ratio Rzy is smaller than that of T.p versus the dimensional ratio Rxy. Thus, it is substantially understood that a turn over temperature point T.p varies, although there exists its large and small quantity of variation by changing a dimensional ratio of the length extensional mode quartz crystal resonator. This is due to the fact that deformation dependent upon Poisson's ratioes and so on of the width x_0

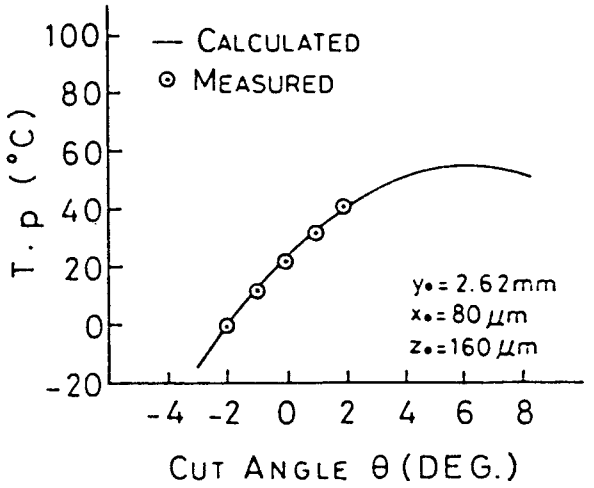


Fig. 8 Relationship of turn over temperature point T.p to cut angle θ.

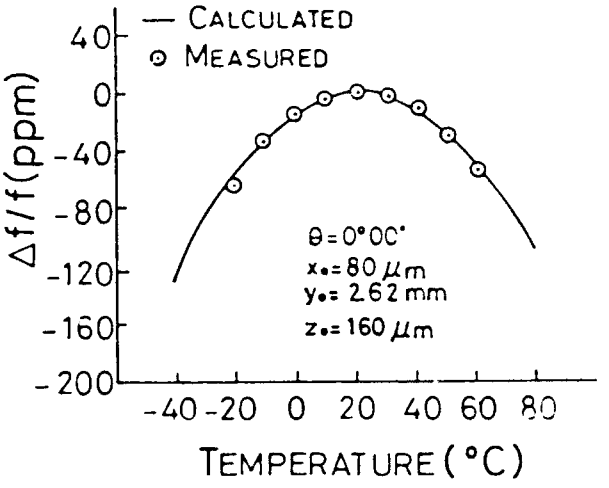


Fig. 9 Frequency temperature characteristics.

and thickness z_0 directions is generated by following the vibration in the length y_0 direction.

Figure 8 shows a relationship between the turn over temperature point and a cut angle θ of a resonator with length $y_0=2.62$mm ($f=1.056$MHz), width $x_0=80\,\mu$m and thickness $z_0=160\,\mu$m. As the cut angle θ is increased, T.p rises and reaches to the maximum value at the cut angle θ of about 6.3°C, and as the cut angle θ is further increased, T.p lowers. The signs ○ are the measured values, which exhibits the values at the cut angle θ of -2° to +2°. It is essentially understood from Fig.8 that the calculated results are in very good agreement with the measured data, and also that the cut angle θ for the present resonator is to be in close vicinity to zero degree so as to establish T.p in the vicinity of ambient temperature.

3.3 Frequency temperature characteristics

Figure 9 shows the frequency temperature characteristics of a resonator with the cut angle $\theta=0\,°00'$, length $y_0=2.62$mm($f=1.056$MHz), width $x_0=80\,\mu$m ($Rxy=3.05\times10^{-2}$) and thickness $z_0=160\,\mu$m ($Rzy=6.11\times10^{-2}$). T.p shows 21.8°C in the calculation, while it shows ??.8°C in the measured values, these results sufficiently agree well. In addition, the second order temperature coefficient β has a value of -3.50×10^{-8}/°C^2 in the calculation, while it has a value of -3.52×10^{-8}/°C^2 in the measured values, both results are in very good agreement. Hence, as shown in Fig.9, even though being apart from the turn over temperature point gradually, the measured values agree well with the calculated ones.

Figure 10 shows another frequency temperature characteristics of a resonator with the cut angle $\theta=0\,°00'$, length $y_0=2.50$mm($f=1.1$MHz) and thickness $z_0=160\,\mu$m, taking width x_0 as a parameter. It has been shown in Fig.6 that the change of width x_0 causes the shift of the turn over temperature point T.p. In case of the present resonator, width $x_0=60\,\mu$m gives T.p=21.1°C in the calculation and T.p=20.8°C in the measured data, while width $x_0=200\,\mu$m increased gives T.p=31.0°C in the calculation and T.p=31.6°C in the measured values. From these results, T.p rises 9.9°C in

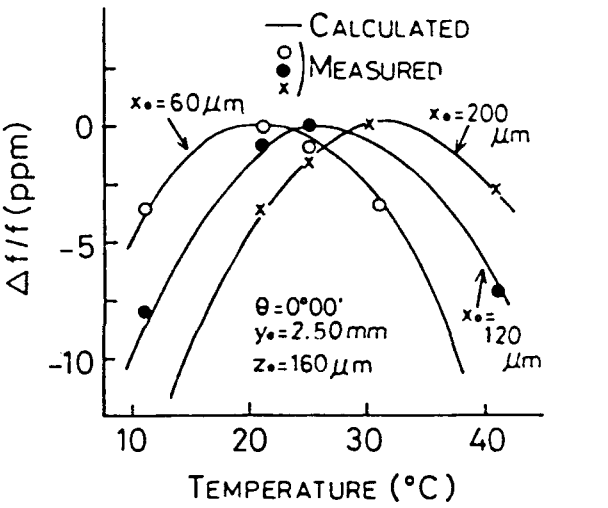


Fig. 10 Frequency temperature characteristics when taking width x_0 as a parameter.

the calculation and 10.8°C in the measured values for the change of width $x_0=60\,\mu$m to $200\,\mu$m. These calculated and measured results are also in close agreement. Thus, it is easily understood that the turn over temperature point T.p of the length extensional mode quartz crystal resonator extremely miniaturized, remarkably shifts by not only the cut angle θ, but also width x_0. Accordingly, T.p can be established to an arbitrary temperature by combining the cut angle θ and particularly, the dimensional ratio Rxy (x_0/y_0).

3.4 Motional inductance L_1

Motional inductance L_1 can be calculated by taking kinetic energy to be equal to electromagnetic energy in the equivalent circuit [5]. Therefore, taking into account that the displacements u, w very dependent upon a Poisson's ratio is so small as to be able to disregard it, compared with the displacement v, the following relationship is given as

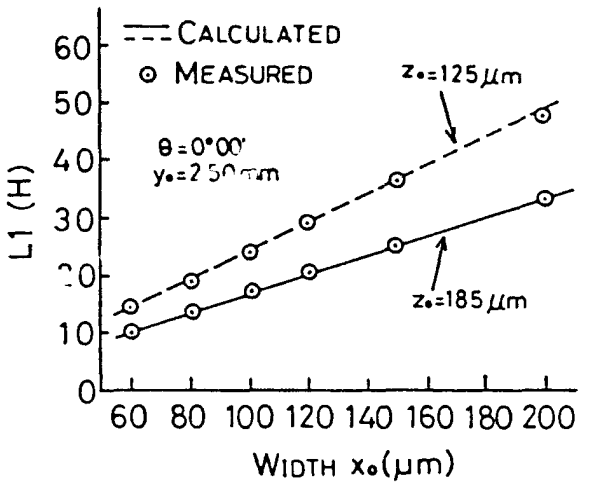


Fig. 11 Motional inductance L_1 to width x_0 when taking thickness z_0 as a parameter.

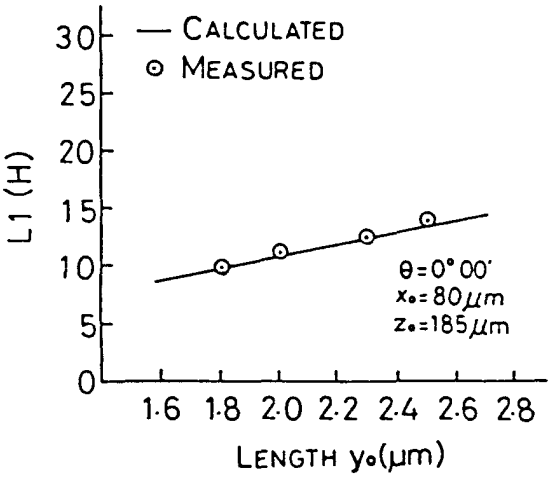


Fig. 12 Motional inductance L_1 to length y_0.

$$L_1 = \frac{\rho \int_{V_1} \left(\frac{\partial v}{\partial t}\right)^2 dv}{e_{12}^2 \left\{ \frac{\partial}{\partial t} \int_{S_e} (\partial v/\partial y)\, dS \right\}^2} \qquad (25)$$

Where density ρ of quartz, volume V_1 of the vibrational portion, piezoelectric constant e_{12} [6] and excitation electrode area Se. Now, let us take the center of the vibrational portion to be an original point o, and when the electrodes for excitation are disposed on the entire surfaces of the vibrational portion parallel to the z axis, not including the surfaces perpendicular to the y axis and, also, its position which are disposed in the y axis direction, have a value of $y_b/2$ and $y_e/2$, as a result of calculation, L_1 is given as

$$L_1 = \frac{1}{8} \frac{\rho}{e_{12}^2} \frac{x_0 y_0}{z_0} \frac{1}{G} \qquad (26)$$

Where G is a constant determined by excitation electrode area and expressed as

$$G = \{\sin(\pi y_e/2y_0) - \sin(\pi y_b/2y_0)\}^2 \qquad (27)$$

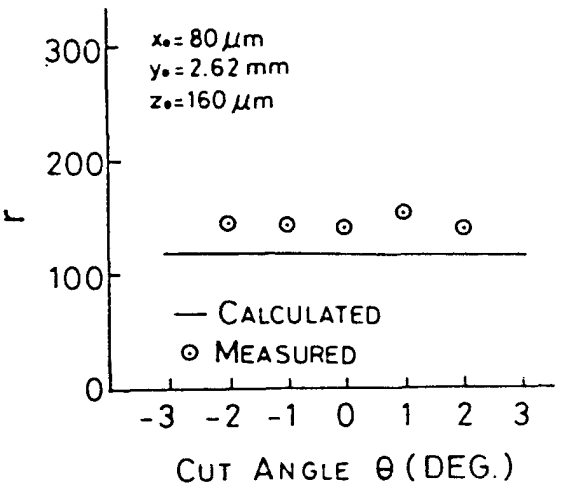


Fig. 13 Capacitance ratio r to cut angle θ.

Needless to say, it goes without saying that G=1 when the electrodes are disposed on the entire of surfaces perpendicular to x axis.

Figure 11 shows a relationship between motional inductance L_1 and width x_0 of a resonator with length y_0=2.50mm, the cut angle θ=0°00', y_b=50 μm and y_e/y_0=0.84, taking thickness z_0 as a parameter. As is apparent from equation (26), L_1 is proportional to x_0 and Fig. 11 also shows the same results. The solid line and the broken line exhibit the calculated results at thickness z_0=185 μm and z_0=125 μm, respectively, both results agrees well with the measured data. Moreover, as thickness z_0 is increased, L_1 is decreased because L_1 is inversely proportional to the thickness z_0.

Figure 12 shows a relationship between motional inductance L_1 and length y_0 of a resonator with width x_0=80 μm, thickness z_0=185 μm, the cut angle θ=0°00', y_b=50 μm and y_e/y_0=0.84. Since L_1 is proportional to length y_0, L_1 shows a marked variation of about 10 to 14 (H), according as length y_0 changes from 1.8mm (f=1.5MHz) to 2.5mm (f=1.1MHz). It is readily apprehended from Fig. 12 that the calculated and measured results are in good accordance.

3.5 Capacitance ratio r

In designing a resonator, it is of great importance what value the resonator has in practice. A capacitance ratio r is defined as a ratio of shunt capacitance C_0 and motional capacitance C_1 of the resonator, namely, $r=C_0/C_1$. Therefore, taking electrode length z_e of the thickness direction and dielectric constant ε, the capacitance ratio r is given as

$$r = \frac{n^2 \pi \varepsilon}{32 G h} \frac{\gamma_{22}}{e_{12}^2} \frac{(y_e - y_b) z_e}{y_0 z_0} \qquad (28)$$

where $h = 1 + k^2 (n\pi/y_0)^2$

From equation (28), it is readily understood that the capacitance ratio r is dependent upon the excitation electrode area, the cut angle θ, and the order of overtone, but independent upon width x_0, due to $k \fallingdotseq 0$ substantially. In addition, r for fundamental vibration is determined by only the cut angle θ, in case of disposing the excitation electrodes on the entire of yz surfaces of the vibrational portion.

Figure 13 shows a relationship between the capacitance ratio r and the cut angle θ of a resonator with width $x_0=80\mu m$, length $y_0=2.62mm$ and thickness $z_0=160\mu m$ (f=1.056MHz), which shows the calculated values at the cut angle θ of -3° to +3° and the measured ones in several points. As is illustrated in Fig. 13, the capacitance ratio has a very small value such as about 120 in the calculation and about 140 in the experiments, hence, the present resonator is a resonator which is extremely excellent in an electromechanical transformation ratio. But, an error of about 14 per cent between the measured and calculated values takes place, this seems to be due to the fact that stray capacitance of the supporting portions for the present resonator is ignored in the calculation of r. In addition, since the piezoelectric constant e_{12} hardly varies between the cut angles of -3° and +3°, and can be almost regarded as a constant value, the capacitance ratio r also shows the constant value between them.

3.6 Series resistance R_1 and quality factor Q

Figure 14 shows a relationship of series resistance R_1 and a quality factor Q versus resonant frequency f of resonators with the cut angle $\theta=0°00'$, which are very suitably designed in the dimensions, that is, such as $x_0=106\mu m$, $y_0=4.6mm$, and $z_0=200\mu m$ at 0.6MHz, $x_0=100\mu m$, $y_0=2.5mm$, and $z_0=160\mu m$ at 1.1MHz, $x_0=80\mu m$, $y_0=2.3mm$, and $z_0=185\mu m$ at 1.2MHz and $x_0=80\mu m$, $y_0=1.8mm$, and $z_0=185\mu m$ at 1.5MHz. For example, when f=0.6MHz, R_1 has a value of 269Ω and, even though the resonant frequency f is markedly varied like f=1.1, 1.2 and 1.5MHz, R_1 has also a very small value such as 313, 262 and 280Ω, respectively. Therefore, these results come to a conclusion that the present resonator's shape and the determination of its dimension are all right, for, it is substantially impossible to obtain the very small values of R_1 like this, in particular, when the shape of the supporting portions and the dimension for a resonator are not basically designed very well. In still more detail, there exist a few factors to determine R_1 of the resonator and, of the factors, it goes without saying that the supporting portion is the greatest factor, because it becomes a big cause to bring about energy losses by vibration. As described until now, since the present length extensional mode quartz crystal resonator has a large value of L_1 and a small one of R_1, it is readily forecast that it is a resonator with a high quality factor. As is obvious from Fig. 14, the extremely high quality factor such as 44×10^4, 43×10^4, 37×10^4 and 35×10^4 can be, respectively, obtained at the resonant frequencies of f=0.6, 1.1, 1.2 and 1.5 MHz. A series of these electrical equivalent circuit parameters are sufficiently satisfactory to a resonator.

Table 1 shows the typical values of electrical parameters for the new shape length extensional mode quartz crystal resonator, when the resonator with the resonant frequency of 0.5MHz to 1.5MHz is very suitably designed, where the capacitance ratio r is a value including stray capacitance of an unit which houses the present resonator, and stray capacitance of the supporting portions for this resonator.

Figure 15 shows the new shape length extensional mode quartz crystal resonators developed by us. A great number of miniaturized quartz crystal resonators are disposed in lines and formed in a piece of wafer by a chemical etching process. This process is, therefore, superior in mass production.

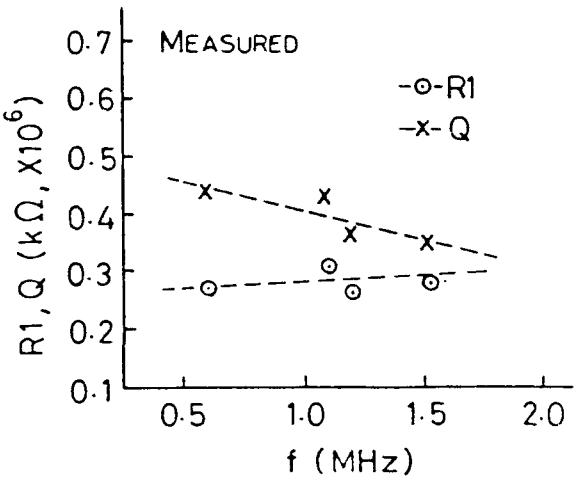


Fig. 14 Relationship of series resistance R_1 and quality factor Q to resonant frequency f.

Table 1 Electrical parameters.

f(MHz)	L_1(H)	R_1(Ω)	Q(×10⁴)	r
0.5	42.2	247	54	252
0.6	31.5	269	44	286
1.1	19.3	313	43	402
1.2	12.8	262	37	358
1.5	10.2	280	35	405

Figure 16 shows so called, a quartz crystal unit of a surface mounting type, housing the said resonator in a ceramic case with the thickness of 1mm. As illustrated in Fig. 16, the present resonator is, also, strong against shock, because it is incorporated with the vibrational portion and the supporting portions by an etching method. Thus, we have succeeded in developing the ultra-miniatured quartz units.

Figure 17 shows another quartz crystal unit housing the said resonator in a tubular case.

§ 4 Conclusions

For the purpose of application to portable products, consumer products and communication equipment, this paper has proposed the new shape length extensional mode quartz crystal resonator consisting of the vibrational portion and the supporting portions, formed by an etching method and has, theoretically and experimentally examined frequency temperature behavior and electrical characteristics of its resonator. In analysis procedure, first, a vibration analysis including the vibrational portion and the supporting portions, and taking lateral motion into account has been performed by an energy method, so that the frequency equation has been derived.

Next, the frequency temperature character-

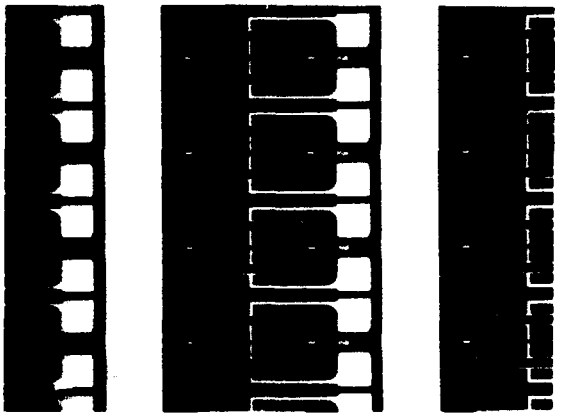


Fig. 15 New shape length extensional mode quartz crystal resonators formed by etching method.

Fig. 16 Quartz crystal unit of surface mounting type.

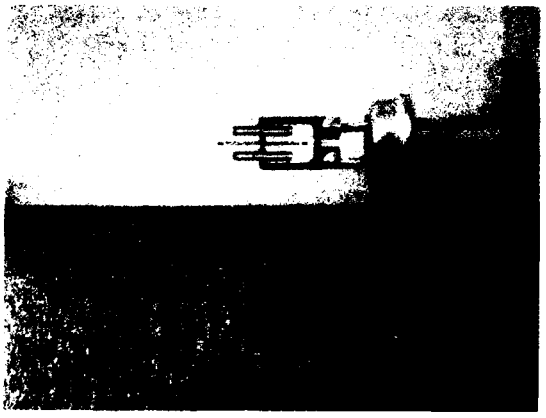


Fig. 17 Quartz crystal unit of tubular type.

istics have been clarified from this equation, the analysis results have been compared and studied with the experimental data. As a result, it has been shown that the cut angles which give a zero temperature coefficient, exist and the excellent frequency temperature characteristics are actually obtained, even if the supporting portions are added to the vibrational portion like the present resonator. In addition, it has been also shown that the turn over temperature point T.p varies by a width-length ratio and a thickness-length ratio, in particular, remarkably by the width-length ratio. Furthermore, as the result of examining the electrical characteristics, the resonators with a small value of R_1 and a high quality factor have been obtained in the frequency range of 0.5 to 1.5MHz, in other words, it has been readily understood that the shape of the supporting portions and the dimension for this resonator are basically designed very well.

Thus, the present resonator is superior in various points, and then, very suitable for a surface mounting type being eagerly required at present, because it is extremely miniaturized and thinly formed by an etching method. Simultaneously, since the resonant frequency and the frequency temperature characteristics of the present resonator are exactly analyzed by three dimensions taking width and thickness into consideration, the turn over temperature point for various frequencies can be, theoretically, established to an arbitrary temperature, so that the developing term is substantially shortened. Finally, for the coming subject, the possibility to higher frequency will be clarified theoretically and experimentally.

References

[1] H. Kawashima:" A variational analysis of a GT cut quartz crystal resonator with the supporting portions at the ends", Trans. IECE(A), J68-A, 8, pp. 771-778(1985-08)(in Japanese)

[2] 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)

[3] M. Aruga:" On elastic constants of quartz and its temperature behavior", Journal of Tokyo institute of technology, A-2, pp. 88-182(1956)

[4] H. Kawashima:" Vibrational analysis of GT cut quartz crystal resonator taking account of dissipation term", Trans. IEICE(A), J72-A, 4, pp. 633-639(1989-04) (in Japanese)

[5] H. Kawashima and M. Matsuyama: " An analysis of electrical equivalent circuit parameters of GT cut quartz crystal resonators with the supporting portions at the ends", Trans. IEICE(A), J71-A, 6, pp. 1217-1224(1988-06)(in Japanese)

[6] R. Bechmann: "Elastic and piezoelectric constants of alpha quartz", Phy. review, 110, 5, pp. 1060-1061 (June 1, 1958)

[7] For example, R. J. Dinger:"A miniature quartz resonator vibrating at 1 MHz", Proc. the 35th Annual Symp. on Frequency Control, pp. 144-148(1981)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Simulation of Noise Processes in Thickness-Shear Resonators Caused by Multilayer Adsorption and Desorption of Surface Molecules.

Yook-Kong Yong

Dept. of Civil/Environmental Engineering, Rutgers University, P.O. Box 909, Piscataway, NJ 08855-0909

ABSTRACT.

Effects of multilayer contamination on mean resonant frequency and frequency fluctuations in thickness-shear resonators are studied. A model based on mass-loading of contaminant molecules with adsorption and desorption rates is developed. Equations relating change in mean frequency and frequency fluctuations to adsorption and desorption rates are derived. Since the adsorption and desorption rates are functions of pressure and temperature, change in mean frequency and spectral density of frequency fluctuations are studied with respect to pressure and temperature. Calculations are performed for a 10 MHz thickness-shear resonator. Frequency-temperature and frequency-pressure curves are plotted for the 10 MHz resonator. The curves do not follow a cubic polynomial function and have a magnitude in the range of 10 ppm. The mean square of frequency fluctuations under multilayer contamination is significantly greater than that under monolayer contamination. The spectral density of frequency fluctuations at 1 Hz is quite constant in a wide range of temperatures (-50 to 100°C) when the values of heat of adsorption for the second and subsequent layers is close to that for the first layer. The magnitude of spectral density of frequency fluctuations is about -120 dBc (Hz^2/Hz).

I. Introduction.

The effect on resonant frequency due to a very thin mass layer, such as one or more layers of molecules, adsorbed on a thickness-shear quartz resonator is measurable, and is employed in quartz microbalances for surface adsorption studies[1]. Since the adsorbed molecules are known to have finite lifetimes on the surface, the number of adsorbed molecules at a given instant will fluctuate. Hence, short-term frequency instabilities will result if the resonator is sufficiently sensitive to the fluctuations. This was shown to be theoretically possible for monolayer contamination in ultra-high-frequency thickness-shear resonators[2]. We examine in this paper the effects of temperature and pressure on changes in mean resonant frequency and frequency fluctuations due to multilayer adsorption of contaminant molecules on the surfaces of a 10 MHz thickness-shear resonator.

II. Adsorption of Gases and Vapors on a Resonator Surface.

For multilayer adsorption of gases and vapors on the surfaces of a thickness-shear resonator, we employ a BET model which was proposed by Brunauer, Emmett, and Teller[3]. The reader is also referred to a text by Adamson[4]. The assumptions are: (1) The number of molecules adsorbing on a given surface is equal the number of molecules desorbing during a steady state equilibrium, (2) the heat of adsorption for the first layer, E_1, has some special value, whereas for all succeeding layers, it is equal to the heat of condensation of the liquid adsorbate, E_v, and (3) the adsorption and desorption can occur only from or on exposed surfaces. Figure 1 from reference [4] shows portions of surfaces S_i covered by i ($i = 0, 1, 2,...,$) layers of contaminant molecules. The adsorbed molecules are assumed to have no lateral interactions with their neighbors.

a) Adsorption and Desorption Rates of Contaminant Molecules.

The rate of arrival of molecules at a surface can be evaluated readily from simple kinetic theory of gases involving the pressure and kinetic energy of the molecules. If a site spacing of 0.5 nm is assumed, the rate of arrival of molecules at a site is

$$r = 87.8\text{x}10^6 P/\sqrt{MT} \text{ molecules/site/s}, \qquad (1)$$

where P is pressure in torrs, M is molecular weight of molecules, and T is temperature in kelvins. Not all the arriving molecules which impinge on the surface stick to it. The rate of contamination of molecules at a site is

$$\lambda_i = s_i r \text{ molecules/site/s}, \qquad (2)$$

where s_i is the sticking coefficient which is the probability of adsorption of a molecule impinging at the surface S_i.

CH2818-3/90/0000-387 $1.00 © 1990 IEEE

An adsorbed molecule will reside on a site for a mean time of stay, and subsequently desorb from the surface. The rate of desorption per site from surface S_1 is

$$\mu_1 = Ae^{-E_1/\sqrt{RT}} \text{ molecules/site/s,} \quad (3)$$

where A is assumed to have a value of $1x10^{13}$, E_1 is the heat of adsorption in kcal/mole, and R is the gas constant. For surfaces S_i where i is greater than one, the rate of desorption per site is

$$\mu_2 = Ae^{-E_v/\sqrt{RT}} \text{ molecules/site/s,} \quad (4)$$

where E_v is the heat of condensation in kcal/mole.

b) Dynamic Equilibrium Between the Resonator Surface and the Contaminant Gas.

The derivation of the steady-state probabilities of a site being uncontaminated, or contaminated with one or more molecules, follows essentially that given by the BET model. We presently assume that the adsorbed mass layer will consists of no more than three layers of molecules. Hence, the sticking coefficient s_3 is given a value of zero. Derivations for models with more than three layers are similar to what is given here. When the system of contaminant gas and resonator surface settles down to a steady state dynamic equilibrium, the states of a site, which is either uncontaminated (state 0), or contaminated with i molecules in a layerwise manner, can be described by steady-state probabilities p_i $(i = 0, 1, 2, 3)$. The rate at which a site enters state i is equal to the rate at which it leaves, that is,

$$\begin{aligned} p_0\lambda_0 &= p_1\mu_1 \\ p_1\lambda_1 &= p_2\mu_2 \\ p_2\lambda_2 &= p_3\mu_2. \end{aligned} \quad (5)$$

The steady-state probabilities must sum to one, since the site must at any given time be in one of the states,

$$p_0 + p_1 + p_2 + p_3 = 1. \quad (6)$$

The steady-state probabilities p_i can be expressed in terms of the adsorption and desorption rates using Eqs.(5) and (6):

$$\begin{aligned} p_0 &= \frac{1}{D} \\ p_1 &= \frac{\lambda_0}{\mu_1 D} \\ p_2 &= \frac{\lambda_0\lambda_1}{\mu_1\mu_2 D} \\ p_3 &= \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2^2 D} \end{aligned} \quad (7)$$

where $D = 1 + \frac{\lambda_0}{\mu_1} + \frac{\lambda_0\lambda_1}{\mu_1\mu_2} + \frac{\lambda_0\lambda_1\lambda_2}{\mu_1\mu_2^2}$.

III. Changes in Mean Resonant Frequency.

From Eq.(7), we observe that the steady-state probabilities p_i, being functions of the adsorption and desorption rates, are functions of temperature and pressure for a given absorbate-contaminant system. The sensitivity of the resonator to mass loading is confined mainly to the electrode patch. A monolayer of contaminant molecules covering the electrode patch will change the resonant frequency by the amount

$$\Delta_f = -f_0\frac{m'}{m} \quad (8)$$

where m' and m are the mass per unit area of contaminant molecules and mass per unit area of resonator plate, respectively. The resonator frequency is f_0. The negative sign signifies that the resonant frequency decreases with mass loading. Eq.(8) is valid only when the mass per unit area of contaminant molecules is much smaller than the mass per unit area of resonator plate. The resonator is assumed to vibrate in a pure thickness-shear mode. The actual change in frequency will depend on the number of molecular layers and fraction of surface covered, namely,

$$\frac{\Delta f}{f_0} = 2\Delta_f\left((p_1 - p_{1_0}) + 2(p_2 - p_{2_0}) + 3(p_3 - p_{3_0})\right) \quad (9)$$

where p_{i_0} $(i = 1, 2, 3)$ are the steady-state probabilities at the reference temperature and pressure, and the factor two on the right hand side accounts for contamination on the top and bottom surface.

a) Changes in Mean Resonant Frequency for a 10 MHz Thickness-Shear Resonator.

Simple calculations are performed for a 10 MHz thickness-shear resonator to examine the characteristics and magnitude of changes in mean frequency as a function of temperature and pressure. The resonator thickness is 0.165 mm and the area of circular electrode patch is 10 mm^2. A 0.5 nm thick monolayer of molecules covering the top electrode and having a density equivalent to quartz will yield

$$\begin{aligned} \Delta_f &= 30 \text{ Hz,} \\ \text{or } \frac{\Delta_f}{f_0} &= 3 \text{ ppm.} \end{aligned} \quad (10)$$

The molecular weight of contaminant molecules is taken as equal to 28. For our calculations of the adsorption rates, we take values of sticking coefficients s_i $(i = 0, 1, 2)$ equal to 0.1.

Figure 2 shows the effects of surface contamination with temperature on the mean resonator frequency in parts per million. Three frequency-temperature curves are given corresponding to heats of adsorption 15, 18, and 21 kcal/mole. The heats of condensation are taken to be 75 percent of their respective heats of adsorption. The pressure is kept constant at 0.001 torr. We observe that the surface contamination as a function of temperature can cause the mean frequency to change as much as 10 ppm. The f-T curves do not follow a cubic polynomial function. The characteristics and magnitudes of curves will be affected by initial pressures, heats of adsorption and condensation, and sticking coefficients. The mean frequency increases with increasing temperature.

Figure 3 shows the effects of surface contamination with pressure on the mean resonator frequency in parts per million. The three frequency-pressure curves have the same corresponding heats of adsorption and condensation as in Fig.2. The temperature is kept constant at 25°C. We observe that the mean frequency is also sensitive to pressure changes, and is of a higher order function than a cubic polynomial. The mean frequency decreases with increasing pressure.

IV. Theoretical Developments for Frequency Fluctuations.

Although there is a steady-state equilibrium in the adsorption and desorption of contaminant molecules, there are instantaneous fluctuations in the number of adsorbed molecules. Each contaminant site on the electrode patch of a thickness-shear resonator exerts a small but finite mass loading effect on the resonant frequency. Hence, the fluctuations will be manifested in short term frequency instabilities. We investigate the magnitude and spectral characteristics of short term frequency fluctuations as caused by spatial and temporal fluctuations in the number of adsorbed contaminant molecules.

a) Birth and Death Processes.

The stochastic analysis of fluctuations in the adsorption and desorption of contaminant molecules follows the usual developments for birth and death processes[5]. The table below gives the birth(adsorption) and death(desorption) rates which follow from Eqs.(2) to (4):

$$\begin{aligned}
&\text{Birth rate in } S_0 \text{ sites} = \lambda_0 \\
&\text{Birth rate in } S_1 \text{ sites} = \lambda_1 \\
&\text{Birth rate in } S_2 \text{ sites} = \lambda_2 \\
&\text{Death rate in } S_1 \text{ sites} = \mu_1 \\
&\text{Death rate in } S_2 \text{ sites} = \mu_2 \\
&\text{Death rate in } S_3 \text{ sites} = \mu_2
\end{aligned} \tag{11}$$

The birth and death process is a continuous-time Markov chain which is more generally an exponential model. Hence, the state transition rate for a S_i site is simply the sum of birth and death rates associated with the state of the site, that is,

$$\begin{aligned}
&S_0 \text{ site: } \lambda_0 \\
&S_1 \text{ site: } \lambda_1 + \mu_1 \\
&S_2 \text{ site: } \lambda_2 + \mu_2 \\
&S_3 \text{ site: } \mu_3
\end{aligned} \tag{12}$$

b) Frequency Effect of a site $f_k(t)$.

Since each contamination site, say site k, exerts a finite effect on the resonant frequency, the frequency effect can be described by a continuous function $f_k(t)$ which is a continuous-time Markov chain. The function has specific values depending on the state of site:

$$\begin{aligned}
f_k(t) &= a_0 = 0 \text{ for } S_0 \text{ site,} \\
f_k(t) &= a_1 = \frac{\Delta_f}{N_r} \text{ for } S_1 \text{ site,} \\
f_k(t) &= a_2 = \frac{2\Delta_f}{N_r} \text{ for } S_2 \text{ site,} \\
f_k(t) &= a_3 = \frac{3\Delta_f}{N_r} \text{ for } S_3 \text{ site.}
\end{aligned} \tag{13}$$

where N_r is the number of sites on one electrode patch. The magnitude of thickness-shear vibrations is maximum at the center of electrode patch with the vibrations decaying to zero near the edges of electrode. A graph of the magnitude of thickness-shear vibrations will exhibit a Gaussian profile. This implies that the mass loading effect of the contaminant molecules is exerted predominantly near the center third of electrode. Hence the effective N_r value is smaller than the number of sites in one electrode patch. Subsequent analysis will show that a smaller N_r value leads to larger magnitudes of root-mean-square frequency fluctuations.

c) Conditional Probability Function $P_{ij}(t)$ of $f_k(t)$.

The probability of $f_k(t)$ at a subsequent time t having a value of a_j, given an initial value of a_i at t equal to zero,

is defined by a conditional probability function

$$P_{ij}(t) = P\{f_k(t) = a_j \mid f_k(0) = a_i\} \quad i,j = 0,1,2,3. \tag{14}$$

For a continuous-time Markov chain, Eq.(14) must satisfy Kolmogorov's backward equations which are a system of first order differential equations:

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{\Lambda P}(t), \tag{15}$$

where $\mathbf{P}(t) =$

$$\begin{bmatrix} P_{00}(t) & P_{01}(t) & P_{02}(t) & P_{03}(t) \\ P_{10}(t) & P_{11}(t) & P_{12}(t) & P_{13}(t) \\ P_{20}(t) & P_{21}(t) & P_{22}(t) & P_{23}(t) \\ P_{30}(t) & P_{31}(t) & P_{32}(t) & P_{33}(t) \end{bmatrix}, \tag{16}$$

and $\mathbf{\Lambda} =$

$$\begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 \\ 0 & 0 & \mu_2 & -\mu_2 \end{bmatrix}. \tag{17}$$

The initial conditions for Eq.(15) is an identity matrix since at zero time there are no transitions,

$$\mathbf{P}(0) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{18}$$

The Kolmogorov's backward equations can be solved[6] to yield

$$\mathbf{P}(t) = \mathbf{VEV}^{-1} \tag{19}$$

where

$$\mathbf{E}(t) = \begin{bmatrix} e^{-\omega_0 t} & 0 & 0 & 0 \\ 0 & e^{-\omega_1 t} & 0 & 0 \\ 0 & 0 & e^{-\omega_2 t} & 0 \\ 0 & 0 & 0 & e^{-\omega_3 t} \end{bmatrix} \tag{20}$$

is a diagonal matrix of exponential time functions with constants ω_m *(m=0, 1, 2, 3)* which are the absolute values of eigenvalues of $\mathbf{\Lambda}$ in Eq.(17). Since $\mathbf{\Lambda}$ is singular, a zero eigenvalue exists, and we can define ω_0 to be the zero eigenvalue. All the nonzero eigenvalues are negative. The column vectors $\mathbf{v}_m$ of $\mathbf{V}$ and eigenvalues $-\omega_m$ form the eigenpairs $\{-\omega_m, \mathbf{v}_m\}$ of $\mathbf{\Lambda}$. Eq.(19) and (20) show that the matrix of conditional probabilities are exponential functions of elapsed time t, namely,

$$P_{ij}(t) = \sum_{m=0}^{3} v_{im} v_{mj}^{-1} e^{-\omega_m t} \tag{21}$$

where v_{im} and v_{mj}^{-1} are elements of matrices $\mathbf{V}$ and $\mathbf{V}^{-1}$, respectively.

d) Autocorrelation Function of Frequency Fluctuations about Mean Resonant Frequency.

The autocorrelation function of the continuous-time Markov chain $f_k(t)$ in Eq.(13) is defined as the expected value of the product of $f_k(t)$, and $f_k(0)$, that is,

$$R_{f_k}(t) = E[f_k(t) f_k(0)], \tag{22}$$

which by definition can be written as

$$R_{f_k}(t) = \sum_{i=0}^{3}\sum_{j=0}^{3} a_i a_j P\{f_k(t) = a_j, f_k(0) = a_i\}, \tag{23}$$

where the term $P\{f_k(t) = a_j, f_k(0) = a_i\}$ is the joint probability distribution function of $f_k(t)$, and $f_k(0)$. The joint probability distribution function can be calculated by employing the conditional probabilities in Eq.(19), and steady-state probabilities in Eq.(7). Hence,

$$P\{f_k(t) = a_j, f_k(0) = a_i\} = P_{ij}(t)\, p_i. \tag{24}$$

Using Eqs.(24) and (21), Eq.(23) can be written as

$$R_{f_k}(t) = \sum_{m=0}^{3} \sigma_m^2 e^{-\omega_m t} \tag{25}$$

$$\text{where } \sigma_m^2 = \sum_{i=0}^{3}\sum_{j=0}^{3} a_i a_j v_{im} v_{mj}^{-1} p_i. \tag{26}$$

Since ω_0 is the zero eigenvalue, σ_0^2 is a constant dc noise which is the square of mean value of $f_k(t)$. In our study of short term instabilities in frequency, we are interested in frequency fluctuations about a mean resonant frequency. Hence, we define a continuous-time Markov chain of frequency fluctuations

$$\Delta f_k(t) = f_k(t) - \sigma_0 \tag{27}$$

The autocorrelation function of frequency fluctuations about a mean frequency for site k is

$$\begin{aligned} R_{\Delta f_k}(t) &= E[\{\Delta f_k(t)\}\{\Delta f_k(0)\}] \\ &= E[f_k(t) f_k(0)] - \sigma_0^2 \\ &= \sum_{m=1}^{3} \sigma_m^2 e^{-\omega_m t} \end{aligned} \tag{28}$$

Note that since a_0 is equal to zero, Eq.(26) can also be written as

$$\sigma_m^2 = \sum_{i=1}^{3}\sum_{j=1}^{3} a_i a_j v_{im} v_{mj}^{-1} p_i. \tag{29}$$

We observe that the sum of σ_m^2 over $m = 1,2,3$ is the mean square of frequency fluctuations due to site k.

There are $2N_r$ sites in the resonator, which accounts for contamination sites on the top and bottom electrodes. Hence, the resonator frequency fluctuation about a mean frequency at any time t is

$$\Delta F(t) = \sum_{k=1}^{2N_r} \Delta f_k(t), \tag{30}$$

whose autocorrelation function is

$$\begin{aligned} R_{\Delta F}(t) &= E[\Delta F(t)\,\Delta F(0)], \\ &= E\left[\left(\sum_{k=1}^{2N_r} \Delta f_k(t)\right)\left(\sum_{l=1}^{2N_r} \Delta f_l(0)\right)\right]. \\ &= \sum_{k=1}^{2N_r}\sum_{l=1}^{2N_r} E[\{\Delta f_k(t)\}\{\Delta f_l(0)\}] \end{aligned} \tag{31}$$

Since the effects of contaminant sites are mutually independent, the expected values of cross product terms are zero, and Eq.(31) becomes

$$\begin{aligned} R_{\Delta F}(t) &= \sum_{k=1}^{2N_r} E[\{\Delta f_k(t)\}\{\Delta f_k(0)\}] \\ &= \sum_{k=1}^{2N_r} R_{\Delta f_k}(t) \end{aligned} \tag{32}$$

The generic site k is indistinguishable from other sites on the electrode surfaces, therefore, the autocorrelation function in Eq.(32) is

$$R_{\Delta F}(t) = 2N_r R_{\Delta f_k}(t). \tag{33}$$

Using the last equation of Eq.(28), we can write Eq.(33) as

$$\begin{aligned} R_{\Delta F}(t) &= \sum_{m=1}^{3} \bar{\sigma}_m^2 e^{-\omega_m t}, \\ \text{where } \bar{\sigma}_m^2 &= 2N_r\sigma_m^2. \end{aligned} \tag{34}$$

We observe that the autocorrelation function in Eq.(34) is made up of three Lorentzian autocorrelation functions. The reciprocals of ω_m are the correlation times of the autocorrelation function. The term

$$\bar{\sigma} = \sqrt{\sum_{m=1}^{3} \bar{\sigma}_m^2} \tag{35}$$

is the root-mean-square of resonator frequency fluctuations, and from Eqs.(13) and (29),

$$\bar{\sigma} = \frac{\Delta_f}{\sqrt{N_r}}\sqrt{\sum_{m=1}^{3}\sum_{i=1}^{3}\sum_{j=1}^{3} 2ij v_{im} v_{mj}^{-1} p_i}. \tag{36}$$

Hence the root-mean-square of frequency fluctuations is proportional to the frequency change of resonator to a monolayer of contaminant molecules, and inversely proportional to the square root of the number of contaminant sites in electrode patch.

e) One-Sided Spectral Density of Frequency Fluctuations about Mean Resonant Frequency.

Wiener-Khintchine relation is employed to obtain the one-sided spectral density of frequency fluctuations from the autocorrelation function of Eq.(34); which yields

$$S_{\Delta F}(f) = \sum_{m=1}^{3} \frac{4\bar{\sigma}_m^2 \omega_m}{\omega_m^2 + 4\pi^2 f^2}, \tag{37}$$

where f is the Fourier frequency. Eq.(37) reveals all the essential characteristics of frequency fluctuations induced by multilayer surface contamination. The spectral density function is a superposition of three Lorentzian spectra with a distribution of corner frequencies ω_m. The function is similar to that given by McWhorter[7]. Over a certain range of Fourier frequencies, 1/f noise can be generated if ω_m is uniformly distributed with at least one corner frequency per decade of Fourier frequency[8], and the associated $\bar{\sigma}_m^2$ having a significant magnitude. The magnitude of spectral density is proportional to the square of Δ_f, and inversely proportional to N_r.

V. Noise Calculations for a 10 MHz Thickness-Shear Resonator.

Noise calculations for the 10 MHz thickness-shear resonator previously defined in section III(a) is performed. In addition, an average site spacing of 0.5 nm is assumed for the contaminant sites on electrode patches. Hence, the number of sites in one electrode patch is

$$N_r = 4\text{x}10^{13} \text{ sites.} \tag{38}$$

Calculations for both monolayer and multilayer contaminations are performed. The model for monolayer contamination was previously reported in reference [2], and is actually a specialized case of the present model.

Figure 4 shows the spectral density curves of frequency fluctuations plotted as a function of Fourier frequency. Four curves are shown. The solid curve represents monolayer contamination, while the three other curves are for multilayer contamination using values of heat of condensation E_v which are 0.5, 0.75 and 1.0 of E_l, the heat of absorption with a value of 21 kcal/mole. We observe that

the magnitude of spectral density for E_v greater than 0.75 E_l is about 50 dBc greater than the curve with monolayer contamination. For the particular pressure (0.001 torr), temperature (25°C), and Fourier frequency range of less than 100 Hz, the spectral density curves with E_v less than 0.5 E_l resemble the monolayer curve.

The spectral density of frequency fluctuations at 1 Hz offset is useful for comparing data of noise magnitudes resulting from different initial pressures, temperatures, and heats of absorption and condensation. This is obtain from Eq.(37) by setting f equal to one,

$$S_{\Delta F}(1) = \sum_{m=1}^{3} \frac{4\bar{\sigma}_m^2 \omega_m}{\omega_m^2 + 4\pi^2}. \qquad (39)$$

Figure 5 shows the spectral density of frequency fluctuations at 1 Hz plotted as a function of pressure. We observe that for wide ranges of pressures (10^{-5} to 10 torr) the spectral density with multilayer contamination and $E_v > 0.75E_l$ has a magnitude greater than monolayer contamination. The magnitude difference is greater at higher pressures. The curve with $E_v = 0.5E_l$ follows the monolayer curve up to a pressure of about 0.002 torr, and diverges from it at higher pressures.

Figure 6 shows the spectral density of frequency fluctuations at 1 Hz plotted as a function of temperature. Calculations show that for multilayer contamination with $E_v > 0.9E_l$, the magnitude of spectral density at 1 Hz remains remarkably constant in a wide range of temperature (-50 to 100°C), as opposed to the monolayer curve. With $E_v = 0.75E_l$, the multilayer contamination curve remains constant in a shorter range of temperatures (-50 to 35°C). The curve with $E_v = 0.5E_l$ follows the monolayer curve in temperatures greater than 10°C.

VI. Summary.

A model based on mass-loading of contaminant molecules having adsorption and desorption rates is developed. Calculations are made for a 10 MHz thickness-shear resonator. The model brings out some interesting aspects of contamination in resonators which may be tested empirically:

1. The change in surface contamination layers due to a change in temperature affects the mean resonant frequency of thickness-shear resonators. This frequency change may be masked by changes in material properties, such as elastic stiffnesses and density. Calculations for the 10 MHz resonator yield mean frequency changes of as much as 10 ppm.
2. The mean resonant frequency is also affected in the same order of magnitude by the ambient pressure which influences the characteristics of the surface contamination layers.
3. Magnitude of spectral density of frequency fluctuations for multilayer adsorption and desorption is larger than for monolayer adsorption and desorption. This is especially true when the value of E_v is close to E_l. Calculations for the 10 MHz resonator show that the magnitude of mean square frequency fluctuations is in the measurable range of about -120 dBc (Hz^2/Hz).
4. Spectral density of frequency fluctuations at 1 Hz for multilayer adsorption and desorption, when the value of E_v is close to E_l, is remarkably constant with changes in temperature. This is not so for monolayer adsorption and desorption.

Acknowledgement

The author gratefully acknowledge the support by the U.S. Army Research Office, contract no. DAAL03-87-K-0107. He thanks John R. Vig at the U.S. Army ETDL for his encouragement and support.

References.

1. Lu, C., and Czanderna, A.W., /Editors, Applications of Piezoelectric Quartz Crystal Microbalances, Methods and Phenomena, Their Applications in Science and Technology, Vol. 7, Elsevier (Amsterdam), 1984.
2. Yong, Y-K, and Vig, J.R., "Resonator Surface Contamination - A Cause of Frequency Fluctuations?" IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vol. 36, No. 4, 1989, pp452-458.
3. Brunauer, S., Emmett, P.H., and Teller, E., Journal of the American Chemical Society, Vol. 60, 1938, pp309.
4. Adamson, A.W., Physical Chemistry of Surfaces, 4th edition, John Wiley, New York, 1982, pp533-539.
5. Ross, S.M., Introduction to Probability Models, 3rd edition, Academic Press, Orlando, 1985, chapt. 6 pp233-246.
6. Potter, M.C., and Goldberg, J., Mathematical Methods, 2nd edition, Prentice-Hall, Englewood Cliffs, 1987, pp293-295.
7. McWhorter, A.L., "1/f Noise and Germanium Surface Properties," Semiconductor Surface Physics, edited by R.H. Kingston, 1957, pp207-228.
8. Keshner, M.S., "1/f Noise," Proceedings of the IEEE, Vol. 70, No. 3, March 1982, pp212-218.

Fig.1 Adsorption of gases and vapors on solid surface[4], the BET model[3].

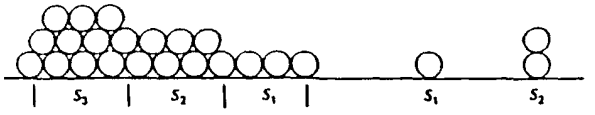


FIG.2 f–T curves for a 10 MHz thickness-shear resonator due to surface contamination. (Reference temperature=25 deg.C, pressure= 001 torr, Ev= 75 E1)

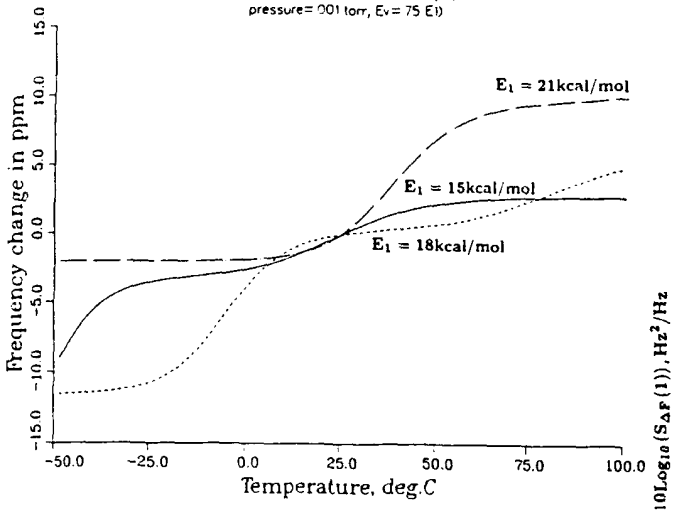


FIG.3 f–P curves for a 10 MHz thickness-shear resonator due to surface contamination. (Reference pressure= 001 torr, Ev=.75 E1 temperature=25 deg.C)

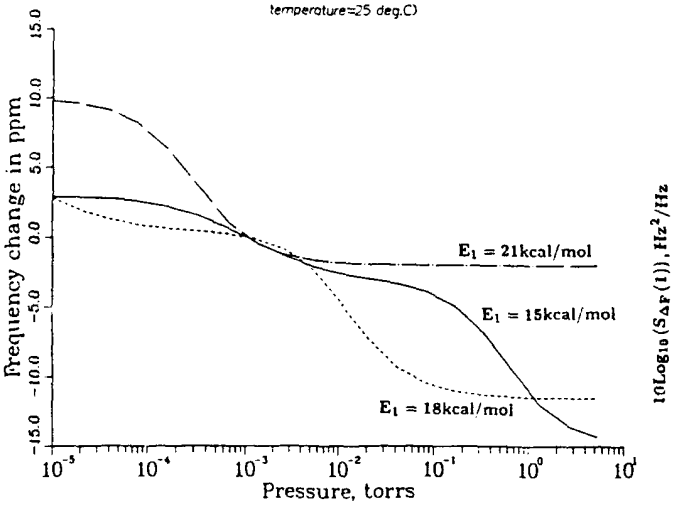


FIG.4 Spectral density of freq. fluctuations vs Fourier frequency. Pressure = E-3 torr Temperature = 25 deg.C E1 = 21 kcal/mol

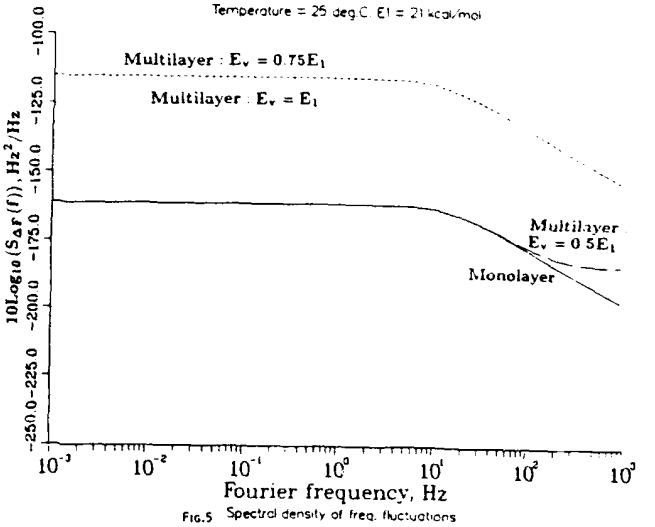


FIG.5 Spectral density of freq. fluctuations at 1 Hz vs pressure. Temperature = 25 deg.C. E1 = 21 kcal/mol

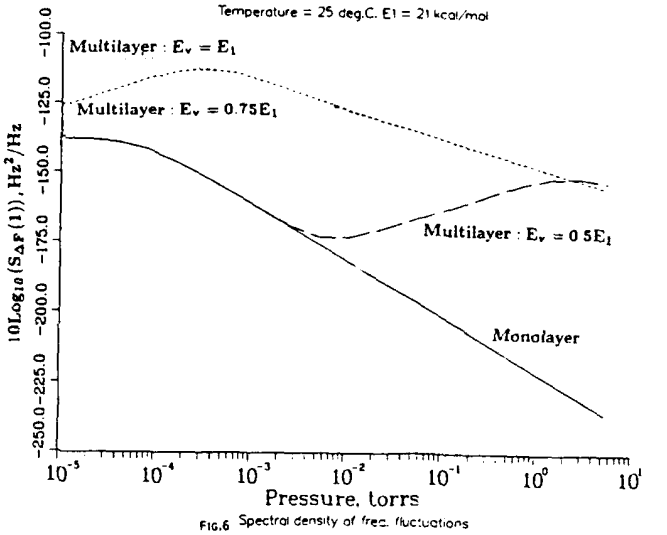
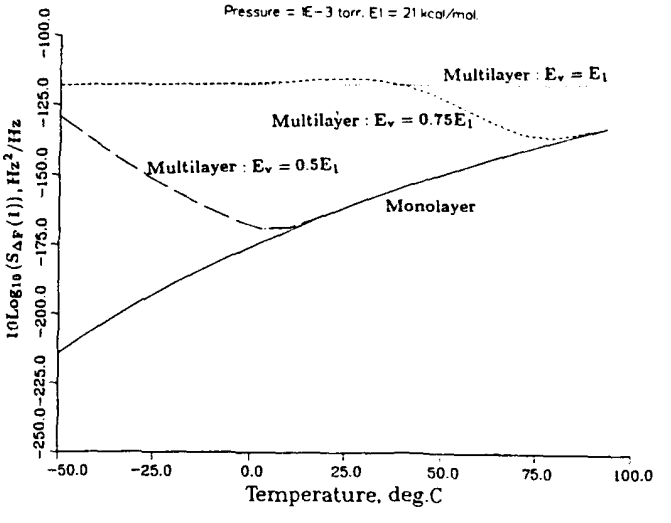


FIG.6 Spectral density of freq. fluctuations at 1 Hz vs temperature Pressure = E-3 torr. E1 = 21 kcal/mol



FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

DYNAMICS OF PIEZOELECTRIC LAMINAE UNDER A BIAS

M. Cengiz Dökmeci

İstanbul Teknik Üniversitesi, P.K. 9, Taksim, 80191 İstanbul

Abstract

This paper is addressed to the macromechanical analysis of dynamics of a piezoelectric laminae under a mechanical bias within the effective stiffness concept of laminated composites. The piezoelectric laminae consists of arbitrary numbers of perfectly bonded layers, each with a distinct but uniform thickness, curvature and electromechanical properties, and it is coated with very thin electrodes on both its faces. First,the fundamental equations of piezoelectric strained medium are expressed by the Euler-Lagrange equations of a unified variational principle. Secondly, a set of two-dimensional, approximate equations of the piezoelectric laminae is consistently established. Thirdly, a direct method of solution is indicated for the macromechanical analysis and certain special cases are considered. The governing equations are derived in invariant Lagrangian form and accommodate all the types of motions of the biased piezoelectric laminae. All the significant effects both mechanical and electrical, are taken into account.

1. Introduction

Laminae or multilayer type of structural elements was appreciated only relatively recently due to their significant improvements in piezoelectric properties for ultrasonic technology. The features and applications of piezoelectric layered and/or composite elements and the basic ideas underlying their sum and product properties are available[1-4]. To predict dynamic response of this type of structural elements, there basically exist two types of macromechanical models: the effective modulus model and the effective stiffeness model. The former model[5,6] replaces a laminae by a representative homogeneous medium with the aid of averaged material constants of laminae constituents. This model, although it is relatively simple, omits the mutual coupling of layers, and it is generally suitable for a rather broad class of static response of laminae. The effective stiffeness model combines both the physical and geometrical properties of laminae constituents and incorporates all their essential electromechanical features, and it accounts for dynamic response of laminae. Within the frame of this model[7], this paper describes a macromechanical analysis of piezoelectric laminae under a state of mechanical bias.

Extension of classical models (for instance, Lagrange's or Kármán's model of plates and Love-Kirchhoff's model of shells) to piezoelectric laminae leads always to an effective modulus model, and hence it disregards the electromechanical interactions between adjacent layers. On the basis of classical models, the macroscopic relations of electroelasticity were derived for multilayer piezoelectric plates and shells; their steady-state vibrations were reported in some special cases[8-10] as well. Parton and Senik[8] derived the macroscopic equations of multilayer piezoceramic shells with thickness polarization of the layers. Likewise, Karnaukhov and Kirichok[10-12] constructed the governing equations of laminated piezoelectric plates and shells by taking into account the geometrical nonlinearity and, in particular, the effect of viscosity and temperature. Evseichik, Rudnitskii and Shul'ga[13] derived the electroelastic equations for the vibrations of a shell that is inhomogeneous in thickness and has piezoelectric layers. Moreover, the thermomechanical behavior of multilayered piezoceramic shells with thickness polarization was treated under harmonic excitation by Motovilovets and Gololobov[14]. Mention should also be made of a theory of vibrations of coated, thermopiezoelectric laminae in which the effects of elastic stiffness of, and the interactions between, layers of the laminae and its electrodes were all included[7]. On the other hand,Holland and EerNisse[15] described the design and analysis of laminae types of piezoelectric bars, disks and plates by means of Green's function technique. Auld and his co-workers[16,17] developed a Floquet theory of wave propagation in periodic composites that was shown to agree well with experiment. Buğdaycı and Bogy[18,19] derived a theory for high frequency motions of piezoelectric layers, including some applications, as did Lee and Moon[20] for low frequency motions of piezoelectric laminated plates. Moreover, a general transfer matrix description of arbitrarily layered piezoelectric structures was obtained[21].

Biasing stress or strain and/or electric field is a new design feature and demand in piezoelectric

CH2818-3/90/0000-394 $1.00 © 1990 IEEE

devices for ultrasonic application in control engineering. The presence of a biasing state induced by external perturbations like thermal, mechanical and electrical fields and even magnetic fields can significantly affect the dynamic response of structural elements (e.g., rods[22], plates[23,24] and shells[25,26]) and the characteristics of BAW and SAW (e.g., [27-31] and references therein). In a biased solid medium, the linear theory of electroelasticity is in need of modification so as to govern its motions. This fact was widely recognized, and taken up by many investigators in electroelasticity. Tiersten[32] derived a properly invariant set of the nonlinear fundamental equations, including thermal effects by means of a systematic use of the axioms of continuum physics. From these general equations, he and Baumhauer[33] established the differential electroelastic equations for small dynamic fields superposed on a static biasing state of solid medium, and also, for intrinsically nonlinear fields. Moreover, the fundamental equations of a biased piezoelectric medium were expressed as the Euler-lagrange equations of some variational principles[34,35].

The aim of the present paper is (i) to present a variational formulation of the fundamental equations of piezoelectric medium under a mechanical bias, and using this together with a linear representation for the field variables, (ii) to derive the two-dimensional, approximate governing equations for all the types of incremental motions of piezoelectric laminae under a bias, and then (iii) to describe a direct method of solution for the incremental motions, to indicate some special cases and also to consider the fully linearized equations of piezoelectric laminae.

Specifically, a definition of the notation to be used herein is given in the rest of this section and the content of the paper is as follows. In the first part of the paper, a unified variational principle is formulated by extending the principle of virtual work through Friedrichs's transformation in Section 2. In the second part of the paper, presented in Sections 3-6, by use of Mindlin's method of reduction, the set of two-dimensional, approximate equations is consistently derived for the incremental motions of piezoelectric laminae under a static, finite, mechanical bias. The geometry of piezoelectric laminae region is described in Section 3. In Section 4, a linear representation in the thickness coordinate of piezo electric laminae is introduced for the fields of incremental mechanical displacements and electric potential which are chosen as a starting point of derivation. Also, in accordance with the linear representation, various resultant quantities averaged over the thickness of laminae are defined. The distributions of incremental strain and quasi-static electric field are given and the macroscopic constitutive equations of piezoelectric laminae are obtained in Section 5. The two-dimensional, approximate governing equations and the associated boundary and initial conditions for the motions of piezoelectric laminae are deduced from the three-dimensional equations of piezoelectricity by use of the unified variational principle together with the series representation of field variables in Section 6. Alternatively, a direct method of solution is indicated in investigating the incremental motions of piezoelectric strained laminae in Section 7. Special motions, geometry and material are treated and the fully linearized governing equations, including the uniqueness of their solutions, are pointed out in Section 8. The last section is devoted to the concluding remarks.

N o t a t i o n - In the paper, standard tensor notation is freely used in a Euclidean 3-space E. Accordingly, Einstein's summation convention is implied over all repeated Latin (1,2,3) and Greek indices (1,2) that stand for space and surface tensors, respectively, unless they are put within parantheses. In the space E, a fixed, right-handed system of geodesic normal convected coordinates is identified by the x^i-sytem. All the field quantities are used in Lagrangian formulation, and a quantity in the initial state is designated by a zero index and a prescribed quantity by an asterisk. A superposed dot stands for time differentiation, a comma for partial differentiation with regard to the indicated space coordinate, and a semicolon and a colon for covariant differentiation with respect to the indicated coordinate, using the space and surface metrics, respectively. The index (m) which takes the values 1,2,...,N refers to the m-th constituent from the lower face of piezoelectric laminae, and, for instance, m=1 (or a prime ') to the lower face electrode, m=2,3,...,N-1 to the layers and m=N (or a double prime ") to the upper face electrode of laminae. Moreover, B(t) represents a regular, finite and bounded region B of the space E at time t, $\bar{B}$(=BU∂B)the closure of B, with its boundary surface ∂B, $\bar{B}$XT the domain of definition for the functions of the space coordinates and time, T=$[t_0,t_1)$ the time interval, and Z=[z-h,z+h] the thickness interval of a constituent.

2. Principle of Virtual Work for the Piezoelectric Medium under a Bias

To derive, in a systematic and consistent manner, lower order field equations and to directly provide their approximate solutions, variational principles were primarily developed by Mindlin, Tiersten and EerNisse for a piezoelectric medium, by Mindlin, Nowacki and the author for a thermopiezoelectric medium, and by the author for a piezoelectric medium with small piezoelectric coupling and/or an internal surface of discontinuity and that under a mechanical bias. Hamilton's principle, the principle of virtual work and an experienced guess work were used in deducing the variational principles of piezoelectricity; a review of the subject was given in Refs. 2, 35. In order to render this paper to be self-contained, a unified variational principle is reformulated by extending the principle of virtual work through the dislocation potentials and Lagrange undetermined multipliers.

In the space E, referring to the x^i-system of general convected coordinates, a regular, finite and

bounded region of piezoelectric elastic medium, B_o, with its boundary surface ∂B_o, under a state of mechanical static stresses is considered at its initial unperturbed or reference state at time $t=t_o$. This initial state which is taken to be self-equilibrating acquires its spatial (perturbed or final) state $B+\partial B$ by a small motion superposed onto the finite, static deformation of piezoelectric region $B_o+\partial B_o$ at the time interval $T=[t_o,t_1)$. Now employing Lagrangian approach, the principle of virtual work is stated for the piezoelectric strained region at its spatial state as an assertation in the form

$$- \int_B (T^{ij}\delta S_{ij} - D^i\delta E_i)dV + 1/2\delta\int_B \rho\dot{U}^i\dot{U}_i dV + \int_{\partial B}(T^i_*\delta U_i + \sigma_*\delta\Phi)dS = 0 \qquad (2.1)$$

Here, $T^{ij}(=t_o^{ij}+t^{ij})$, t_o^{ij} and t^{ij} are the total, initial and incremental stress tensors; $S_{ij}=(s^o_{ij}+s_{ij})$, s^o_{ij} and s_{ij} the total, initial and incremental strain tensors; ρ the mass density of undeformed piezoelectric medium; $U_i(=u_i^o+u_i)$, u_i^o and u_i the total, initial and incremental displacement vectors, $a^i(=\ddot{u}_i)$ the Lagrangian acceleration vector; T^i $(=t_o^i+t^i)$, t_i^o and t^i the total, initial and incremental stress vectors; D^i the electric displacement vector, E^i the quasi-static electric field vector, $\sigma(=n_iD^i)$ the surface charge, Φ the electric potential and n_i the unit outward vector normal to a surface element. Substituting the gradient equations of the form

$$S_{ij} = E_{ij} + 1/2\, U^k{}_{;i}U_{k;j} \quad \text{in } \bar{B}XT \qquad (2.2)$$

$$E_{ij} = 1/2\,(U_{i;j} + U_{j;i}), \quad E_i = -\Phi_{,i}$$

into (2.1), applying the Green-Gauss transformation of integrals for the regular region $B+\partial B$, carrying out the indicated variations, and then integrating over T, one finally obtains a two-field variational principle for the piezoelectric biased medium as

$$\delta L\{u_i,\Phi\} = \int_T dt\int_B (L^i\delta u_i + L\delta\Phi)dV + \int_T dt\int_{\partial B}(L^i_*\delta u_i + L_*\delta\Phi)dS = 0 \qquad (2.3)$$

with the divergence equations of incremental motion by

$$L^j = (t^{ij} + t_o^{ik}u^j{}_{;k})_{;i} - \rho a^j = 0 \quad \text{in } \bar{B}XT \qquad (2.4)$$

$$L = D^i{}_{;i} = 0 \quad \text{in } \bar{B}XT \qquad (2.5)$$

and the associated natural boundary conditions by

$$L^i_* = \tau^j_* - n_i(t^{ij} + t_o^{ik}u^j{}_{;k}) = 0 \text{ on } \partial BXT \qquad (2.6)$$

$$L_* = \sigma_* - n_iD^i = 0 \text{ on } \partial BXT \qquad (2.7)$$

as its Euler-Lagrange equations. In deriving (2.3) the familiar relations between the stress tensors and the stress vectors, the stress equations of equilibrium and the associated boundary conditions, namely,

$$L^j_o = [t_o^{ik}(\delta^j_k + u^j_{o\,;k})]_{;i} = 0 \quad \text{in } \bar{B}XT$$
$$L^j_{o*} = t^j_{o*} - n_it_o^{ik}(\delta^j_k + u^j_{o\,;k}) = 0 \text{ on } \partial BXT \qquad (2.8)$$

are considered, the usual arguments are implied on the increments of field variables[36] and the axiom of conservation of mass is employed. Also, the constraint conditions of the form

$$\delta u_i = \delta\Phi = 0 \quad \text{in } B(t_o) \;\&\; B(t_1) \qquad (2.9)$$

are imposed.

To describe fully the motions of piezoelectric strained medium, the two-field variational principle (2.3) should be supplemented by the gradient equations (2.2), the constitutive equations in the form

$$t^{ij} = 1/2\,\left(\frac{\partial\P}{\partial s_{ij}} + \frac{\partial\P}{\partial s_{ji}}\right), \quad D^i = -\frac{\partial\P}{\partial E_i} \qquad (2.10)$$

where $\P(s_{ij},E_i,t_o^{ij})$ stands for an electric enthalpy function which contains the initial stresses as parameters[37], in addition to (2.6), the boundary conditions as

$$u_i - u_i^* = 0 \text{ on } \partial B_u XT,\ \Phi - \Phi_* = 0 \text{ on } \partial B_\Phi XT \qquad (2.11)$$

the initial conditions of the form

$$u_i(x^j,t_o) - v_i^*(x^j) = 0,\ \dot{u}_i(x^j,t_o) - w_i^*(x^j) = 0,\ \Phi(x^i,t_o) - \Omega(x^i) = 0 \quad \text{in } B(t_o) \qquad (2.12)$$

and the constraint conditions (2.9). These constraint conditions prevent a free choice of trial functions, and hence, variational principles with as few constraints as possible become desirable in computation. Thus, all the constraint conditions but the initial conditions are relaxed through Friedrichs's transformation[34] and the initial conditions following Tiersten's[38] approach. The result is a unified variational principle by

$$\delta L\{\Lambda\} = \delta J^{\alpha i}_{\alpha i} + \delta I^{\alpha\beta}_{\alpha\beta} + \delta I^i_i = 0 \qquad (2.13a)$$

with the admissible state

$$\Lambda = \{u_i,\ s_{ij},\ t^{ij};\ \tau^i,\ \Phi,\ E_i,\ D^i,\ \sigma\} \quad (2.13b)$$

and the denotations by

$$(\delta J^{11}_{11},\ \delta J^{12}_{12},\ \delta J^{13}_{13}) = \int_T dt \int_B (L^1 \delta u_i,\ L^{ij}\delta s_{ij},\ J_{ij}\delta t^{ij})dV$$

$$(\delta J^{21}_{21},\ \delta J^{22}_{22},\ \delta J^{23}_{23}) = \int_T dt \int_B (L\delta\Phi,\ I^i\delta E_i,\ J_i\delta D^i)dV \quad (2.14)$$

$$(\delta I^{11}_{11},\ \delta I^{12}_{12}) = \int_T dt \left(\int_{\partial B_t} L^i_{\star}\delta u_i dS,\ \int_{\partial B_u} J_{\star i}\delta\tau^i dS\right)$$

$$(\delta I^{21}_{21},\ \delta I^{22}_{22}) = \int_T dt \left(\int_{\partial B_\sigma} L_{\star}\delta\Phi dS,\ \int_{\partial B_\Phi} J_{\star}\delta\sigma dS\right)$$

and

$$[\delta I^1_1,\ \delta I^2_2] = \int_B [\lambda^i \delta u_i(x^j,t_0),\ \mu^i \delta\dot{u}_i(x^j,t_0)]dV$$

$$\delta I^3_3 = \int_B \lambda\delta\Phi(x^j,t_0)dV \quad (2.15)$$

Here,

$$L^{ij} = t^{ij} - 1/2\ (\frac{\partial\pi}{\partial s_{ij}} + \frac{\partial\pi}{\partial s_{ij}})$$

$$J_{ij} = s_{ij} - 1/2\ (u_{i;j} + u_{j;i}) \quad (2.16)$$

$$J_i = -\ (E_i + \Phi_{,i}),\quad I^i = -\ (D^i + \frac{\partial\pi}{\partial E_i})$$

and

$$J^{\star}_i = u_i - u^{\star}_i\ ,\quad J_{\star} = \sigma_{\star} - n_1 D^i \quad (2.17)$$

and

$$\lambda^i = \rho[\dot{u}^i(x^j,t_0) - w^i_{\star}(x^j)]$$

$$\mu^i = \rho[u^i(x^j,t_0) - v^i_{\star}(x^j)] \quad (2.18)$$

$$\lambda = \Phi(u^j,t_0) - \Omega_{\star}(x^j)$$

and also those in (2.4)-(2.7) are defined. The variational principle should be modified for the linearized constitutive relations which implies the dislocation potentials by

$$L^{ij} = t^{ij} - (C^{ijkl}s_{kl} - C^{kij}E_k)$$

$$I^i = -\ D^i + (C^{ijk}s_{jk} + C^{ij}E_j) \quad (2.19)$$

in lieu of those defined in (2.16). In (2.19), C^{ijkl}, C^{ijk} and C^{ij} are the elastic and piezoelectric strain constants and the dielectric permittivity of piezoelectric medium, with their usual symmetry properties [35].

Evidently, the unified eight-field variational principle (2.13) yields, as its Euler-Lagrange equations, all the fundamental differential equations of motion of piezoelectric strained media but the symmetry of stress tensor; and conversely, if the fundamental differential equations are satisfied, the unified variational principle is definitely verified.

The unified variational principle (2.13) recovers those[26,34] deduced from Hamilton's principle and the principle of virtual work, and it includes several earlier variational principles as special cases; the generation of the initial conditions is the novelty of this unified principle.

3. Geometry of Laminae

In the space E, consider a piezoelectric laminae composed of two perfectly conducting lower and face electrodes and (N-2) piezoelectric layers between them. Each constituent may possess distinct but uniform thickness $2h_m$, curvature and electromechanical properties. The regular region of laminae, V+S, with its boundary surface S is referred to the x^i-system of geodesic normal convected coordinates, the midsurface A of first layer $x^3=0$ being taken as the reference surface, such that the corresponding metric tensors of the undeformed laminae are given by

$$g_{\alpha\beta} = \mu^{\lambda}_{\alpha}\mu^{\sigma}_{\beta}a_{\lambda\sigma}\ ,\quad g^{\alpha\beta} = (\mu^{-1})^{\beta}_{\sigma}(\mu^{-1})^{\alpha}_{\lambda}a^{\lambda\sigma}$$
$$g_{\alpha 3} = 0\ ,\quad g_{33} = 1 \quad (3.1)$$

with the shifters of the form

$$\mu^{\alpha}_{\beta} = \delta^{\alpha}_{\beta} - x^3 b^{\alpha}_{\beta},\quad \mu^{\alpha}_{\sigma}(\mu^{-1})^{\sigma}_{\beta} = \delta^{\alpha}_{\beta} \quad (3.2)$$

where $a_{\alpha\beta}$ and $b_{\alpha\beta}$ denote the first and second fundamental forms of the reference surface A, and $c_{\alpha\beta}(=b_{\alpha\sigma}b^{\sigma}_{\beta})$ its third fundamental form. By use of the shifters, the components of a vector field, (X^i,X_i) and $(\bar{X}^i,\bar{X}_i)$, which are respectively referred to the base vectors of laminae space and those of reference surface are associated with one another as

$$X_{\alpha} = \mu^{\beta}_{\alpha}\bar{X}_{\beta}\ ,\quad X^{\alpha} = (\mu^{-1})^{\alpha}_{\beta}\bar{X}^{\beta}\ ;\quad \bar{X}_{\alpha} = (\mu^{-1})^{\beta}_{\alpha}X_{\beta}$$
$$\bar{X}^{\alpha} = \mu^{\alpha}_{\beta}X^{\beta};\quad X_3 = X^3 = \bar{X}_3 = \bar{X}^3 \quad (3.3)$$

Besides, the equations of the form

$$x^3 = -\ h_1,\quad x^3 = 2h - h_1,\quad f(x^1,x^2) = 0 \quad (3.4)$$

define the lower and upper faces, S_{lf} and S_{uf}, and the edge boundary surface S_e of laminae. The reference surface A intersects the edge boundary surface along a Jordan curve C. The bonding surface

between the m-th and (m+1)-th constituent is denoted by $A_{m,m+1}$, and the outward unit vector normal to S_e by ν_i and that to S_f by n_i. In addition to the x^i-system of coordinates, a system of local coordinates x_m^i is introduced which is situated on the midsurface A_m of the m-th constituent. Thus, one reads

$$x_m^\alpha = x^\alpha, \quad x_m^3 = x^3 - z_m \; ; \; m=1,2,\ldots,N \tag{3.5}$$

where z_m is the distance between the surfaces A_m and A. Also, the equations as

$$x_m^3 = 0 \; , \quad x^3 - z_m = 0 \tag{3.6}$$

clearly define the surface A_m and those by

$$\begin{gathered} x_m^3 - h_m = 0 \; , \quad x_{m+1}^3 + h_{m+1} = 0 \\ x^3 - (z_m + h_m) = 0, \quad x^3 - (z_{m+1} - h_{m+1}) = 0 \end{gathered} \tag{3.7}$$

with

$$z_m = \sum_{r=1}^{m} (2 - \delta_{1r} - \delta_{mr})h_r \tag{3.8}$$

the bonding surface $A_{m,m+1}$.

In the region of piezoelectric laminae,

$$2h/R_{min} \ll 1 \tag{3.9}$$

where R_{min} is the least principal radius of curvature of the midsurface A, and the elements of volume dV, of surface dS on S, of area dA on A and of line ds along C are given by

$$\begin{gathered} dV = \sqrt{g}dx^1dx^2dx^3 = dSdx^3 = \mu\, dAdx^3 \\ n_\alpha dS = \mu\nu_\alpha dsdx^3 \end{gathered} \tag{3.10}$$

with

$$\begin{gathered} \mu = |\mu_\beta^\alpha| = (g/a)^{1/2} = 1 - 2cx^3 + (x^3)^2 b \\ a = |a_{\alpha\beta}|, \; b = |b_\beta^\alpha|, \; c = 1/2\, b_\alpha^\alpha, \; g = |g_{ij}| \end{gathered} \tag{3.11}$$

where b and c denote the Gaussian and mean curvature of A, respectively; a more elaborate account of the results recorded can be found (e.g., [35]).

4. Mechanical Displacements, Electric Potential and Resultants for the Piezoelectric Strained Laminae

In mathematical terms, the regular, finite and bounded region of piezoelectric laminae is defined by the fundamental assumption (3.9) which allows one to treat the laminae region as a two-dimensional medium. In addition to (3.9), all the field variables together with their derivatives are assumed to exist, to be single-valued and continuous functions of (x^1,t) in the closure of laminae region with no singularities of any kind, and not to vary widely across the thickness of layers. Accordingly, the fields of incremental mechanical displacements and electric potential which are chosen as a starting point of derivation are expressed by

$$\bar{u}_i^m = v_i^m + x^3 w_i^m \; ; \quad m=1,2,\ldots,N \tag{4.1}$$

and

$$\Phi^m = \varphi^m + x^3\psi^m \; ; \quad m=2,3,\ldots,N-1 \tag{4.2}$$

where $(v_i, w_i, \varphi, \psi)^m$ are unknown a priori, independent and functions of x^α and t, only. In (4.1), v_α^m, v_3^m and w_α^m, and w_3^m represent, in this order, the extensional, flexural and thickness motions of the m-th constituent. Also, the use of (4.1) was shown to account for the coupled motions of laminae, as indicated already by Drumheller and Kalnins[39] and the author[7]. Moreover, in (4.2), m=1 and m=N are excluded, since the electrodes are perfectly conducting. When an alternative potential difference is applied to the electrodes, one reads

$$\Phi^1 = \Phi' \; , \quad \Phi^N = \Phi'' \; , \quad \Phi' = -\Phi'' = \Phi_0 \cos\omega t \tag{4.3}$$

where Φ_0 is a constant and ω the circular frequency.

In the piezoelectric laminae, the constituents are adhered one another and no relative deformation are permitted at their interfaces. Thus, the continuity of mechanical displacements and electric potential on, and that of tractions and surface charge across, the bonding interfaces $A_{m,m+1}$ are maintained. First, using (4.1), the continuity of incremental mechanical displacements is written as

$$\begin{gathered} v_i^m + (z_m + h_m)w_i^m = v_i^{m+1} + (z_{m+1} - h_{m+1})w_i^{m+1} \\ \text{no sum over m; } m=1,2,\ldots,N \text{ ; on } A_{m,m+1}XT \end{gathered} \tag{4.4}$$

This represents 3(N-1) constraints and reduces the number of the independent functions of displacements, 6N, in (4.1) to 3(N+1). The independent functions are chosen as

$$v_i^1, \; w_i^m \; ; \; m=1,2,\ldots,N \tag{4.5}$$

and the rest of the displacement functions is expressed by

$$v_i^m = v_i^1 + \sum_{r=1}^{m} z_{rm} w_i^r \; ; \quad m=2,3,\ldots,N \tag{4.6}$$

with

$$z_{rm} = (2 - \delta_{1r} - \delta_{mr})h_{(r)} - \delta_{mr} z_{(m)} \tag{4.7}$$

in terms of them.

Next, the continuity of electric potential is similarly expressed by

$$\varphi^m + (z_m + h_m)\psi^m = \varphi^{m+1} + (z_{m+1} - h_{m+1})\psi^{m+1} \tag{4.8}$$

no sum over m; m=2,3,...,N-1,
on $A_{m,m+1}{}^{XT}$ (4.8)

and

$$\Phi' = \varphi^2 + (z_2 - h_2)\psi^2 \text{ on } A_{1,2}{}^{XT}$$
$$\Phi'' = \varphi^{N-1} + (z_{N-1} + h_{N-1})\psi^{N-1} \text{ on } A_{N-1,N}{}^{XT} \quad (4.9)$$

In view of the constraints (4.8), the (N-3) independent functions of electric potential are chosen as

$$\psi^m \; ; \quad m=2,3,\ldots,N-2 \quad (4.10)$$

The dependent functions of electric potential are expressed by

$$\varphi^2 = \Phi' - (z_2 - h_2)\psi^2$$
$$\varphi^m = \Phi' + \sum_{r=2}^{m} z_{rm}\psi^r \; ; \quad m=2,3,\ldots,N-2$$
$$\varphi^{N-1} = -(z_{N-1}/h_{N-1})\Phi'' + (1 + z_{N-1}/h_{N-1})\sum_{r=2}^{N-2} h_r\psi^r$$
$$\psi^{N-1} = (1/h_{N-1})(\Phi'' - \sum_{r=2}^{N-2} h_r\psi^r) \quad (4.11)$$

in terms of the independent functions (4.10).

Evidently, the linear representation (4.1) and (4.2) and the gradient equations (2.2) imply a distribution of mechanical strain for each constituent as

$$s_{ij} = \sum_{r=0}^{R} (x^3)^r [{}_r s_{ij}(x^\alpha, t)] \quad (4.12)$$

and that of electric field for each layer as

$$E_i = \sum_{r=0}^{R} (x^3)^r [{}_r E_i(x^\alpha, t)] \quad (4.13)$$

of which the explicit expressions are obtained in the next section.

Now, in accordance with the linear representation above, various field quantities are averaged over the thickness interval of each constituent for the subsequent development. Thus, the two-dimensional incremental resultants of stress are defined by

$$[N^{\alpha\beta}, M^{\alpha\beta}, K^{\alpha\beta}] = \int_Z [1, x^3, (x^3)^2] t^{\alpha\beta} \mu dx^3 \quad (4.14)$$
$$[(Q^\alpha, R^\alpha); N^{33}] = \int_Z [(1, x^3)t^{\alpha 3}; t^{33}] \mu dx^3$$

and those of initial stress by

$$[N_o^{\alpha\beta}, \ldots, R_o^\alpha, N_o^{33}] = \int_Z [t_o^{\alpha\beta}, \ldots, t_o^{33}] \mu dx^3 \quad (4.15)$$

those of acceleration by

$$A^i = \mu_o \ddot{v}^i + \mu_1 \ddot{w}^i \; , \quad B^i = \mu_1 \ddot{v}^i + \mu_2 \ddot{w}^i \quad (4.16)$$

with

$$\mu_n = I_n - 2cI_{n+1} + bI_{n+2} \quad (4.17)$$

where

$$I_n = \int_Z (x^3)^n dx^3 = [(z+h)^{n+1} - (z-h)^{n+1}]/(n+1)$$
$$n=0,1,\ldots, \quad (4.18)$$

those of traction by

$$(q^\alpha, p^\alpha) = (\mu\mu_\beta^\alpha t^{3\beta}), \quad (q^3, p^3) = \mu t^{33}$$
$$\text{at } (x^3 = z+h, z-h) \quad (4.19)$$

and

$$(r_o^\alpha, s_o^\alpha) = \{\mu t_o^{3\beta}[v^\alpha{}_{:\beta} - b_\beta^\alpha v_3 + x^3(w^\alpha{}_{:\beta} - b_\beta^\alpha w_3)$$
$$+ \mu t_o^{33} w^\alpha\} \quad \text{at } (x^3 = z+h, z-h) \quad (4.20)$$
$$(r_o^3, s_o^3) = \{\mu t_o^{3\alpha}[v_{3,\alpha} + b_\alpha^\beta v_\beta + x^3(w_{3,\alpha} + b_\alpha^\beta w_\beta)]$$
$$+ \mu t_o^{33} w_3\} \quad \text{at } (x^3 = z+h, z-h)$$

those of loads by

$$(N_\star^\alpha, M_\star^\alpha) = \int_Z \tau_\star^\beta \mu_\beta^\alpha (1, x^3)\mu dx^3$$
$$(N_\star^3, M_\star^3) = \int_Z \tau_\star^3 (1, x^3)\mu dx^3 \quad (4.21)$$

and

$$(S_\star^\alpha, P_\star^\alpha) = (\mu\mu_\beta^\alpha \tau_\star^\beta) \text{ and } (S_\star^3, P_\star^3) = (\mu\tau_\star^3)$$
$$\text{at } (x^3 = z+h, z-h) \quad (4.22)$$

and

$$l^i = q^i - p^i \; , \quad l_o^i = r_o^i - s_o^i \quad (4.23)$$
$$m^i = (z+h)q^i - (z-h)p^i; \quad m_o^i = (z+h)r_o^i - (z-h)s_o^i$$

Besides, the two-dimensional resultants of electrical displacements in the form

$$(F^i, G^i) = \int_Z (1, x^3) D^i \mu dx^3 \quad (4.24)$$

those of surface charge by

$$(d, f) = (\mu D^3) \text{ at } (x^3 = z+h, z-h) \quad (4.25a)$$

and

$$D = (z + h)d, \quad F = (z - h)e \quad (4.25b)$$

and those of edge-surface charge by

$$(F, G) = \int_Z (1, x^3)\sigma\mu dx^3 \quad (4.26)$$

are introduced. In (4.14)-(4.27), the resultants of stress, initial stress and electric displacements are measured per unit length of the coordinate curves on A, those of acceleration, surface load and surface charge per unit area of A, and those of edge-load and edge-surface charge per unit length of C. Moreover, in terms of the foregoing definitions, the continuity of tractions and that of surface charge by

$$(q^i + r_o^i)^m - (p^i + s_o^i)^{m+1} = 0 \; ; \quad d^m - f^{m+1} = 0$$
$$\text{on } A_{m,m+1}{}^{XT} \quad (4.27)$$

are given; the resultants can be similarly referred to the A_m of each constituent in place of A,[35].

5. Distributions of Strain and Electric Field. Macroscopic Constitutive Equations

The components of incremental strain of order (r) are obtained by use of the appropriate term of the unified principle (2.13), namely,

$$\delta J^{13}_{13} = \int_T dt \int_A \sum_{r=1}^{N} \{ \int_Z [s_{ij} - 1/2\,(u_{i;j} + u_{j;i})] \delta t^{ij} \}^{(r)} \mu dA dx^3 = 0 \tag{5.1}$$

By inserting (4.1) into this equation, and then performing the integrals over the entire thickness of piezoelectric laminae, recalling the resultants of stress (4.14), one finally obtains the distribution of incremental strain in a variational form as

$$\begin{aligned} \delta J^{13}_{13} = \int_T dt \int_A \sum_{r=1}^{N} [& ({}_0S_{\alpha\beta} - e_{\alpha\beta})\delta N^{\alpha\beta} \\ & + ({}_1S_{\alpha\beta} - \varepsilon_{\alpha\beta})\delta M^{\alpha\beta} + ({}_2S_{\alpha\beta} - \gamma_{\alpha\beta})\delta K^{\alpha\beta} \\ & + ({}_0S_{\alpha 3} - e_{\alpha 3})\delta Q^{\alpha} + ({}_1S_{\alpha 3} - \varepsilon_{\alpha 3})\delta R^{\alpha} \\ & + ({}_0S_{33} - e_{33})\delta N^{33}]^{(r)} dA = 0 \end{aligned} \tag{5.2}$$

This equation leads, as its Euler-Lagrange equations, to

$${}_0S_{ij} = e_{ij}\,, \quad {}_1S_{ij} = \varepsilon_{ij}\,, \quad {}_2S_{ij} = \gamma_{ij} \tag{5.3a}$$

where

$$\begin{aligned} e'_{\alpha\beta} &= 1/2\,(v_{\alpha:\beta} + v_{\beta:\alpha} - 2b_{\alpha\beta}v_3)' \\ e^{(m)}_{\alpha\beta} &= e'_{\alpha\beta} + 1/2 \sum_{r=1}^{m} z_{rm}(w_{\alpha:\beta} + w_{\beta:\alpha} - 2b_{\alpha\beta}w_3)^{(r)} \\ e'_{\alpha 3} &= 1/2\,(v_{3,\alpha} + b^{\sigma}_{\alpha}v_{\sigma} + w_{\alpha})'\,; \quad m=2,3,\ldots,N \\ e^{(m)}_{\alpha 3} &= e'_{\alpha 3} + 1/2\,[w^{(m)}_{\alpha} + \sum_{r=1}^{m} z_{rm}(w_{3,\alpha} + b^{\sigma}_{\alpha}w_{\sigma})^{(r)}] \\ e^{(m)}_{33} &= w^{(m)}_3\,; \quad m=1,2,\ldots,N \\ \varepsilon'_{\alpha\beta} &= 1/2\,(-b^{\sigma}_{\alpha}v_{\sigma:\beta} - b^{\sigma}_{\beta}v_{\sigma:\alpha} + 2c_{\alpha\beta}v_3 + w_{\alpha:\beta} \\ &\qquad + w_{\beta:\alpha} - 2c_{\alpha\beta}w_3)' \\ \varepsilon^{(m)}_{\alpha\beta} &= 1/2[(-b^{\sigma}_{\alpha}v_{\sigma:\beta} - b^{\sigma}_{\beta}v_{\sigma:\alpha} + 2c_{\alpha\beta}v_3)' \\ &\qquad + (w_{\alpha:\beta} + w_{\beta:\alpha} - 2c_{\alpha\beta}w_3)^{(m)} \\ &\qquad + \sum_{r=1}^{m} z_{rm}(-b^{\sigma}_{\alpha}w_{\sigma:\beta} - b^{\alpha}_{\beta}w_{\sigma:\alpha} + 2c_{\alpha\beta}w_3)^{(r)}] \\ \varepsilon^{(m)}_{\alpha 3} &= 1/2\,w_{3,\alpha}\,; \quad \varepsilon^{(m)}_{33} = 0\,; \quad m=1,2,\ldots,N \\ \gamma^{(m)}_{\alpha\beta} &= 1/2\,(-b^{\sigma}_{\alpha}w_{\sigma:\beta} - b^{\sigma}_{\beta}w_{\sigma:\alpha} + 2c_{\alpha\beta}w_3)^{(m)} \end{aligned} \tag{5.3b}$$

$$\gamma^{(m)}_{\alpha 3} = \gamma^{(m)}_{33} = 0\,; \quad m=1,2,\ldots,N \tag{5.3b}$$

In deriving (5.3), the covariant derivatives of the displacement vector are expressed with respect to surface metrics by means of the identities as

$$u_{\alpha;\beta} = u^{\sigma}_{\alpha}(\bar{u}_{\sigma:\beta} - b_{\sigma\beta}\bar{u}_3), \ldots, \; u_{3;3} = \bar{u}_{3,3} \tag{5.4}$$

Here, an overbar indicates the displacement components, as defined in (3.3).

In a similar manner, the distribution of electric field is found by use of the part of (2.13) in the form

$$\delta J^{23}_{23} = \int_T dt \int_A \sum_{r=2}^{N-1} \{ \int_Z [\delta D^i (E_i + \Phi_{,i})] \}^{(r)} \mu dA dx^3 = 0 \tag{5.5}$$

This yields the distribution of electric field in a variational form by

$$\begin{aligned} \delta J^{23}_{23} = \int_T dt \int_A \sum_{r=2}^{N-1} [& ({}_0E_i - e_i)\delta F^i \\ & + ({}_1E_i - \varepsilon_i)\delta G^i]^{(r)} dA = 0 \end{aligned} \tag{5.6}$$

and

$${}_0E_i = e_i\,, \quad {}_1E_i = \varepsilon_i \tag{5.7}$$

Here, the denotations by

$$\begin{aligned} e^{(2)}_{\alpha} &= (z_2 - h_2)\phi^{(2)}_{,\alpha} \\ e^{(m)}_{\alpha} &= -\sum_{r=1}^{m} z_{rm}\phi^{(r)}_{,\alpha}\,; \quad m=3,4,\ldots,N-2 \\ e^{(N-1)}_{\alpha} &= -(1 + z_{N-1}/h_{N-1})\sum_{r=2}^{N-2} h_r\phi^{(r)}_{,\alpha} \\ e^{(m)}_3 &= \phi^{(m)}\,; \quad m=2,3,\ldots,N-2 \\ e^{(N-1)}_3 &= 1/h_{N-1}(\Phi'' - \sum_{r=2}^{N-2} h_r\phi^{(r)}) \\ \varepsilon^{(m)}_{\alpha} &= -\phi^{(m)}_{,\alpha}\,; \quad m=2,3,\ldots,N-2 \\ \varepsilon^{(N-1)}_{\alpha} &= 1/h_{N-1}\sum_{r=2}^{N} h_r\phi^{(r)}_{,\alpha} \\ e'_i &= e''_i = \varepsilon'_i = \varepsilon''_i = \varepsilon^{(m)}_3 = 0\,; \quad m=2,3,\ldots,N \end{aligned} \tag{5.8}$$

are introduced.

Now, the distributions (5.3) and (5.7) are substituted into the constitutive part of (2.13), and then the integrations are carried out with respect to the thickness coordinate and the resultants of stress and electric displacements are used wherever feasible, with the results in variational form by

$$\begin{aligned} \delta J^{12}_{12} = \int_T dt \int_A \sum_{r=1}^{N} [& (N^{\alpha\beta} - N^{\alpha\beta}_C)\delta e_{\alpha\beta} + (M^{\alpha\beta} - M^{\alpha\beta}_C)\delta\varepsilon_{\alpha\beta} \\ & + (K^{\alpha\beta} - K^{\alpha\beta}_C)\delta\gamma_{\alpha\beta} + (Q^{\alpha} - Q^{\alpha}_C)\delta e_{\alpha 3} + (R^{\alpha} - R^{\alpha}_C)\delta\varepsilon_{\alpha 3} \\ & + (N^{33} - N^{33}_C)\delta e_{33}]^{(r)} dA = 0 \end{aligned} \tag{5.9}$$

and

$$\delta J^{22}_{22} = \int_T dt \int_A \sum_{r=1}^{N} [(F^i - F^i_c)\delta e_i + (G^i - G^i_c)\delta \epsilon_i]^{(r)} dA = 0 \tag{5.10}$$

The Euler-Lagrange equations of (5.9) and (5.10) are the macroscopic constitutive relations in the form

$$N^{\alpha\beta} - N^{\alpha\beta}_c = 0, \quad M^{\alpha\beta} - M^{\alpha\beta}_c = 0, \quad K^{\alpha\beta} - K^{\alpha\beta}_c = 0$$
$$Q^{\alpha} - Q^{\alpha}_c = 0, \quad R^{\alpha} - R^{\alpha}_c = 0, \quad N^{33} - N^{33}_c = 0 \tag{5.11}$$

and

$$F^i - F^i_c = 0, \quad G^i - G^i_c = 0 \quad \text{on AXT} \tag{5.12}$$

Here,

$$(N^{\alpha\beta}_c, M^{\alpha\beta}_c, K^{\alpha\beta}_c) = (\underset{\sim}{C}_0, \underset{\sim}{C}_1, \underset{\sim}{C}_2)^{\alpha\beta kl}(\underset{\sim}{s}_{kl})^T - (\underset{\sim}{C}_0, \underset{\sim}{C}_1, \underset{\sim}{C}_2)^{k\alpha\beta}(\underset{\sim}{E}_k)^T$$

$$(Q^{\alpha}_c, R^{\alpha}_c) = (\underset{\sim}{C}_0, \underset{\sim}{C}_1)^{\alpha 3kl}(\underset{\sim}{s}_{kl})^T - (\underset{\sim}{C}_0, \underset{\sim}{C}_1)^{k\alpha 3}(\underset{\sim}{E}_k)^T$$

$$N^{33}_c = C^{33kl}_0(\underset{\sim}{s}_{kl})^T - C^{k33}_0(\underset{\sim}{E}_k)^T \quad \text{on AXT} \tag{5.13a}$$

and

$$(F^i_c, G^i_c) = (\underset{\sim}{C}_0, \underset{\sim}{C}_1)^{ijk}(\underset{\sim}{s}_{jk})^T + (\underset{\sim}{C}_0, \underset{\sim}{C}_1)^{ik}(\underset{\sim}{E}_k)^T$$

In the above equations,

$$(\underset{\sim}{s}_{kl}) = (e_{kl}, \epsilon_{kl}, \gamma_{kl}), \quad (\underset{\sim}{E}_k) = (e_k, \epsilon_k, 0) \tag{5.14}$$

are defined. Also, the elastic stiffenesses by

$$\underset{\sim}{C}^{ij\ldots k}_n = (C_n, C_{n-1}, C_{n-2})^{ij\ldots k} \tag{5.15a}$$

with

$$C^{ij\ldots k}_n = \mu_n C^{ij\ldots k} \tag{5.15b}$$

are introduced.

6. Governing Equations of Incremental Motion

In this section, within the order of approximation of the linear representation (4.1) and (4.2), the macroscopic stress equations of incremental motion, the macroscopic charge equations of electrostatics and the associated natural boundary and initial conditions are systematically derived. In the derivation, Mindlin's method of reduction is followed (see, e.g., [40]), and the results are expressed in both variational and differential forms. Also, the governing equations of piezoelectric biased laminae are fully stated.

To begin with, the first term of (2.13) is written in the form

$$\delta J^{11}_{11} = \int_T dt \int_A \sum_{r=1}^{N} \{ \int_Z [(t^{ij} + t^{ik}_o u^j{}_{;k})_{;i} - \rho a^j]\delta u_j]^{(r)} \} \mu dA dx^3 = 0 \tag{6.1}$$

for all the constituents of piezoelectric laminae. By substituting (4.1) into this variational integral, using various relations between space and surface tensors and their derivatives, performing integrations with respect to the thickness coordinate and recalling the resultants of stress, acceleration and load in Section 4, one finally arrives at the variational equation of incremental motion as

$$\delta J^{11}_{11} = \int_T dt \int_A \sum_{r=1}^{N} [(V^i + U^i_o + l^i + l^i_o - \rho A^i)\delta v_i + (W^i + T^i_o + m^i + m^i_o - \rho B^i)\delta w_i]^{(r)} dA = 0 \tag{6.2}$$

where

$$V^{\alpha} = V^{\beta\alpha}{}_{:\beta} - b^{\alpha}_{\sigma} Q^{\sigma}, \quad V^3 = V^{\alpha 3}{}_{:\alpha} + b_{\alpha\beta} V^{\alpha\beta}$$
$$W^{\alpha} = W^{\beta\alpha}{}_{:\beta} - Q^{\alpha}, \quad W^3 = W^{\alpha 3}{}_{:\alpha} - N^{33} + b_{\alpha\beta} W^{\alpha\beta} \tag{6.3}$$

and

$$U^{\alpha}_o = U^{\beta\alpha}_o{}_{:\beta} - b^{\alpha}_{\sigma} U^{\sigma 3}_o, \quad U^3_o = U^{\alpha 3}_o{}_{:\alpha} + b_{\alpha\beta} U^{\alpha\beta}_o$$
$$T^{\alpha}_o = T^{\beta\alpha}_o{}_{:\beta} - b^{\alpha}_{\beta} T^{\beta 3}_o - Q^{\beta}_o (v^{\alpha}{}_{:\beta} - b^{\alpha}_{\beta} v_3) + (R^{\beta}_{o:\beta} - N^{33}) w^{\alpha}$$
$$T^3_o = T^{\alpha 3}_o{}_{:\alpha} + b_{\alpha\beta} T^{\alpha\beta}_o - (R^{\alpha}_{o:\alpha} + N^{33}_o) w^3 - Q^{\alpha}_o (v_{3,\alpha} + b^{\beta}_{\alpha} w_{\beta}) \tag{6.4}$$

In these equations, the denotations of the form

$$V^{\alpha\beta} = N^{\alpha\beta} - b^{\beta}_{\sigma} M^{\alpha\sigma}, \quad W^{\alpha\beta} = M^{\alpha\beta} - b^{\beta}_{\sigma} K^{\alpha\sigma}$$
$$V^{\alpha 3} = Q^{\alpha}, \quad W^{\alpha 3} = R^{\alpha} \tag{6.5a}$$

and

$$U^{\alpha\beta}_o = N^{\alpha\sigma}_o (v^{\beta}{}_{:\sigma} - b^{\beta}_{\sigma} v_3) + M^{\alpha\sigma}_o (w^{\beta}{}_{:\sigma} - b^{\beta}_{\sigma} w_3) + Q^{\alpha 3}_o w^{\beta}$$
$$U^{\alpha 3}_o = Q^{\alpha 3}_o w_3 + N^{\alpha\beta}_o (v_{3,\beta} + b^{\sigma}_{\beta} v_{\sigma}) + M^{\alpha\beta}_o (w_{3,\beta} + b^{\sigma}_{\beta} w_{\sigma})$$
$$T^{\alpha\beta}_o = M^{\alpha\sigma}_o (v^{\beta}{}_{:\sigma} - b^{\beta}_{\sigma} v_3) + K^{\alpha\sigma}_o (w^{\beta}{}_{:\sigma} - b^{\beta}_{\sigma} w_3)$$
$$T^{\alpha 3}_o = M^{\alpha\beta}_o (v_{3,\beta} + b^{\sigma}_{\beta} v_{\sigma}) + K^{\alpha\beta}_o (w_{3,\beta} + b^{\sigma}_{\beta} w_{\sigma}) \tag{6.5b}$$

are introduced in terms of (4.14)-(4.27). As a last step, using (4.6) and considering the continuity conditions of tractions (4.27), the macroscopic stress equation of incremental motion is expressed in variational form by

$$\delta J^{11}_{11} = \int_T dt \int_A \{ [\sum_{r=1}^{N} (r^i)^{(r)} + b^{i\alpha} - c^{i,}]\delta v_i + [\Pi^{i,} + h' \sum_{r=2}^{N} (r^i)^{(r)} + h'(c^{i,} + b^{i\alpha}{}_{,\alpha}) w^{i,} + \sum_{m=2}^{N-1} [(\Pi^i)^{(m)} + \sum_{r=m}^{N} z_{mr} (r^i)^{(r)} + 2h_m b^{i\alpha}]\delta w^{(m)}_i + [\Pi^{i\alpha} - (2h - h' - 2h'') r^{i\alpha} + 2h'' b^{i\alpha}]\delta w^{\alpha}_i \} dA = 0 \tag{6.6a}$$

with

$$r^i = V^i + U_o^i - \rho A^i, \quad \Pi^i = W^i + T_o^i - \rho B^i$$
$$b^i = q^i + r_o^i, \quad c^i = p^i + s_o^i \qquad (6.6b)$$

in terms of the variations of independent displacement functions (4.1). From (6.6), the macroscopic stress equations follow in differential form as

$$\sum_{r=1}^{N}(V^i + U_o^i)^{(r)} + b^{i}{}'' - c^{i}{}' = \sum_{r=1}^{N}(\rho A^i)^{(r)}$$

$$(W^i + T_o^i)' + h'\sum_{r=2}^{N}(V^i + U_o^i)^{(r)} + h'(c' + b'') = (\rho B^i)' + h'\sum_{r=2}^{N}(\rho A^i)^{(r)}$$

on AXT

$$(W^i + T_o^i)^{(m)} + \sum_{r=m}^{N} z_{mr}(V^i + U_o^i)^{(r)} + 2h_m b^{i}{}'' = (\rho B^i)^{(m)} + \sum_{r=m}^{N} z_{mr}(\rho A^i)^{(r)}$$

$$(W^i + T_o^i)'' - (2h - h' - 2h'')(V^i + U_o^i)'' + 2h''b'' = (\rho B^i)'' - (2h - h' - 2h'')(\rho A^i)'' \qquad (6.7)$$

This equation or (6.6) represents 3(N+1) equations for the piezoelectric biased laminae. In deriving (6.6) and (6.7), details of lengthy computations are omitted; they are, however, given in a recent report[35].

In a similar manner, to derive the macroscopic charge equations of electrostatics, from (2.13), the variational volume integral of the form

$$\delta J_{21}^{21} = -\int_T dt \int_A \sum_{r=2}^{N-1}[\int_Z D^i{}_{;i}\delta\Phi]^{(r)}\mu dA dx^3 = 0 \qquad (6.8)$$

is evaluated for all the layers of piezoelectric laminae. In doing so, (4.11) is inserted into this equation and the integrations are carried out with respect to the thickness coordinate, and then the variational integral (6.8) is expressed by

$$\delta J_{21}^{21} = \int_T dt \int_A \sum_{r=1}^{N}[(F^{\alpha}{}_{:\alpha} + d - f)\delta\varphi + (G^{\alpha}{}_{:\alpha} - F^3 + D - F)\delta\psi]^{(r)} dA = 0 \qquad (6.9)$$

with $\delta\Phi' = \delta\Phi'' = 0$. This equation is now written with respect to the variations of independent functions of electric potential (4.10) as

$$\delta J_{21}^{21} = \int_T dt \int_A \{\sum_{m=2}^{N-3}[x_m^{\alpha}{}_{:\alpha} + x_m^3 + \sum_{r=m}^{N-1} x_{mr}F_{r:\alpha}^{\alpha}]\delta\psi^{(m)} + [x_{N-2}^{\alpha}{}_{:\alpha} - (z_{N-2} - h_{N-2})F^{(N-2)\alpha}{}_{:\alpha} + (1 + z_{N-1}/h_{N-1})h_{N-2}F_{N-1}^{\alpha}{}_{:\alpha} + x_{N-2}^3]\delta\psi^{(N-2)}\}dA = 0 \qquad (6.10a)$$

with the denotations by

$$x_{mr} = [2 - \delta_{mr} + (-1 + z_{N-1}/h_{N-1})\delta_{N-1,r}]h_m - \delta_{mr}z_m$$

$$x_m^{\alpha} = G_m^{\alpha} - h_m/h_{N-1}G_{N-1}^{\alpha}, \quad x_m^3 = -F_m^3 + h_m/h_{N-1}F_{N-1}^3 \qquad (6.10b)$$

where the continuity of surface charge (4.27) is considered. The Euler-Lagrange equations of (6.10) are readily written as

$$x_m^{\alpha}{}_{:\alpha} + x_m^3 + \sum_{r=m}^{N-1} x_{mr}F_r^{\alpha}{}_{:\alpha} = 0 \ ; \ m=2,3,\ldots,N-3$$

$$x_{N-2}^{\alpha}{}_{:\alpha} - (z_{N-2} - h_{N-2})F_{N-2}^{\alpha}{}_{:\alpha} + x_{N-2}^3 + (1 + z_{N-1}/h_{N-1})h_{N-2}F_{N-1}^{\alpha}{}_{:\alpha} = 0 \quad \text{on AXT} \qquad (6.11)$$

which represent (N-3) equations. Thus, the macroscopic charge equations of electrostatics are expressed by (6.10) in variational form and by (6.11) in differential form.

Paralleling the derivation of the macroscopic divergence equations (6.1)-(6.11), the mechanical and electrical, natural boundary conditions of piezoelectric biased laminae are obtained by use of the variational principle (2.13). The tractions of biased laminae are prescribed on a part $S_t(=C_t XT)$ of S_e and S_{lf} and the surface charges on only S_e. To begin with, consider the pertinent term of (2.13) for the mechanical boundary conditions, namely,

$$\delta I_{11}^{11} = \int_T dt \oint_{C_t}\{\sum_{r=1}^{N}\int_Z[\tau_*^j - \nu_{\alpha}(t^{\alpha j} + t_o^{\alpha k}u^j{}_{;k})]\delta u_j\}^{(r)}\mu\, ds dx^3 + \int_T\int_{S_{lf}}\{[\tau_*^j - n_3(t^{3j} + t_o^{3k}u^j{}_{;k})]\delta u_j\}'\mu dA = 0 \qquad (6.12)$$

After evaluation as before, this equation leads to the natural boundary conditions of tractions in variational form as

$$\delta I_{11}^{11} = \int_T dt \oint_{C_t}\{\sum_{r=1}^{N}[N_*^j - \nu_{\alpha}(V^{\alpha j} + U_o^{\alpha j})]^{(r)}\delta v_j' + \{[M_*^j - \nu_{\alpha}(W^{\alpha j} + T_o^{\alpha j})]' + h'\sum_{r=2}^{N}[N_*^j - \nu_{\alpha}(V^{\alpha j} + U_o^{\alpha j})]^{(r)}\}\delta w_j' + \sum_{m=1}^{N-1}\{[M_*^j - \nu_{\alpha}(W^{\alpha j} + T_o^{\alpha j})]^{(m)} + \sum_{r=m}^{N} z_{mr}[N_*^j - \nu_{\alpha}(V^{\alpha j} + U_o^{\alpha j})]^{(r)}\}\delta w_j^{(m)} + \{[M_*^j - \nu_{\alpha}(W^{\alpha j} + T_o^{\alpha j})]'' - (2h - h' - h'')[N_*^j - \nu_{\alpha}(V^{\alpha j} + U_o^{\alpha j})]''\}\delta w_j\}ds + \int_T dt\int_{S_{lf}}[(p_*^j{}' - c^{j\prime})\delta v_j' + h'(p_*^j{}' - c^{j\prime})\delta w_j']dA = 0 \qquad (6.13)$$

which yields readily the boundary conditions of tractions as follows

$$\sum_{r=1}^{N}[N_*^j - \nu_{\alpha}(V^{\alpha j} + U_o^{\alpha j})]^{(r)} = 0 \qquad (6.14a)$$

$$[M_*^j - \nu_\alpha(W^{\alpha j} + T_0^{\alpha j})]' + h_1\sum_{r=2}^{N}[N_*^j - \nu_\alpha(V^{\alpha j} + U_0^{\alpha j})]^{(r)}$$

$$= 0\ ; \qquad [M_*^j - \nu_\alpha(W^{\alpha j} + T_0^{\alpha j})]^{(m)}$$

$$+ \sum_{r=m}^{N} z_{mr}[N_*^j - \nu_\alpha(V^{\alpha j} + U_0^{\alpha j})]^{(r)} = 0 \qquad m=2,3,\ldots,N-1$$

$$[M_*^j - \nu_\alpha(W^{\alpha j} + T_0^{\alpha j})]''$$

$$- (2h - h' - 2h'')[N_*^j - \nu_\alpha(V^{\alpha j} + U_0^{\alpha j})]'' = 0 \qquad \text{on } C_t XT \qquad (6.14a)$$

and

$$P_*^j - (p^j + s_0^j)' = 0 \quad \text{on } S_{1f}XT \qquad (6.14b)$$

in differential form as well. Besides, the natural boundary conditions of mechanical displacements by

$$v_i' - v_i^{*\prime} = 0\ , \quad w_i^m - w_{i*}^m = 0 \text{ on } C_u XT$$
$$v_i'' - \delta_i^{*\prime\prime} = 0\ , \quad w_i'' - \gamma_i^{*\prime\prime} = 0 \text{ on } S_{uf}XT \qquad (6.15)$$

are recorded.

In like manner, substituting (4.9) into the variational surface integral of the form

$$\delta I_{21}^{21} = \int_T dt \oint_C \sum_{r=2}^{N-1}[\int_Z(\nu_\alpha D^\alpha - \sigma_*)(\delta\varphi + x^3\delta\psi)]^{(r)}\mu ds dx^3 = 0 \qquad (6.16)$$

and evaluating it, one reads

$$\delta I_{21}^{21} = \int_T dt \oint_C \sum_{r=2}^{N-1}[(\nu_\alpha F^\alpha - F_*)\delta\varphi + (\nu_\alpha G^\alpha - G_*)\delta\psi]^{(r)} ds = 0 \qquad (6.17)$$

in terms of the resultants (4.24)-(4.26). By use of (4.11), the natural boundary conditions of surface charge are expressed in variational form by

$$\delta I_{21}^{21} = \int_T dt \oint_C \{\sum_{m=2}^{N-3}[H_*^{(m)} - \nu_\alpha(x_m^\alpha + \sum_{r=m}^{N-1} x_{mr}F_r^\alpha)]\delta\varphi^m + \{H_*^{N-2} - \nu_\alpha[x_{N-2}^\alpha + (z_{N-2} - h_{N-2})F_{N-2}^\alpha + (1 + z_{N-1}/h_{N-1})h_{N-2}F_{N-1}^\alpha]\}\delta\psi^{(N-1)}\}ds = 0$$

and those in differential form by

$$H_*^{(m)} - \nu_\alpha(x_m^\alpha + \sum_{r=m}^{N-1} x_{mr}F_r^\alpha) = 0\ ; \quad m=2,3,\ldots,N-3$$

$$H_*^{N-2} - \nu_\alpha[x_{N-2}^\alpha + (z_{N-2} - h_{N-2})F_{N-2}^\alpha + (1 + z_{N-1}/H_{N-1})h_{N-2}F_{N-1}^\alpha] = 0$$

$$\text{along } CXT \qquad /6.19)$$

Here,

$$H_*^m = G_*^m - (h_m/h_{N-1})G_*^{N-1} + \sum_{r=m}^{N-1} x_{mr}F_*^r\ ; \quad m=2,3,..N-3 \qquad (6.20)$$

$$H_*^{N-2} = G_*^{N-2} - (h_{N-2}/h_{N-1})G_*^{N-1} - (z_{N-2} - h_{N-2})F_*^{N-2} + (1 + z_{N-1}/h_{N-1})h_{N-2}F_*^{N-1} \qquad (6.20)$$

with

$$(F_*, G_*)^m = \int_{Z_m}(1, x^3)\sigma_*^{(m)}\mu dx^3 \qquad (6.21)$$

are introduced.

Lastly, an evaluation of the volume integrals I_i^i of (2.13) yields the natural initial conditions of mechanical displacements and electric potential by

$$v_i^!(x^\alpha,t_0) - \alpha_i^{*\prime}(x^\alpha) = 0,\ \dot{v}_i^!(x^\alpha,t_0) - \beta_i^{*\prime}(x^\alpha) = 0$$
$$w_i^m(x^\alpha,t_0) - \xi_i^{*m}(x^\alpha) = 0,\ \dot{w}_i^m(x^\alpha,t_0) - \xi_i^{*m}(x^\alpha) = 0$$
$$m=1,2,\ldots,N \qquad (6.22)$$

$$\psi^m(x^\alpha,t_0) - \eta_i^{*m}(x^\alpha) = 0;\ m=2,3,\ldots,N-1$$
$$\text{on } A(t_0) \qquad (6.23)$$

where $\alpha_i^*,\ldots,\eta_i^*$ are given functions.

Up to now, the set of two-dimensional, approximate equations of piezoelectric biased laminae is systematically and consistently derived by means of the unified variational principle (2.13) together with the linear representation (4.1) and (4.2). The electroelastic equations are given both in variational and differential forms at the perturbed state. Similarly the governing equations can be derived at the unperturbed state for the static behavior of laminae at the unperturbed state. This is recorded in[35].

7. A Direct Method of Solution

In this section a general algorithm is pointed out which is based upon Kantorovich's method for the fields of mechanical displacements and electric potential, as an alternative of the macromechanical analysis of piezoelectric biased laminae. The algorithm starts from the integral principle of (2.3) in lieu of its Euler-Lagrange differential equations and it rests entirely upon a selection of the fields for each constituent under the ad hoc assumptions for the piezoelectric region in Section 4. The method can be readily and succesfully employed by means of high-speed digital computers for the macromechanical analysis.

To begin with, the fields of incremental mechanical displacements and electric potential are expressed by

$$\bar{u}_i^m(x^j,t) = \sum_{p+q+r=0}^{R}[f_i^{pqr}(x^\alpha,t)\ f_r(x^3)]^m$$
$$\Phi^m(x^i,t) = \sum_{p+q+r=0}^{R}[g_{pqr}(x^\alpha,t)\ g_r(x^3)]^m \qquad (7.1a)$$

with

$$[f_i^{pqr},\ g_{pqr}]^m = [\alpha_{pqr}(t)u_i^{pq}(x^\alpha),\ \beta_{pqr}(t)\Phi_{pq}(x^\alpha)]^m \qquad (7.1b)$$

$$[f_r(x^3),\ g_r(x^3)] = (x^3)^r \qquad (7.1b)$$

Here, (α_{pqr} and β_{pqr}) are the functions to be determined, whereas (u_i^{pq}, Φ_{pq}) are the approximating functions to be chosen appropriately in order to satisfy all or some of the given boundary conditions; the rest of constraint conditions can be taken into account through Lagrange multipliers as illustrated by the author[35]. The approximating functions should be selected as simple as possible so that operations involving them can be carried out easily.

With the help of (7.1), the evaluation of the variational integral (2.3) leads readily to a system of ordinary differential equations in terms of $\alpha_{rqr}(t)$ and $\beta_{pqr}(t)$. The system of equations can be reduced to that of nonlinear algebraic equations for the case when vibrations and wave propagation are considered in the piezoelectric biased laminae. The results with some applications are presented in detail in a forthcoming report[35].

8. On Special Cases

Various cases involving special geometry, material and incremental motion of piezoelectric laminae may be readily investigated with the help of the general results derived in invariant form in the previous sections. Here attention is first limited to the case of piezoelectric plates in which the curvature effect vanishes, namely, $b_\beta^\alpha = \mu_\beta^\alpha = 0$ (cf.,[24]). The results for one layer (N=1) agree with those[26]. A complete linearization in the results leads to the linear theory of piezoelectric laminae. In such a case, it is shown by logarithmic convexity argument that the conditions (6.13)-(6.23) are suffient to ensure the uniqueness in solutions of the electroelastic equations of laminae. This and a variety of applications of the general results to particular problems are given in a recent report[35]. Further, special classes of materials for the constituent of piezoelectric laminae may be considered in the macroscopic constitutive relations (5.9)-(5.14),and also special kinematics may be introduced in (4.1), (4.2) and (7.1).

9. Summary and Conclusions

Established here in is a systematic and consistent derivation of the two-dimensional electroelastic equations of piezoelectric laminae under a mechanical bias by means of the unified variational principle (2.13) together with the linear representation (4.1) and (4.2). The electroelastic equations given in both differential and variational forms govern all the incremental types of laminae motions. The variational principle generates all the fundamental equations of piezoelectric strained media. The results contain some of earlier results as special cases[35]. Lastly, an extension of the present results to viscoelastic and electromagnetic layers will be reported elsewhere.

Acknowledgements

The author wishes to extend his gratitude to Dr. Solomon J.Becerano for his continuous guidance and enthusiasm in his academic activities, and to Dr. Arthur Ballato and Dr. Julian J. Wu for their kind interest and encouragement for the research, and also, he acknowledges the support in part by the U. S. Army through its European Research Office.

References

1. M.C. Dökmeci, "Recent Advances: Vibrations of Piezoelectric Crystals," Int. J. Eng. Sci., vol. 18, pp.431-448, March 1980.

2. M.C. Dökmeci, "Recent Progress in the Dynamic Applications of Piezoelectric Crystals," Shock Vib. Dig., vol.20, pp.3-20, Feb. 1988.

3. J.R.H. Black, "Technology and Market Trends in Multilayer Ceramic Devices," in Advances in Ceramics, vol. 19: Multilayer Ceramic Devices. Ohio: Am. Ceramic Soc., 1986, pp.3-11.

4. A.J. Moulson and J.M. Herbert, Electroceramics, Materials,Properties, Applications. Cambridge: University Press, 1990, ch.6, pp.265-317.

5. T.C.T. Ting, "Dynamic Response of Composites," Appl. Mech. Rev., vol. 33, pp.1629-1635, Dec. 1980.

6. Z. Hashin, "Analysis of Composite Materials - A Survey," J. Appl. Mechs., vol. 50, pp.481-505, Sept. 1983.

7. M.C. Dökmeci, "Theory of Vibrations of Coated, Thermopiezoelectric Laminae," J. Math. Phys., vol. 19, pp.109-126, Jan. 1978.

8. V.Z. Parton and N.A. Senik, "Relations of Electroelastcity for Multilayer Piezoceramic Shells with Thickness Polarization of the Layers," Mech. Solids, vol. 18(3), pp. 124-129 1983; Mekh. Tverd. Tela, vol. 18(3), pp. 124-129, 1983.

9. V.I. Gololobov, "Elasticity Relations for Multilayer Piezoceramic Plates," Dopov. Akad. Nauk UkRSR, Ser. A(USSR), No. 11, pp. 37-39, 1983.

10. V.G. Karnaukhov, I.F. Kirichok, and A.M. Novikova, "Geometrically Nonlinear Theory of Viscoelastic Multilaminate Piezoelectric Plates and Shells," Sov. Appl. Mech., vol. 19(10), pp. 891-897, 1984; Prikl. Mekh., vol. 19(10), pp.71-77, 1983.

11. V.G. Karnaukhov and I.F. Kirichok, "Refined Theory of Laminated Viscoelastic Piezoceramic Shells with Allowance for Heat Generation," Sov. Appl. Mech., vol. 21(6), pp.574-580, 1985; Prikl. Mekh., vol. 21(6), pp. 53-60.

12. V.G. Karnaukhov and I.F. Kirichok, "Thermomechanical Theory of Layered Viscoelastic Piezoelectric Shells Polarized in one Coordinate Direction," Sov. Appl. Mech., vol. 22(11),

pp.1074-1086, 1987; Prikl. Mekh., vol. 22(11), pp.71-78.

13. Yu. B. Evseichik, S.I. Rudnitskii, and N.A. Shul 'ga, "Equations of Vibrations of Multilayer Pieceramic shells with Tangential Polarization," Sov. Appl. Mech., vol. 24(8), pp.758-763, 1989; Prikl. Mekh., vol. 24(8), pp.36-41.

14. I.A. Motovilovets and V.I. Gololobov, "Thermomechanical Behavior of Multilayered Piezoceramic Shells with Harmonic Excitation," Sov. Appl. Mech., vol. 21(7), pp.661-666, 1986; Prikl. Mekh. vol. 21(7), pp.43-50.

15. R. Holland and E.P. EerNisse, Design of Resonant Piezoelectric Devices. Cambridge: MIT Press, No.56, 1969.

16. B.A. Auld, H.A. Kunkel, Y.A. Shul, and Y. Wang, "Dynamic Behavior of Periodic Piezoelectric Composites," in 1983 Ultrasonics Symposium Proc., pp.554-558. New York: IEEE Press.

17. B.A. Auld and Y. Wang, "Acoustic Wave Vibrations in Periodic Composite Plates," in 1984 Ultrasonics Symposium Proc., pp.528-532. New York: IEEE Press.

18. N. Buğdaycı and D.B. Bogy, "A Two-Dimensional Theory for Piezoelectric Layers Used in Electromechanical Transducers - 1: Derivation," Int. J. Solids Structs., vol. 17, pp.1159-1178, 1981.

19. N. Buğdaycı and D.B. Bogy, "A Two-Dimensional Theory for Piezoelectric Layers Used in Electromechanical Transducers - II: Application," Int. J. Solids Structs., vol. 17, pp.1179-1202, 1981.

20. C.K. Lee and F.C. Moon, "Laminated Piezopolymer Plates for Torsion and Bending Sensors and Actuators," J. Acoust. Soc. Am., vol. 85(6), pp. 2432-2439, 1989.

21. H. Nowotny and E. Benes, "General One-Dimensional Treatment of the Layered Piezoelectric Resonator with Two Electrodes," J. Acoust. Soc. Am., vol. 82(2), pp.513-521, 1987.

22. M.C. Dökmeci and E.E. nanI'D.,"Dynamics of Piezo electric Rods Under Initial Stresses," J. Acoust Soc. Am., vol. 81, sup. 1, p. S73, 1987.

23. H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Doubly-Rotated Quartz Thickness Mode Resonators," J. Applied Physics, vol. 50(12), pp.8038-8051, Dec. 1979.

24. M.C. Dökmeci, "Vibrations of Piezoelectric Discs Under Initial Stresses," in Proc. 39th Annu. Freq. Contr. Symp. pp.431-435. New York: IEEE Press, 1985.

25. M.C. Dökmeci, "Dynamics of Piezoceramic Shells Under Initial Stress," presented at 41st Annu. Freq. Contr. Symp., Philadelphia, PA., May 1987.

26. M.C. Dökmeci, "Shell Theory for Vibrations of Piezoceramics Under a Bias," in 1989 Ultrasonics Symposium Proc., pp.561-566. New York: IEEE ; to appear in IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol.37, 1990.

27. B.K. Sinha, W.J. Tanski, T. Lukaszek, and A. Ballato, "Influence of Biasing Stresses on the Propagation of Surface Waves," J. Appl. Phys., vol. 57(3), pp. 767-776, 1985.

28. A.L. Nalamwar and M. Epstein, "Surface Acoustic Waves in Strained Media," J. Appl. Phys., vol. 47(1), pp.43-48, Jan. 1976.

29. B.K. Sinha and H.F. Tiersten , "On the influence of a Flexural Biasing State on the Velocity of Piezoelectric Surface Waves," Wave Motion, vol. 1(1), pp.37-51, 1979.

30. S.G.Joshi, "Surface Acoustic Wave Propagation in a Biasing Electric Field," J. Acoust. Soc. Am., vol. 72(6), pp. 1872-1878, 1982.

31. M. Planat and D. Hauden, "Nonlinear Properties of Bulk and Surface Aocustic Waves in Piezoelectric Crystals," Ferroelectrics, vol. 42, pp.117-136, 1982.

32. H.F. Tiersten, "On the Nonlinear Equations of Thermoelectroelasticity," Int. J. Eng. Sci., vol. 9, pp.587-604, 1971.

33. J.C. Baumhauer and H.F. Tiersten, "Nonlinear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., vol.54, pp.1017-1034, 1973.

34. M.C. Dökmeci, "Certain Integral and Differential Types of Variational Principles in Nonlinear Piezoelectricity," IEEE Trans. Ultrason. Ferroelec. Freq. Contr., vol. 35(6), pp.775-787,1988.

35. M.C. Dökmeci, "A Dynamic Analysis of Piezoelectric Strained Elemenets," TR-1/90, İstanbul Teknik Üniversitesi, Dec. 1990.

36. V.V. Bolotin, Nonconservative Problems of the Theory of Elastic Stability. London: Pergamon Press, 1962.

37. K. Washizu, Variational Methods in Elasticity and Plasticity. London: Pergamon Press, 1983.

38. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., vol. 9, pp.1445-1451, 1968.

39. D.S. Drumheller and A. Kalnins, "Dynamic Shell Theory for Ferroelectric Ceramics," J. Acoust. Soc. Am., vol. 47, pp.1343-1353, 1970.

40. H.F. Tiersten, Linear Piezoelectric Plate Vibrations. New York: Plenum Press, 1969.

41. M.C. Dökmeci, "A General Method of Solution for Composite Laminates," in Composite Materials in Engineering Design, pp.587-599. Metals Park,OH: Am. Soc. Metals, 1973.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PIEZOELECTRIC BEAMS AND ANGULAR RATE SENSORS

Jan Söderkvist

Institute of Technology, University of Uppsala,
Box 534, S-751 21 Uppsala, Sweden

Abstract

A tuning fork angular rate sensor made out of a single piece of quartz has been studied. The piezoelectric effect is used both to excite a reference vibration in the plane of the tuning fork and to detect a vibration normal to this plane. The amplitude of the second vibration is directly proportional to the applied angular velocity. The structure is made rigid in order for it to survive in a harsh environment. This implies that the only vibrationally active areas are the tines of the tuning fork. The performance of the sensor is predicted with the help of a phenomenological piezoelectric beam theory. This theory shows that it suffices to study the two-dimensional dielectric field in the cross-sections of the beams in order to obtain the values of the piezoelectric equivalent components. Estimates of these values can be obtained without the use of special computer programs. The predictions are shown to be in agreement with measurements.

I. Introduction

The market for inexpensive, small sensors is expanding steadily. It is becoming more and more important to be able to reduce price (and size) since sensors and actuators are becoming one of the more expensive parts of computerized systems. One of the few techniques that can be used in producing inexpensive, small structures is micromachining (including batch fabrication). The fabrication techniques, including anisotropic etching and bonding, have recently been developed extensively for both quartz and silicon. True three-dimensional structures can now be designed in both materials [1,2,3]. However, it is still difficult to etch in some directions. A number of quartz and silicon sensors have recently been designed [4,5] and commercialized successfully using this technique.

Now, the design of micromechanical sensors is no longer restricted by fabrication difficulties. The complexity of the structures can be expected to increase. This implies that more experiments are needed in order to optimize performance. It will thus become more and more important to be able to predict the behavior of the sensors without having to perform expensive and time-consuming experiments. This points out the importance of accurate analytical tools. These tools can also be helpful in explaining anomalies in experiments as well as to stimulate new experimental ideas and identify important design parameters of the structure.

CH2818-3/90/0000-406 $1.00 © 1990 IEEE

A number of computerized methods of calculating interesting properties exist. The most well known of these is the finite element method (FEA or FEM). This method is general and can handle almost any differential equation describing the physical situation, including piezoelectricity [6]. It is important that the analytic methods are based on simplicity and understanding in order for the user to gain a better sense of what parameters the interesting properties depend on. There is thus a large demand for simple phenomenological methods in addition to the more general, detailed and complicated methods.

This article will address the problem of flexural vibrations of piezoelectric beams. These vibrations are normally described using the electric equivalent circuit in Fig. 1. Methods on how to predict the values of the components in this circuit have been presented previously [7,8]. These methods are not entirely based on a phenomenological description of the piezoelectric effect. They do not, for instance, clearly indicate the role of the different electric fields. It is thus difficult, using these methods, to visually understand the performance of different electrode configurations, as well as to predict the performance if only simple analytic tools are available.

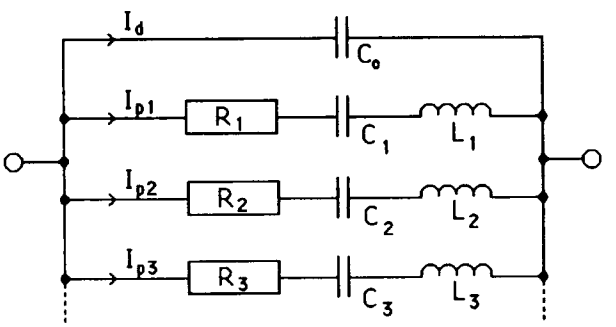


Fig. 1. *An electric equivalent circuit of a vibrating structure. Each of the $R_nC_nL_n$-branches represents the behavior close to a resonance frequency ($1=L_nC_n\cdot\omega^2$, $R_nQ_n=L_n\omega_n$). The dielectric capacitance is denoted C_o.*

The obtained knowledge has been used to study angular rate sensors made of quartz. These sensors are based on the principle of conservation of linear momentum [9,10]. This specific sensor application uses all three of the properties of well-defined resonance frequencies, of electrically activated mechanical vibrations and of the conversion of a mechanical deflection into an electric signal. It is thus a well-chosen example of how flexural vibrations can be used

in different sensor and actuator applications. Piezoelectric angular rate sensors will therefore be discussed in more detail in this article. Predictions for this sensor will be compared with experiments.

IIa. Piezoelectrically excited flexural vibrations electric field considerations

The piezoelectric effect is due to the atomic dipole structure of the material. The value of the polarizing effect of each dipole is proportional to the local stress, **T**. The charges of neighboring dipoles will partly neutralize each other. This will result in a global distribution of bound piezoelectric volume and surface charges, ρ_{vol} and ρ_{surf}. This distribution is given by (cf. [11])

$$\begin{cases} \rho_{vol} = -\mathrm{div}\mathbf{P} \\ \rho_{surf} = \mathbf{P}\cdot\mathbf{n}_{out} \\ \mathbf{P} = \mathbf{d}\cdot\mathbf{T} \end{cases} , \qquad (1)$$

where **P** is the piezoelectric polarization, **d** the piezoelectric coupling coefficient and $\mathbf{n}_{out}$ the normal out of the material. The last equation is based on the well-known piezoelectric equation $\mathbf{D}=\boldsymbol{\varepsilon}^T\cdot\mathbf{E}+\mathbf{d}\cdot\mathbf{T}$ in which **D** is the electric displacement vector and $\boldsymbol{\varepsilon}^T$ the permittivity tensor at constant stress (e.g. [12]).

The stress-induced bound charges in Eq. (1) will create a piezoelectric field, $\mathbf{E}^p$, in addition to the dielectric field, $\mathbf{E}^d$, created by the potential of the electrodes

$$\mathbf{E} = \mathbf{E}^d(\text{no deformation}) + \mathbf{E}^p(\text{electrodes grounded}) . \qquad (2)$$

These fields will in reality be coupled to each other since the dielectric field piezoelectrically introduces stress and/or strain in the material. Poisson's equation is used to find the piezoelectric field due to the bound charges, while the well-known Laplace's equation is used to find the dielectric field.

The two fields will give rise to I, a current of free charges to and from the electrodes ($\int I dt = \oint \mathbf{D}d\mathbf{A}$). The dielectric field and current will give the value of the C_o-capacitance in Fig. 1, and the piezoelectric field and current the values of the components in the $R_nC_nL_n$-branches of the well-known electric equivalent circuit [7].

Beams are defined as being long compared to the dimensions of the cross-section. This implies that stress in all other directions than in the direction of the beam can be neglected [13]. This transforms the mechanical part of the problem into a one-dimensional one. The strain, **S**, in the direction of the beam in Fig. 2 can then be described by a radius of curvature, $R_{cv}(y)$, ($=-\zeta_v''(y)^{-1}$ for small deflection), according to

$$S_y(x,y,z) = \frac{v}{R_{cv}} + S_{y,o} = d_{ky}E_k + s^{\varepsilon}_{yy}T_y , \qquad (3)$$

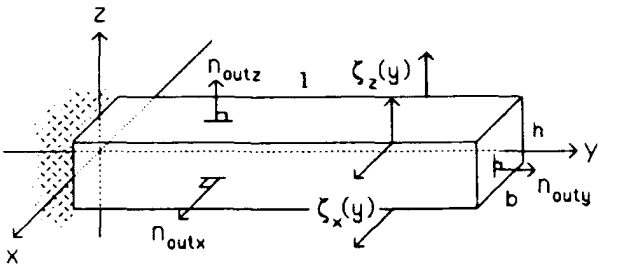


Fig. 2. *Geometry of a simplified beam. The coordinate system does not need to coincide with the crystallographic system.*

in which v denotes the direction of vibration (e.g. = x or z) and $S_{y,o}$ the strain at the origin of the coordinate system (the mass center). The last equality in Eq. (3) indicates that the strain can be described with one of the well-known piezoelectric equations ($\mathbf{s}^{\varepsilon}$ is the compliance tensor at constant electric field) (e.g. [12]).

The beam approximation can also motivate the neglect of electric field components in the direction of the beam. This is not a necessary restriction but it simplifies the calculation of the electric fields by reducing the electric part of the problem into a two-dimensional one. Experiments indicate that the electric field in the surrounding of the piezoelectric material does not noticeably influence the piezoelectric properties. It can thus also be neglected if desired.

One can define the following integrals once the electric fields are known:

$$\int_{d,v}(y) = \frac{d_{ky}}{s^{\varepsilon}_{yy}}\cdot\iint v\cdot E^d_k(x,y,z)\,dxdz , \qquad (4a)$$

$$\int_{p,v}(y) = \frac{d_{ky}}{I_n(y)}\cdot\iint v\cdot U^p_k(x,y,z)\,dxdz \qquad (4b)$$

and

$$\oint_{+,v}(y) = \oint_{+}\left(\varepsilon_{k1}U^{pv}_1(x,y,z)+\frac{v\,d_{ky}}{s_{yy}}\right)d\Gamma_k , \qquad (4c)$$

where $\mathbf{U}^{pv}$ equals $\mathbf{E}^pR_{cv}$, and I_n the second moment of area with respect to an axis normal to v and y. The third integral is to be taken over the part of the surface of the cross-section that is covered by positive electrodes. The $_k$ denotes the direction normal to the line element, $d\Gamma_k$. It will be shown that these integrals contain all information of interest regarding the electric fields in the beam. Some interesting properties of the integrals are as follows:

- The first integral determines the vibration amplitude due to an applied electric field (see Eq. (7)). This integral can be estimated accurately without using a computer.
- The second integral causes the effective value of s^{ε}_{yy}, for mechanical properties, to depend on the shape of the electrodes ($\propto[1-\int_{p,v}]^{-1}$, see Eqs. (6) to (8)). This alteration should be compared with that of the effective value of ε^T ($\propto 1-d^2/\varepsilon s$) which

is independent of the electrode configuration used. The second integral can normally be neglected ($|\int_{\rho,v}| \leq d^2/\varepsilon s$).

- The third integral determines the piezoelectric current due to a known vibration. This integral is useful in understanding the efficiency of an electrode configuration. In most cases it is proportional to the first integral ($\int_{d,v} = -\Delta V \oint_{+,v}$, see Eq. (10b)).

It is only the first *or* the third of these integrals that needs to be considered if field components in the direction of the beam are negligible (true for a beam). *One can thus choose whether to treat the problem purely dielectrically or purely piezoelectrically.*

IIb. Piezoelectrically excited flexural vibrations mechanical considerations

The mechanical deflection due to an applied electric field is determined by a force, **F**, and a bending moment, **M**. Neutrality in each cross-section of the beam is determined by

$$F_y(y) = M_x(y) = M_z(y) = 0 \quad \Leftrightarrow \quad \iint T_y dxdz = \iint z \cdot T_y dxdz = \iint x \cdot T_y dxdz = 0 \qquad (5)$$

This gives the following value for the static radius of curvature in the absence of external forces, q(y) (T_y is given by Eq. (3)),

$$\zeta''(y,\omega=0) = \frac{-1}{R_{cv}(y,\omega=0)} = \frac{-s^E_{yy}}{1-\int_{\rho,v}(y)} \cdot \frac{\int_{d,v}(y)}{I_n(y)} \qquad (6)$$

Dynamic effects can be taken into account by deriving the piezoelectric version of the differential equation for beams [14]

$$\frac{\partial^2 \zeta_v}{\partial t^2} + 2\delta \cdot \frac{\partial \zeta_v}{\partial t} + \frac{1}{A\rho} \cdot \frac{\partial^2}{\partial y^2}\left[\frac{I_n}{s^E_{yy}} \cdot \left[1-\int_{\rho,v}\right] \cdot \frac{\partial^2 \zeta_v}{\partial y^2}\right] = \frac{1}{A\rho} \cdot \left[q(y) - \frac{\partial^2}{\partial y^2}\int_{d,v}\right] , \qquad (7)$$

in which δ is a general damping term, ρ the density and A the area of the cross-section. It is worth noting in Eq. (7) that it is only the dielectric part of the electric field that acts as a driving force. The piezoelectric part affects the effective value of S^E_{yy}, which, in turn affects, for instance, the values of the resonance frequencies, ω_n.

The solution to Eq. (7) is best obtained via an extended eigenmode decomposition of the static deflection in Eq. (6). For a harmonic voltage with frequency ω (q(y)=0, $w=\omega/\omega_n$), this gives

$$\zeta_v(y,t,\omega) = \sum D_n \cdot \xi_n(y) \cdot \frac{\sin[\omega t + \varphi_n(w)]}{\left\{\left[1-w^2\right]^2 + \left[\frac{w}{Q_n}\right]^2\right\}^{1/2}} \qquad (8a)$$

where

$$D_n = \frac{\int \frac{-\xi_n(y)}{R_{cv}(y,\omega=0)} dm}{\int \xi_n(y) \cdot \xi_n(y) dm} \qquad (8b)$$

and $A\rho$ in $dm = A\rho \cdot dy$ is the weight function of the integration. The n^{th} eigenmode and the quality factor are denoted ξ_n and Q_n. The factor D_n describes how much the static deflection contains of each eigenmode. It depends on $\int_{d,v}$, on the location of the electrodes and on the geometry on the beam. The dependence of the vibrational energy ($\propto D_n^2$) on the location of the electrodes for the beam in Fig. 2 is shown in Fig. 3. The last part of Eq. (8a) describes the increased response at frequencies close to the resonance frequencies. It is normally desirable to keep the energy losses small ($Q_n \gg 1$). It can then be assumed that the n^{th} term in the sum in Eq.(8) dominates completely close to the n^{th} resonance frequency.

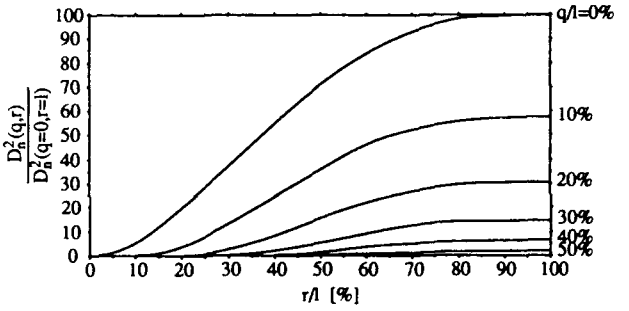


Fig. 3. *The dependence of the vibrational energy ($\propto D_n^2$) on the location of the electrodes. The electrodes are assumed to cover the distance $y \in \{q,r\} \subseteq \{0,l\}$ in Fig. 2. The beam has a constant cross-section area, is supported at one end and vibrates at the fundamental resonance frequency.*

IIc. Piezoelectrically excited flexural vibrations equivalent circuit

The piezoelectric current, given by

$$I^P = \frac{d}{dt}\int\int D_k^P d\Gamma_k dy = \frac{d}{dt}\int \frac{\oint_{+,v}(y)}{R_{cv}(y,\omega)} dy , \qquad (9)$$

will, via R_{cv}, have the same frequency dependence as the piezoelectrically induced mechanical vibration. Thus both can be described by the same second order system. The scale factor between the electric and the mechanical properties is obtained by equating the power dissipation in the two systems ($UI^P = \int M^P(y) d\dot{\varphi}$, where $\varphi = \zeta'(y)$). This gives the following important relation ($\Delta V(\omega,t)$ is the voltage applied to the electrodes)

$$\int \zeta''(y)\cdot\left[\int_{d,v}+\Delta V\cdot \oint_{+,v}\right]dy = 0 \ , \qquad (10a)$$

which is transformed into

$$\int_{d,v} = -\Delta V\cdot \oint_{+,v} \qquad (10b)$$

when field components in the direction of the beam are neglected. The fact that $\omega_n^2\int\zeta\xi_n dm=-\int\xi_n''\int_{d,v}dy$ has been used in deriving Eq. (10). Equation (10b) can also be proven using vector analysis. Note that the piezoelectric current is proportional to the velocity of the beam, and that the force acting on the beam is proportional to the applied voltage.

An electric second order system can be represented by an $R_nC_nL_n$ circuit. This explains the correctness and popularity of the equivalent circuit in Fig. 1. Comparing the currents through the $R_nC_nL_n$-branches in Fig. 1 with the mechanical amplitude of the n^{th} term of the sum in Eq. (8) [15] gives the following measurable quantity:

$$C_n = \frac{-1}{\Delta V\omega_n^2 m_o}\cdot\left[\int\int_{d,v}(y)\cdot\xi_n''(y)dy\right]\left[\int\oint_{+,v}(y)\cdot\xi_n''(y)dy\right], \qquad (11a)$$

where $m_o=\int\xi_n^2 A\rho dy$. The electrodes are allowed to cover any part of the beam. Equation (10a) can be used to simplify the expression for C_n according to

$$C_n = \frac{1}{\omega_n^2 m_o}\cdot\left[\int_q^r \oint_{+,v}(y)\cdot\xi_n''(y)dy\right]^2 =$$

$$= \frac{1}{\Delta V^2\omega_n^2 m_o}\cdot\left[\int_q^r \int_{d,v}(y)\cdot\xi_n''(y)dy\right]^2 =$$

$$= \frac{\omega_n^2}{\Delta V^2 m_o}\cdot\left[\int_0^1 \zeta(y,\omega=0)\cdot\xi_n(y)dy\right]^2 . \qquad (11b)$$

The electrodes have been assumed to cover the area $y\in\{q,r\}\subseteq\{0,1\}$. Equation (11b) is further simplified if it is assumed that the shape of the electrodes and the geometry do not change in the interval $y\in\{q,r\}$ and if field components in the direction of the beam are negligible:

$$C_n = \frac{\int_{d,v}^2}{\Delta V^2\omega_n^2 m_o}\cdot\left[\xi_n'(r)-\xi_n'(q)\right]^2 =$$

$$= \frac{\oint_{+,v}^2}{\omega_n^2 m_o}\cdot\left[\xi_n'(r)-\xi_n'(q)\right]^2 . \qquad (11c)$$

Equation (11) contains a number of unknown quantities. These can either be measured (C_n and $\xi_n(y)$), determined analytically ($\xi_n(y)$, $\int_{d,v}$ and $\oint_{+,v}$) or calculated via Eq. (11).

The integral $\int_{d,v}$, which is better suited for numerical calculations than the integral $\oint_{+,v}$, is based on the dielectric field in the cross-sections of the beam (Laplace's equation). Accurate estimates of the value of $\int_{d,v}$ can be obtained by first dividing the cross-section into a small number of area elements, ΔA_m. The appropriate component of the electric field intensity, $E_{v,m}$, at the center of the elements, is then estimated by roughly sketching the equipotential and the field lines between the electrodes (see Fig. 4 for an advanced sketch). The estimated values are then inserted into the discrete form of the integral definition of $\int_{d,v}$

$$\int_{d,v} \approx \frac{d_{kv}}{s_{yy}^{\xi}}\cdot\sum_m v_m E_{k,m}\cdot\Delta A_m \ , \qquad (12)$$

in which v_m denotes the distance between the center of the m^{th} element and the center of bending. A fairly rough guess as to the shape of the dielectric field gives surprisingly accurate estimates without the need of advanced calculation tools.

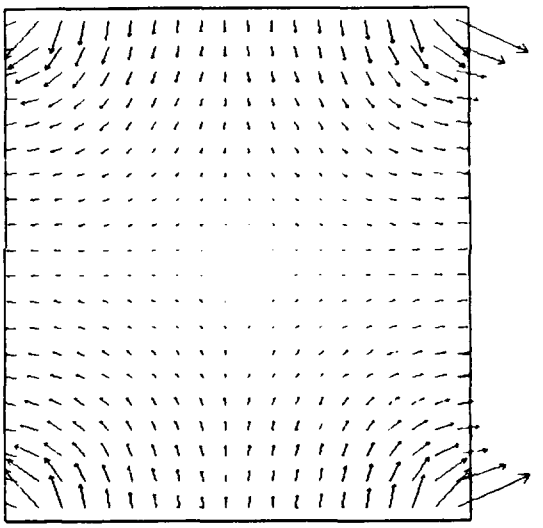


Fig. 4. *Electric field intensity in the cross-section for the drive electrode configuration in Fig. 8. The size of the arrows is proportional to the field intensity. The direction of the arrows indicates the direction of the electric field.*

The integral $\oint_{+,v}$ is better suited for understanding how different electrode configurations work. Its value can be estimated by first plotting the distribution of bound charges (Eq. (1)) in the cross-section. Smoothly curved electric field lines are then drawn between these charges under the condition that the electrodes are short-circuited. The lines should be approximately normal to the surfaces. Let k_e denote the fraction of lines that enters one electrode (free charges are induced) and exits another that dielectrically has opposite polarity. The value of $\oint_{+,v}$ is then given by

$$\oint_{+,v} = -k_e\rho_p R_{cv} \ , \qquad (13)$$

where ρ_p denotes the amount of bound positive piezoelectric charge in the cross-section. The effect of, for instance, dimensional as well as electrode configuration changes and strange cross-section geometries can easily be studied by observing changes in k_e and ρ_p from a rough sketch of the charge distribution. This method is well suited for rapid determination of the most efficient electrode configurations.

Different aspect ratios of the cross-section can easily be studied using the integral $\oint_{+,v}$. An example on how this is done using the estimate in Eq. (13) is shown in Fig. 5. The estimates of $\oint_{+,x}$ obtained in this way can be compared with the values obtained from Eq. (4) (see Fig. 6). The deviation is found to be small.

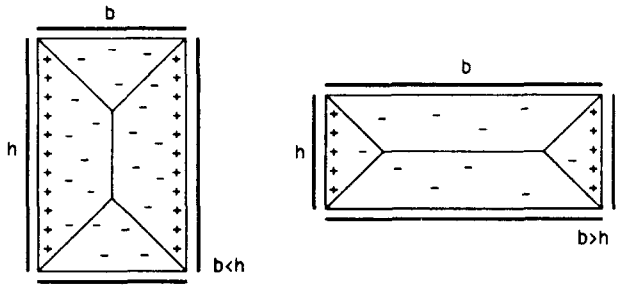


Fig. 5. *The value of $\oint_{+,x}$ for the electrode configuration in Fig. 8a can be estimated in a simple way using Eq. (13) by dividing the cross-section into four areas according to the figure. The bound charges attract/-repel free charges on the electrodes. The charge on each electrode can be said to depend only on the amount of charge within the adjacent area. This results in k_e= b/2h and =1-h/2b respectively. The amount of bound positive charge equals $\rho_p R_{cx} = d_{xy} bh / s^E_{yy}$.*

The slope, $\zeta_n'(y)$, at the electrode boundaries can be determined from a rough sketch of the eigenmodes or by using numerical methods (e.g. perturbation theories or ordinary mechanical FEA).

The expressions for C_n can be used in order to determine the values of $\int_{d,v}$ and $\oint_{+,v}$ from measurements of C_n. The values of these integrals can be used to determine actual vibrational amplitudes (Eq. (8)) and piezoelectric currents (Eq. (9)). The normalized shape of the eigenmodes can also be obtained from Eq. (11) and measurements of the motional capacitance by varying the location of the electrodes. This is done most easily if the electrodes are not in physical contact with the beam. The value of the integral $\int_{d,v}$ can be assumed to be constant during these measurements if the electrodes are located close to the surface of the beam (the field in the surroundings is not negligible in this case). The amplitude of the vibration can also be determined once the value of $\int_{d,v}$ is known.

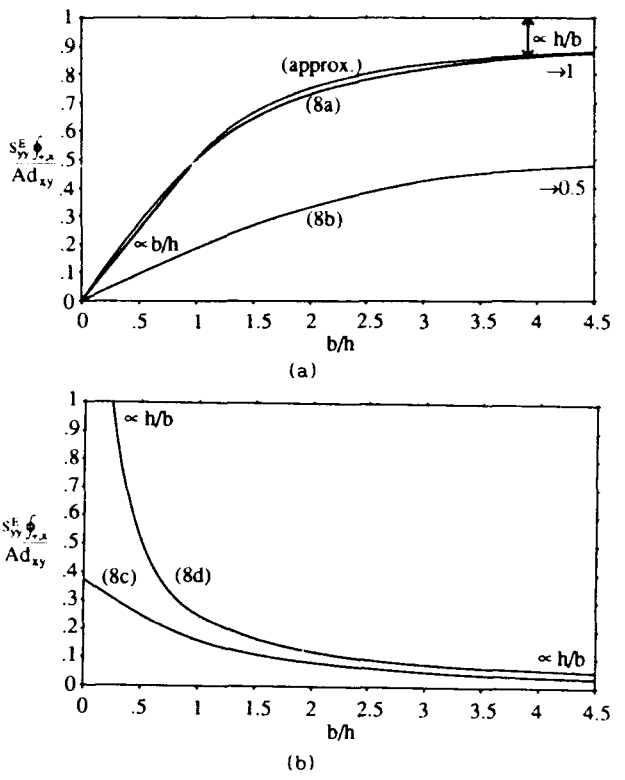


Fig. 6. *The dependence of $\oint_{+,v}$ on the aspect ratio of the cross-section for the electrode configurations in Figs. 8a-8d (s=1/2). Curves (8a) and (8b) are related to a vibration in the x-direction and curves (8c) and (8d) to one in the z-direction. The curve '(approx.)' is obtained from the estimate made in Fig. 5. The separation between neighboring electrodes has been assumed to be infinitely small. It has also been assumed that d_{xy} is much larger than d_{zy}.*

IIIa. Tuning fork angular rate sensors general overview

Piezoelectric beams and membranes/plates are common in sensor and actuator applications. These applications use either the properties of well-defined resonance frequencies, of electrically activated mechanical vibrations or of the conversion of a mechanical deflection into an electric signal. A sensor that uses all three of these principles, a piezoelectric angular rate sensor, has been studied. An overview of rate sensors will be given in this section in order for the reader to fully appreciate the advantages of the studied sensor concept.

An angular rate sensor can be based on a number of different principles. Most well-known are those based on a rotating mass. The conservation of angular momentum implies that it is very difficult to alter the direction of rotation. This memory function will be observed as a change in direction of rotation of the rotor with respect to the mounting if the mounting, e.g. an aircraft, is turning (cf. a compass). The rotating movement can be replaced by a vibrating one. The memory function will then be based on conservation of the linear momentum. The most well-known example of

this second type of rate sensor is Foucault's pendulum. The vibrating structure has the advantage of not using bearings and other parts that can wear out. Other examples of rate sensors are laser and fiber-optical gyroscopes based on the Sagnac effect [16].

Most effort has so far been concentrated on conventional rotating gyroscopes and on laser gyroscopes. It is thus natural that these sensors show the best performance. Unfortunately, they are also the most expensive. Vibrating rate sensors will in the future show competitive performance. It was shown as early as 25 years ago that it is possible to obtain a resolution below 1°/h [17] by using a vibrating steel tuning fork (electromagnetic drive and capacitive pick-off). The cost of vibrating rate sensors can be expected to be substantially lower than for other types of rate sensors if solid-state technology and ASIC circuits are used. Other advantages with vibrating solid-state rate sensors are that they have an 'infinite' lifetime, a rapid response, a wide range of measurements, are simple and can withstand a harsh environment.

A vibrating angular rate sensor is preferably shaped like a tuning fork. The necessary vibration will then be localized to the tines at the same time as external vibrations are unable to reach the tines. The symmetry of the fork structure reduces the acceleration sensitivity, since linear acceleration deflects both tines in the same direction, while the Coriolis force deflects them in the opposite direction. Signal processing can then be used to suppress the linear acceleration dependence. The sensor will withstand a harsh environment if the stem and the mounting are designed to be rigid. The only vibrationally active areas will then be the tines. This is in contrast to older types of tuning fork rate sensors that used a fragile mounting. This type of mounting enabled a transfer of tine vibration energy into a torsional movement of the lower part of the stem, but it also implied difficulties in finding a proper mounting area. The strength and the acceleration insensitivity is further increased if a fairly large resonance frequency is used.

The principle of a vibrating tuning fork rate sensor is as follows. The linear momentum that is to be conserved is created by exciting the tines piezo-electrically in an oscillation in the plane of the fork (D_1-D_2 in Fig. 7). This is the vibration used by musical tuning forks as well as by most watch crystals. The reference vibration is restricted to the plane that D_1-D_2 defines due to conservation of momentum. However, this plane is altered if the tuning fork is rotated around an axis parallel with the tines. The tines compensate this by initiating a vibration (S_1-S_2) out of the plane of the tuning fork. The amplitude of this second vibration, which is said to be induced by the Coriolis force, is directly proportional to the speed of rotation, Ω. This second vibration can be detected piezoelectrically.

A large output is obtained if a large drive amplitude is used. This is obtained by utilizing one of the drive resonance frequencies. It is also important that this drive frequency is almost equal to one of the resonance frequencies of the sense vibration. An amplitude amplification according to the last factor in

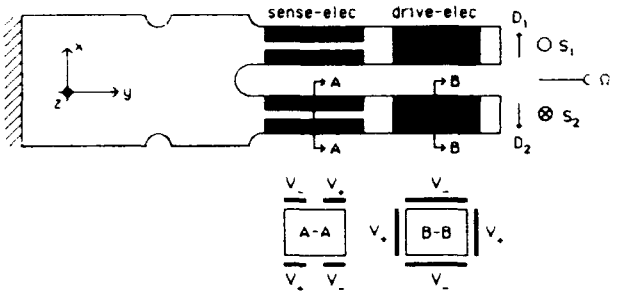


Fig. 7. *Simplified geometry used for a tuning fork angular rate sensor. A suggested electrode configuration is also shown. The drive electrodes excite a vibration in the x-direction and the sense electrodes detect a vibration proportional to Ω in the z-direction.*

Eq. (8a) is then obtained. However, thermal stability considerations imply that it is not desirable to have the two resonance frequencies coincide.

It is important that the location of the electrodes and the geometry of the tines are chosen carefully in order to obtain a large output (see Fig. 8 for some useful electrode configurations). A compromise between the location of the drive and the sense electrodes is necessary since they must both be located on the tines if a rigid mounting is used. One suggested location is shown in Fig. 7. The drive and the sense electrodes can also be interchanged. It is also possible to use one tine to drive and the other to sense the resulting vibrations.

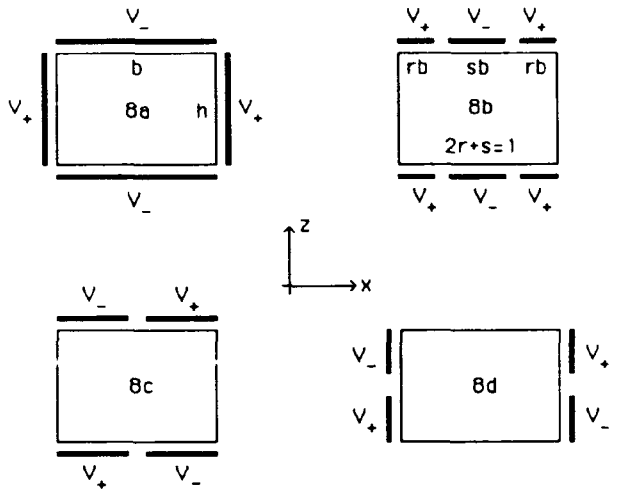


Fig. 8. *Four common electrode configurations used for quartz. Configurations 8a and 8b are used to excite a vibration in the x-direction and the remaining two in the z-direction. The equivalent component efficiencies of the four configurations are related according to 1.0 : 0.13 : 0.10 : 0.25.*

It is important to reduce the zero rate output of the sensor. It is easy, if one is not careful, to obtain a zero rate signal that is more than a factor 1000 times larger than the desired resolution. A number of different causes for the unwanted signal exist [18]. The major ones are capacitive coupling between the

drive and sense electrodes and mechanical coupling between the drive and sense vibrations [19]. The effect of the first can be reduced by a careful optimization of the electronics, and the second by a careful control of the etching and by balancing the tines.

The piezoelectric tuning fork is preferably made out of quartz. Conventional wafer technology, etching and evaporation can then be used.

IIIb. Tuning fork angular rate sensors predictions/experiments

The theory of predicting piezoelectric performance presented above can be used in order to predict the performance of various sensors and actuators. The predictions are best performed by using the electric equivalent circuit in Fig. 1. The components in this circuit easily represent measurable quantities. The theory has been used in order to study the tuning fork angular rate sensor described in the last section.

The electric input to the sensor element is the voltage of the drive signal. This voltage is transformed into a mechanical amplitude. The amplitude of this vibration is given by Eqs. (8) and (11). A faster way of converting the voltage into an amplitude is obtained by equating the piezoelectric and the mechanical energies. Assume that the frequency used equals the drive resonance frequency ω_D, with eigenmode $\xi_D(y)$. This gives the drive amplitude

$$\zeta_{drive}(y) = \xi_D(y)\cdot\frac{\Delta V Q_D}{\omega_D}\cdot\left[\frac{C_{1,D}}{\int\xi_D^2 dm}\right]^{1/2} . \qquad (14)$$

A constant amplitude is ensured if the drive current is monitored ($I\propto d\zeta/dt$). The amplitude chosen should not be larger than the strength of the material permits.

The amplitude in Eq. (14) is converted into a sense amplitude via the Coriolis force:

$$dF_c(y) = -2A\rho\Omega\cdot\dot{\zeta}_{drive}(y)dy . \qquad (15)$$

The amplitude of the resulting sense vibration, $\zeta_{sense}(y)$, depends strongly on the geometry of the tines. A square beam supported at one end with sense resonance frequency ω_S ($\omega_D/\omega_S=w\approx 1$, $\xi_S(y)=\xi_D(y)$) has

$$\zeta_{sense}(y) = \zeta_{drive}(y)\cdot\frac{2\Omega}{\omega_D}\cdot\left\{\left[1-w^2\right]^2 + \left[\frac{w}{Q_S}\right]^2\right\}^{-1/2} . \qquad (16)$$

It can be seen that the sense vibration is directly proportional to the applied rotation. The sense vibration is not allowed to be large compared to the drive vibration since that could introduce non-linearity (the sense vibration 'steals' energy from the drive vibration). It is realistic to maximize the sense amplitude to 1/10 of the drive amplitude as well as to assume a dynamic range of 1000. This implies that the sense electrodes should be able to detect the amplitude 1 Å if the drive amplitude is in the order of 1 μm.

The sense amplitude is most easily converted into a detectable current via

$$I_{sense} = \omega_D^2\cdot\left(C_{1,S}\cdot\int\zeta_S^2 dm\right)^{1/2} . \qquad (17)$$

The output from the sensor should not only respond to a static rotation. It should also respond well to a harmonic rotation ($\Omega\propto\cos\omega_{rot}t$) if ω_{rot} is below the bandwidth, BW, of the system. The harmonic rotation together with the harmonic drive vibration ($\zeta_{drive}\propto\cos\omega_D t$) implies that the frequency dependence of the Coriolis force is altered slightly:

$$dF_c(y) = A(y)\cdot\sin(\omega_D+\omega_{rot})t + B(y)\cdot\sin(\omega_D-\omega_{rot})t , \qquad (18)$$

where A(y) and B(y) are time independent constants. The sense vibration is given by the superposition of the responses of the two parts of the Coriolis force. These responses are obtained by replacing ω_D in the definition of w in Eq. (16) by $\omega_D+\omega_{rot}$ and $\omega_D-\omega_{rot}$ respectively. The dependence of ζ_{sense} on ω_{rot} can be described by a correction factor, C_{rot}. The shape of this factor depends strongly on the principle used in the sense electronics. It is advisable to make the sense electronics sensible to only one phase of the sense signal, the phase at $\omega_{rot}=0$. This can be achieved by first multiply the sense signal with a signal with known phase and frequency ω_D, and then low-pass filter the resulting signal. This gives the stationary solution

$$\zeta_{sense} \propto C_{rot} \approx \frac{(1+v^2)^2 + u^2(1-v^2)}{(1+v^2)^2 + 2u^2(1-v^2) + u^4} , \qquad (19a)$$

where

$$\begin{cases} u = \dfrac{2\omega_{rot}}{\Delta\omega_{3dB}} \\ v = \dfrac{2(\omega_D-\omega_S)}{\Delta\omega_{3dB}} \end{cases} . \qquad (19b)$$

It has been assumed that $\omega_{rot}\ll\omega_D\approx\omega_S$. The 3dB-bandwidth of the sense resonance peak is approximately equal to $\omega_S/Q_S\equiv\Delta\omega_{3dB}$. The value of C_{rot} has been plotted in Fig. 9 for appropriate values of v.

The bandwidth, which is defined as the value of ω_{rot} at which C_{rot} deviates with more than 3dB from the value 1, can be found from Eq. (19):

$$\left(\frac{2BW}{\Delta\omega_{3dB}}\right)^2 = (1-v^2)\cdot\left(\frac{\alpha^{\pm 1}}{2} - 1\right) \pm$$

$$\pm\left[(1-v^2)^2\cdot\frac{\alpha^{\pm 2}}{4} + 4v^2\cdot(\alpha^{\pm 1}-1)\right]^{1/2} , \qquad (20)$$

where α equals $10^{0.15}$ (3dB). The sign (+ or −) should

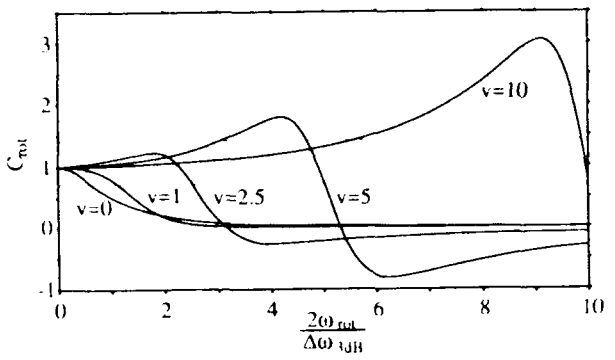


Fig. 9. *The dependence of the amplitude in the sense direction on the frequency of the applied rotation. Various values of the frequency difference has been assumed.*

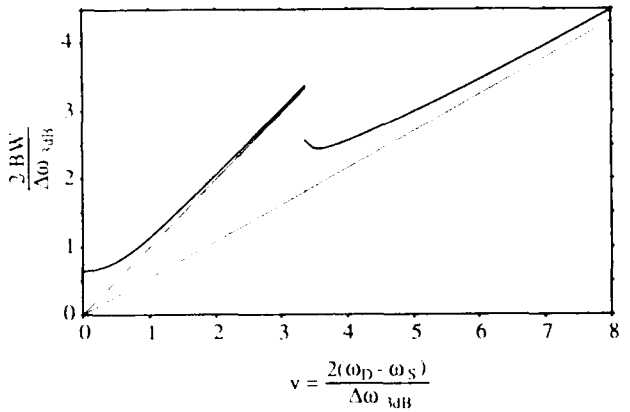


Fig. 10 *The 3dB-bandwidth as a function of the frequency difference. The bandwidth approaches* $0.54\cdot|\omega_D-\omega_S|$ *at large values of v and* $|\omega_D-\omega_S|$ *at values just below v=3.35.*

be chosen so as to minimize BW. The bandwidth, given by Eq. (20) is plotted in Fig. 10. The limiting cases are found to be $BW=0.54\cdot|\omega_D-\omega_S|$ if $v\gg1$ (a large separation between the resonance peaks) and $BW=0.32\cdot\Delta\omega_{3dB}$ if $v\approx0$ (wide resonance peaks). The abrupt change in bandwidth at v=3.35 is due to C_{xys} having values 3dB greater than 1 above this value of v. It can be seen that having $\omega_D=\omega_S$ gives a low value of the bandwidth. This choice of resonance frequency separation normally also implies problem with temperature stability. Best performance is obtained if the separation between the two resonance frequencies is chosen so large that temperature drift of the resonance frequency separation does not affect the responces in Eqs. (16) and (19) noticeably.

A number of tuning forks have been tested (e.g. that in Fig. 11) in order to verify the sensor concept and the described method of prediction. Various locations of the electrodes and different electrode configurations have been tested. The electric equivalent components of the crystals were measured and compared with predictions. The measured values of the impedance were between 0 and 7% higher than those predicted. This implies a maximum error in the values of $\int_{d,v}$ and $\int_{s,v}$ of 3.5%. This is comparable to the inaccuracy of the impedance measurements. The small difference might also be due to treating the geometry

Fig. 11. *One of the geometries tested. The opaque areas consist of evaporated gold on top of a thin layer of chrome. The drive electrodes are located near the stem and the sense electrodes near the tip of the tines. The large number of wire bonding pads on the stem are useful in testing different electrode configurations.*

as an ideal mechanical beam. Some of the results are shown in Table 1. The efficiency factors in the table indicate the reduced efficiency of the equivalent components due to the electrodes not covering the entire length of the tines (location efficiency, see Fig. 3) and the entire cross-section (electrode efficiency, the electrodes are smaller than those shown in Fig. 8). The obtained deviations between measurements and predictions are equal to or smaller than those derived using other theories [e.g. 8]. The advantage of the presented mathematical tool is that it is phenomenological and simpler.

Some of the tuning forks have been connected to a simple electronic circuit (15 active components) according to Fig. 12a. The electronics consists of four parts (see Fig. 12b) [20]:

- A drive loop that activates the reference vibration with a well-defined amplitude (AGC=-automatic gain control),
- a preamplifier that picks up and band-pass filters the small signal from the sense electrodes,
- a demodulator that blocks signals 90° out of phase with the useful signal and

Case:	1	2	3	4
b/h	1.0	0.83	1.0	0.83
l [mm]	2.4	3.5	2.4	3.5
electrode config.	Fig. 8a		Fig. 8c	
location eff. [%]	50	9	6	28
electrode eff. [%]	60	75	85	45
predicted C_1 [fF]	1.1	0.29	0.018	0.074
measured C_1 [fF]	1.1	0.27	0.017	0.07

Table 1. *A comparison between predicted and measured values of the equivalent component* C_1 *of tuning forks with constant cross-section area.*

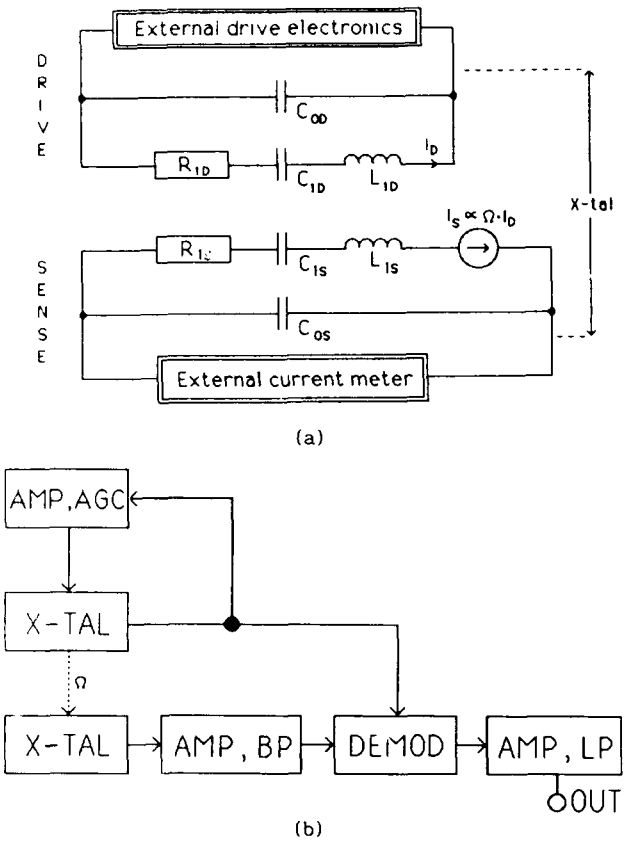


Fig. 12. *The piezoelectric angular rate sensor as an electric component. Figure 12a shows the electric equivalent circuit representation of the sensor and Fig. 12b the electronics it is connected to.*

- a postamplifier that low-pass filters and amplifies the final signal.

Note that the frequency of the angular rate signal from the sensor (see Eq. (18) implies that low frequency acceleration dependent signals can be cancelled via filtering. The purpose of the electronics is to show that it is possible to construct a rate sensor based on the described concept and to identify difficulties. The design criterion of this generation of electronics has not been that of optimizing performance.

Best performance was obtained with a 600 μm thick tuning fork having a resonance frequency of 37 kHz (cases 2 and 4 in Table 1). The crystal was encapsulated in air giving a quality factor of $Q_O \approx Q_S \approx 13\ 700$. This gives a 3 dB bandwidth of the peak of 2.7 Hz. The output from the sensor is 30 pA/°/s at a drive level of 5μW. This is well in agreement with predictions made using Eqs. (14)-(17).

The output noise for the crystal together with the first generation electronics corresponds to 0.4°/s (0.03 seconds integration time). Most of this is due to phase noise together with a large zero rate output. The noise can be expected to be substantially reduced with the next generation of electronics, due to the introduction of a capacitive coupling compensation together with the absence of cancellation of large signals.

The large zero rate output together with the unsophisticated electronics implied that the output signal was temperature sensitive (1°/s/°C).

Experiments have been performed on reducing the mechanical coupling by depositing material on selected areas of the tines. It has been shown that successful results can be obtained if gold is used.

Hysteresis and ordinary non-linearity are two quantities that are normally negligible for a solid-state sensor based on crystalline material. It has not been possible to distinguish any hysteresis and non-linearity of the sensor element from that of the electronics in the experiments. This indicates that these are not limiting parameters and are thus less interesting properties to study at this point.

IV. Discussion

In this article it has been described how predictions of the performance of piezoelectric beams can be made in a simple way. It was also shown how the vibration amplitude could be measured by varying the location of the electrodes. The method of prediction can be extended to plates/membranes as well as to more complicated geometries. However, the simplicity of the method will not come to its right in plates/membrane applications since the electric fields in such applications are much simpler than those used in beams.

It has also been discussed how a tuning fork angular rate sensor can be designed. It should be noted that beams are not the only geometry that can be used in vibrating rate sensors. Vibrating plates/membranes, cylinders and other structures can also be used as long as it is possible to activate one large vibration and to detect another vibration (e.g. shear) perpendicular to the first vibration.

Piezoelectricity is advantageous in this sensor application since the sensor element can be manufactured out of one single piece of material in a simple way. Other ways of activating and detecting a vibration exist also. Electrostatic and optothermal effects as well as piezoresistivity and changes in capacitance due to a deflection can be used.

V. Acknowledgement

This work is supported by a grant from the Swedish Board for Technical Development (STU).

VI. References

[1] J.S. Danel, F. Michel and G. Delapierre, Micromachining of quartz and its applications to an acceleration sensor, *Presented at 5th Int. Conf. on Solid-State Sensors and Actuators (Transducers '89)*, Montreux, Switzerland, June 25-30 (1989).

[2] L. Tenerz and B. Hök, Silicon microcavities fabricated with a new technique, *Electronic Letters*, 22(11) (1986) 615-616.

[3] H. Seidel, The mechanism of anisotropic silicon

etching and its relevance for micromachining, 4th *Int. Conf. on Solid-State Sensors and Actuators (Transducers '87)*, Tokyo, Japan, June 2-5 (1987) 120-125.

[4] E.P. EerNisse, R.W. Ward and R.B. Wiggins, Survey of quartz bulk resonator sensor technologies, *IEEE Tr. UFFC.*, **35**(3) (1988) 323-330.

[5] K.E. Petersen, Silicon as a mechanical material, *Proc. IEEE*, **70**(5) (1982) 420-457.

[6] M. Rodamaker, Finite element analysis of a quartz angular rate sensor, *1989 ANSYS Conf. Proc.*, Swanson Analysis System, Inc., (1989) 3.35-48.

[7] W.G. Cady, *Piezoelectricity, Vol. 1*, Dover, New York, 1964.

[8] J. Hermann, Determination of the electromechanical coupling factor of quartz bars vibrating in flexure or length-extension, *Proc. 29th symposium on frequency control*, (1975) 26-34.

[9] J.B. Chatterton, Some general comparisons between the vibratory and conventional rate sensors, *J. Aeronaut. Sci*, **22** (1955) 633-638.

[10] E.C. Alsenz, W.F. Juptner and D.F. Macy, *Angular rate sensor system*, U.S. Patent 4,654,663, Mar. 31 (1987).

[11] B.D. Popovič, *Introductory engineering electromagnetics*, Addison-Wesley Publishing Company, London, 1971, 80-100.

[12] J. Tichý and G. Gautschi, *Piezoelectrische Meßtechnik*, Springer-Verlag, New York, 1980, 54-73, 100-117.

[13] S. Timoshenko, D.H. Young and W. Weaver Jr., *Vibration problems in engineering*, Willey, New York, 1974.

[14] J. Söderkvist, Dynamic behavior of a piezoelectric beam, *submitted to J. Acoust. Soc. Am.*.

[15] J. Söderkvist, Electric equivalent circuit for flexural vibrations in piezoelectric materials, *submitted to IEEE Tr. UFFC.*.

[16] B.Y. Kim and H.J. Shaw, Fiber-optic gyroscopes, *IEEE Spectrum*, March (1986) 54-60.

[17] G.H. Hunt and A.E.W. Hobbs, Development of an accurate tuning-fork gyroscope, *Proc. Instn. Mech. Engrs.*, **179** (1964-65) 129-139.

[18] C.T. Morrow, Zero signals in the Sperry tuning fork Gyrotron, *J. Acoust. Soc. Am.*, **27**(3) (1955) 581-585.

[19] J. Söderkvist, An equivalent circuit description of two coupled vibrations, *submitted to J. Acoust. Soc. Am.*.

[20] W.S. Watson, *Curcuit for tracking and maintaining drive of actuator/mass at resonance*, U.S. Patent 4,479,098, Oct. 23, (1984).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Mode shape analysis techniques using synchrotron X-ray topography

B. Capelle, J. Détaint*, A. Zarka, Y. Zheng and J. Schwartzel*

Lab. de Minéralogie-Cristallographie, Universités P. & M. CURIE (Paris VI) et Paris VII, CNRS UA09.
4 place Jussieu, 75252 Paris Cedex 05, France

*Centre National d'Etudes des Télécommunications, PAB/BAG/MCT, 92220 Bagneux, France

Abstract : X-ray topography using synchrotron radiation was used to investigate vibrational states in quartz resonators. After a brief review of different methods and possibilities obtained with synchrotron radiation, some observations on vibration modes, especially on coupled piezo-electric components in quartz AT and BT resonators, are reported. The experiments performed revealed time-progressive components and several complex coupled components in AT and BT cuts.

Introduction

Laboratory X-ray topography [1] is an usual technique to characterise vibration modes in quartz [2]. The synchrotron radiation from a storage ring provides new facilities to this technique : the high flux of synchrotron X-ray sources and the large range of X-ray wavelengths permit to get Laue patterns of the resonators, thus to analyse the spatial structure of vibration modes; the pulsed synchrotron radiation allows the use of stroboscopic technique and the study of the time structure of vibration modes. Quartz resonators have been analyzed using synchrotron radiation at the D.C.I. storage ring (Orsay, France) and specific results have been obtained concerning coupled piezo-electric components.

I. Principles of the experimental techniques

The principle of the Lang topography is illustrated in figure 1. The drawing is made in one incidence plane. The incident beam coming from a point focus of an X-ray tube is collimated by a first slit to reduce the beam divergence to about one minute of arc. The crystal is placed on a goniometer head at the exact Bragg condition for a given set of reflecting planes characterized by a normal diffraction vector **g**. Inside the crystal waves are diffracted in all the Borrmann fan. Outside the crystal the transmitted beam is stopped by a second slit and the reflected beam falls on a photographic plate. In the Borrmann fan the wavefields belonging to both branches of the dispersion surface interfere and give rise for a crystal of constant thickness a set of fringes parallel to the edges of the image, called *Pendellösung* fringes. A traverse topograph is obtained when the crystal and the photographic plate are scanned together. It corresponds to the integration of a series of section topographs along the scanning direction. In the case of a perfect crystal no contrast exists on a traverse topograph since *Pendellösung* fringes on section topographs are integrated. In presence of crystal defects different contrasts appear and the fringes on section topographs are perturbed. The analysis of these contrasts permit to determine defect characteristics.

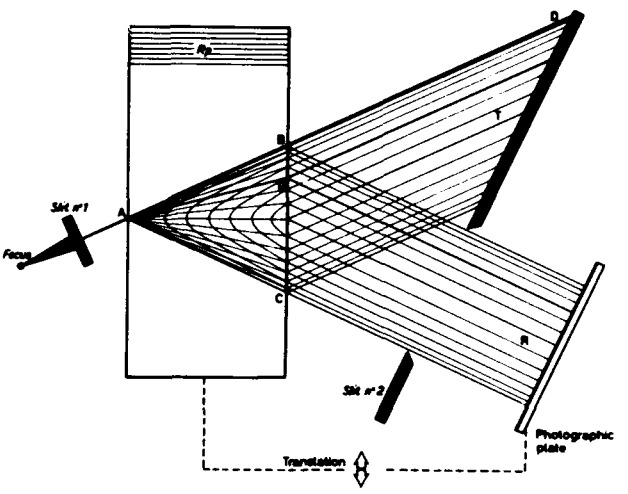


Fig.1. Principle of the Lang X-ray topography.

CH2818-3/90/0000-416 $1.00 ©1990 IEEE

Using synchrotron radiation it is possible to get Laue patterns with the white X-ray beam. A transmission Laue setting is shown in figure 2 where the crystal is simultaneously imaged with different diffraction vectors **g** and all the diffraction spots are recorded on a photographic plate. Each diffraction spot corresponds to a topograph made with the corresponding diffraction vector **g**. As like the laboratory topography two types of topographs are possible. A traverse topograph is obtained using a large incident X-ray beam (typically $15x7mm^2$) and a large incident X-ray beam with a low divergence is made possible in this case because the X-ray source is at a large distance of the sample (typically 23m). When the incident beam is limited by a fine slit (about 15µm), a section topograph is obtained.

The contrasts observed are related to both the diffraction vector **g** and the displacement vector **u** in the crystal. In particular for an acoustic wave the contrasts due to the vibration vanish when these two vectors are perpendicular (**g.u**=0) and the vibration is the best visualized when the two vectors are parallel. For each type of resonators one has to set the crystal in such a way that the suitable diffraction patterns **g** can be recorded. For quartz AT and BT resonators the three vectors **g** which separate at best the three displacements (u_1,u_2,u_3) are listed below (Tab.1). The two components u_1 and u_3 can be simultaneously visualized in rotating the crystal about respectively the x_3 and x_1 axis of the resonator on transmission Laue setting and the component u_2 on reflection Laue setting. In figure 3 is shown an example of the transmission Laue patterns obtained with a quartz AT plano-convex resonator operating at the 7th thickness-shear overtone and the components of u_1 and u_3 can be seen respectively on the spots $\mathbf{g}=2\bar{1}0$ and $\mathbf{g}=0\bar{1}2$.

components **u**	diffraction vectors **g**	
	AT cut	BT cut
u_1	$2\bar{1}0$	$2\bar{1}0$
u_2	011	$0\bar{1}2$
u_3	$0\bar{1}2$	011

Tab.1. Diffraction vectors g separating the three displacements for quartz AT and BT resonators.

To obtain classic (time-integrated) topographs vibrations in resonators are simply driven by a frequency generator and different resonances can be analyzed. In the case of

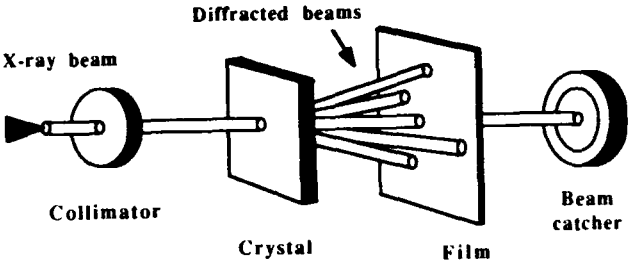


Fig.2. Principle of the transmission Laue setting.

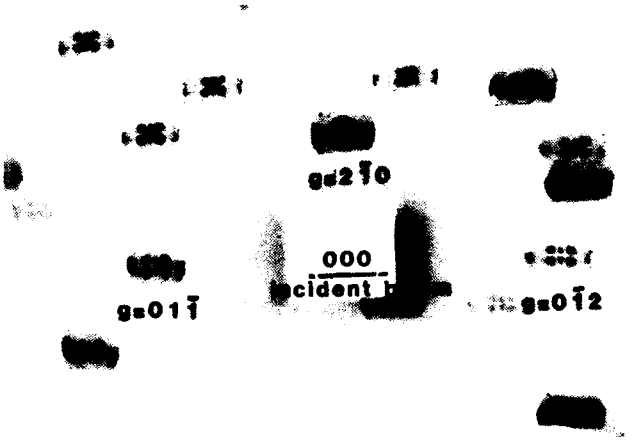


Fig.3. Transmission Laue patterns of a quartz AT plano-convex resonator where the resonator was rotated about 8 degrees around the both x_1 and x_3 axis to get the spots $\mathbf{g}=2\bar{1}0$ *and* $\mathbf{g}=0\bar{1}2$.

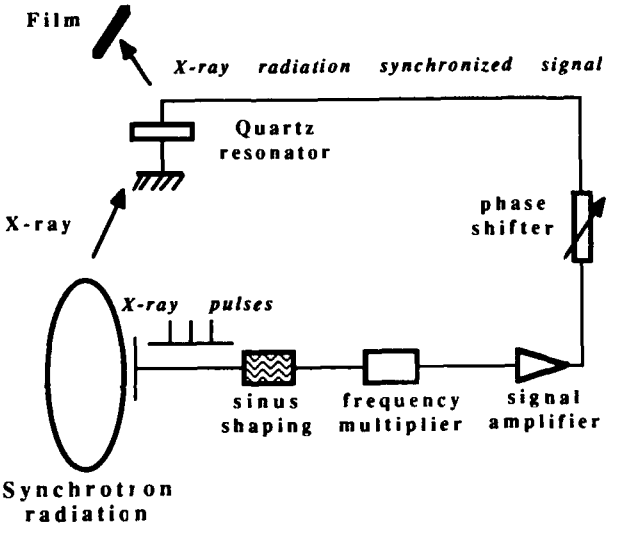


Fig.4.Principle of the stroboscopic X-ray topography.

stroboscopic (time-resolved) topography only one resonant vibration whose frequency is n-times of that of pulsed X-ray beam (about 3MHz) can be analyzed. The driving electrical signal on resonators is directly piloted by a signal taken by a capacitive probe in the storage ring, so that the exact synchronization between X-ray beam and vibrations is assured (Fig.4) [3].

II. Characterization of vibrations and crystal defects

X-ray dynamic contrasts were evidenced by stroboscopic X-ray topography in connection to vibration amplitudes in resonators. On section topographs extra *Pendellösung* fringes appear with increasing vibration amplitudes and on traverse ones *elliptical* fringes, resulted from the integration of the previous extra *Pendellösung* fringes over the whole resonator surface, constitute equal-level curves of vibration amplitudes. It was established that the fringe number is related to vibration amplitudes within a 0.3Å precision in usual cases [4,5].

In presence of crystal defects the modification of these typical X-ray contrasts characterizing vibrational states in resonators provide indications on the defects influences on acoustic vibrations. Concerning dislocations and etch channels it was observed that *elliptical* fringes are mostly disturbed in the vicinity of defects suggesting a perturbation of acoustic spatial coherences in these zones and the shape of the resonator vibrating zone is affected by defect bunches [6]. Similar results were obtained in presence of growth bands and a series of stroboscopic X-ray traverse topographs of a growth band in function of vibration amplitudes is shown in figures 5a-e. It can be clearly seen that *elliptical* fringes are also perturbed by the presence of the growth band.

Observations by X-ray topography revealed several interesting aspects of resonators which need still further investigations in order to get a better understanding. Two samples presenting different electrical phase noises were analyzed and a correlation between dislocations densities and electrical phase noises might be drawn from these observations. The sample of lower phase noise shown in figure 6a has a low density of dislocations and its vibration shape seems to be regular. In contrary for the sample of higher phase noise shown in figure 6b, the dislocation density is higher and the vibration shape is disturbed by dislocations. This point has to be detailed by further observations. Resonators under diametrical

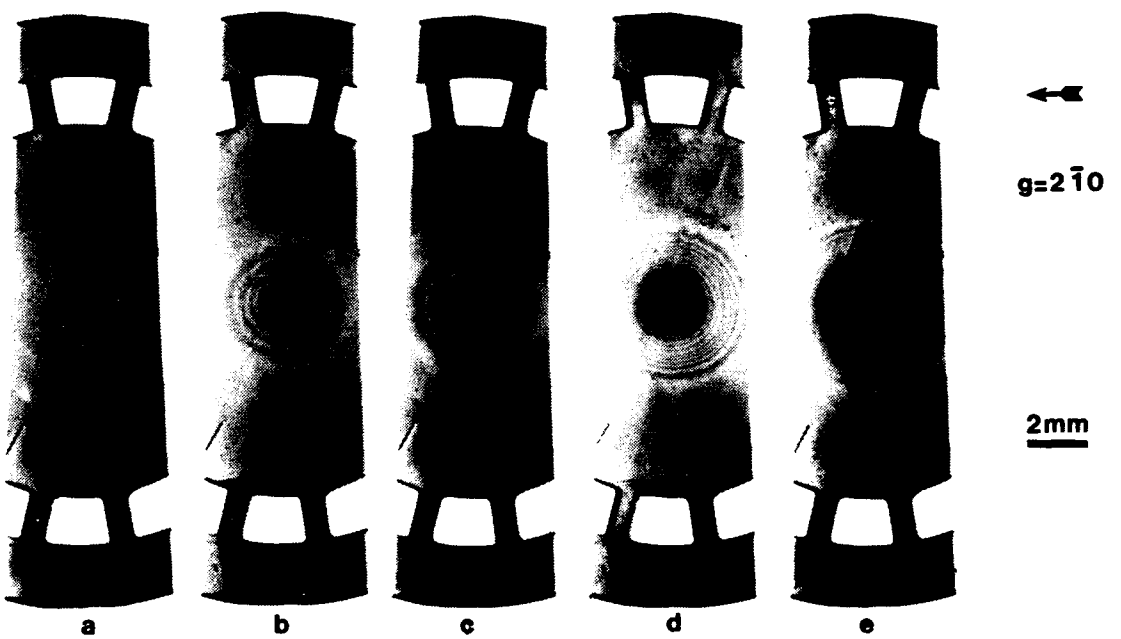


Fig.5a-e. Stroboscopic X-ray traverse topographs of a quartz AT resonator containing a growth band. a) without vibration; b)-e) with increasing vibration amplitudes.

Fig.6. Classic X-ray topographs of two quartz AT resonators presenting different electrical phase noises. a) sample of lower electrical phase noise; b) sample of higher electrical phase noise

compression were also analyzed in a preliminary study. Figure 7a shows a stress-free sample where *elliptical* fringes can be seen and the shape of the central vibrating zone is slightly affected by dislocations. When a diametrical compression along the x_1 axis was applied to the resonator under the same electrical excitation, the vibration contrast disappeared apparently due to the frequency shift induced by the stress and the dislocation images became blurred (Fig.7b). To get again the vibration contrast it was necessary to increase the level of the electrical excitation (ten times for the present case) on the resonator (Fig.7c). Further studies on the stress effect in connection with the mode shapes and resonator impedances will be carried out.

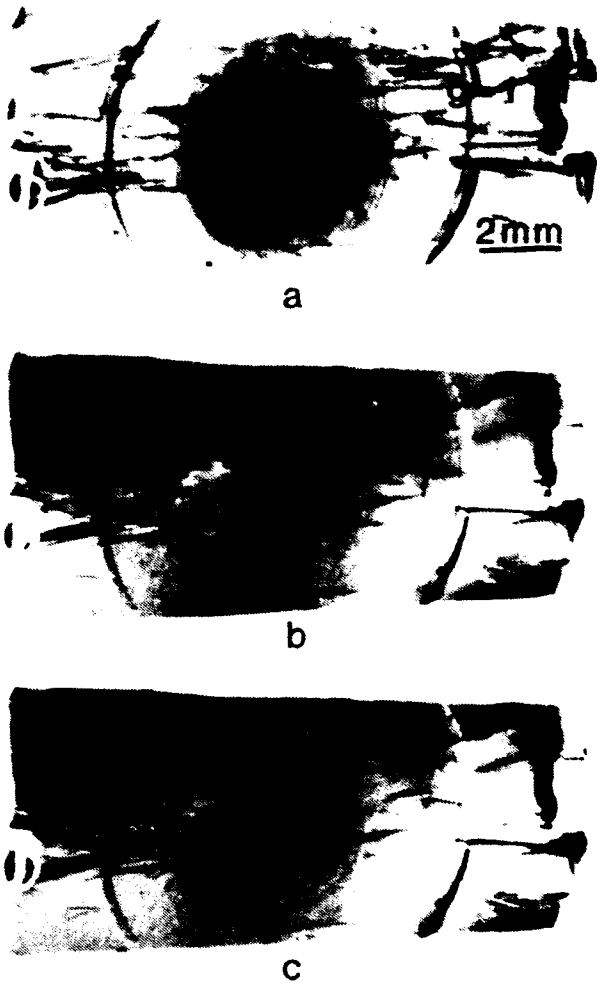


Fig.7. Classic X-ray topographs of a vibrating quartz AT resonator under a diametrical compression along the x_1 axis. a) stress-free; b) under a diametrical compression with the same electrical excitation as in a); c) under a diametrical compression with ten time higher electrical excitation than in a)

III. Characterization of vibration mode shapes

A mostly interesting aspect revealed by stroboscopic X-ray topography is the existence of time progressive acoustic component beside the classical acoustic standing waves. Observations made in high quality synthetic quartz AT resonators suggest that these time progressive components are expected to be untrapped and related to the energy dissipation in the resonators, since the intrinsic dissipation is very low in the material. The progressive components appear to couple all the three displacements (u_1,u_2,u_3) [4].

More classically for untrapped modes, coupled u_2 and/or u_3 standing components have always been observed in plano-plane resonators or in plano-convex resonators operating at low overtone modes. u_2 and/or u_3 components extend up to the resonator edges and they have a very complex spatial form. Figures 7a-b show a case of a quartz AT resonator operating at an anharmonic of the fundamental thickness shear mode (nmp=120).

Fig 8. Coupled standing and untrapped u_2 and/or u_3 components observed by classic X-ray topographs in a quartz AT plano convex resonator operating at on anharmonic of the fundamental thickness shear mode. a) u_1 component imaged with $\mathbf{g}=2\bar{1}0$; *b) u_2 and/or u_3 components imaged with* $\mathbf{g}=01\bar{3}$.

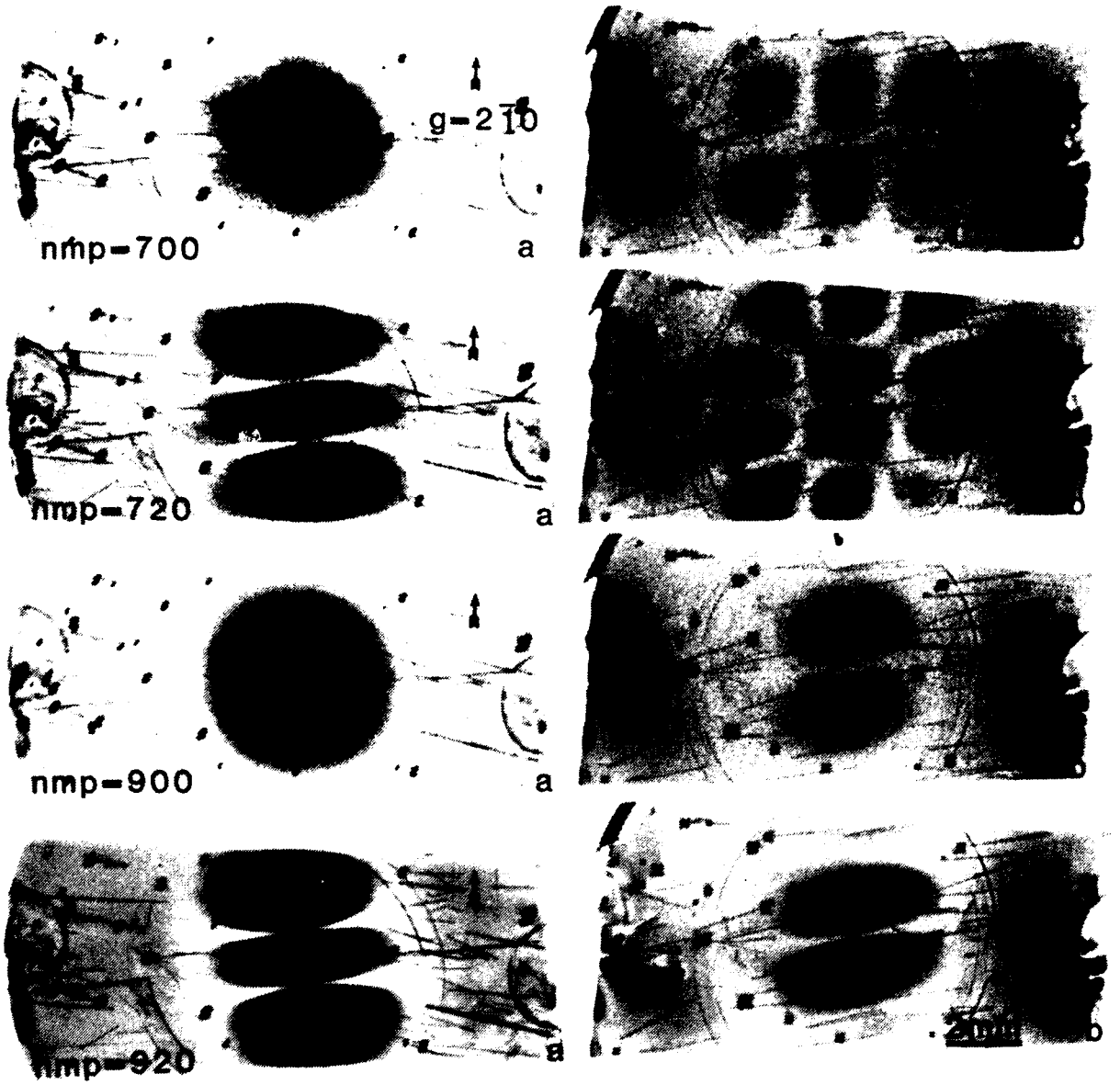


Fig.9 - Fig.12. Coupled u_1 and u_3 modes observed by classic X-ray topographs in a quartz AT-cut plano-convex resonator operating at respectively 7th and 9th overtone thickness shear modes. a) main u_1 modes imaged with $\mathbf{g}=2\bar{1}0$; *b) coupled u_3 modes imaged with* $\mathbf{g}=01\bar{3}$.

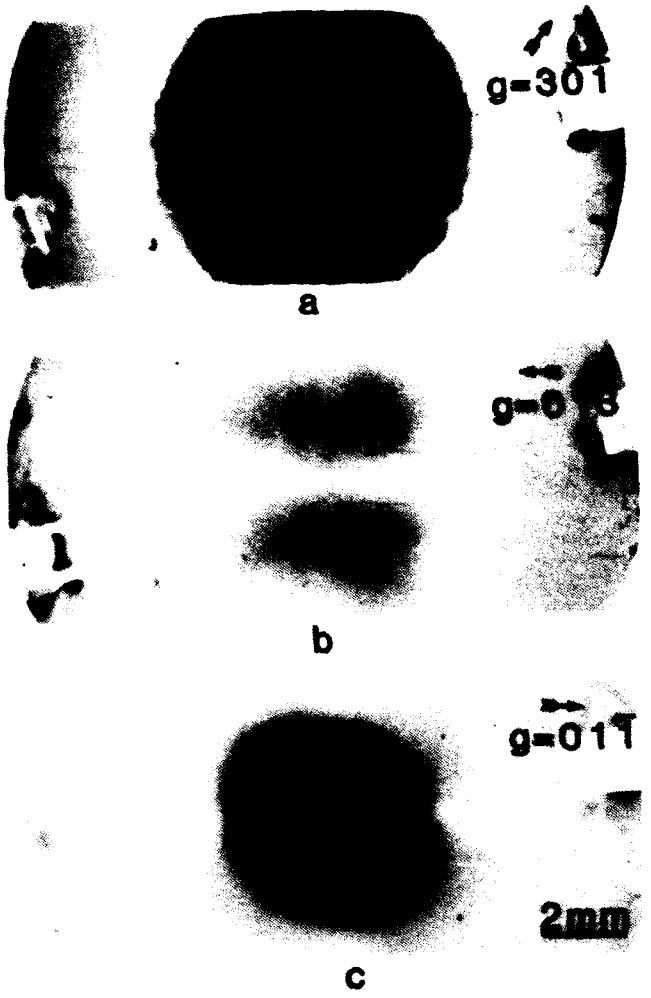


Fig.13. Three mode shapes observed in a quartz AT resonator. a) u_1 mode imaged with g=301*; b) u_3 mode imaged with* g=0$\bar{1}$3*; c) a mixture of u_2 and u_3 modes imaged by* g=01$\bar{1}$.

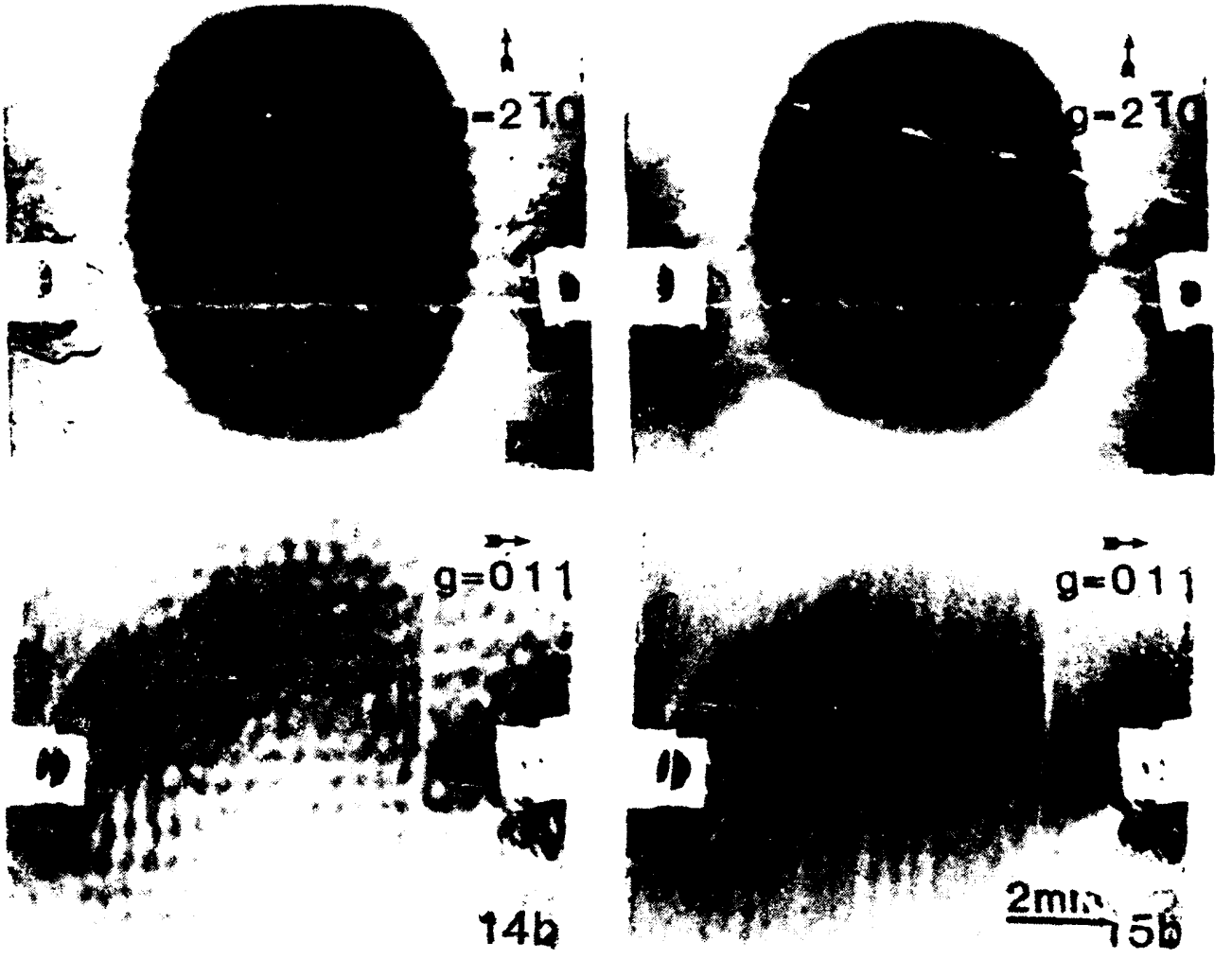
For trapped modes, strong mode coupling may occur with standing waves. It was observed that the two u_1 and u_3 thickness-shear modes are coupled particularly in the range of 7th (and 9th) overtones. Figures 9a-12b are examples of observed coupled modes in quartz AT plano-convex resonators. This coupling may be due to the overlapping of u_1 and u_3 mode frequency spectra in this range [7,8]. An other interesting case is shown in figures 13a-c where three different mode shapes were observed. The u_1 shape is represented in figure 13a, while the u_3 shape is approximately visualized in figure 13b and the third topograph (Fig.13c) presents a mixture of u_2 and u_3 components. The three distinct mode shapes observed might suggest a mode coupling three basic eigenmodes. This point has to be verified more precisely.

An interesting case of mode coupling was observed in highly excited resonators. It concerned a coupling between two u_1 modes which have a relative time phase :

$$u_{1\,(total)} = u_1^{(1)} e^{i\omega t} + u_1^{(2)} e^{i(\omega t+\phi)}$$

where $u_1^{(1)}$ represents the main mode, $u_1^{(2)}$ a secondary one and ϕ the relative time phase between two modes. Using the stroboscopic technique and by varying the relative time phase between X-ray beam and vibrations in such a way that a mode disappears from the observation, the modes were separated. While the main $u_1^{(1)}$ mode excited was the 5th overtone thickness-shear one, the secondary $u_1^{(2)}$ mode observed might be assigned to the closest antisymmetric anharmonic. This observation, available only by the stroboscopic technique, shall be instructive concerning non-linear effects in highly excited resonators. For the present case a correlation was observed between the detection of the secondary mode and the electrical hysteresis phenomenon related to the anisochronism effect.

Plano-plane BT cut resonators were designed with rectangular electrodes and analyzed by synchrotron X-ray topography. In figure 14a the u_1 component is seen. It can be noticed the presence of horizontal fringes which are due to the u_1 component of the coupled flexure. This component stays relatively confined. The u_2 component of the coupled flexure is seen in figure 14c with almost regular horizontal fringes, whereas the u_3 component appears also with the presence of vertical and horizontal fringes (Fig.14b). Corresponding time integrated topographs show that the u_1 component including the coupled flexure contribution is standing (Fig.15a), but for the u_3 component the part which disappears is progressive (Fig.15b). The vertical fringes observed in figure 15b are rather difficult to interpret, but they correspond to standing part of the u_3 component. This example showed that, for plano-plane BT rectangular resonators, the thickness-shear u_1 modes are coupled to the flexure modes as it can be expected. More surprising is the u_3 component including not only standing and progressive parts but also two sets of fringes suggesting several coupled modes in the crystal.


g=2 1 0
g=2 1 0
g=0 1 1
g=0 1 1
14b
2mm
15b
2

Conclusion

In this paper examples were given to illustrate the use of synchrotron X-ray topography to analyse different aspects of resonators, namely the coupled mode shapes. For quartz AT and BT resonators under thickness-shear (u_1) excitation several coupled components can be induced and they may be trapped or untrapped and standing or progressive. These observations provide interesting data to further theoretical calculations of the acoustic displacement field in resonators.

Acknowledgement : The authors would like to acknowledge SICN, CEPE and CQE companies and Dr. Euler of RADC (USA) for kindly providing samples studied here. This work was supported by D.R.E.T. (France) through the contract no. 87 34 13 80 04 70 75 01.

References :

[1] A.R. Lang (1959). Acta Crystallogr., **12**, 249.

[2] W.J. Spencer (1972). Phys. Acoust., **9**, 167.

[3] A. Zarka, B. Capelle, J. Détaint and J. Schwartzel (1988). J. Appl. Cryst., **21**, 967-971.

[4] A. Zarka, B. Capelle, Y. Zheng, J. Détaint and J. Schwartzel (1988). Proc. of the 42nd Annual Symp. on Freq. Control, p.85-92, 1-3 June 1988, Baltimore, USA.

[5] Y. Zheng, A. Zarka, B. Capelle, J. Détaint and J. Schwartzel (1989). Acta Cryst., A**45**, 275-285.

[6] B. Capelle, A. Zarka, Y. Zheng, J. Détaint and J. Schwartzel (1989). Proc. of the 43rd Annual Symp. on Freq. Control, p.470-476, 2 May - 2 June 1989, Denver, USA.

[7] R. Bourquin, B. Dulmet, G. Genestier (1985). Proc. of the 39th Annual Symp. on Freq. Control, p.405-411, 29-31 May 1985, Philadelphia, USA.

[8] B. Capelle, J. Détaint, Y. Zheng, J. Schwartzel and A. Zarka (1990). Proc. of the 4th European Time and Frequency Forum, 13-15 March 1990, Neuchatel, Switzerland.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

NON-DESTRUCTIVE MEANS FOR DETERMINING MODE SHAPE IN THICKNESS-SHEAR RESONATORS

Roger J Williamson

STC Technology Ltd
London Road, Harlow, Essex, England

Abstract

A number of methods of plotting the vibrational distribution of quartz resonators have been developed. Probably the first attempt at any vibrating structure was by Chladni in 1787 using fine powder. Many of the methods used light as a probe, perhaps the most well known being Lang topography using X-rays. Others using interferometry are only really applicable to modes with a predominantly normal component, such as flexural modes. A review of the different methods of plotting the vibrational distribution of resonators will be given. This will include a new method that is applicable to thickness-shear modes at high frequencies.

The method uses light as a probe in a similar way as a toothpick is commonly used in the quartz industry. There the toothpick is gently tracked across the surface of the resonator whilst the electrical response of the crystal is monitored. A mixture of damping and loading by the toothpick alters the crystal in both the frequency and amplitude domains. Using light, however, the effect is to locally perturb the standing wave of the resonating area. This perturbation manifests itself as a transitory change in the frequency of the resonance. This change is proportional to the magnitude of the standing wave. Thus by rastering the pulsed light beam across the area of the resonator and monitoring the change in its resonant frequency, the vibrational distribution can be mapped.

1.REVIEW OF OTHER METHODS

1.1 Introduction

A number of methods of plotting the vibrational distribution of quartz resonators have been developed. There are a few non-optical methods using either powders or electrical probes, but most use light. Many of the methods use the wave properties of photons in that interference patterns are produced. The most well known method is Lang topography using X-rays. The earlier methods using interferometry were only really applicable to modes with a predominantly normal component, such as flexural modes, and at low frequencies. Latterly, since the advent of the laser, a number of methods have been developed which are capable of probing the in-plane vibrations of thickness-shear resonators. Other fundamental particles, such as neutrons and electrons, have also been used in probing quartz vibrations. For instance, very interesting results have been produced from SEM investigations.

1.2 Non-optical Methods

1.2.1 Powder Methods

The earliest method for visualising a resonating structure is attributed to Chladni [1] in 1787. Here it was assumed that the powder settled along the nodes. However, Savart [2] in 1820 showed that there were more nodes indicated than existed. Wood [3] analysed the problem taking into account both horizontal and vertical displacements and showed that where both displacements occurred, the pattern became more complex. Furthermore, the method requires large displacements to move the powder and is hence unsuitable for thickness-shear resonators.

1.2.2 Electrical Probe Methods

Two probe methods have been described. One method, due to van Dyke [4], maps the strain across the crystal, whereas the other, due to Koga [5], maps charge polarisation. The equipment used by both is similar, in that a small electrical probe protrudes through an earth plane, both probe and plane being close to the surface of the crystal (figure 1). In the method of van Dyke, the opposite face of the crystal also has an earth plane and the probe forms one arm of an admittance bridge. In the Koga and Fukuyo method the face opposite the probe is connected to a frequency source which drives the resonator, and the probe is connected to a vacuum tube voltmeter. Recently, Dworsky [6] has applied the van Dyke method to

CH2818-3/90/0000-424 $1.00 © 1990 IEEE

investigate contoured blanks. The main disadvantage of these methods is that they are not readily applicable to normal electroded devices.

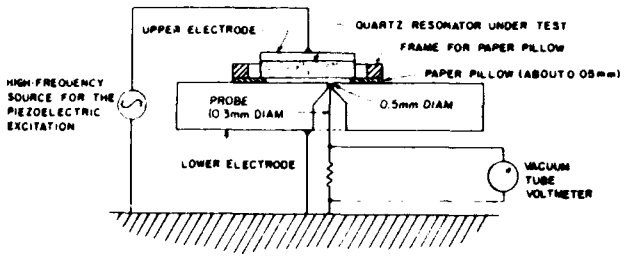


Figure 1. Schematic Diagram of Electrical Probe Method [Ref 5]

1.2.3 SEM Methods

In an SEM (scanning electron microscope) a focussed electron beam is rastered across the surface being investigated. Ordinarily, the specimen is coated with a thin conductive film and the secondary electron emission corresponds to the topological features of the specimen surface. When used to probe resonators, however, it is the charge distribution on the surface due to the vibration that is imaged. Thus only unelectroded parts of the resonator can be plotted. The acceleration potential must be kept between 1 and 2 kV in order to prevent charge build-up on the surface of the quartz.

Gerdes and Wagner [7] optimised the conditions for imaging. In order to image both inside and outside the resonator they used annular electrodes, and wavy patterns appeared in both areas. Initially, it was thought that the patterns inside the annulus were due to thickness-shear vibrations. However, these patterns did not correlate with X-ray topographs. Subsequent work by Bahadur et al. [8,9] demonstrated that these patterns were probably due to flexural waves spreading out from the plated area and setting up standing waves. The use of annular electrodes had significantly reduced the energy trapping of the thickness-shear waves. The SEM method is therefore best used for investigating energy trapping and mode conversion with any resultant plate modes.

Thickness-shear vibrations can, however, be imaged by SEM by using very high magnifications, of about 2000. The electron images of a vibrating crystal show an apparent smearing effect due to the in-plane vibrations of thickness-shear. With a few well-defined points like topological features to focus on, it is possible to measure both the magnitude and direction of the vibrations.

1.2.4 Neutron Diffraction Topography

For completeness, the use of the wave characteristics of slow neutrons to investigate the vibrations in resonators is included [10,11]. The principle is much the same as using X-rays in that the diffraction intensity is increased as a result of crystal vibrations and is many times greater than that from X-rays. However, the amplitude of vibrations required is very high, the experimental conditions are complex and the exposure time is very long.

1.3 Optical Methods

1.3.1 Glow-Discharge Method

This method is of historical interest only and is described in Ref. [12]. Giebe and Scheibe were the first to show that luminous effects can be exhibited around a vibrating piezoelectric crystal. As in the case of the SEM method, surface charges are produced under the antinodes of standing waves. If the device is in an argon atmosphere of about 10 torr and one of the driving electrodes is about 0.5 mm away, then glows appear between the crystal and this electrode when it is driven at resonance.

1.3.2 Sauerbrey's Method

This is an elegant method [13] whereby a focussed light beam is reflected normally from the surface of a vibrating resonator whose electrodes are given a reflective pattern (figure 2). As a result of the lateral vibration of thickness-shear

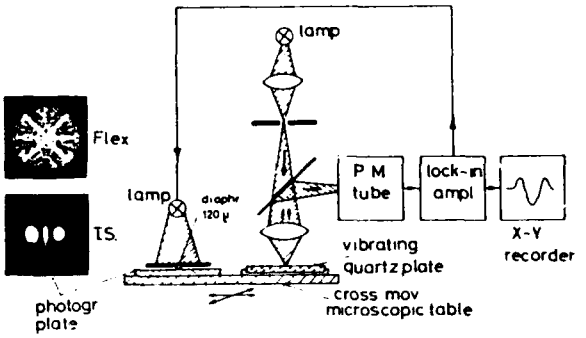


Figure 2. Schematic Diagram of Sauerbrey's Method [Ref 13]

resonators, the focussed beam passes over the edge of the pattern and hence the reflected beam is amplitude-modulated. By plotting this modulation against position, a map of the vibration can be obtained with a resolution of 1Å. Figure 3 shows some typical results.

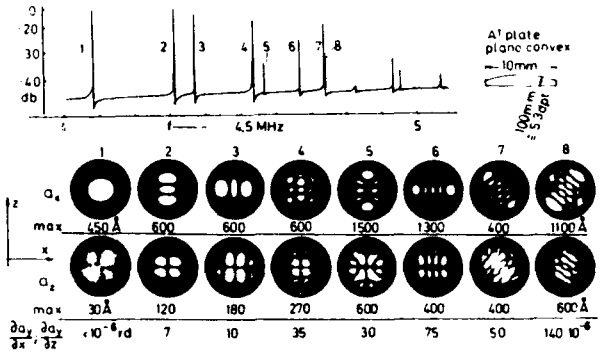


Figure 3. Plots from Sauerbrey's Method [Ref 13]

1.3.3 Laser Speckle Method

In principle this uses the same idea as Sauerbrey's method except that it uses the long coherence length of modern lasers. If a laser beam such as that from a HeNe laser is reflected from a surface which has a surface roughness greater than the wavelength, the resultant wavefront consists of a speckle pattern. This is caused by the differing path lengths from the rough surface interfering with one another. This speckle pattern is very sensitive to changes of the path length

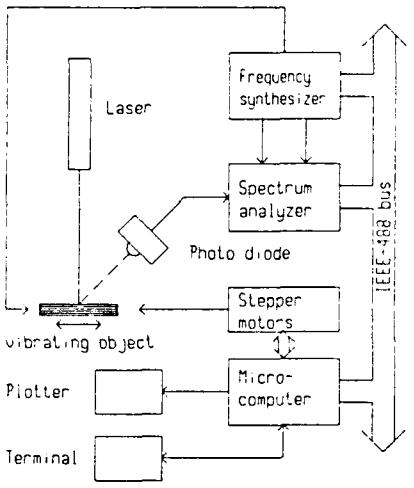


Figure 4. Schematic Diagram of Laser Speckle Method [Ref 14]

such as that resulting from the vibration of the surface. Hertl et al [14] detected the movement of the speckle by means of a photo-diode. By rastering the laser across the resonator surface (figure 4) a plot of the in-plane vibrational distribution can be formed to a resolution of 1 nm.

1.3.4 Michelson Interferometry

Osterberg [15] used a modified Michelson interferometer to observe normal vibrations of crystal resonators (X-cut). Here, one of the mirrors is replaced by the quartz crystal. The effect of the vibration of the crystal was to smear the interference patterns. Disadvantages of the method are its inability to measure in-plane vibrations and the need for an excellent optical finish on the crystal surface.

1.3.5 Multiple Beam Interferometry

In multiple beam interferometry, the interference fringes are formed by many passes of the light through the interferometer. Advantages of this method are that the sharpness (finesse) of the fringes can be enhanced by using appropriate end-mirror reflection, and the path length is increased by the multiple passes. Tolansky [16] made the crystal itself the interferometer and, as in the case of the Michelson interferometer, the resultant fringes are smeared due to the vibrations. Also, the method is only capable of measuring normal vibration.

1.3.6 Michelson/Mach-Zehnder Interferometry

This method as originally developed by a team from Yokohama National University, used a HeNe laser as the light source [17,18]. Recently [19] they have extended it by the use of an optical fibre. In the Michelson configuration, it can measure the normal vibration of a resonator as in the Osterberg method and in the Mach-Zehnder configuration it can measure in-plane vibrations. The modified method differs from the earlier method in that a low frequency plus DC optical path length modulation is placed on the reference arm of the interferometer. This adds a phase modulation to that signal when it is remixed with the light reflected from the crystal, which is itself phase-modulated at the crystal resonator frequency. By ratioing the amplitudes of these two components of the signal received by the photodetector, the amplitudes of the two beams cancel out, leaving the magnitude of the crystal vibration with respect to the

known low frequency modulation. Figure 5 shows the arrangement for measuring the in-plane displacements of a crystal resonator. For this to work, it is necessary for the resonator to have a diffuse surface to scatter the light into the detector. The fibre optic version is claimed to be able to measure displacements in all three directions to 0.1Å with a spot size of 10 μm up to 150 MHz.

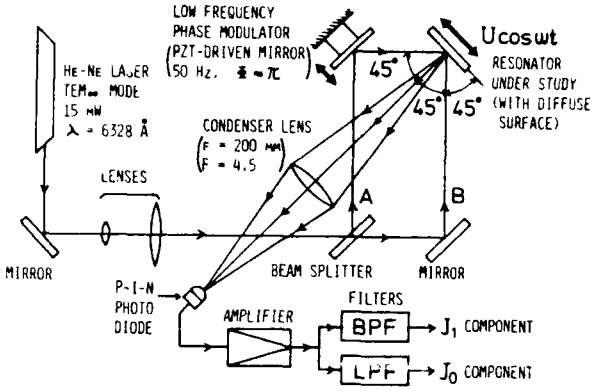


Figure 5. Schematic Diagram of Mach-Zehnder Interferometer Method [Ref 17]

1.3.7 Holographic Method

Tsuzuki et al. [20] managed to create holograms of flexural and extensional-mode resonators onto a photographic plate. The reference beam from the laser is made incident on the photographic plate by specular reflection from a mirror. The surface of the crystal resonator is made diffuse so that the laser beam incident on it reaches the photographic plate by scattering. The two beams interfere to make the hologram which is a space-modulated optical grating. The hologram is reconstructed by placing the hologram where the plate was and illuminating it by the reference light in the absence of the resonator. Apart from the difficulty of interpreting the resultant pattern, the method also needs amplitude displacements of several wavelengths in order to work.

1.3.8 X-Ray Topography

This method has been used more than any other. It was originated by Lang for the study of silicon and germanium. The method is illustrated in figure 6. An X-ray beam is first collimated by two sets of slits, one near the X-ray tube and the other about a metre away. The collimated ribbon of X-rays is diffracted from the planes of the crystal and allowed to pass through a beam stop onto an X-ray sensitive film.

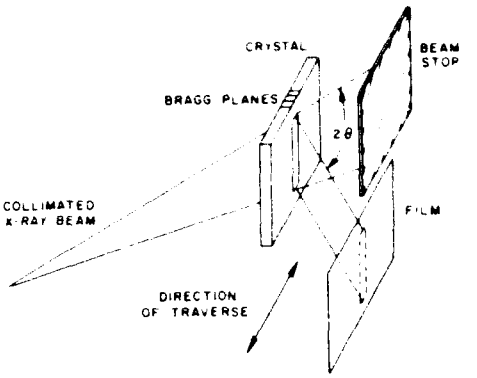


Figure 6. Schematic Diagram of X-ray Topographic Method [Ref 22]

The crystal and film are locked together at the Bragg angle and translated to obtain a photographic record of the diffracted intensity over a large area of the plate. For a perfect crystal, the lattice planes are perfectly parallel and light satisfying the Bragg condition, and being diffracted from one plane, is diffracted back by the next plane. However, strained regions around defects diffract a much higher percentage of the incident X-ray beam than the perfect portion of the crystal.

The intensity of the diffracted X-ray is also very sensitive to acoustic displacements so that vibrations can also be observed. Fox and Carr [21] were the first to observe the change in intensity of X-rays diffracted from vibrating quartz plates. Spencer [22,23] took up the technique and produced some of the most well known pictures (figure 7).

The topographs produced by the above method are time-integrated. Recently, Zarka et al [24,25,26] have enhanced the sensitivity of the method by using synchrotron radiation from a storage ring. Time-resolved pictures of vibrating resonators have been obtained by synchronising the X-ray pulses with the vibrations of the resonators.

One other X-ray method, the Source Image Distortion (SID) method of Wagner and Young [27], is also worth mentioning. In this method the line X-ray source is in front of Soller slits. These consist of 150 mm long sheets, 0.315 mm apart, and help to remove non-parallel components from the X-ray beam. Thus each slit forms a separate virtual source of parallel X-rays which are diffracted from the crystal. The resultant diffracted image of the whole perfect crystal will consist of parallel lines, which become distorted and intensified as a result of strain in the

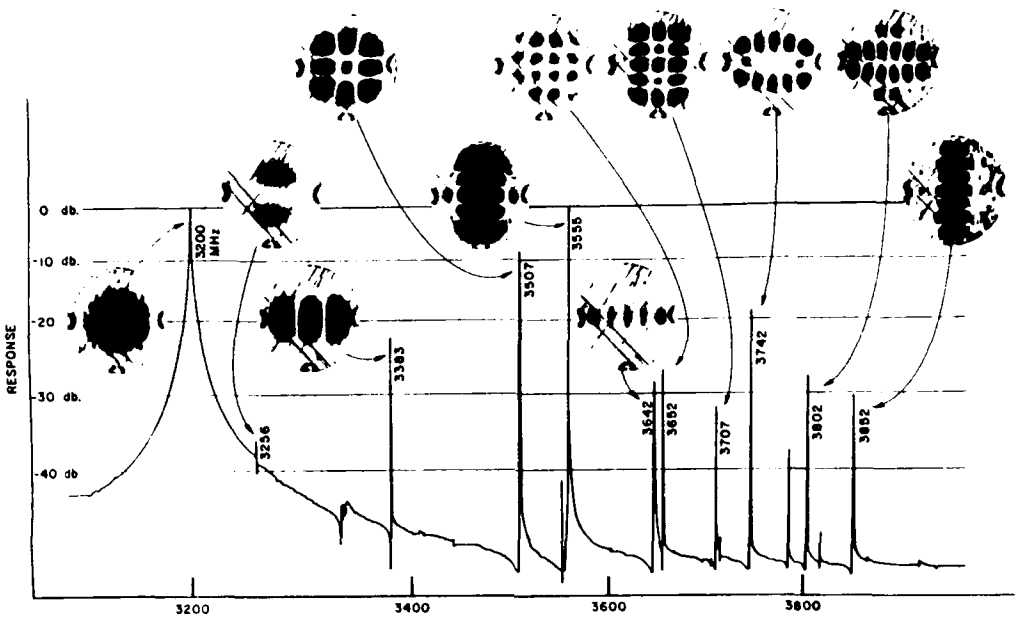


Figure 7. X-ray Topographs of an AT-cut resonator [ref 22]

lattice. Thickness-shear vibrations show up as bright areas. The method has the advantage that scanning techniques are not required and the exposure time is less than for traditional topography.

1.4 Conclusion

Some of the above methods are described in a review article by Bahadur and Parshad [28]. In terms of the present discussion, only a few can measure the in-plane vibration of a thickness-shear resonator. There is no ideal method yet devised, they all impose constraints on the type of device. In the case of the electrical probe method, for instance, the device has to be unelectroded. Sauerbrey's method requires a grid drawn over the surface of the resonator. The recent laser methods using either speckle or interferometry require a rough surface. Finally the X-topographic methods are constrained in the thickness of the quartz and hence the frequency of operation.

2. LASER PROBE METHOD

2.1 Introduction

The laser probe method [29] is applicable to thickness-shear modes at high frequencies and can be likened to the classical toothpick method. There the toothpick is gently tracked across the surface of the resonator whilst the electrical response of the crystal is monitored. A mixture of dampening and loading by the toothpick alters the crystal in both the frequency and amplitude domains. Great skill is needed to use this method since the likelihood of breaking the crystal is high. In the laser probe method the laser beam replaces the toothpick and the effect is only to modulate the crystal frequency. It is proposed that the mechanism for this modulation is the force-frequency effect whereby the light-induced heat pulse creates a localised compression in the quartz due to thermal expansion.

2.2 Equipment

The laser used for the current work was a Questek 2040 excimer laser. In operation, excited dimers are formed between a noble gas and a halogen such as argon fluoride which emits at 193 nm. The output energy of the laser can be varied between 20 and 200 mJ per 15 ns pulse with a repetition rate up to 100 Hz. The output beam is about 10 by 20 mm^2 so that the fluence is between 10 and 100 mJ/cm^2. The large beam size enables a very simple means of beam placement to be used, as shown in figure 8. A computer-controlled X-Y table moves a brass foil containing a number of apertures, thus enabling different areas to be irradiated but with the same fluence. In front of the foil is a beam expander which has a twofold purpose: it reduces the fluence available from the laser and evens out the intensity distribution across the beam. After the foil a simple lens system is positioned at twice the focal length from both the foil and the crystal. This forms an inverted image (magnitude -1) of the foil and reduces the diffraction effects from the apertures.

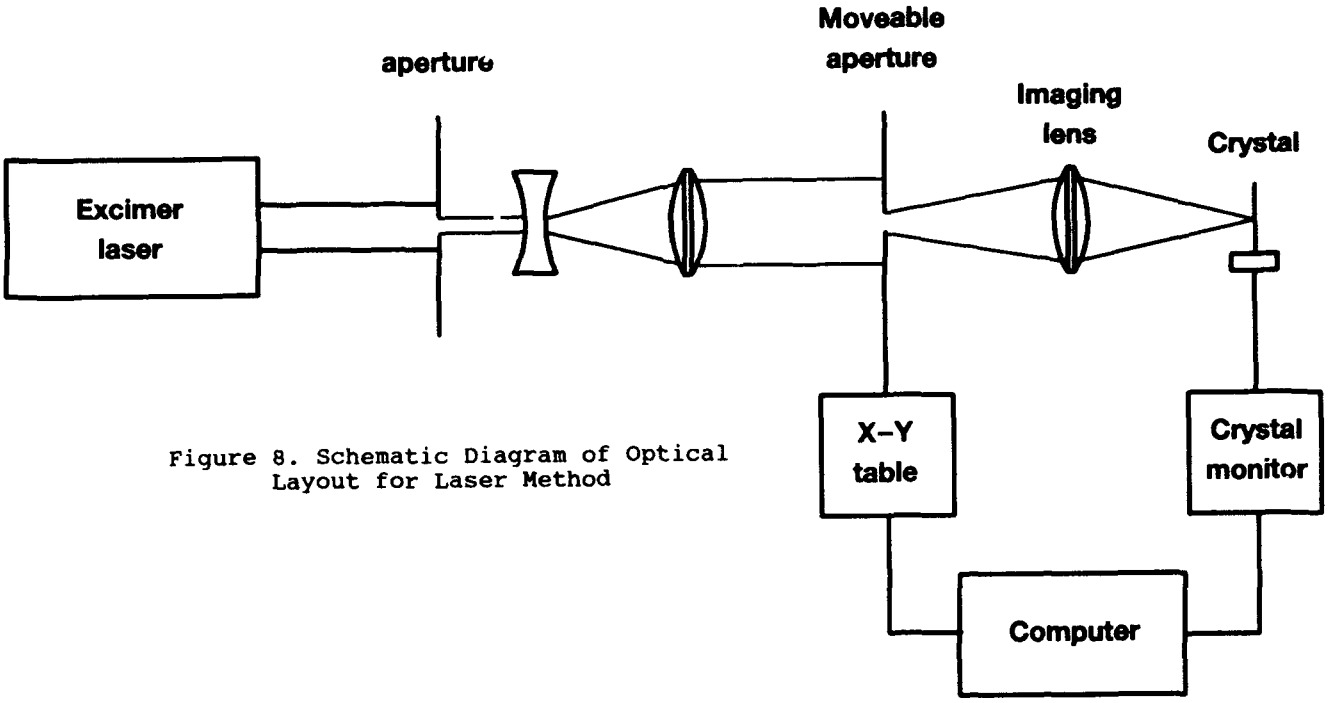


Figure 8. Schematic Diagram of Optical Layout for Laser Method

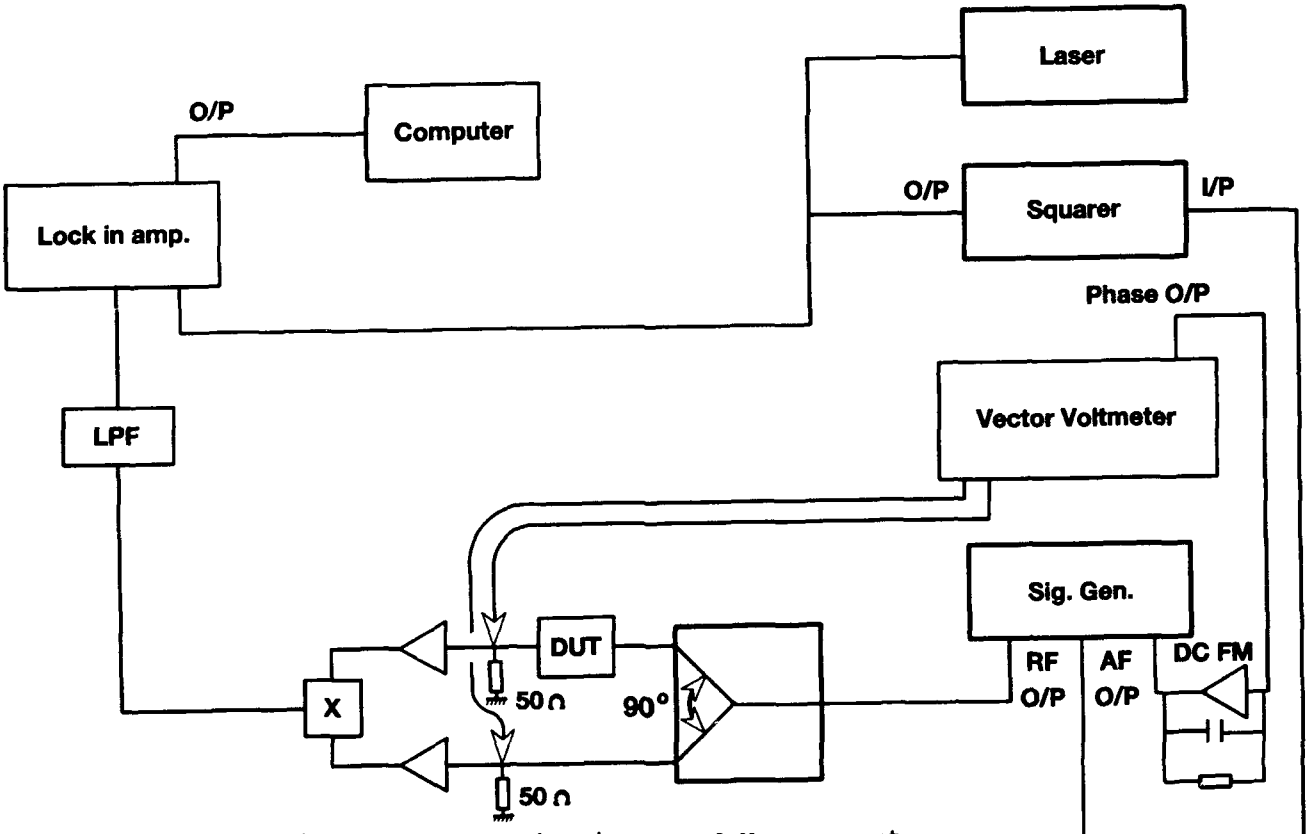


Figure 9. Schematic Diagram of Measurement Layout for Laser Method

The layout of the crystal measurement equipment is shown in figure 9. The signal generator used was a Rohde and Schwarz SMG since it has extensive modulation capabilities. The vector voltmeter was the HP 8405A and was used to find the resonance. Unique to what was in essence a modified IEC 444 crystal measurement circuit was the quadrature hybrid from Anzac (JH 131). This power splits with a ninety degree phase shift within a few degrees over the range 20 MHz to 200 MHz. This considerably simplified the circuit since line stretchers were not required when the frequency was changed. The crystal was measured in a 100 ohm circuit made from SMA coaxial attenuators. This is not ideal from the point of view of sensitivity, since it substantially increases the loaded Q of typical resonators. However, it made for a much simpler jig. Two identical amplifiers from Minicircuits (ZFL500LN) were used to bring the signal levels up to about 7 dBm to maximise the sensitivity of the HP 10514 double-balanced mixer used for phase detection. The output of the mixer was low-pass filtered to remove the sum component and then measured using a Stanford Research SR510 lock-in amplifier. This was also fed from the same signal source that was used to trigger the laser (the AF output of the SMG). The lock-in amplifier is in essence a narrow band detector so that the wanted signal can be measured buried in noise. The phase output from the vector voltmeter can be used instead of the mixer, in which case both the quadrature hybrid and the amplifiers are unnecessary. However, the rise time of the phase modulation will be slugged. Furthermore, that output is noisier due to supply interference as well as the 20 kHz IF remnants. Nonetheless, that output has been successfully used for plotting.

2.3 Method

For plotting the vibrational distribution of a quartz resonator, the laser beam was rastered across the electrode of the crystal whilst it was sitting on resonance. The laser spot sizes used were 0.15 mm and 0.3 mm, depending on the size of the resonator electrode and the degree of detail required. The laser fluence used was about 3 mJ/cm^2 since this gave adequate response without damage to the electrode. Repetition rates used were typically 40 Hz since this was a compromise between avoiding both 1/f noise and supply interference whilst maintaining reasonable gas lifetimes for the laser.

To avoid any drift in the crystal frequency whilst plotting, the output from the mixer can be fed via a very-low-pass filter into the DC FM input of the signal generator. However, since a synthesised source was used and the ambient was well controlled, this proved unnecessary. The rastering and voltage measurement were under the control of a computer using the IEEE interface. It proved necessary to take an average of several measurements at each point since the pulse to pulse variation of the laser energy was at least 20%. The data were then used to generate contours which were drawn on the screen of the computer.

2.4 Results

This method has been used to plot the vibrational distribution of both overtones and inharmonics. Figure 10 gives a sequence of plots of a 50 MHz fundamental and its overtones up to the 11th at 550 MHz, illustrating that the method is very sensitive. The circle in the background corresponds to the position of the 1 mm electrodes and the dots are the measurement positions. In this case, a 0.3 mm laser spot was used with no overlap of the spots. Despite the large spot used, the main characteristics of this resonator

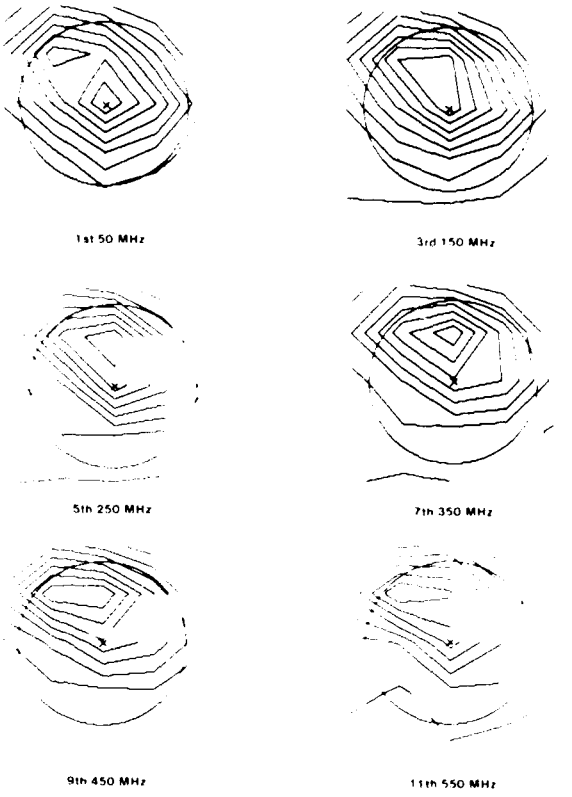


Figure 10. Plots of 50 MHz Fundamental and Overtones

can be seen. There is a wedge in the quartz towards the top-left-hand corner, the effect of which gets more pronounced with higher overtones.

Figure 11 gives a sequence of a 64.3 MHz fundamental with some of its inharmonics. In this case, the laser spot diameter was 0.15 mm so that greater detail is revealed. Resolution is limited by the spot size and lateral thermal diffusion, both of which will smooth out detail. This will affect the plots of the inharmonics most where the nodal lines are suppressed. However, the general shape is still evident. The crystal used was of the standard key-hole design with the tabs positioned on both sides of the electrode (the circle). It can be seen how the vibrational distribution extends down the tabs for the higher inharmonics.

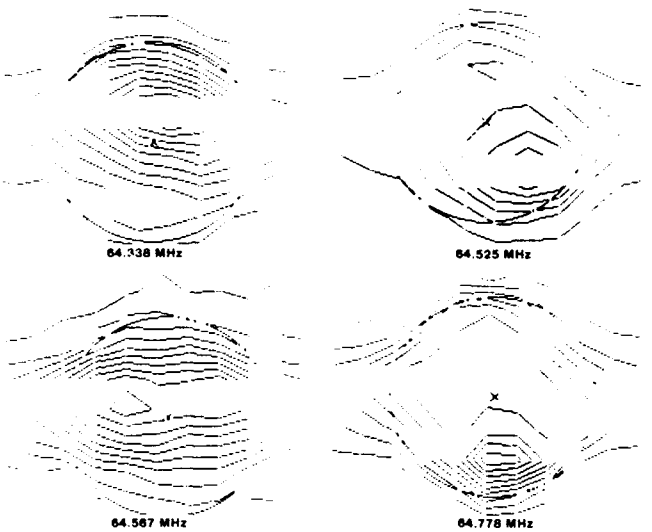


Figure 11. Plots of 64 MHz Fundamental and Inharmonics

2.5 Mechanism

In order to elucidate the mechanism responsible for the observed phase shifts used for the plotting process, the time constants and the magnitudes of the phase shifts have been investigated. It is assumed that a fraction of the laser energy from each 15 ns pulse is absorbed by the electrode and converted into heat which then diffuses into the quartz. It is also assumed that for beam sizes large compared with the thickness of the quartz, lateral diffusion may be ignored initially.

2.5.1 Time Domain

The output of the mixer from a 50 MHz fundamental crystal irradiated by a single laser pulse is given in figures 12 and 13.

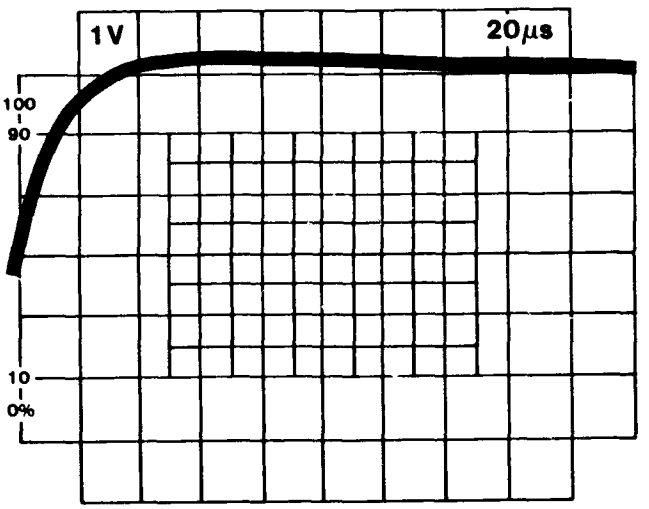


Figure 12. Rising Edge of Laser Induced Frequency Modulation

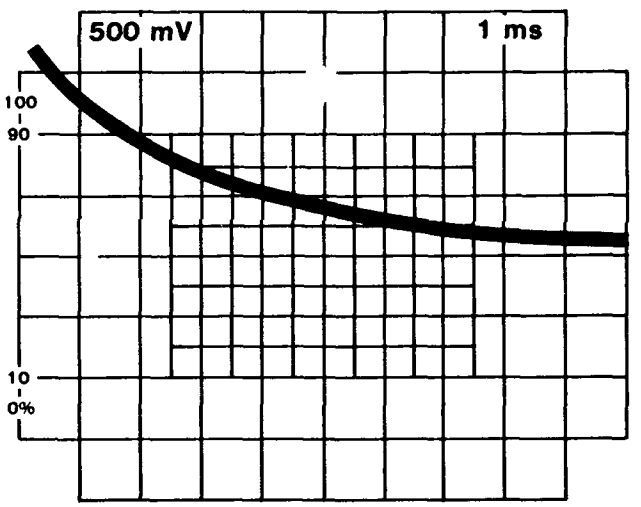


Figure 13. Falling Edge of Laser Induced Frequency Modulation

This shows that the rise time of the phase shift is exponential and is about three orders of magnitude faster than the fall time. Figure 14 plots this rise time (measured assuming an exponential form) for a number of crystals between 50 MHz and 114 MHz, and shows that it is proportional to $1/f^2$. This relationship suggests that the phase pulse may be the result of a laser-induced heat pulse through the thickness of the quartz.

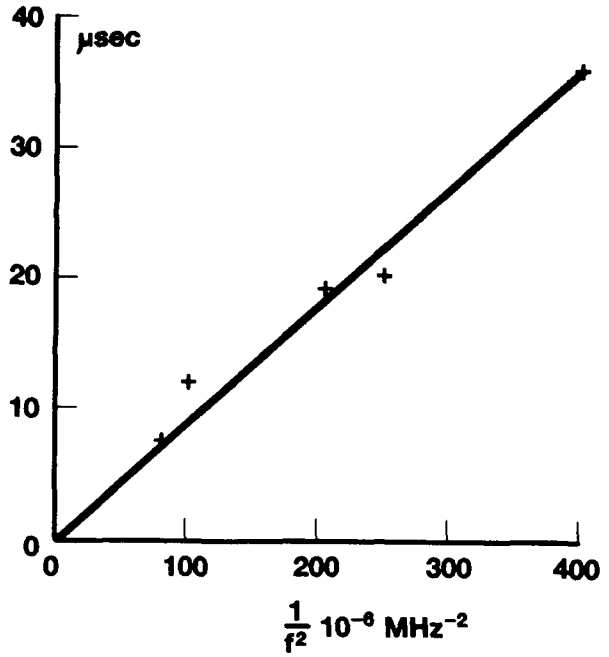


Figure 14. Plot of Rise Time with $1/f^2$

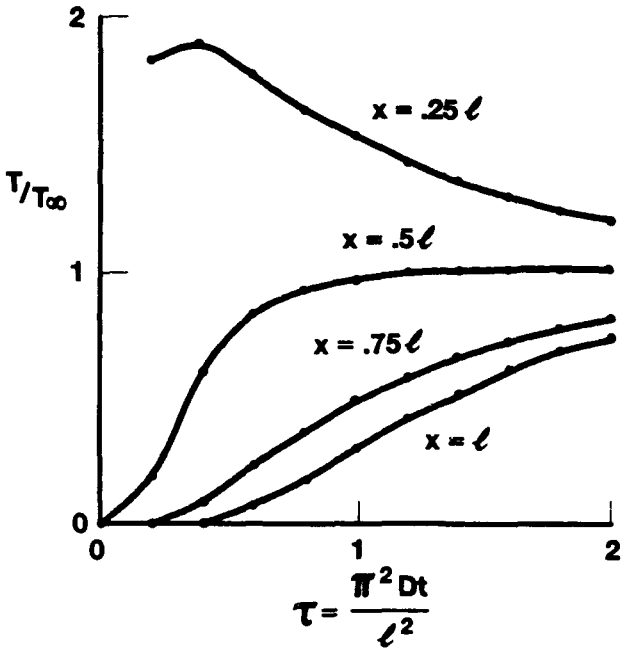


Figure 15. Plot of Heat Transfer into Substrate

A mathematical model of a short heat pulse on one side of a thin substrate is given by Danielson and Sidles [30]. This includes the optical absorption depth of the substrate. In the present case, the absorption in the quartz surrounding the electrode is small. For the metallic electrode, however, the absorption depth is typically less than 100 Å. This is much less than the thickness of electroding of a crystal so that the heat from the light will be generated in the electrode and no light will penetrate into the quartz. This simplifies the heat equation for a short laser pulse on one side (x = 0) of a substrate of thickness l to

$$T(x,t)/T_\infty = 1+2\sum_{n=1}^{\infty} \cos(n\pi x/l)\cdot\exp(-n^2\tau) \qquad 1$$

where $T_\infty = Q/\rho c l$ is the uniform temperature after a long time, Q is the absorbed fluence, ρ is the density of the substrate and c is its specific heat. $\tau=\pi^2 Dt/l^2$ is a dimensionless time with D the thermal diffusivity. This is presented graphically in figure 15.

For a quartz crystal resonator vibrating in thickness-shear, its frequency is given by

$$f = C/l \qquad 2$$

where C is the appropriate frequency constant for the cut of quartz. Thus the dimensionless time becomes

$$\tau = \pi^2 Dtf^2/C^2 \qquad 3$$

or

$$t = (\tau C^2/\pi^2 D)\cdot 1/f^2 \qquad 4$$

Thus the actual time is proportional to this dimensionless time and $1/f^2$. For AT quartz, D = 4.3 mm^2/s in the thickness direction and C = 1660 kHz.mm. Taking t to be the rise time of the phase pulse for a crystal at frequency f, we get τ = 1.3. From figure 15 this corresponds to the time for the back face of the crystal to have reached half its final temperature, thus confirming that heat transport is the mechanism.

2.5.2 Amplitude Domain

It is concluded that the time response of the phase pulse is governed by the thermal diffusion time of the laser-induced heat pulse. It is now proposed that the mechanism responsible for the frequency modulation is the force frequency effect, with the force supplied by the thermal

expansion of the quartz. This is given by [31,32]

$$df/f = Kt_q \quad (5)$$

where K is a constant called the stress coefficient for a given crystallographic orientation and t_q is the lateral stress across the whole resonator. For AT and BT-cuts, K is $-2.75*10^{-11}$ and $2.65*10^{-11}$ m^2/N, respectively. Thus these two cuts are expected to show opposite phase shifts. This change of sign has been demonstrated.

We can test this hypothesis further by comparing the measured frequency shift for a crystal with theory. The maximum frequency deviation occurs once the heat pulse has equilibrated through the thickness of the crystal but before lateral diffusion becomes significant. At this point the temperature increase of the quartz is T_∞. The resulting thermally induced lateral stress will be

$$t_q = \alpha C T_\infty \quad (6)$$

where α is the lateral expansion coefficient and C is the lateral elastic constant. This gives

$$df/f = K\alpha CQ/\rho cl \quad (7)$$

For a 64 MHz crystal with aluminium electrodes irradiated by 3.4 mJ, and assuming an absorption at 193 nm to be about 0.3, this should give a frequency shift of 6.6 ppm. To measure this, it is necessary to calculate the mixer conversion factor and the Q of the crystal. This can conveniently be achieved by applying a known frequency modulation and measuring the output from the mixer. This gave a measured frequency shift of 10.4 ppm, which,in view of the assumptions made, is remarkably close to the theoretical value.

2.5.3 Vibrational Distribution

It is important to remember that equation (7) holds only for a uniform stress, that is where the laser beam covers the whole resonator area. To use this mechanism to plot the vibrational distribution, it is necessary to show that the phase and hence frequency deviation resulting from a laser pulse at a particular point are proportional to the thickness-shear amplitude at that point. A similar problem is that of proving that the mass sensitivity of such a resonator is proportional to the amplitude of vibration. This is a proof that is often evaded in books on microbalances [33]. A solution to both problems lies in the variational method used in quantum mechanics [34] but first applied by Lord Rayleigh in 1873 [35].

Consider that the resonator is composed of n unit volumes k = 1...n. Rayleigh's Principle states that the frequency of resonance of a freely vibrating system is found by equating the time-averaged potential energy V of the system with the time-averaged kinetic energy T:

$$\langle V \rangle = \langle T \rangle \quad (8)$$

The potential energy for the k^{th} unit volume is given by:

$$V_k = {}^1/_2\, C_k . S_k^2 \quad (9)$$

where C_k is the appropriate elastic modulus and S_k the strain.
But $S_k = U_k \sin \omega t$ for a unit volume;

$$\text{so } V = \Sigma_k V_k = {}^1/_2 \Sigma_k U_k^2 . C_k . \sin^2 \omega t \quad (10)$$

$$\text{and } \langle V \rangle = {}^1/_4 \Sigma_k U_k^2 . C_k \quad (11)$$

$$\text{Similarly } T_k = {}^1/_2 \rho (\dot{u}_k)^2 \quad (12)$$

$$\text{where } u_k = U_k . \sin \omega t \quad (13)$$

$$\text{and } T = {}^1/_2 \rho\omega^2 \Sigma_k (U_k)^2 . \cos^2 \omega t \quad (14)$$

$$\text{and } \langle T \rangle = {}^1/_4 \rho\omega^2 \Sigma_k (U_k)^2 \quad (15)$$

Equating the two energies and rearranging:

$$\omega^2 = \frac{\Sigma_k (U_k)^2 . C_k}{\rho \Sigma_k (U_k)^2} \quad (16)$$

As described above,the effect of the laser beam is to alter the elastic constant of the quartz within the area of the beam. Suppose therefore that one element (j) has C_j changed to $C_j + \partial C$

$$\omega'^2 = \frac{\Sigma_k C_k . (U_k)^2 + \partial C_j (U_j)^2}{\rho \Sigma_k (U_k)^2} \quad (17)$$

$$\text{and } \frac{\partial\omega}{\omega} = \frac{\partial C_j (U_j)^2}{2\omega^2 \rho \Sigma_k (U_k)^2} \quad (18)$$

In integral form this becomes:

$$\frac{\partial\omega}{\omega} = \frac{1}{2\omega^2\rho} \frac{\int_v C'_{ijkl}\, U_{l,k}\, U_{j,i}\, dv}{\int_v (U_i)^2\, dv} \quad (19)$$

where C'_{ijkl} is the increment of the effective elastic coefficient under stress.

It is clear that for a change in elastic constant at a particular location, the frequency change is proportional to the square of the vibrational distribution at that location. In other words the frequency shift is proportional to the intensity of the standing wave at the point of laser irradiation.

A similar line of reasoning, but invoking kinetic energy, can be used to explain the mass sensitivity of a thickness-shear resonator.

2.5.4 Effect of Anisotropy

The force-frequency effect for quartz is caused by elastic nonlinearities (third-order elastic constants). The direction of the force is important because of the anisotropy of quartz. The coefficient described above is the stress coefficient. The force-frequency coefficient, however, is defined by:

$$K_f(\Psi) = \frac{\partial f}{f_o}\,\frac{2hD}{F}\,\frac{1}{N} \qquad 20$$

where $\partial f/f$ is the fractional change in resonant frequency due to the application of a pair of in-plane forces F at the edge at an azimuthal angle of Ψ to the rotated X-axis. N is the frequency constant and 2h and D are the thickness and diameter of the plate resonator, respectively. For the BT-cut, this coefficient varies slightly with angle; however, for many of the other cuts [36] including the AT-cut (figure 16), the coefficient goes through zero and hence changes sign in some directions.

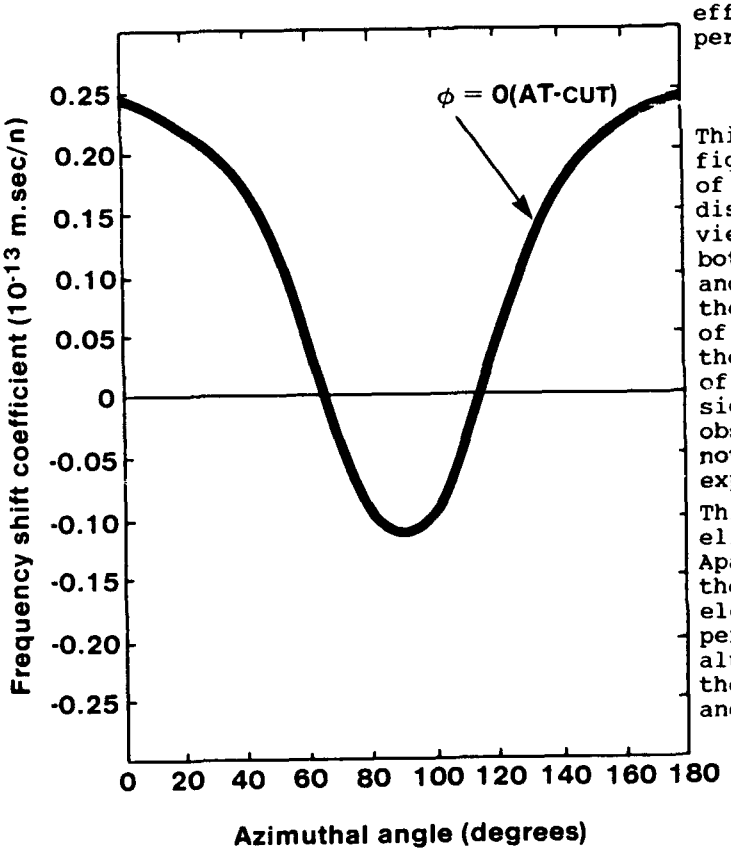


Figure 16. Force Frequency Effect for AT Quartz

This anisotropy revealed itself when plotting AT-cut resonators. It was observed that in some areas of the resonator, always near the edge and often in two places nearly opposite one another, the phase pulse changed sign. This happened when the maximum of the slope of the amplitude was about ±90° to the X-axis. Since the laser spot in the present investigation had a size only about an eighth of the resonator diameter, there could be an appreciable difference in amplitude across the spot on the slope of the amplitude distribution. The change in the elastic coefficient shown in equation (18) is proportional to the force-frequency coefficient in the direction of the force. The effect of the laser spot is to cause the quartz under that spot to expand and, because it is constrained by the surrounding quartz, it goes into compression. From equation (7), this compressive force is proportional to the product of the in-plane expansion coefficient with the elastic coefficient. For AT-cut crystals this is about 1 $N/m^2/°C$ irrespective of azimuth angle Ψ. It follows, therefore, that the resulting frequency change is proportional to the intensity of the vibration times the angular value of the force frequency effect $K'(\Psi)$, integrated around the perimeter of the laser spot:

$$df/f \;\propto\; \int U^2(x_o,y_o,r,\Psi)\cdot K'(\Psi)d\Psi \qquad 21$$

This is illustrated geometrically in figure 17 which shows a side and plan view of a resonator with the vibrational distribution and a laser spot. An expanded view of the spot shows polar diagrams of both the force-frequency coefficient $K'(\Psi)$ and the vibrational intensity. Finally, the resultant contour map across the area of the resonator is illustrated, showing the distortion of the map towards the edge of the vibrating area with a reversing of sign in the direction of the Z' axis as observed. This reversing of sign was noticed by Oura et al. [37] without any explanation.

This anisotropy can be diminished but not eliminated by reducing the beam diameter. Apart from signal-to-noise considerations, the lateral diffusion through the electrode will extend the area of quartz perturbed. The thermal diffusivity of aluminium is 81 mm^2/s and the ratio of the thermal diffusion lengths in the electrode and the quartz is given by

$$l_{Al}/l_Q = \sqrt{D_{Al}/D_Q} = 4.35 \qquad 22$$

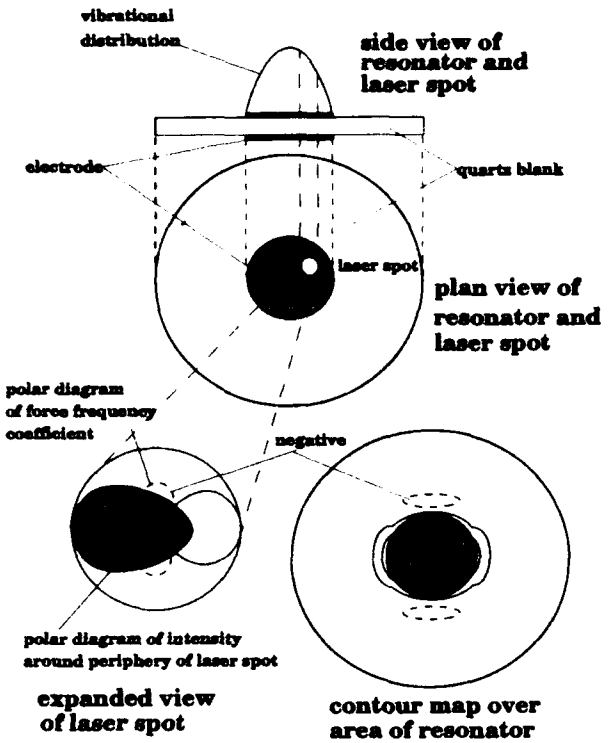


Figure 17. Effect of Anisotropy on Contour Map

Therefore, if the laser spot diameter is similar to the quartz thickness the lateral diffusion will dominate. The result of the anisotropy in the force-frequency effects for some cuts is a differing sensitivity along the two axes. For instance a circular vibrational distribution will appear elliptical. In principle this can be taken into account when forming the contours if the orientation is known. Alternatively, this property can be used to measure the orientation of the device.

2.6 Conclusions

The method described here is capable of plotting the vibrational distribution of fundamental thickness-shear resonators, their overtones and their inharmonics. It uses a pulsed mode laser to probe the surface of the resonator in a raster pattern, thus building up a contour map of the distribution. The parameter measured is the magnitude of the induced phase modulation as the crystal is momentarily knocked off frequency. It is proposed that this perturbation is caused by thermally-induced compressive stress through the thickness of the quartz under the laser beam. Because of the elastic nonlinearities in quartz this gives rise to a frequency shift.

As mentioned above, no method is ideal for plotting the vibrational distribution of a crystal. The limitation of the present method is that it relies on the azimuth-averaged value of the force-frequency coefficient being non-zero. This is clearly not the case with SC-cut resonators. However, this disadvantage can be turned into an advantage by making use of the method to investigate the force-frequency effect with both crystal orientation and vibrational distribution [38]. Another use envisaged for this method is the investigation of activity dips when plate modes couple into the thickness-shear resonance. Wilson [39] showed that X-ray topography imaged the resultant changes to the mode shape. The current method should be as sensitive.

The basic principle of the method is applicable to any device whose operating parameter has a temperature coefficient. For example, the distribution of current density in a circuit or IC could be plotted by monitoring the modulation in the current through it at the modulation frequency. The sensitivity of the method is obtained by the use of a lock-in amplifier or wave analyser to measure this modulation in noise.

4 ACKNOWLEDGEMENT

The author wishes to thank Dr. A.F.B. Wood of STC Quartz Crystal Division and R.A. Heinecke of STC Technology Ltd for valuable discussions. This work has been carried out with the support of the Procurement Executive, Ministry of Defence (Royal Signals and Radar Establishment), and sponsored by Admiralty Research Establishment (Portsdown) and also by STC Quartz Crystal Division.

5 REFERENCES

[1] E.F.F.Chladni, "Entdeckungen uber die Theorie des Klanges." Leopzig 1787.

[2] F.Savart, Ann. Chim. Phys. [2] 14,p113

[3] A.F.B.Wood, Proc. Phys. Soc.,London 77,1961,p 1076.

[4] K.S.van Dyke, Proc. 11th AFCS.

[5]I.Koga and H.Fukuyo J.Inst.Electr. Commun.Eng.Jpn. 36, 1953,p59.

[6] L.Dworsky and G.Kennedy, Proc. 35th AFCS,p237.

[7] R.J.Gerdes and C.E.Wagner, Proc. 25th AFCS,p118.

[8] H.Bahadur, A.Hepworth, V.K.Lall, R.Parshad, Proc. 32nd AFCS,p207.

[9] H.Bahadur and R.Parshad, Proc.39th AFCS,p544.

[10] T.F.Parkinson and M.W.Moyer, Nature(London) 211,1966,p400.

[11] A.G.Klein,P.Prager,H.Wagenfeld, P.J.Ellis,T.M.Sabine, Appl.Phys.Lett.10 1967,p293.

[12] P.Vigoureux and C.F.Booth, "Quartz Crystals and Their Applications",Chapter 15,H.M.Stationery Office,London.

[13] G.Sauerbrey, Proc. 21st AFCS,p63.

[14]S.Hertl,E.Benes,L.Wimmer,M.Schmid, Proc. 39th AFCS,p535.

[15]H.Osterberg,J.Opt.Soc.Am. 23,1933,p30.

[16] S.Tolansky "Microstructure of Surfaces Using Interferometry", Arnold,London,p30.

[17] K.Iijima,Y.Tsuzuki,Y.Hirose,M.Akiyama, Proc.30th AFCS,p65.

[18] T.Adachi,M.Okazaki,Y.Tsuzuki, Proc. 37th AFCS,p187.

[19] Y.Hirose and Y.Tsuzuki, Proc.43rd AFCS,p360.

[20]Y.Tsuzuki,Y.Hirose,K.Iijima, Proc.25th AFCS,p113.

[21]G.W.Fox and P.H.Carr, Phys.Rev. 37,1931,p1622.

[22] W.J.Spencer, Ch.3, Vol.V, Physical Acoustics, Academic Press 1968.

[23] R.A.Sykes,W.L.Smith,W.J.Spencer, Proc. 17th AFCS,p4.

[24] A.Zarka,B.Capelle,J.Detaint, J.Schwartzel, Proc. 41st AFCS,p236.

[25] A.Zarka,B.Capelle,J.Detaint, J.Schwartzel,2nd E.F.T.F. 1988,p651.

[26] A.Zarka,B.Capelle,J.Detaint, J.Schwartzel,Proc. 42nd AFCS.p85.

[27] G.E.Wagner and R.A.Young, Proc.21st AFCS,p72.

[28] H.Bahadur and R.Parshad,Ch.2,Vol.XVI, Physical Acoustics, Academic Press 1982.

[29] R.J.Williamson, 3rd EFTF 1989, p22.

[30] G.C.Danielson and P.H.Sidles, "Thermal diffusivity and other non-steady state methods", Thermal Conductivity, Vol.2, Ed. R.P.Tye Acedemic Press Inc. (London) Ltd.

[31] E.P.EerNisse, "Quartz resonator frequency shifts arising from electrode stress", Proc. 29th AFCS, pp 1-4.

[32] E.P.EerNisse, "Stress effects in quartz crystal microbalances", Methods and Phenomena, Vol.7, Ed. C.Lu and A.W.Czanderna, Elsvier Scientific Publishing.

[33] H.K.Pulker and J.P.Decosterd, "Applications of quartz crystal microbalances for thin film deposition process control",Methods and Phenomena, Vol.7, Ed. C.Lu and A.W.Czanderna, Elsvier Scientific Publishing.

[34] L.I.Schiff, "Quantum Mechanics", p171, McGraw-Hill Book Company.

[35] Lord Rayleigh, "Theory of Sound", 2nd rev. ed., vol.1, Sec.88, reprinted by Dover, New York.

[36] A.Ballato, E.P.EerNisse and T.Lukaszek, "The force frequency effect in doubly rotated quartz resonators", Proc 31st AFCS, pp 8-16.

[37] N.Oura, N.Kuramochi, J.Nakamura, T.Ogawa, Proc 36th AFCS, pp133-139

[38] P.E.Morley and R.J.Williamson, this proceedings.

[39] C.J.Wilson, Proc 27th AFCS, p35.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ADJUSTMENT OF RESONATOR G-SENSITIVITY BY CIRCUIT MEANS

R.C. SMYTHE & W.H. HORTON

Piezo Technology Inc.
Orlando, Florida

Abstract

Novel, multi-electroded resonators are described which allow flexible control of acceleration sensitivity using simple external circuit techniques. By varying the relative excitations of the electrodes, the mode shape, and hence the acceleration sensitivity of the resonator, can be altered. These "Gamma tuning" adjustments are external to the resonator; hence, they do not require the introduction of new resonator manufacturing processes or the modification of existing ones.

Experimental efforts to date have shown that individual components of the G-sensitivity vector can be nulled, and the magnitude of the vector can be minimized. G-sensitivities as low as 1.5×10^{-10}/G and reductions by a factor of 5 have been obtained. A qualitative theory is presented.

Introduction

Currently, there is an urgent need for crystal oscillators and crystal resonators having low acceleration sensitivity [1]. For accelerations which are not too large, the frequency change is the scalar (dot) product of two spatial vector quantities, the acceleration sensitivity of the resonator, Γ, and the acceleration [1]. The acceleration sensitivity depends upon two factors -- the deformation produced by the acceleration and the mode shape. While the mathematical analysis is extremely complex, the basic idea is simple [2]. At each point in the resonator, acceleration-induced deformation alters the effective elastic stiffness of the resonator, thereby incrementally affecting the resonance frequency of each mode of vibration by an amount that depends upon the mode amplitude and sign and upon the amplitude and sign of the deformation at that point.[1] The total effect on the resonance frequency of a particular mode is just the algebraic sum of the incremental effects taken over the entire volume of the resonator.[2] For regions of the resonator where the amplitude of vibration is small, the incremental effect is small. Similarly, where the deformation is relatively small the effect will be relatively small. An extremely important aspect of the summation is that a high degree of cancellation takes place; that is, the sum of the positive increments is very nearly equal to the sum of the negative increments [2,3,4]. Consequently, quite small changes in either the mode shape or the deformation result in large changes in Γ.

In a conventional thickness-mode quartz crystal unit, the resonator plate is supported at two or more points on its periphery by metal ribbons or clips, which in turn are fastened to a header or base. Under acceleration, the body forces on the blank are balanced by reaction forces of the support structure. Thus the support structure determines the acceleration-induced deformation of the blank. The mode shape of the resonator is determined by an energy-trapping mechanism. For a plano-plano resonator, the primary controls are the lateral dimensions of the electrodes, including the tabs, the electrode mass loading, and the piezoelectric loading. However, the mode shape, and hence the acceleration sensitivity, will also unavoidably be affected by point-to-point variations in the thickness of the blank (i.e., non-parallelism) or even of the electrode, as well as by material defects such as inclusions and etch channels. For contoured resonators, mode shape is primarily controlled by contour, but will also be affected to some degree by the electrodes and by material defects.

[1] Although, in the interest of simplicity, we speak of the mode amplitude as if it were a single quantity, it should be recognized that the mode of vibration employed may have more than one component, each of which may contribute to the total acceleration sensitivity. Similarly, the acceleration-induced deformation will, in general, have more than one component.

[2] It follows that the acceleration sensitivities of different modes of the same resonator are, in general, different from one another.

CH2818-3/90/0000-437 $1.00 © 1990 IEEE

Because of fabrication limitations, variations from resonator to resonator in support geometry, in electrode dimensions, and in blank contour are unavoidable. Because the resonator acceleration sensitivity is the difference of two nearly equal quantities, small changes in the resonator, representing normal manufacturing tolerances, can cause large changes in resonator acceleration sensitivity. This is borne out by experience [1].

Consequently, it is desirable to have a means of adjusting, or trimming, the acceleration sensitivity in order to minimize its magnitude or the magnitude of one or more of its vector components. In principle, this may be accomplished by altering either the mode shape or the acceleration-induced deformation. The latter might be accomplished, say, by modifying the mounting structure. One method by which the mode shape may be altered is to add or remove mass from selected regions of the blank [5,6,7].

It is the purpose of this paper to describe resonators for which the mode shape at resonance may be modified by circuit techniques in order to minimize or otherwise control the acceleration sensitivity. An advantage of this approach over those which depend upon the addition or removal of mass or upon the alteration or adjustment of the mounting clips, or upon other alterations to the resonator structure is that the circuit adjustments to be described are external to the resonator and may be carried out after its manufacture is complete. Techniques for trimming acceleration sensitivity by modifying the resonator, on the other hand, must be carried out as part of the resonator manufacturing process, while the resonator is in a partially completed state. Thus they may be cumbersome and costly. In addition, the techniques to be described are reversible, further facilitating the adjustment process.

The devices to be described are novel, multi-electroded resonators. With these, simple circuit adjustment techniques, which we call "Gamma tuning," are used for the reduction and control of gamma.

Principle

The principle behind the multi-electrode approach can be understood from figures 1 and 2. Figure 1 shows a thickness-mode resonator, such as an AT-cut resonator. Axes x_1 and x_3 are a pair of coordinate axes lying in the plane of the plate. They may correspond to rotated crystallographic axes or may be otherwise selected, as desired. For the device shown, one electrode has been divided in a direction normal to x_1 so that the device can now be considered as two resonators, R1 and

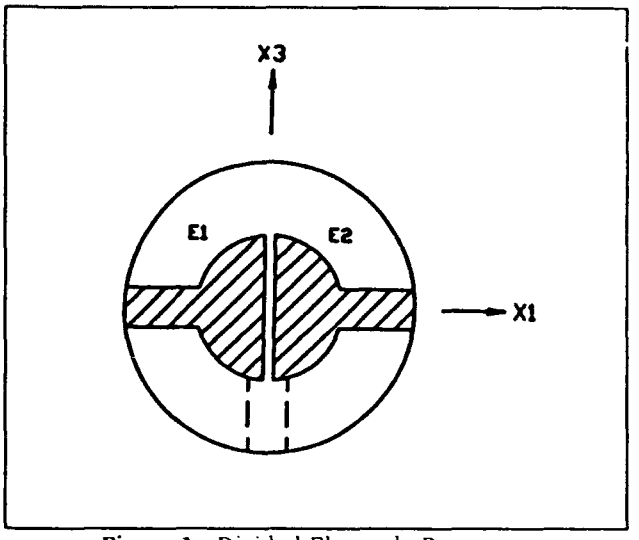


Figure 1. Divided-Electrode Resonator

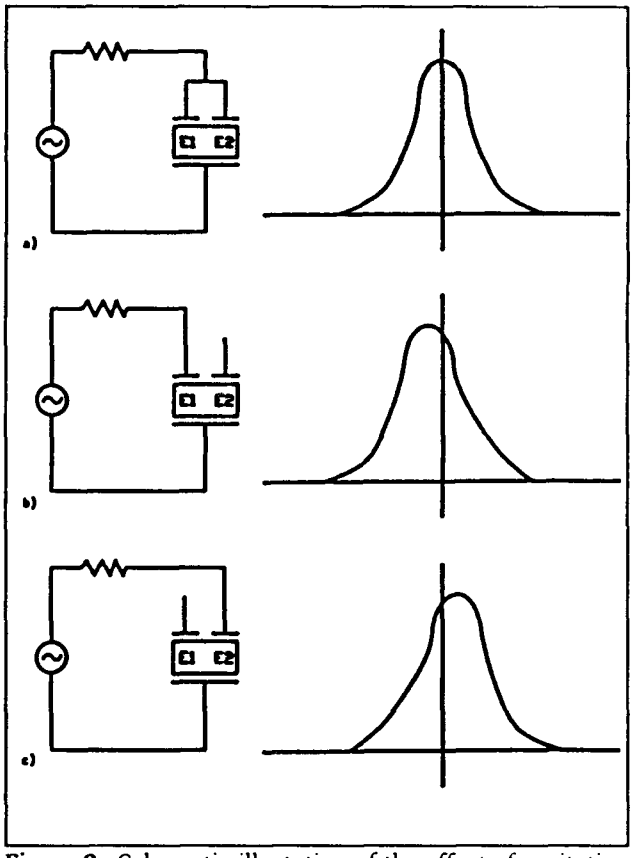


Figure 2. Schematic illustation of the effect of excitation on mode shape for a symmetric resonator structure, such as Fig. 1: a)Equal excitation of E1 & E2; b) E1 excited; c) E2 excited.

R2, with associated electrodes E1 and E2 and a common counter-electrode. We note that R1 and R2 are, in general, acoustically coupled. For purposes of exposition, consider the device to be perfectly symmetrical. When E1 and E2 are connected together and driven electrically, figure 2a, only the modes symmetric about the x_3 axis are excited. The amplitude along the x_1 axis of one such mode is sketched in figure 2a. When only E1 is electrically excited, figure 2b, the mode shifts to the left, toward R1; however, R2 is also excited to some degree since it is acoustically coupled to R1. At the same time, the mode shape changes and is no longer symmetrical, and the resonance frequency changes. Similarly, when only E2 is electrically excited, figure 2c, the mode shifts to the right, toward R2. It follows that by controlling the relative excitation of E1 and E2 in magnitude and phase, the position of the mode along x_1 can be controlled. Since the acceleration sensitivity, Γ, depends strongly upon the mode shape and location, varying the relative excitation of E1 and E2 will cause the acceleration sensitivity to vary.

The effect of exciting E1 differently from E2 can also be explained from another viewpoint. In figure 2a, the excitation is symmetric. Since the device is also symmetric, only symmetric modes are excited. In figures 2b and 2c, the excitation is no longer symmetric, but can be considered as the sum of a symmetric excitation and an antisymmetric excitation. Hence, both symmetric and antisymmetric modes are excited, although unequally, so that for these connections the device will have resonances which are neither symmetric nor antisymmetric.

Figure 3 shows, in general form, some simple circuit techniques by which the relative excitation of E1 and E2 may be controlled. X1 and X2 are reactances;

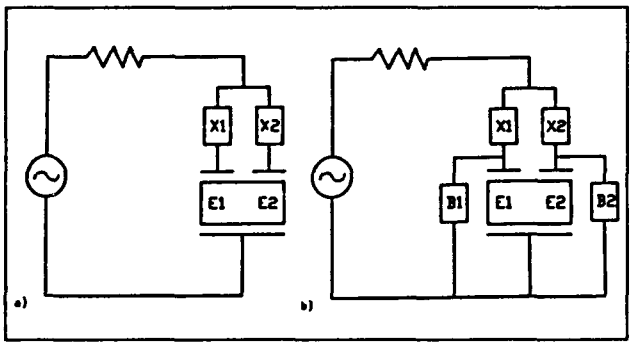


Figure 3. Some simple circuit arrangements for varying the relative excitation of E1 & E2.

they may be simply capacitors or inductors, or they may be more complex L-C networks. Using X1 and X2, the relative amplitudes and phases of the currents into E1 and E2 can be varied, thereby changing the position of the mode in the x_1 direction. In figure 3b, susceptances B1 and B2 have been added. These might be used, for example, to extend the range of control obtainable with X1 and X2. Other circuit techniques can also be applied, for example, E1 and E2 might be driven by separate amplifiers, or other active networks, having variable gain and phase shift. In any case, the circuit composed of the resonator and its additional circuitry has a resonance with which is associated an acceleration sensitivity, which we can now control by suitable circuit adjustments. This is the Gamma tuning concept. (Throughout, we speak of acceleration sensitivity at a resonance frequency. Clearly, the concepts and techniques discussed apply equally in the case of operation with a load capacitance.)

The principle may be extended to two dimensions, figure 4. In figure 4a, four electrodes E1, ..., E4 have been formed on one face of the blank. By controlling the relative excitation of the four, the position of the mode with respect to both x_1 and x_3 can be varied. This can be done using the circuit techniques of figure 3, as illustrated, for example in figure 5. Still other electrode configurations are possible, such as figure 4b, in which three electrodes have been formed on one face.

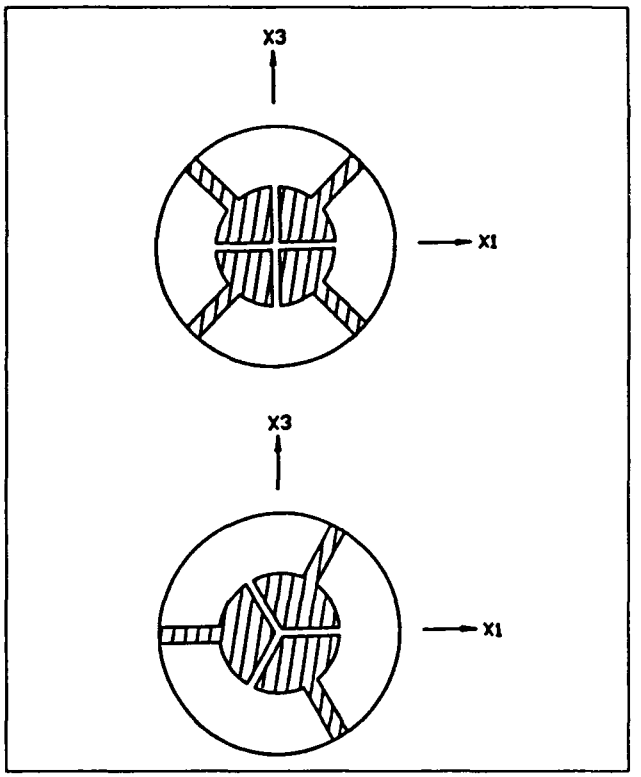


Figure 4. Two electrode configurations which allow two-dimensional control of the mode shape.

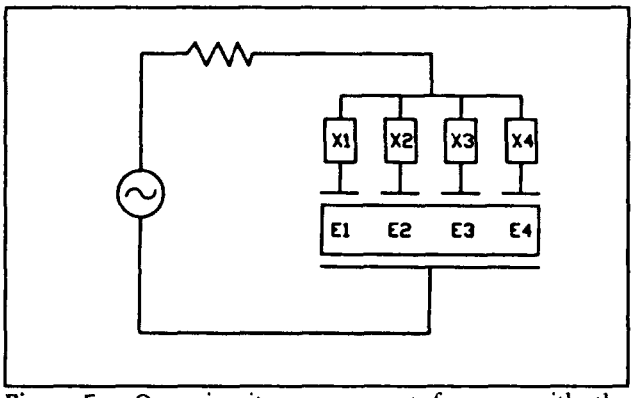


Figure 5. One circuit arrangement for use with the electrode configuration of Fig. 4.

Yet another configuration is shown in figure 6a, in which the electrodes of both faces of the blank are divided. This configuration provides a simple means of independently controlling the x_1 and x_3 position of the mode. In figure 6b reactances X1 and X2 control the x_1 position, while reactances X3 and X4 control the x_3 position.

While the foregoing discussion assumed a symmetrical device, the techniques described do not require that the device be symmetrical. To the contrary, it is a purpose of these techniques to correct for the unavoidable deviations from symmetry found in actual devices. For example, the locations of the mounting ribbons are often not accurately reproduced from unit to unit. While improved manufacturing methods may reduce such variations, it is nevertheless desirable to have a means for trimming or adjusting the acceleration sensitivity. Current research is expected to yield resonator designs having lower theoretical values of acceleration sensitivity than conventional configurations. Since these will also be subject to manufacturing variations, it seems likely that an adjustment method will still be useful.

Experimental Results

Preliminary measurements on crystal units similar to figure 1 confirmed the validity of the technique. Units were then fabricated having an electrode configuration similar to figure 6a. Crystal units to be tested are mounted in a test fixture having adjustable reactances corresponding to X1 - X4, figure 6b. Initially, these were multi-turn air trimmer capacitors; later, varactor diodes were used. The varactor diodes allow convenient, remote adjustment of the capacitance in series with each electrode. Forward biasing provides a low impedance condition, while reverse biasing provides a variable capacitance which can be determined from the bias voltage. The reference state was taken with all diodes forward biased, so that E1 and E2 are at the same r-f potential, as are E3 and E4.

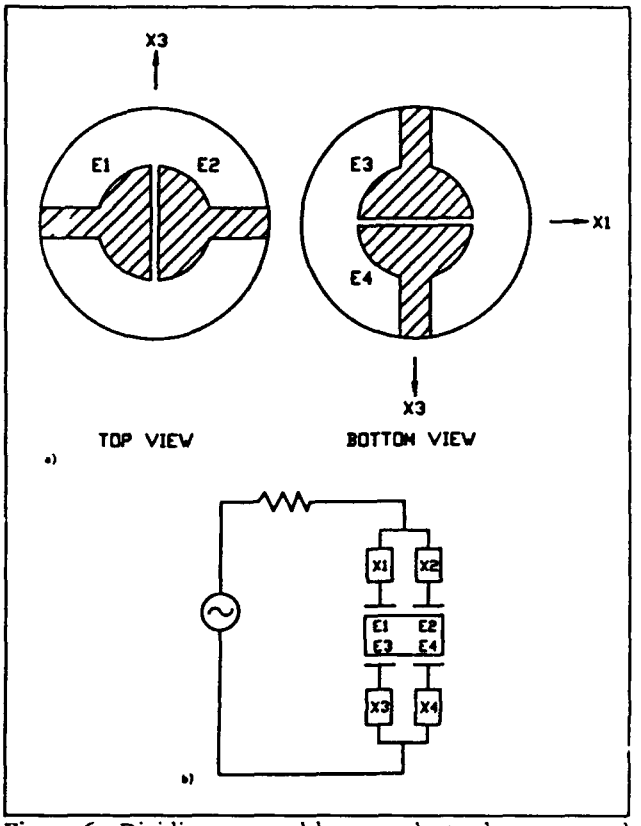


Figure 6. Dividing top and bottom electrodes to control excitation and mode shape in two dimensions: a) electrodes; b) series reactances provide one means for adjusting relative excitation.

The crystal test fixture was mounted on a standard vibration exciter ("shaker"). A single coaxial cable connects the fixture to a standard PTI test oscillator. In this oscillator the crystal is in one arm of a four-port hybrid network which is located near the shaker. A balancing capacitor is in the opposing arm. This arrangement provides a simple means of separating the crystal fixture from the oscillator circuitry by any desired distance. A separate, shielded, multiconductor cable carries the varactor bias voltages, which are provided by a remote control box. Acceleration sensitivity is determined by measuring the vibration-induced FM sidebands [1]. A block diagram of the measurement system is shown in figure 7. The sidebands can be measured directly, using an RF spectrum analyzer,

such as the H-P 3585A. Better sensitivity was obtained, however, using a phase noise bridge, and most measurements were made in this manner. Comparison of the two methods showed good agreement. Most measurements were made with 5 G's peak acceleration at 90 Hz.

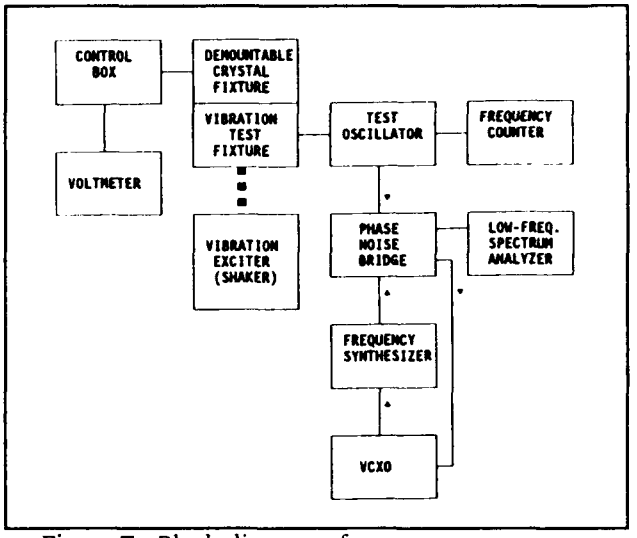


Figure 7. Block diagram of measurement system.

The mechanical configuration of the fixture, figure 8, consists of an aluminum cube mounted on a base plate which can be bolted to the shake table. A coaxial line passes through the base and terminates in pin sockets on the top face and one side face of the cube. A crystal carrier, with a socket and retaining bracket for the crystal unit, can be bolted to either face, and can be rotated in 90° steps about the normal to the face, allowing the crystal unit to be vibrated in three orthogonal directions. For two of the three directions, the carrier can be inverted, allowing measurements to determine errors due to cable effects, if any. The carrier also contains the variable reactance circuitry.

Table 1 presents selected results obtained on 30 MHz, 3rd overtone, AT-cut resonators. The electrode arrangement is similar to figure 6a, except that a square array, rather then a circular, array was used. The electrode array measures 0.140 inch (3.56 mm) on a side, with a gap of 0.020 inch (0.51 mm). Except for the electrodes, construction of the crystal units was conventional. Blanks having a diameter of 0.310 inch (7.9 mm) were mounted at four points with 50 x 4 mil nickel ribbons in coldweld TO-8 enclosures. Mounts were aligned with the crystallographic x- and z'-axes of the blanks.

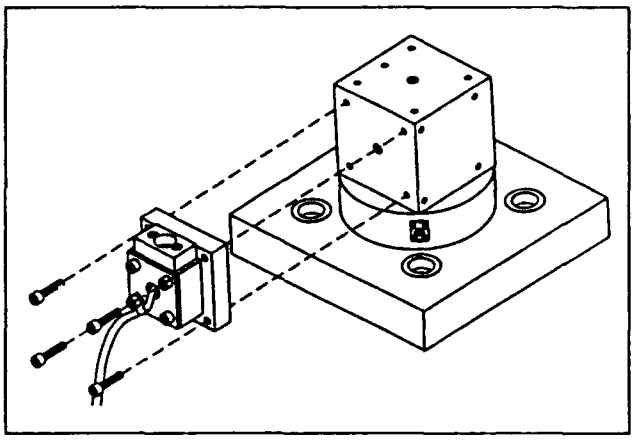


Figure 8. Vibration test fixture. The fixture shown is for a TO-8 size crystal unit.

For each resonator, capacitive reactances X1,..., X4 were adjusted iteratively to minimize $|\Gamma|$, obtaining reduction factors of 2.3 to 5.3. The mean value of $|\Gamma|$ was reduced from 7.99 to 2.46 x 10-10. The change in frequency was between 0.41 and 1.30 kHz. Loaded Q in the test oscillator was typically 50 x 10^3, and improved somewhat with Gamma tuning.

For some units, $|\Gamma|$ could not be reduced significantly. For example, No. 1-4, with $|\Gamma|$ = 5.43 x 10^{-10}, and No. 4-6 with $|\Gamma|$ = 2.58 x 10^{-10}, could not be significantly improved with the simple capacitive tuning circuits used in these experiments.

For most units, x-, y-, and z-components of Γ could be nulled, although not simultaneously.

Discussion

Some of the ways in which these techniques can be employed include:

1. Minimization of one component of the g-sensitivity vector.

2. Minimization of the magnitude of the g-sensitivity vector.

3. Adjustment of the orientation and/or magnitude of the g-sensitivity vector to facilitate the use of active or passive compensation.

4. Separate adjustment of the G-sensitivity vector of two modes, as might be desirable in an MCXO application, for example.

Table 1. Acceleration Sensitivity Before & After Gamma Tuning

Unit Number	Frequency (kHz)	Acceleration Sensitivity x 10^{10} (/G)				
		In-Plane			Normal	Total
		x-axis	z-axis	Total In-Plane	y-axis	$\|\Gamma\|$
1-3 Init	30,309.41	0.74	13.79	13.81	0.74	13.83
Final	30,311.71	0.26	1.17	1.20	2.32	2.61
2-4 Init	30,296.39	0.75	9.44	9.47	0.48	9.48
Final	30,297.03	0.17	0.95	0.97	3.27	3.41
3-1 Init	30,011.79	1.70	4.97	5.25	2.10	5.66
Final	30,012.20	2.24	0.22	2.26	0.95	2.45
4-3 Init	30,277.02	1.16	5.89	6.00	0.62	6.04
Final	30,277.81	0.80	0.51	0.95	2.19	2.39
4-8 Init	30,273.72	0.87	4.79	4.87	0.91	4.95
Final	30,274.55	0.96	0.33	1.01	1.05	1.46

The methods which have been described alter the resonance frequency, as well as the acceleration sensitivity, of the crystal unit and its accompanying Gamma tuning circuitry. If the frequency change cannot be tolerated, there are possible remedies. First, the frequency specification may be taken as a constraint during acceleration tuning. Second, Gamma tuning may be carried out prior to final frequency adjustment of the resonator. After Gamma tuning, final frequency adjustment is performed using a method, such as dielectric overplating, designed to minimize mode shape changes.

It will be understood that the electrode and circuit configurations shown are not the only ones possible; for example, it may be desirable for many applications to include an adjustable load capacitance as a means of frequency adjustment. Finally, although it has not been pursued in the present work, it should be noted that Gamma tuning could also be used with analogous SAW resonator configurations, using in-line and transverse coupling.

Conclusions

A resonator whose G-sensitivity can be adjusted by simple external circuitry has been described. Limited experimental work indicates that the approach has considerable potential. This approach is not intended as a substitute for configurations having inherently low acceleration sensitivity; rather, its purpose is to allow the potential g-sensitivity of a given configuration to be realized.

Acknowledgment

The authors wish to thank Prof. H. F. Tiersten for many helpful discussions with one of us (RCS) concerning the underlying principles of resonator acceleration sensitivity, including the importance of the mode shape.

References

[1] R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: a Review," IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control, v. 35, no. 3, pp.297-305, May, 1988.

[2] H.F. Tiersten & D.V. Shick, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported along the Edges," Proc. IEEE Ultrasonics Symp., pp. 357 363, 1988.

[3] D.V. Shick & H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," Proc. 43rd Ann. Symp. on Freq. Control, pp. 405-412, 1989.

[4] Y.S. Zhou, D.V. Shick, & H.F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Simple Rectangular Supports," Proc. IEEE Ultrasonics Symp., pp. 393-400, 1989.

[5] Errol P. EerNisse, Roger W. Ward, & O. Lew Wood, "Acceleration-Induced Frequency Shifts in Quartz Resonators," Proc. 43rd Ann. Symp. on Freq. Control, pp. 388-395, 1989.

[6] Errol P. EerNisse & Roger W. Ward, "Crystal Resonator with Low Acceleration Sensitivity and Method of Manufacture Thereof," U.S. Pat. 4,837,475, June 6, 1989.

[7] Arthur Ballato, "Method of Making an Acceleration Hardened Resonator," U.S. Pat. 4,836,882, June 6, 1989.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ELECTRONIC DESENSITIZATION OF RESONATORS TO ACCELERATIONS

A. Ballato, J. Kosinski, T. Lukaszek, M. Mizan, & R. McGowan

US Army Electronics Technology & Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

ABSTRACT

A variety of approaches to compensate for acceleration effects in crystal resonators have been used over the years, however these schemes have met with mixed success. A good example is the intensive work on the mounting supports of BAW resonators which has occasionally produced a unit with a very small acceleration sensitivity (Γ), but with very poor reproducibility. Tests performed at ETDL as part of an in-house research program have now confirmed the value of a new method to reduce Γ electronically. The underlying principle of the new method centers on altering Γ by electronically altering the modal distribution of motion. This method is a substantially more direct and cost effective means of adjustment than the others used to date. This paper describes the compensation technique, and gives experimental results for resonator units.

INTRODUCTION

Resonances in structures are frequency regions where the structures exhibit large amplitude vibrations. These vibrations may be mechanical, electromagnetic, or various other types. Resonances are functions of frequency, the material supporting the vibration, the substrate orientation, the mode of vibration, and the device geometry or configuration. Resonances also depend on the supporting structure and on external influences such as temperature and forces. The sensitivity of the resonance frequency to external disturbances is determined by the same general parameters that determine the resonance frequency.

For quartz crystal resonators, it is possible to calculate accurately the actual three dimensional distribution of motion for both bulk (BAW) and surface (SAW) acoustic wave devices. When, for a given crystal substance, cut, mode of motion, crystal geometry, and support structure, the distribution of motion is known, then the resonance frequency, as well as the sensitivity of the resonance frequency to external accelerations, can be computed.

At present, the problem of acceleration sensitivity remains a major concern in the application of quartz crystal resonators. As long ago as the 1940's, "shock and vibration" was one of the biggest problems awaiting solution [1,2]. Through the years, a variety of approaches have been used: mechanical isolation; electrical compensations; determination of mounting support locations; different crystal cuts; different types of vibration; different geometrical shapes and aspect ratios; and the use of two or more crystal resonators connected electrically, but having reversed axial directions with respect to the acceleration vector. These schemes have met with mixed success.

Over the years, intensive work on the mounting supports of BAW resonators occasionally produced a unit (with extremely low yield and reproducibility) with a very small acceleration sensitivity. The acceleration sensitivity vector, Γ, was found on some units to be in the 10^{-10}/g and 10^{-11}/g range, where the norm is $2 * 10^{-9}$/g. The poor reproducibility was attributed to such things as a lack of symmetry of the mounting supports, and the impossibility of attaining such symmetry in the manufacturing process. The problem lies, in part, with the complicated mode spectrum due to the anisotropic nature of the crystal, and to the boundary conditions [3-8]. Figure 1 shows an example of some of the couplings that occur in quartz plates [4]. In Fig. 2 are seen topographs identifying the modal distributions of motion with various resonances in a quartz plate [5].

US GOVERNMENT WORK IS NOT PROTECTED
BY US COPYRIGHT

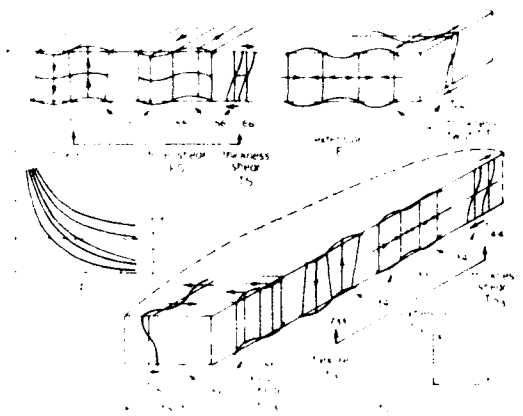


Figure 1. Some of the modes of motion in an anisotropic plate [4]; elastic stiffnesses coupling the motions are also shown.

ANALYTIC WORK

Calculations by P.C.Y. Lee of Princeton University [e.g., 9, 10] and H. F. Tiersten of Rensselaer Polytechnic Institute [e.g., 11-13], and their students, have demonstrated the importance of the mounting supports and resonator configuration on the acceleration sensitivity. In the calculations the induced stress/strain pattern interacts with the mode shape to produce the acceleration sensitivity. The calculations utilize a linearization of the equations for small fields superposed on a bias. Figure 3 shows common electronic devices that have nonlinear characteristics; establishment of an operating point is accomplished by means of a biasing network. About this operating point, small excursions produce linear responses; the linear slope, of course, is a function of the nonlinear device parameters. In like manner, the quasistatic bias of the crystal lattice due to the acceleration field establishes the operating point for the resonator. Small amplitude waves, due to the resonances, propagate with velocities that depend on the biasing state.

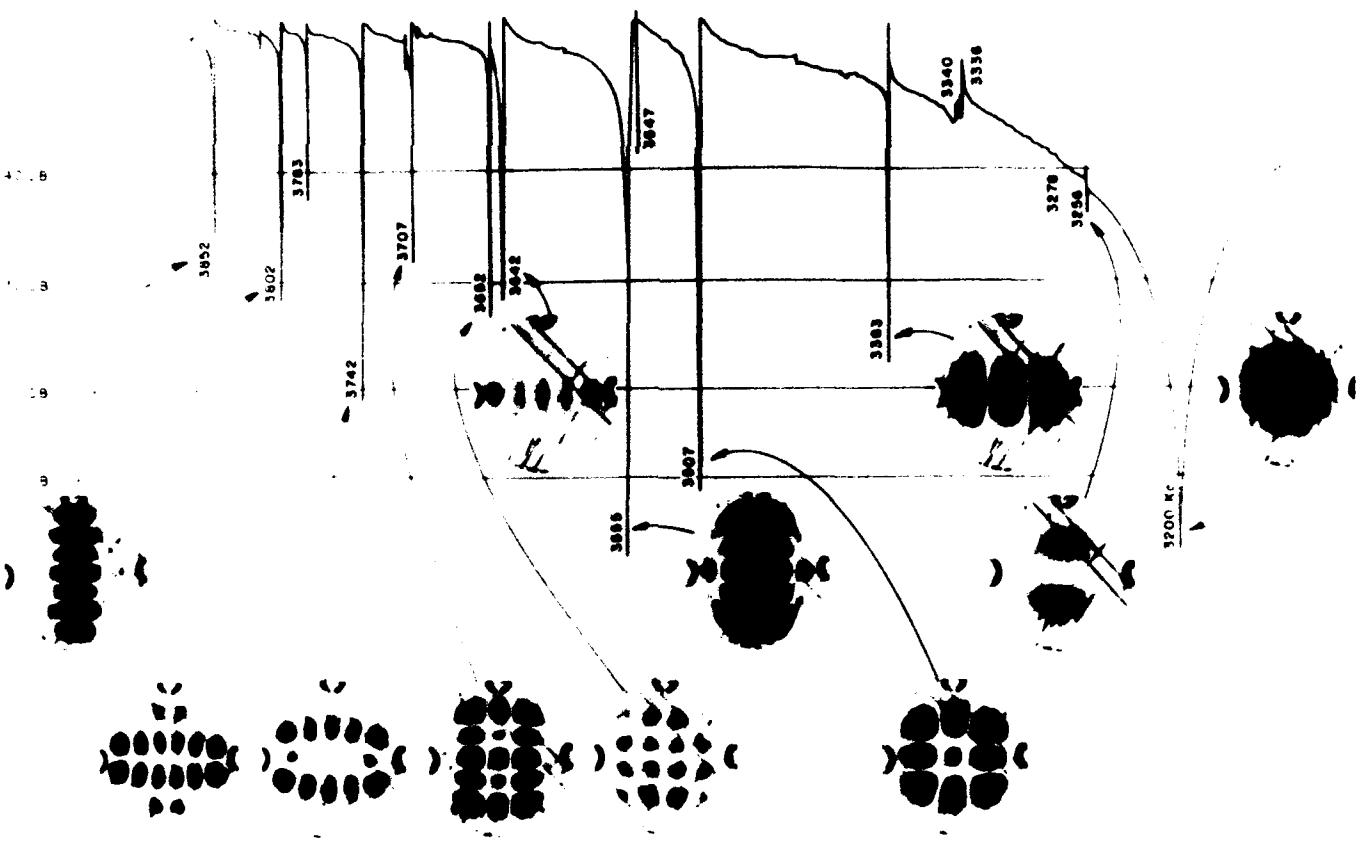


Figure 2. Mode spectrograph of a thickness shear resonator, showing resonance strength (crystal current, proportional to motional amplitude) versus driving frequency [5]. For a number of resonances, X-ray topographs are displayed; these show that each resonance has its own unique distribution of motion.

The frequency perturbation due to the acceleration can be written as

$$\delta f/f = \int (\text{mode shape}) * (\text{stress pattern})\, dV$$

where the integration is over the entire crystal resonator volume. See, e.g., [11].

ACCELERATION COMPENSATION

Since the acceleration sensitivity arises out of the interaction of the stress/strain pattern and spatial distribution of motion, Γ can be altered either by adjusting the stress/strain pattern, via the supports and/or the other factors mentioned, e.g., crystal, cut, geometry, etc., or by altering the mode shape of the vibrator. A bibliography to 1980 is contained in [14]; newer work is detailed in [15-17]; and [18], [19] describe modal effects.

The mode shape changes as a function of frequency, therefore, altering the operating point of the resonator on the immittance circle can change Γ. The latter method is a substantially more direct and cost effective means of adjustment than the others used to date. Figure 4 shows the variation of immittance over a single resonance. As the resonance region is traversed, not only do the electrical characteristics vary, but also the spatial distribution of motion within the plate.

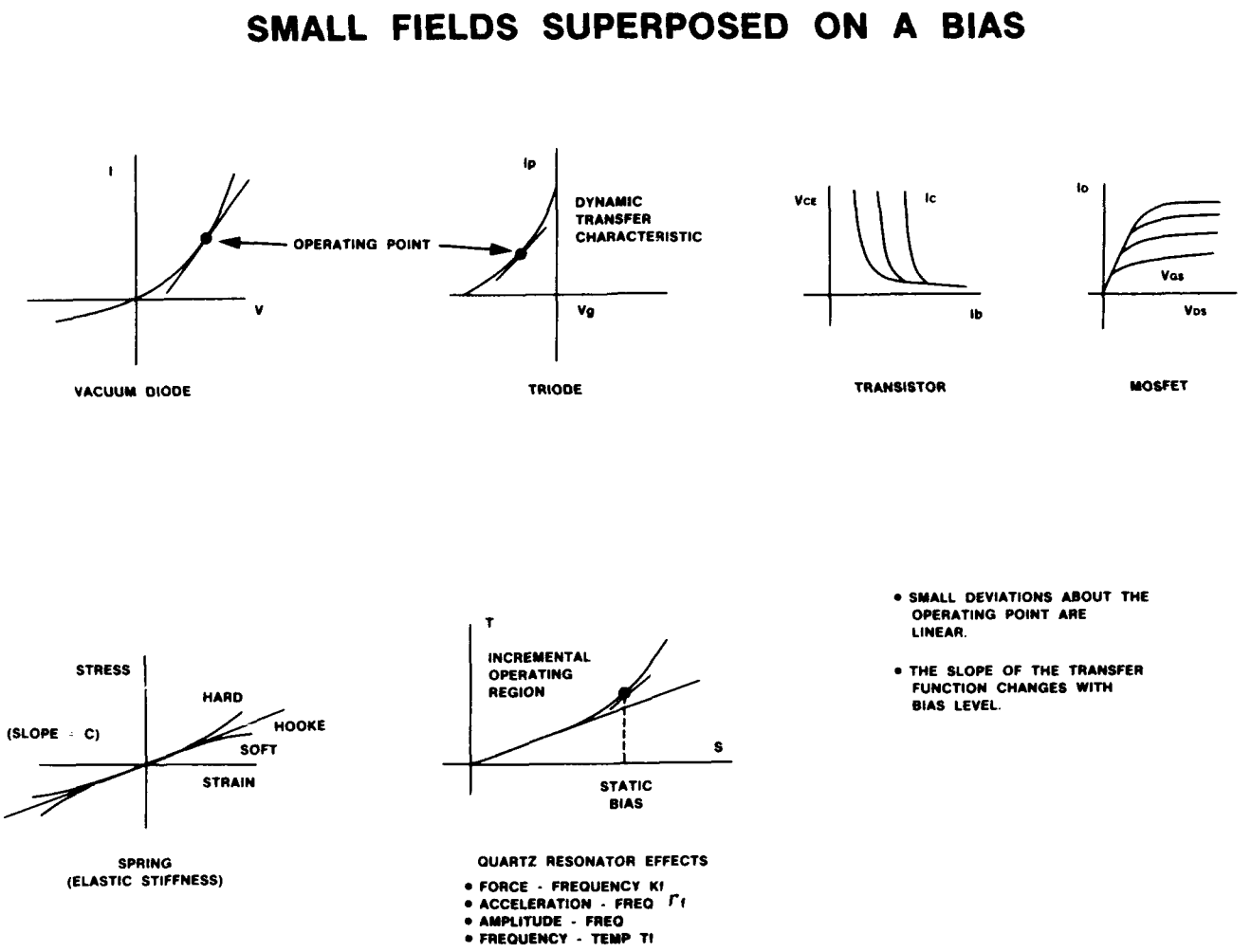


Figure 3. Examples of small fields superposed on a bias. The vacuum diode, triode, transistor, and MOSFET are familiar electronic examples of nonlinear devices. A linear operating point is established by a suitable bias network. The effective elastic constant of quartz is linear for small stress/strain variables, and nonlinear otherwise. The elastic and electronic cases are analogous when the elastic operating point is produced by a quasistatic acceleration field.

Figure 2 shows some of the mode patterns associated with the resonances of a quartz plate. The resonances are the eigenfrequences and the mode patterns are the corresponding eigenfunctions. The eigenfunctions form a complete set. At a resonance, the pattern of motion will be the eigenfunction appropriate to that eigenfrequency. This is normally what is utilized in practice. The slightest departure from an eigenfrequency produces a distribution of motion that is an admixture of all the eigenfunctions [20]. The extent of mixing, and therefore the resulting pattern, depends on the frequency shift with respect to all the other eigenfrequencies, and is calculable.

The above remarks require some modification for nonlinear effects, but the idea is substantially as outlined: mode mixing occurs when one detunes from resonance. The change in mode pattern can be utilized to provide acceleration compensation.

Heretofore many of the attempts at reducing Γ have concentrated on changing the supports, or the overall geometry, or the aspect ratio. The alternative of electronically changing the mode shape is not at all obvious, primarily because changing the mode shape changes the resonant frequency, and this is usually avoided at all costs.

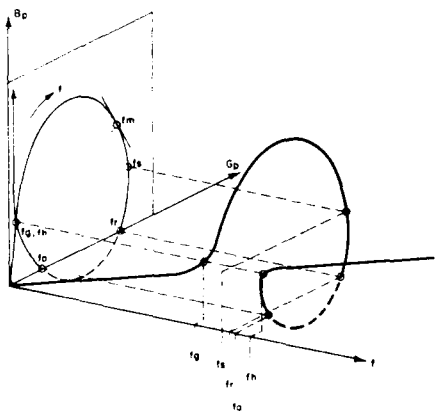


Figure 4. Plot of real and imaginary parts of the crystal admittance versus frequency. As the region of resonance is traversed, the shape of the mode changes. This change is subtle for small frequency shifts near a resonance peak, but the change in mode shape can nevertheless have a dramatic effect on the acceleration sensitivity coefficient (Γ).

Once the resonator is manufactured and mounted, the last stage is typically devoted to frequency trimming or adjustment by depositing a small spot of electrode material (in the case of a BAW resonator) on top of the regular electrode. The nominal frequency of the untrimmed unit is deliberately made too high, and the small trimming spot lowers the frequency to that desired. The unit is then sealed, marked, and shipped. In the process of final frequency adjustment, however, the acceleration sensitivity has likewise been changed, and very often in a capricious way, since the final trimming is usually not applied with care, or with much symmetry. The variation of the sensitivity coefficient during resonator fabrication has escaped notice to date because it has been tacitly assumed either that the acceleration sensitivity is a constant, or that its value is not of interest until manufacture is complete.

By performing acceleration tests to determine Γ after mounting, but before trimming, one has available a very simple method of changing the motional distribution of the resonator in order to change Γ. The method is simply to operate the resonant structure somewhat off-resonance, where the amplitude of vibration is still large, and therefore still well-suited to frequency stabilization, but where the spatial distribution of motion differs from that at resonance.

If the distribution of motion pattern is plotted as a function of driving frequency f, then as the structure approaches resonance, the motional pattern will change, not only in size (amplitude), but also in spatial distribution. Since the motional pattern changes, the acceleration sensitivity also changes as a function of the location of the resonator operating point on the resonance impedance circle. Another manifestation of this is that the acceleration sensitivity is a function of both overtone and mode of vibration. This has been observed experimentally, but the importance of the observation appears not to have been appreciated.

It is possible to obtain information about how to adjust the acceleration sensitivity by measuring Γ as a function of location on the impedance circle. Indeed, if a particular part of the impedance circle yields a low value of Γ, one may fix the resonator/oscillator to operate at this point, even though it is not strictly at the resonance point. The basic oscillator is then operating at a point on the impedance circle where it does not have the desired frequency

output, but does have the proper mode shape to give low values of acceleration sensitivity. The desired frequency can then be derived by frequency synthesis techniques, i.e., by bit-stuffing the output data stream. The acceleration desensitization is electronic, as is the synthesis of the proper output frequency. The resonator units are manufactured conventionally with high yield, not individually trimmed or sorted for low-acceleration sensitivity. Sorting would result in very low yields.

EXPERIMENTS

Tests performed at ETDL as part of an in-house research program have confirmed the ability to reduce the acceleration sensitivity electronically, for both BAW and SAW resonators. For devices where the mode shape would not be expected to change very much with frequency adjustment, e.g., SAW delay lines, changes in the sensitivity coefficient will be minimal. Some representative examples of SAW devices tested are shown on Figs. 5 to 9. A portion of the change in the gamma coefficient seen in Figs. 5 to 8 is due to a structure resonance, but the mode shape change is large. Figure 9 shows a twenty-fold improvement in gamma with electronic tuning. Even greater effects have been seen in BAW resonators at present. It is possible to design resonators of each type so that the mode shape is more sensitive to frequency offsets to maximize the effect.

REFERENCES

[1] R. D. Mindlin, "Dynamics of package cushioning," Bell System Tech. Jour., Vol. 24, 1945, pp. 353-461.

[2] E. A. Gerber and J. Havel, "Precision frequency control for military applications," IRE Trans. on Military Electronics, Vol. MIL-4, October 1960, pp. 424-437.

[3] K. S. Van Dyke, "Strain patterns in thickness-shear resonators," Proc. 11th Annual Frequency Control Symposium, May 1957, pp.41-61. AD-298 323.

[4] R. D. Mindlin and D. C. Gazis, "Strong resonances of rectangular AT-cut quartz plates," Proc. Fourth U.S. National Congress of Appl. Mech., 1962, pp. 305-310. See also The Collected Papers of Raymond D. Mindlin (H. Deresiewicz, M. P. Bieniek, and F. L. DiMaggio, eds.) Vol. 2, Springer Verlag, New York, 1989, pp. 533-538.

[5] K. Haruta and W. J. Spencer, "X-ray diffraction study of vibrational modes," Proc. 20th Annual Frequency Control

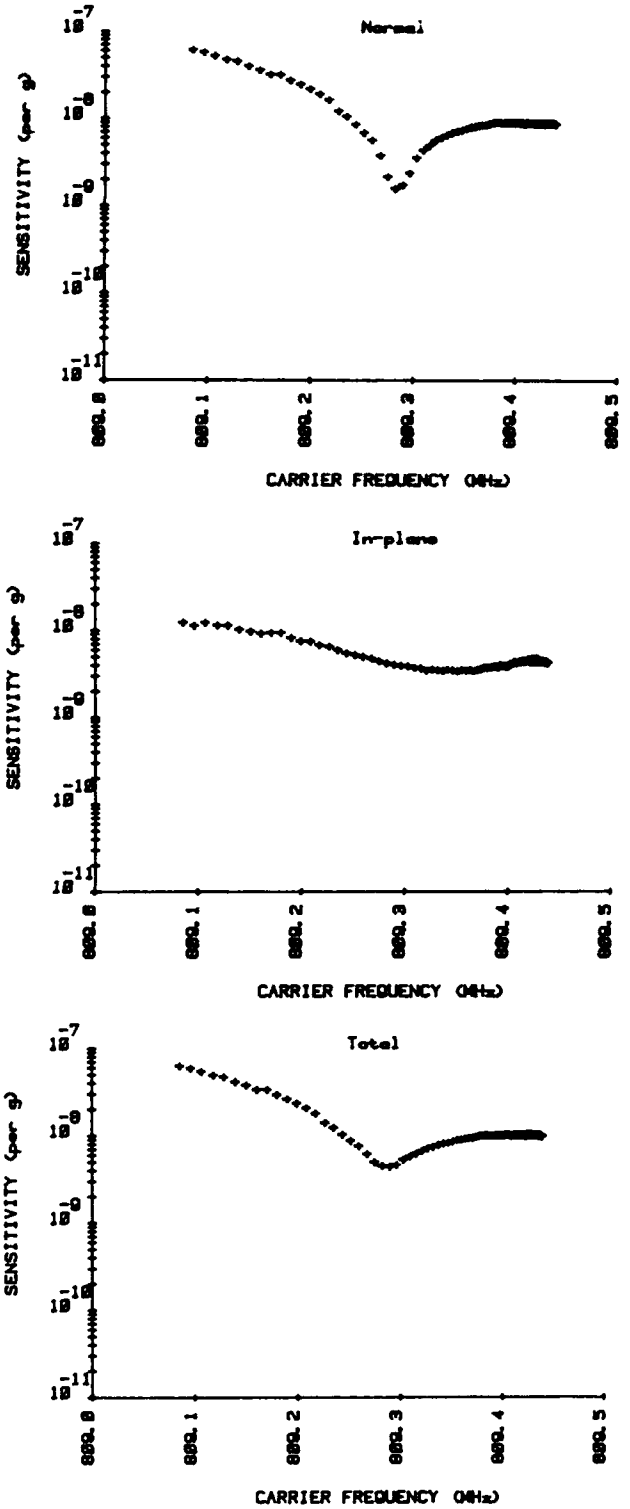


Figure 5. An 809 MHz SAW resonator subjected to an acceleration of 2g (peak-to-peak) at a modulation frequency of 210 Hz. Carrier frequency versus acceleration sensitivity for (a) normal and (b) in-plane accelerations; (c) the total sensitivity is also shown.

Symposium, April 1966, pp. 1-13. AD-800 523.

[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. Am., Vol. 42, December 1967, pp. 1268-1277. See also The Collected Papers of Raymond D. Mindlin (H. Deresiewicz, M. P. Bieniek, and F. L. DiMaggio, eds.) Vol. 2, Springer Verlag, New York, 1989, pp. 694-703.

[7] E. G. Newman and R. D. Mindlin, "Vibrations of a monoclinic crystal plate," J. Acoust. Soc. Am., Vol. 29, No. 11, November 1957, pp. 1206-1218. See also The Collected Papers of Raymond D. Mindlin (H. Deresiewicz, M. P. Bieniek, and F. L. DiMaggio, eds.) Vol. 2, Springer Verlag, New York, 1989, pp. 368-380.

[8] I. Koga and Y. Tsuzuki, "Comments on 'Modes in circular AT quartz plates,'" IEEE Trans. Sonics Ultrason., Vol. SU-13, No. 2, July 1966, p.78.

[9] P. C. Y. Lee and K.-M. Wu, "In-plane accelerations and forces on frequency changes in doubly rotated quartz plates," J. Acoust. Soc. Am., Vol. 75, No. 4, April 1984, pp. 1105-1117.

[10] P. C. Y. Lee and M. S. H. Tang, "Acceleration effect on the thickness vibrations of doubly rotated crystal resonators," Proc. 41st Annual Frequency Control Symposium, May 1987, pp. 277-281.

[11] D. V. Shick, Y. S. Zhou, and H. F. Tiersten, "An analysis of the in-plane acceleration sensitivity of quartz surface-wave resonators rigidly supported along the edges," J. Appl. Phys., Vol. 65, No. 1, 1 January 1989, pp. 35-40.

[12] H. F. Tiersten and D. V. Shick, "On the normal acceleration sensitivity of contoured quartz resonators rigidly supported along rectangular edges," J. Appl. Phys., Vol. 67, No. 1, 1 January 1990, pp. 60-67.

[13] H. F. Tiersten and D. V. Shick, "On the in-plane acceleration sensitivity of ST-cut quartz surface-wave resonators with interior rectangular supports," J. Appl. Phys., Vol. 67, No. 5, 1 March 1990, pp. 2554-2566.

[14] A. Ballato, "Crystal resonators with increased immunity to acceleration fields," IEEE Trans. Sonics Ultrason., Vol. SU-27, No. 4, July 1980, pp. 195-201.

[15] M. E. Frerking, "Vibration compensated crystal oscillator," U.S. Patent No. 4,891,611, issued January 2, 1990.

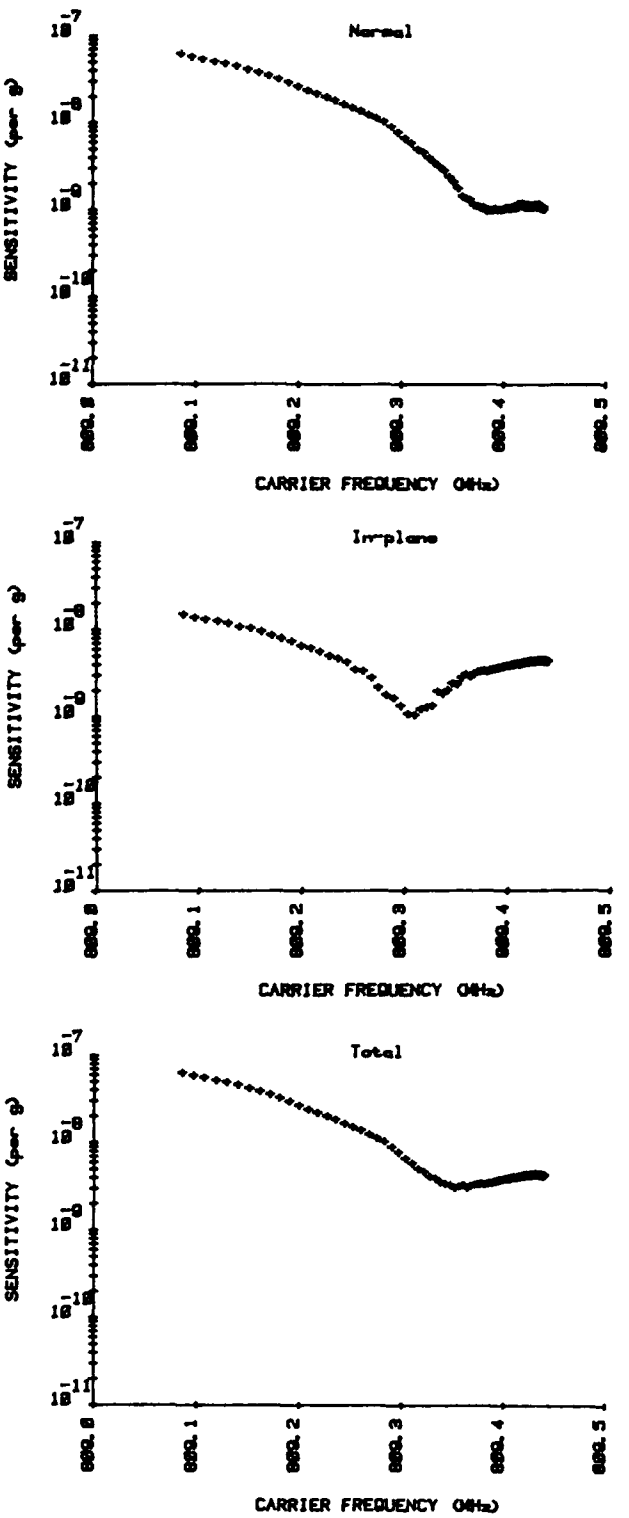


Figure 6. An 809 MHz SAW resonator subjected to an acceleration of 2g (peak-to-peak) at a modulation frequency of 450 Hz. Carrier frequency versus acceleration sensitivity for (a) normal and (b) in-plane accelerations; (c) the total sensitivity is also shown.

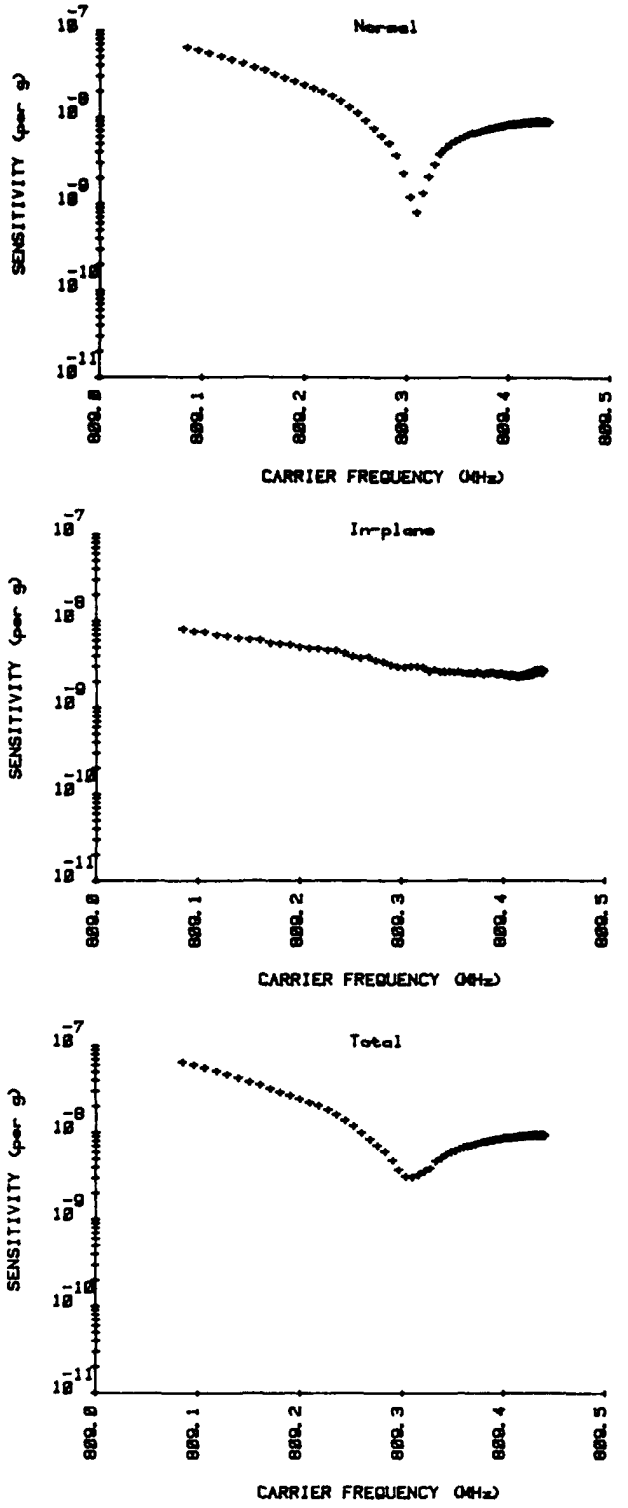


Figure 7. An 809 MHz SAW resonator subjected to an acceleration of 2g (peak-to-peak) at a modulation frequency of 750 Hz. Carrier frequency versus acceleration sensitivity for (a) normal and (b) in-plane accelerations; (c) the total sensitivity is also shown.

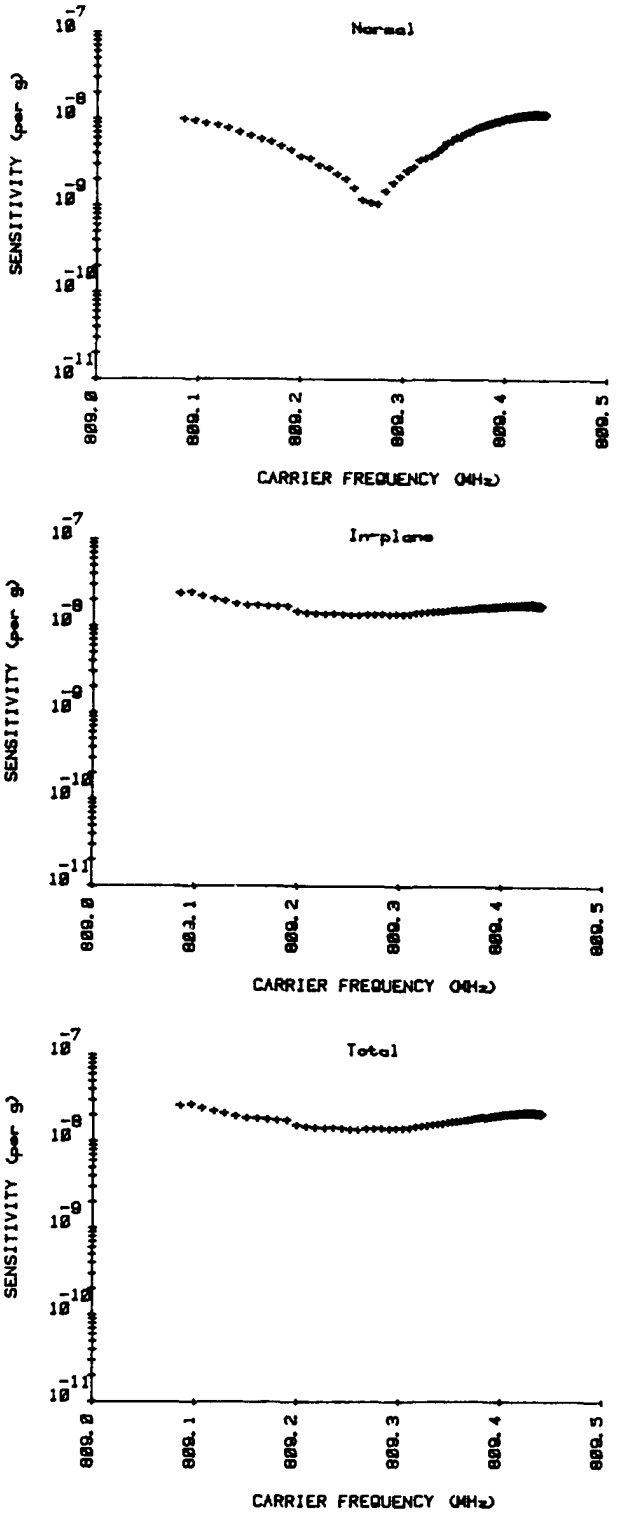


Figure 8. An 809 MHz SAW resonator subjected to an acceleration of 2g (peak-to-peak) at a modulation frequency of 1530 Hz. Carrier frequency versus acceleration sensitivity for (a) normal and (b) in-plane accelerations; (c) the total sensitivity is also shown.

[16] A. Ballato, "Method of making an acceleration hardened resonator," U.S. Patent No. 4,836,882, issued June 6, 1989.

[17] E. P. EerNisse and R. W. Ward, "Crystal resonator with low acceleration sensitivity and method of manufacture thereof," U.S. Patent No. 4,837,475, issued June 6, 1989.

[18] A. Ballato, "Method of making a crystal oscillator desensitized to acceleration fields," U.S. Patent No. 4,871,986, issued October 3, 1989.

[19] R. C. Smythe and W. H. Horton, "Adjustment of resonator G-sensitivity by circuit means," these proceedings.

[20] F. B. Hildebrand, Methods of Applied Mathematics, Prentice-Hall, Inc., Englewood Cliffs, NJ, 1952.

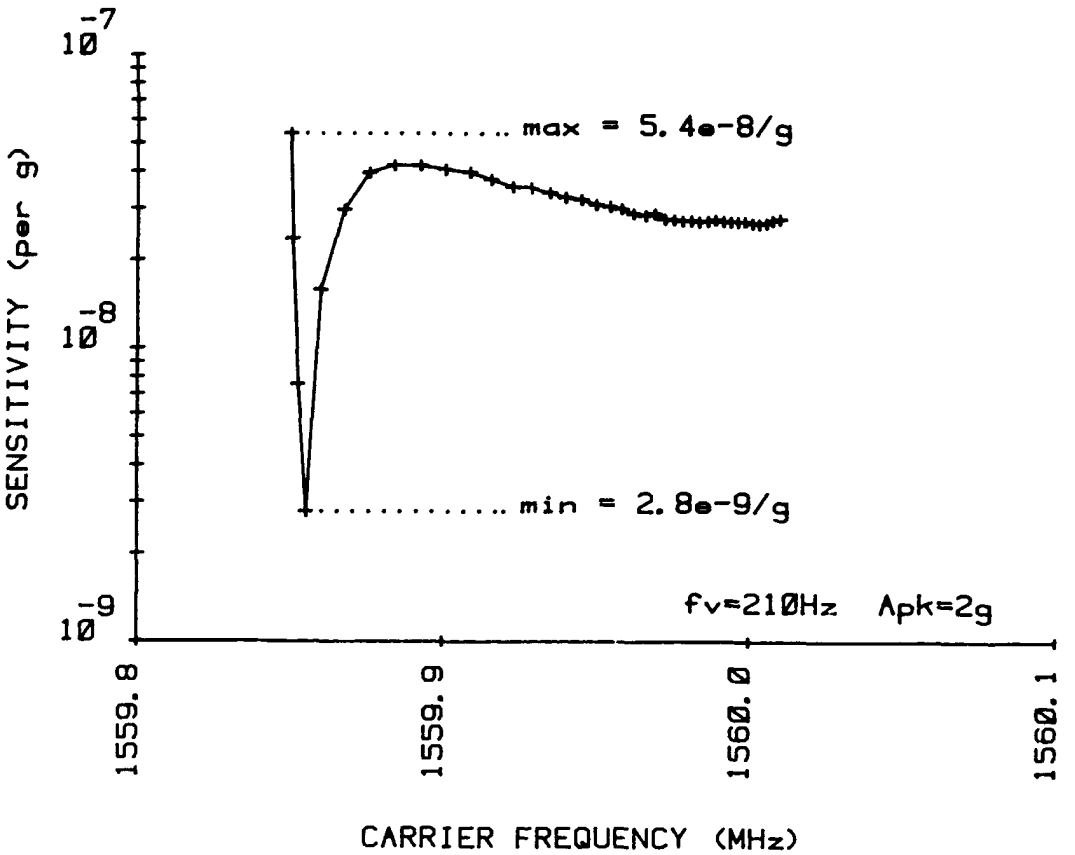


Figure 9. Acceleration sensitivity versus carrier frequency for an L-band SAW resonator at a modulation (shake) frequency of 210 Hz and 2g peak-to-peak amplitude. The graph shows a nearly twenty-fold improvement.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ON THE INFLUENCE OF A FABRICATION IMPERFECTION ON THE NORMAL ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS WITH RECTANGULAR SUPPORTS

Y.S. Zhou and H.F. Tiersten

Department of Mechanical Engineering, Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute
Troy, NY 12180-3590

Abstract

It is shown that the normal acceleration sensitivity of contoured quartz resonators with rectangular supports vanishes when the centers of the mode shape and support rectangle coincide. This result is essentially a consequence of symmetry and applies to many other shapes. Since it is extremely difficult to realize this situation in practice, an analysis of the influence of an offset of the centers on the normal acceleration sensitivity is performed. The biasing deformation is determined by means of a variational approximation procedure using the variational principle with all natural conditions for anisotropic static flexure. The very important accompanying strains varying quadratically across the thickness are determined recursively, as in earlier work. The resulting flexural biasing states are employed in the existing perturbation equation along with the equivalent trapped energy mode shapes of the contoured resonators to calculate the normal acceleration sensitivities. It is shown that for small offsets the acceleration sensitivity increases linearly with offset and orientations for which this effect is minimized are found.

1. Introduction

In recent work[1,2] on the normal acceleration sensitivity of contoured quartz resonators supported along rectangular edges it was shown that the sensitivity is always negligibly small and vanishes for certain cases when the centers of the mode shape and support rectangle coincide. In this work it is shown that, as a consequence of symmetry, the normal acceleration sensitivity always vanishes when the centers coincide. Since this situation is extremely difficult to realize in practice, an analysis of the influence of an offset of the centers on the normal acceleration sensitivity is performed. It is further shown that for small offsets of centers, which is the practical case, the normal acceleration sensitivity increases linearly with offset. This results in the definition of a resonator vector parameter, the minimization of the magnitude of which yields the best orientations for fabrication imperfection. The linear increase with offset underscores the importance of care in fabrication[3] to achieve consistently low acceleration sensitivities.

The biasing deformation fields are calculated by means of a variational approximation procedure, which employs the variational principle for anisotropic static flexure as in earlier work[4,1]. The important quadratically varying strains are then determined recursively as in earlier work[1,2]. The calculated biasing deformation fields are employed in an existing perturbation equation[5] along with the equivalent trapped energy mode shapes[6] for the contoured resonator to calculate the normal acceleration sensitivities. Results are presented for a range of offsets for some orientations and aspect ratios of the rectangular support configuration for both the AT and SC cuts.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency obtained from the perturbation analysis[5] 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

$$H_\mu = - \int_V \hat{K}^n_{L\gamma} g^\mu_{\gamma,L} \, dV, \tag{2.2}$$

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

$$\hat{K}^n_{L\gamma} = \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M} , \tag{2.3}$$

where

$$\hat{c}_{L\gamma M\alpha} = T^1_{LM} \delta_{\gamma\alpha} + c_{3L\gamma M\alpha KN} E^1_{KN} + c_{2L\gamma KM} w_{\alpha,K} + c_{2LKM\alpha} w_{\gamma,K} , \tag{2.4}$$

and

CH2818-3/90/0000-452 $1.00 © 1990 IEEE

$$T^1_{LM} = c_{2LMKN} E^1_{KN}, \quad E^1_{KN} = \frac{1}{2}\left(w_{K,N} + w_{N,K}\right). \tag{2.5}$$

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

$$g^\mu_\gamma \equiv \frac{u^\mu_\gamma}{N_\mu}, \quad \hat{f}^\mu = \frac{\hat{\varphi}^\mu}{N_\mu}, \quad N^2_\mu = \int_V \rho u^\mu_\gamma u^\mu_\gamma \, dV, \tag{2.6}$$

where u^μ_γ and $\hat{\varphi}^\mu$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\hat{K}^\ell_{L\gamma} = c_{2L\gamma M\alpha} u_{\alpha,M} + e_{ML\gamma}\hat{\varphi}_{,M},$$
$$\hat{\mathcal{D}}^\ell_L = e_{LM\gamma} u_{\gamma,M} - e_{LM}\hat{\varphi}_{,M}, \tag{2.7}$$

$$\hat{K}^\ell_{L\gamma,L} = \rho\ddot{u}_\gamma, \quad \hat{\mathcal{D}}^\ell_{L,L} = 0, \tag{2.8}$$

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.5, as is the remainder of the notation in this section.

The substitution of (2.3) in (2.2) yields

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha} g^\mu_{\alpha,M} g^\mu_{\gamma,L} \, dV. \tag{2.9}$$

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. Equations for Anisotropic Static Flexure

With the aid of Mindlin's plate equations[7-9] it has been shown that the equation of anisotropic static flexure with X_2 normal to the major surfaces can be written in the form[10]

$$K^{(1)}_{AB,AB} + F^{(1)}_{B,B} + F^{(0)}_2 = 0, \tag{3.1}$$

where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2 and

$$K^{(1)}_{AB} = \int_{-h}^{h} X_2 K_{AB}\, dX_2, \quad F^{(1)}_B = h[K_{2B}(h) + K_{2B}(-h)], \tag{3.2}$$

and in this instance

$$F^{(1)}_B = 0, \quad F^{(0)}_2 = K_{22}(h) - K_{22}(-h) - 2h\rho a_2 = -2h\rho a_2, \tag{3.3}$$

since K_{2K} vanishes on the major surfaces and where a_2 is the acceleration in the X_2-direction. From Eqs.(4.25) of Ref.10 the constitutive equations for the stress-resultants take the form

$$K^{(1)}_{AB} = \frac{2}{3} h^3 \gamma_{ABCD} E^{(1)}_{CD}, \tag{3.4}$$

where Voigt's anisotropic plate elastic constants are given by

$$\gamma_{RS} = c_{RS} - c_{RW} c^{-1}_{WV} c_{VS}, \quad R,S = 1,3,5; \; W,V = 2,4,6. \tag{3.5}$$

in the compressed notation, and where we have introduced the scheme shown. The plate strains $E^{(1)}_{CD}$ in (3.4) are given by

$$E^{(1)}_{CD} = \frac{1}{2}\left(w^{(1)}_{C,D} + w^{(1)}_{D,C}\right), \tag{3.6}$$

and from the relaxation of the stress resultants $K^{(1)}_{2L}$, the vanishing of the plate shear strains $E^{(0)}_{2A}$ and (3.6) we have the respective relations

$$E^{(1)}_W = -c^{-1}_{WV} c_{VS} E^{(1)}_S, \quad w^{(1)}_A = -w^{(0)}_{2,A},$$
$$E^{(1)}_{CD} = -w^{(0)}_{2,CD}, \tag{3.7}$$

which, respectively, are given in Eqs.(4.24), (3.34) and (3.37) of Ref.10.

It has been shown[1] that the biasing shearing stresses that are determined recursively in the classical theory of the flexure of thin plates may be written in the form

$$\hat{K}_{2B} = (3/4\, h^3)(h^2 - X_2^2)\, K^{(1)}_{AB,A}, \tag{3.8}$$

where the caret indicates that a quantity has been determined recursively and, hence, contains an order of smallness. The associated three-dimensional strain fields are obtained simply by inverting the linear constitutive equations

$$\hat{K}_{LM} = c_{LMKN}\hat{E}_{KN}, \tag{3.9}$$

with the result

$$\hat{E}_{KL} = c^{-1}_{KL2B}\hat{K}_{2B} = {}_o\hat{E}_{KL} + X_2^2\, {}_2\hat{E}_{KL}, \tag{3.10}$$

since all stresses that vary symmetrically with X_2 vanish except $\hat{K}_{2B}$, and ${}_o\hat{E}_{KL}$ and ${}_2\hat{E}_{KL}$ are clear from (3.8) and (3.10). Since from Eqs.(3.31) of Ref.6

we have $E_{KL}^{(0)} = 0$, when the plate deflection $w_2^{(0)}$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = {}_o\hat{E}_{KL} + X_2 E_{KL}^{(1)} + X_2^2 {}_2\hat{E}_{KL}. \tag{3.11}$$

Although we now have the plate strains from (3.11), we cannot yet determine the $\hat{c}_{L\gamma M\alpha}$ from (2.4) because, as noted in Ref.6, we need the displacement gradients $w_{L,K}$ or equivalently the rotations

$$\Omega_{KL} = \frac{1}{2}(w_{L,K} - w_{K,L}). \tag{3.12}$$

As in Ref.6, we realize that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K_{2L}^{(1)}$ and the recursive inversion in (3.10) satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L}, \tag{3.13}$$

where

$$\Omega_{KL} = \sum_{n=0}^{2} X_2^n \Omega_{KL}^{(n)}, \tag{3.14}$$

and the $\Omega_{KL}^{(n)}$ are the plate rotations of nth order. As in the end of Sec.III of Ref.6 [after Eq.(3.32)] from the flexural solution without recursive strains we obtain

$$\Omega_{13}^{(0)} = \text{arbitrary constant} = 0,$$
$$\Omega_{13}^{(1)} = 0, \quad \Omega_{2A}^{(0)} = -w_{2,A}^{(0)}. \tag{3.15}$$

Including recursive strains for n = 1 and 2 and equating like powers of X_2 in the same way we obtain[1]

$$\hat{\Omega}_{13}^{(1)} = {}_o\hat{E}_{23,1} - {}_o\hat{E}_{21,3}, \quad \Omega_{13}^{(2)} = \frac{1}{2}(E_{23,1}^{(1)} - E_{21,3}^{(1)}),$$
$$\Omega_{2A}^{(1)} = E_{2A}^{(1)} - {}_o\hat{E}_{22,A}, \quad \Omega_{2A}^{(2)} = {}_2\hat{E}_{2A} - \frac{1}{2}E_{22,A}^{(1)}. \tag{3.16}$$

The substitution of (3.15) and (3.16) in (3.14) yields Ω_{KL}, which with (3.11) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \tag{3.17}$$

gives us the desired three-dimensional displacement gradients when the plate deflection $w_2^{(0)}$ has been found.

4. Analysis for Flexural Biasing State

A plan view of the plate is shown in Figure 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 d in an arbitrary direction θ. Hence, $\hat{d}$ has arbitrary components d_1 and d_3. In this section we are concerned with the static solutions for the biasing states referred to the unprimed coordinate system only.

The substitution of (3.3), (3.4) and $(3.7)_3$ in (3.1) yields the equilibrium equation for flexure of the thin plate in the form

$$\frac{2}{3}h^3 \gamma_{ABCD} w_{2,CDAB}^{(0)} + 2h\rho a_2 = 0. \tag{4.1}$$

The boundary conditions are given by

$$w_2^{(0)} = 0, \; K_{11}^{(1)} = 0 \text{ at } X_1 = \pm a, \; -b < X_3 < b,$$
$$w_2^{(0)} = 0, \; K_{33}^{(1)} = 0 \text{ at } X_3 = \pm b, \; -a < X_1 < a. \tag{4.2}$$

Since as already indicated the problem defined in (4.1) and (4.2) cannot be solved exactly, a variational approximation procedure is employed[4]. To this end we first transform the inhomogeneities from the differential equation (4.1) into certain of the other boundary conditions by writing

$$w_2^{(0)} = \hat{w}_2^{(0)} + A(X_1^2 - a^2)(X_1^2 - 5a^2), \tag{4.3}$$

which when substituted into (4.1) yields

$$\gamma_{11}\hat{w}_{2,1111}^{(0)} + 4\gamma_{15}\hat{w}_{2,1113}^{(0)} + 2(\gamma_{13} + 2\gamma_{55})\hat{w}_{2,1133}^{(0)} + 4\gamma_{35}\hat{w}_{2,1333}^{(0)} + \gamma_{33}\hat{w}_{2,3333}^{(0)} = 0, \tag{4.4}$$

since A is selected as

$$A = -\rho a_2/8h^2\gamma_{11}. \tag{4.5}$$

The further substitution of (4.3) with (3.4) and (3.7) into (4.2) yields the edge conditions, which for brevity we do not bother to write here[11].

As a solution of (4.4) we write

$$\hat{w}_2^{(0)} = D_o(X_1 - a)\left(X_3 - \frac{\gamma_{15}}{\gamma_{11}}X_1\right) + \hat{D}_o(X_1 - a)\left[3X_3^2 + \frac{\gamma_{13}}{\gamma_{11}}X_1(2a - X_1) - \frac{2\gamma_{15}}{\gamma_{11}}(X_1 + a)\left(3X_3 - \frac{2\gamma_{15}}{\gamma_{11}}X_1\right)\right] + \sum_{n=-N}^{N} D_n e^{i(\eta_n X_1 + \nu_n X_3)}, \tag{4.6}$$

where $\nu_n = n\pi/2pb$ and p is chosen to be an irrational number (here 2π) in order that neither $\cos \nu_n b$ or $\sin \nu_n b$ vanish. Let $\eta_n = h\nu_n$ and substitute from (4.6) into (4.4) to obtain

$$\gamma_{11}\hat{h}^4 + 4\gamma_{15}\hat{h}^3 + 2(\gamma_{13} + 2\gamma_{55})\hat{h}^2) + 4\gamma_{35}\hat{h} + \gamma_{33} = 0, \tag{4.7}$$

which is independent of n. Equation (4.7) yields four independent roots $\hat{h}_p$ (p = 1,2,3,4) independent of n. Hence, the sum in (4.6) can be written in the form

$$\sum_{n=-N}^{N} D_n e^{i(\eta_n X_1 + \nu_n X_3)} = \sum_{n=-N}^{N} \sum_{p=1}^{4} D_n^{(p)} e^{i\eta_n^{(p)} X_1} e^{i\nu_n X_3}, \quad (4.8)$$

where the $D_n^{(p)}$ along with D_o and $\hat{D}_o$ are amplitude coefficients still to be determined. By satisfying the edge conditions $\hat{w}_2^{(0)} = 0$ at $X_1 = +a$ and $\hat{K}_{11}^{(1)} = 0$ at $X_1 = \pm a$ for each n, we obtain

$$D_n^{(p)} = g_n^{(p)} D_n^{(1)}, \quad (4.9)$$

where the $g_n^{(p)}$ are obtained from the resulting linear algebra[11]. This reduces the number of unknown coefficients from 4 to 1 for each n. It should be noted that the terms in (4.6) with the coefficients D_o and $\hat{D}_o$ already satisfy these same conditions. The substitution of (4.8) and (4.9) in (4.6) enables us to write

$$\hat{w}_2^{(0)} = D_o(X_1 - a)\left(X_3 - \frac{\gamma_{15}}{\gamma_{11}} X_1\right) + \hat{D}_o(X_1 - a)\left[3X_3^2 + \frac{\gamma_{13}}{\gamma_{11}} X_1(2a - X_1) - 2\frac{\gamma_{15}}{\gamma_{11}}(X_1 + a)\left(3X_3 - 2\frac{\gamma_{15}}{\gamma_{11}} X_1\right)\right] + \sum_{n=-N}^{N} D_n^{(1)} e^{i\nu_n X_3} \sum_{p=1}^{N} g_n^{(p)} e^{i\eta_n^{(p)} X_1}, \quad (4.10)$$

as the approximate solution function.

Since the solution (4.10) with (4.3) satisfies the differential equation (4.1) exactly and one edge condition is of constraint type, but in natural form, and the other is of natural type, all that remains of the variational principle given in Eq.(44) of Ref.4, in which all conditions appear as natural conditions, is

$$\int_{C^N} (N_A K_{AB}^{(1)} N_B - \bar{m}) \delta \frac{\partial w_2^{(0)}}{\partial n} \, ds - \oint_{C^C} (\delta M_{NS,S} + N_B \delta K_{AB,A}^{(1)})(\bar{w}_2^{(0)} - w_2^{(0)}) \, ds + (\delta M_{NS}^{+} - \delta M_{NS}^{-})(\bar{w}_2^{(0)C_2} - w_2^{(0)C_2}) = 0, \quad (4.11)$$

where

$$M_{NS} = N_D K_{DC}^{(1)} \mathscr{s}_C, \quad (4.12)$$

and N_B and $\mathscr{s}_C$ denote unit vectors, respectively, outwardly directed normal and tangential to C in the counterclockwise direction, as shown in Figure 2 of Ref.4, which defines + and - and the point C_2. The bars over quantities in (4.11) denote that they are prescribed. The variation $\delta M_{NS,S}$ may readily be evaluated from (4.12) and

$$\bar{w}_2^{(0)} = \bar{w}_2^{(0)C_2} = 0, \ \bar{m} = 0 \text{ on } C^C. \quad (4.13)$$

Substituting from (4.3) with (4.10), (4.13) and the homogeneous edge conditions mentioned above which have been satisfied exactly by each term in the solution sum, into (4.11) and employing (4.12), we obtain

$$\int_{-a}^{a} \left[\left(\hat{K}_{33}^{(1)} + \frac{\gamma_{13}}{\gamma_{11}} \rho a_2 h \, (X_1^2 - a^2)\right)\delta\hat{w}_{2,3}^{(0)} + (\hat{w}_2^{(0)} + A(X_1^2 - a^2)(X_1^2 - 5a^2))(\delta\hat{K}_{33,3}^{(1)} + 2\delta\hat{K}_{13,1}^{(1)})\right]_{X_3=-b}^{X_3=b} dX_1 - \int_{-b}^{b} \hat{w}_2^{(0)}(\delta\hat{K}_{11,1}^{(1)} + 2\delta\hat{K}_{13,3}^{(1)})\Big|_{X_1=-a} dX_3 + \hat{w}_2^{(0)}(\delta\hat{K}_{13,1}^{(1)} + \delta\hat{K}_{13,3}^{(1)})\Big|_{\substack{X_1=-a \\ X_3=b}} + \hat{w}_2^{(0)}(\delta\hat{K}_{13,1}^{(1)} - \delta\hat{K}_{13,3}^{(1)})\Big|_{\substack{X_1=-a \\ X_3=-b}} = 0, \quad (4.14)$$

where the notation $[\ ]_{X_3=-b}^{X_3=b}$ means the quantity evaluated at b minus the quantity evaluated at -b and the $\delta\hat{K}_{AB}^{(1)}$ are obtained from (3.4) and $(3.7)_3$ with $\hat{w}_2^{(0)}$ instead of $w_2^{(0)}$ since A is fixed and the variation of prescribed quantities vanishes. Substituting from (4.3) with (4.10) into (4.14) and performing the integrations, we obtain

$$\sum_{n=-N}^{N} \sum_{m=-N}^{N} a_{mn} D_m \delta D_n + \sum_{n=-N}^{N} b_n \delta D_n = 0, \ m,n = 0, \hat{0}, \pm 1, \pm 2, \pm N, \quad (4.15)$$

where 0 and $\hat{0}$ refer to the amplitudes D_o, $\hat{D}_o$ and superscripts have been suppressed. The expressions for a_{mn} and b_n are too lengthy to present here[11]. Since the variations δD_n are arbitrary, we obtain

$$\sum_{m=-N}^{N} a_{mn} D_m = -b_n, \ m,n = 0, \hat{0}, \pm 1, \pm 2, \ldots \pm N, \quad (4.16)$$

which constitute 2N + 2 inhomogeneous linear algebraic equations in the 2N + 2 unknowns D_m, the inversion of which gives the approximate solution.

Convergence is determined by increasing the number N and comparing the solution for N with that for N - 1.

5. The Equivalent Trapped Energy Mode Shape

A schematic diagram of the contoured crystal resonator along with the associated coordinate system is shown in Figure 2. 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[12]

$$u_{1n} = \sin \frac{n\pi X_2}{2h}\, u_{noo} e^{i\omega_{noo} t}, \quad n = 1,3,5, \ldots, \tag{5.1}$$

where

$$u_{noo} = e^{-\alpha_n(\hat{X}_1^2/2) - \beta_n(\hat{X}_3^2/2)}, \tag{5.2}$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2 \hat{c}(1)}{8Rh_o^3 M_n'}, \quad \beta_n^2 = \frac{n^2\pi^2 \hat{c}(1)}{8Rh_o^3 P_n'}, \tag{5.3}$$

and M_n' and P_n' are involved lengthy expressions that are defined in Sec.II of Ref.12 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 R denotes the radius of the spherical contour. From Eq.(97) of Ref.12 in this work we have

$$\hat{X}_A = R_{AB}(X_B - d_B), \tag{5.4}$$

where, from Eq.$(98)_1$ of Ref.12

$$R_{AB} = \begin{Vmatrix} \cos\hat{\beta}_n & -\sin\hat{\beta}_n \\ \sin\hat{\beta}_n & \cos\hat{\beta}_n \end{Vmatrix}, \tag{5.5}$$

and $\hat{\beta}_n$ is defined in Eq.$(98)_2$ of Ref.12. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2 \hat{c}(1)}{4h_o^2 \hat{\rho}} \left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \cdot \left(\sqrt{\frac{M_n'}{\hat{c}(1)}} + \sqrt{\frac{P_n'}{\hat{c}(1)}}\right)\right]. \tag{5.6}$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.12.

In addition to the thickness eigendisplacement u_{1n} given in (5.1) there are thickness eigendisplacements u_{2n} and u_{3n}, which are an order of magnitude smaller than u_{1n} but are required in this work because as shown in earlier work[1] the important biasing shearing deformation is an order of magnitude smaller than the biasing flexural deformation. However, since the u_2 and u_3 displacement components accompanying the larger u_1 component are known only for the flat plate, we fit the Gaussian mode shape given in (5.1) and (5.2) for the contoured resonator to a trapped energy mode in a flat plate in accordance with Sec.V of Ref.1, wherein the wavenumbers $\bar{\hat{\xi}}$, $\bar{\hat{\nu}}$, $\hat{\xi}^{SR}$ and $\hat{\nu}^{TM}$ are obtained from quantities contained in the Gaussian mode shape. As shown in Sec.V of Ref.1, when this has been done it is found from Eqs.(65) of Ref.12 that to the order of interest the $u_{\hat{n}}^{SR}$ displacement field for the SR region may be written

$$u_1^{SR} = A_+^{(1)SR}\left(1 + \frac{(c_{16}\xi^{SR} - c_{56}i\bar{\nu})}{\bar{c}(1)} X_2\right) \cdot \cdot \sin\frac{n\pi X_2}{2h}\, e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\bar{\hat{\nu}}\hat{X}_3},$$

$$u_2^{SR} = \left(\frac{(r_2\xi^{SR} - r_4 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR} \cos\frac{n\pi}{2h} X_2 + iC_+^{(2)SR} \cos\kappa_2 \frac{n\pi}{2h} X_2\right) e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\bar{\hat{\nu}}\hat{X}_3},$$

$$u_3^{SR} = \left(\frac{(r_5\xi^{SR} - r_3 i\bar{\nu})}{n\pi/2h} A_+^{(1)SR} \cos\frac{n\pi}{2h} X_2 + iE_+^{(3)SR} \cos\kappa_3 \frac{n\pi}{2h} X_2\right) e^{-\hat{\xi}^{SR}(\hat{X}_1 - a_1)}\, e^{i\bar{\hat{\nu}}\hat{X}_3}, \tag{5.7}$$

where

$$\xi^{SR} = R_{11}\hat{\xi}^{SR} - R_{31} i\bar{\hat{\nu}}, \quad i\bar{\nu} = R_{33} i\bar{\hat{\nu}} - R_{13}\hat{\xi}^{SR}, \tag{5.8}$$

and from Eqs.(72) of Ref.12, we know $C_+^{(2)SR}$ and $E_+^{(3)SR}$ in terms of $A_+^{(1)SR}$ and as in Ref.1 we do not write the relations of the corresponding displacement fields for the other regions. We note that $A_+^{(1)SR}$ here corresponds to B_S in Sec.V of Ref.6

6. Acceleration Sensitivity for Zero Offset

In this section it is shown analytically that the normal acceleration sensitivity vanishes when the centers of the mode shape and support rectangle coincide. To this end recourse is had to the plan view shown in Figure 3, in which the unprimed coordinate system from Figure 1 is reproduced along with the dotted coordinate system obtained by means of a rotation of π about X_2. The transformation from the solid to the dotted coordinate system is given by

$$S_{KL} = \begin{Vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & & -1 \end{Vmatrix}. \tag{6.1}$$

The argument presented holds for the case of rectangular supports and, in fact, for any symmetric shape which has symmetric boundary conditions.

When the solid coordinate system is employed the differential equation is given by (4.1) and the boundary conditions are given by (4.2). When the dotted coordinate system is employed, the differential equation is given by

$$\frac{2}{3} h^3 \tilde{\gamma}_{ABCD} w^{(0)}_{2,CDAB} + 2h\rho a_2 = 0 , \qquad (6.2)$$

and the differentiation is with respect to the dotted coordinate system. The boundary conditions are still given by (4.2), but now the + and - locations refer to the opposite edges and that has no influence because the boundary conditions are symmetric[13]. Since A,B,C,D take the values 1 and 3 only and the transformation is given by (6.1), have

$$\tilde{\gamma}_{ABCD} = \gamma_{ABCD} , \qquad (6.3)$$

which shows that Eq.(6.2) is exactly the same as Eq.(4.1). Thus we have shown that the differential equations and boundary conditions are exactly the same in either coordinate system shown in Figure 3. Since the solutions will be exactly the same in each coordinate system and the physical result, i.e., $w_2^{(0)}$, must be independent of the coordinate system employed, we have shown that $w_2^{(0)}$ can contain terms only of the forms $E(X_1)E(X_3)$ and $O(X_1)O(X_3)$, where E means even and O means odd, and cannot contain terms of the forms $E(X_1)O(X_3)$ or $O(X_1)E(X_3)$. This means that the amplitude $\hat{D}_o$ in the solution function in (4.6) and (4.10) serves to cancel the influence of the $E(X_1)O(X_3)$ and $O(X_1)E(X_3)$ terms that arise from the complex exponential notation employed in the sums in (4.6) and (4.10). The application of the transformation in (5.5) to the allowed solution functions shows that terms only of the forms $E(\hat{X}_1)E(\hat{X}_3)$ and $O(\hat{X}_1)O(\hat{X}_3)$ arise. Furthermore, an examination of the perturbation integral (2.9) for the equivalent trapped energy mode shape reveals that biasing terms only of the forms $E(\hat{X}_1)O(\hat{X}_3)$ and $O(\hat{X}_1)E(\hat{X}_3)$ survive. Hence we have shown that the normal acceleration sensitivity of the contoured resonator vanishes when the centers of the mode shape and support rectangle coincide.

7. Small Offset of Centers

In this section we obtain a very useful result for the limiting case of small offset of centers. From Figure 1 it is clear that we may write the transformation

$$X_A = d_A + X'_A , \qquad (7.1)$$

and from (4.3) and (4.10), we may write

$$w_2^{(0)} = \overset{*}{w}_2 + \sum_n D_n e^{i\nu_n X_3} \sum_{p=1}^{4} g_n^{(p)} e^{i\eta_n^{(p)} X_1} , \qquad (7.2)$$

where $\overset{*}{w}_2$ is a polynomial in X_1 and X_3. The substitution of (7.1) in (7.2) enables us to write

$$w_2^{(0)} = w_2^*(d_A, X'_A) + \sum_n \hat{D}_n e^{i\nu_n X'_3} \sum_{p=1}^{4} \hat{g}_n^{(p)} e^{i\eta_n^{(p)} X'_1} , \qquad (7.3)$$

where

$$\hat{D}_n = e^{i\nu_n d_3} D_n , \quad \hat{g}_n^{(p)} = e^{i\eta_n^{(p)} d_1} g_n^{(p)} . \qquad (7.4)$$

Since the perturbation integral in (2.9) vanishes for $d_A = 0$, the substitution of (7.3) with (7.4) into (2.9) and expansion to first order in d_A yields, with the aid of (2.1)

$$\frac{\Delta\omega}{\omega} = C_1 d_1 + C_3 d_3 = C_A d_A , \qquad (7.5)$$

where the C_A are complicated expressions which are not terribly revealing and have been calculated.

Clearly, the C_A defined above denote the components of a resonator vector parameter which is of some interest. The orientations and aspect ratios for which the magnitude of C_A is smallest are those for which the normal acceleration sensitivity is least influenced by the type of error in fabrication considered here. Calculations of the magnitude of C_A for a few cases are presented and discussed in the next section.

8. Calculated Results

From Section 5 we now know g_γ^μ and from Sections 3 and 4 we know $\hat{c}_{L\gamma M\alpha}$ for normal acceleration, i.e., flexure with simple rectangular supports. Hence, we can now evaluate H_μ in Eq.(2.9) for non-zero d_A. Such calculations have been performed using the known values of the second order[14] and third order[15] elastic constants of quartz. Since at the time of the work presented in Ref.1 it was not known that the normal acceleration sensitivity always vanishes when the centers of the support rectangle and mode shape coincide, results are presented in this section for the case of rigid supports correctly analyzed in Ref.1 as well as for the case of simple rectangular supports analyzed in this work.

The calculated normal acceleration sensitivities for a rigidly supported AT-cut plate with an aspect ratio of a/b = 2 and a given support orientation are plotted in Figure 4 as a function of offset in different directions. It can be seen from the figures that all curves go through zero at $d_A = 0$, as they should. For reference the a_1 distance to the point of inflection of the

Gaussian in the $\hat{X}_1$-direction is shown in the figure. Since the direction of the relative displacement $\underset{\sim}{d}$ is unknown, only the most rapidly varying curve is significant. It can be seen that as $|\underset{\sim}{d}|$ gets small, the curves all become straight lines in accordance with Section 7. The figure shows that the sensitivity increases very rapidly with small offset of centers, which clearly demonstrates the importance of care in fabrication[3]. Figures 5 - 7 show similar curves for a rigidly supported SC-cut plate, a simply supported AT-cut plate and a simply supported SC-cut plate, respectively. The calculated values of the components C_1 and C_3, which were defined in Section 7 and are indicative of the influence of a relative displacement $\underset{\sim}{d}$ of the centers on the normal acceleration sensitivity, are plotted in Figure 8 as a function of support orientation for a rigidly supported SC-cut square plate for N = 1, 3 and 5. The corresponding values of the resonator parameter C, which is the magnitude of $\underset{\sim}{C}$, are plotted in Figure 9. The figure shows that the values of C are about the same for N = 3 and N = 5 and are about twice as large for N = 1, but do not vary that much for a given N. The calculated values of C_1 and C_3 are plotted in Figure 10 as a function of support orientation for a simply supported SC-cut plate with an aspect ratio a/b = 0.5 for N = 1, 3 and 5. The corresponding values of the resonator parameter C are plotted in Figure 11, which also shows that C is larger for N = 1 than for N = 3 and N = 5. However, this time C shows a strong minimum at a value of β around 40°. This means that for an aspect ratio of a/b = 0.5 that orientation of the support configuration is significantly insensitive to the type of fabrication imperfection considered in this work, i.e., a relative displacement of the centers of the support configuration and the mode shape.

Acknowledgements

We wish to thank Dr. J. Vig for suggesting that we investigate the influence of a relative displacement of the centers of the mode shape and support rectangle on the acceleration sensitivity.

This work was supported in part by the Army Research Office under Contract No.DAAL03-88-K-0065 and the National Science Foundation under Grant No. MSS-8908188.

References

1. 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)

2. Y.S. Zhou, D.V. Shick and H.F. Tiersten, "An Analysis of the Normal Acceleration Sensitivity of Contoured Quartz Resonators with Simple Rectangular Supports," 1989 Ultrasonics Symposium Proceedings, IEEE Cat. No.89 CH 2791-2, Institute of Electrical and Electronics Engineers, New York, 393 (1989).

3. R.J. Besson and U.R. Peier, "Further Advances on B.V.A. Quartz Resonator," Proceedings of the 34th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, 175 (1980).

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, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).

6. D.S. Stevens, H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Contoured AT-Cut Quartz Crystal Resonators," J. Appl. Phys., 54, 1709 (1983).

7. R.D. Mindlin, "An Introduction to the Mathematical Theory of the Vibration of Elastic Plates," U.S. Army Signal Corps Eng. Lab., Fort Monmouth, New Jersey (1955). Signal Corps Contract DA-36-03956-56772.

8. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).

9. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.13.

10. H.F. Tiersten and B.K. Sinha, "Temperature Dependence of the Resonant Frequency of Electroded Doubly-Rotated Quartz Thickness-Mode Resonators," J. Appl. Phys., 50, 8038 (1979).

11. For more detail see 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," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.

12. 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).

13. The demonstration presented in this section holds for an arbitrary symmetric shape to the resonator provided only that the boundary conditions are symmetric with respect to the center of the mode shape.

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).

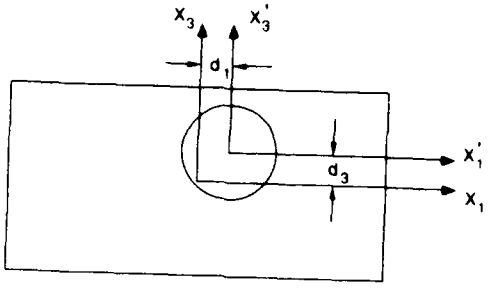


Figure 1 Plan View of Rectangular Quartz Plate Showing Relative Displacement of the Mode Shape and Support Rectangle

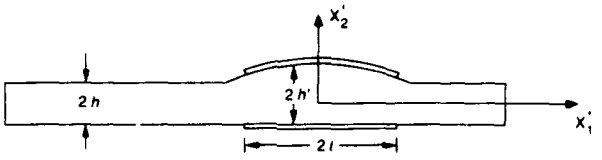


Figure 2 Cross-Section of Plano-Convex Resonator Displaced with Respect to the Supports

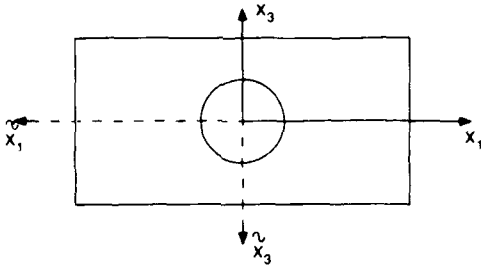


Figure 3 Plan View of Rectangular Quartz Plate with Opposing Planar Coordinate Systems

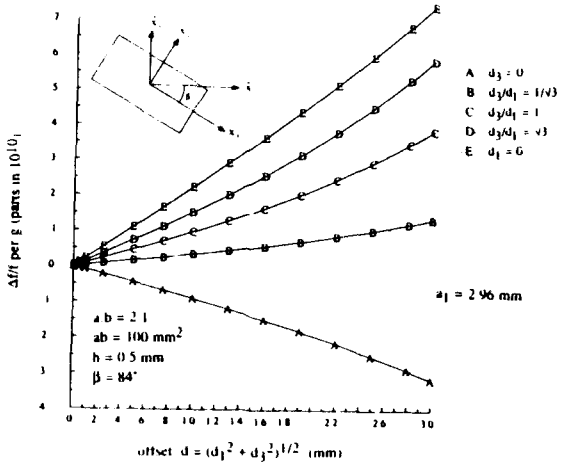


Figure 4 Sensitivity Versus Offset in Different Directions for Fundamental Mode of Rigidly Supported AT-Cut Plate

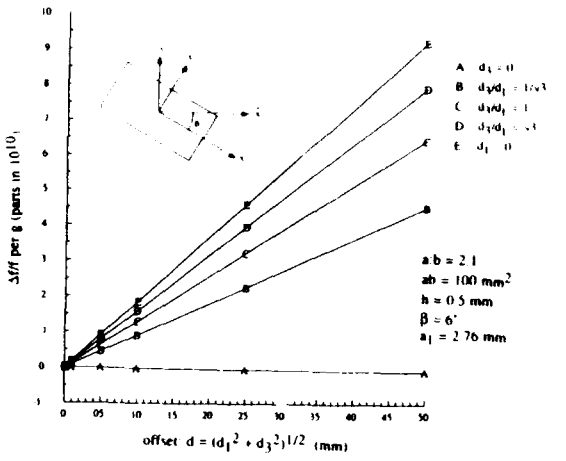


Figure 5 Sensitivity Versus Offset in Different Directions for Fundamental Mode of Rigidly Supported SC-Cut Plate

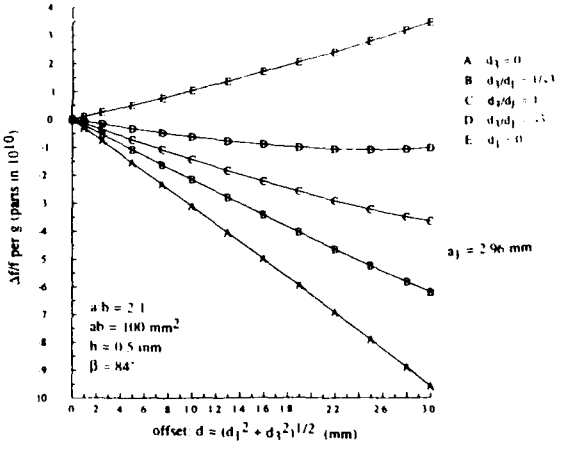


Figure 6 Sensitivity Versus Offset in Different Directions for Fundamental Mode of Simply Supported AT-Cut Plate

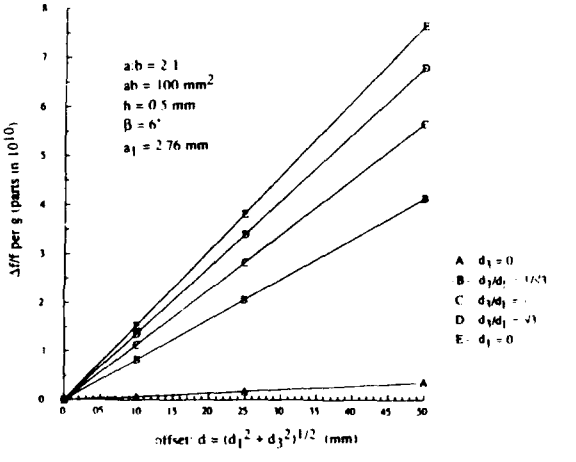


Figure 7 Sensitivity Versus Offset in Different Directions for Fundamental Mode of Simply Supported SC-Cut Plate

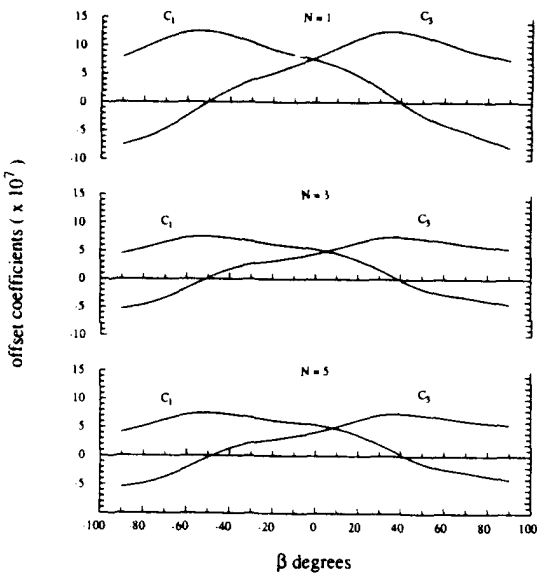


Figure 8 Components of Influence of Offset Vector on Normal Acceleration Sensitivity Versus Orientation for Rigidly Supported SC-Cut Plate

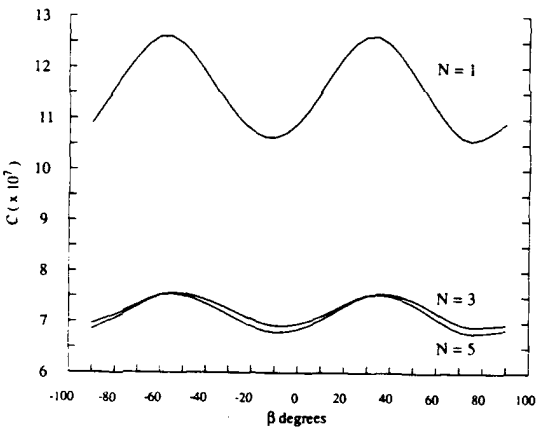


Figure 9 Magnitude of Influence of Offset Vector Obtained from Figure 8

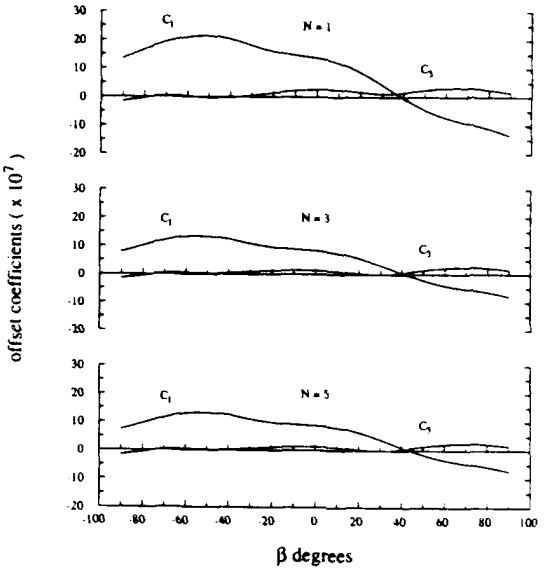


Figure 10 Components of Influence of Offset Vector on Normal Acceleration Sensitivity Versus Orientation for Simply Supported SC-Cut Plate

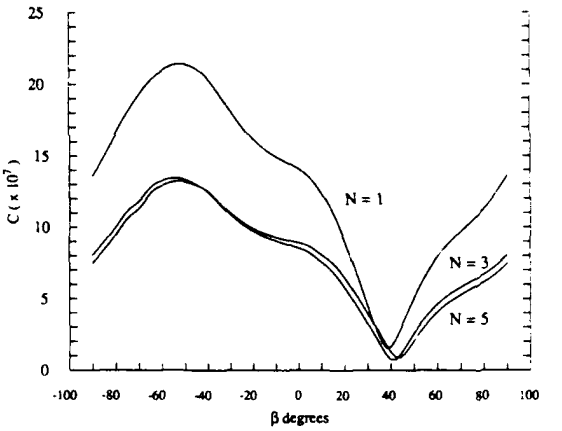


Figure 11 Magnitude of Influence of Offset Vector Obtained from Figure 10

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN ANALYSIS OF THE IN-PLANE ACCELERATION SENSITIVITY OF CONTOURED QUARTZ RESONATORS WITH RECTANGULAR SUPPORTS

H.F. Tiersten and Y.S. Zhou

Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Rensselaer Polytechnic Institute, Troy, NY 12180-3590

Abstract

It is shown that the resultant in-plane acceleration sensitivity of contoured quartz resonators symmetrically supported along rectangular edges always vanishes. This result is essentially a consequence of symmetry and applies to any shape for which the support system is symmetric with respect to the center of the mode shape. When the active plate is supported on one side only, a state of flexure is induced in the resonator which degrades the in-plane sensitivity. This emphasizes the importance of a symmetric support system to achieve low acceleration sensitivities. The extensional and induced in-plane flexural biasing deformations are determined by means of the variational approximation procedure using the variational principles in which all conditions appear as natural conditions. The resulting biasing states are employed in the existing perturbation equation along with the mode shapes of the contoured resonators to calculate the non-vanishing acceleration sensitivities when the resonator is supported on one side only.

1. Introduction

In earlier work[1] on the in-plane acceleration sensitivity of contoured resonators rigidly supported along rectangular edges an error was made in the calculations. The error resulted in a severe overestimate of the in-plane acceleration sensitivity. The error was found while performing the calculations presented in this paper. In this work it is shown that when the centers of the support rectangle and the mode shape coincide and the active resonator region is supported symmetrically on both sides, the in-plane acceleration sensitivity vanishes for any orientation and aspect ratio of the plate. This is shown analytically and is a consequence of the symmetry of the biasing state and the mode shape. The demonstration holds for any support configuration which exhibits the required symmetry with respect to the center of the mode shape. If the active region is supported on one side only, in-plane flexure is induced in the plate and this results in an in-plane acceleration sensitivity of a few parts in 10^{10} per g. This degradation is, of course, eliminated simply by supporting the plate symmetrically on both sides.

When the active plate is supported on one side only, an analysis of the in-plane acceleration sensitivity is performed. The biasing state is determined by decomposing the problem into an extensional and flexural part. First the extensional biasing state is determined by means of our variational approximation procedure which has been used in earlier work[2,3]. The resulting solution for the extensional biasing state defines the associated flexural problem. The flexural biasing state is obtained by means of our variational approximation procedure using the variational principle for anisotropic static flexure, in which all conditions appear as natural conditions[4]. The calculated biasing deformation fields are employed in an existing perturbation equation[5] along with the mode shape for the contoured resonator[6] to calculate the in-plane acceleration sensitivity. Results are presented for a few cases when the active plate is supported on one side only.

2. Preliminary Considerations

For purely elastic nonlinearities the equation for the perturbation in eigenfrequency[5] may be written in the form given in Eq.(2.1) of the previous paper (Ref.7), which we reproduce here for completeness

$$\Delta_\mu = H_\mu/2\omega_\mu, \quad \omega = \omega_\mu - \Delta_\mu, \tag{2.1}$$

where ω_μ and ω are the unperturbed and perturbed eigenfrequencies, respectively. It has further been shown that H_μ may be written in the form given in Eq.(2.10) of Ref.7, which we also reproduce here for completeness

$$H_\mu = -\int_V \hat{c}_{L\gamma M\alpha} g^\mu_{\gamma,L} g^\mu_{\alpha,M}\, dV, \tag{2.2}$$

where $\hat{c}_{L\gamma M\alpha}$ is defined in Eq.(2.4) of Ref.7 and all other quantities are defined in Sec.2 of Ref.7. Since g^μ_α denotes the normalized mode shape in the contoured resonator and from Eq.(2.4) of Ref.7 $\hat{c}_{L\gamma M\alpha}$ depends on the biasing state, H_μ can be

CH2818-3/90/0000-461 $1.00 © 1990 IEEE

evaluated when the mode shape in the resonator and biasing state are known.

Since the plate is 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

$$K_{AB,A} - \rho a_B = 0, \tag{2.3}$$

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

$$K_{AB} = \gamma_{ABCD} E_{CD}, \quad E_{CD} = \frac{1}{2}(w_{C,D} + w_{D,C}), \tag{2.4}$$

and γ_{ABCD} are Voigt's anisotropic plate elastic constants, which are defined in Eq.(3.5) of Ref.7. From the relaxation of the stresses K_{2L}, we have

$$E_W = - C_{WV}^{-1} C_{VS} E_S, \tag{2.5}$$

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

$$\Omega_{13} = \frac{1}{2}(w_{3,1} - w_{1,3}). \tag{2.6}$$

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

$$\Omega_{2A} = 0, \tag{2.7}$$

and now with (2.4) - (2.7) and

$$w_{K,L} = E_{KL} + \Omega_{LK}, \tag{2.8}$$

we have the three-dimensional displacement gradients for the extensional portion of the biasing deformation field, which is required for the evaluation of $\hat{c}_{L\gamma M\sigma}$ of Ref.7.

3. Decomposition into Extensional and Flexural Biases

A plan view and cross-section of the plate when supported on one side only are shown in Fig.1 along with the coordinate system. Since the plate is thin, the equations of anisotropic plane stress introduced in Sec.2 apply in the extensional portion of the problem, for which we take the resisting forces due to the frits to lie in the center-plane of the plate rather than below the plate as shown in the cross-section in Fig.1. As a consequence of this the plan view shown in Fig.2 refers to the extensional portion of the problem only and we must add a flexural portion to the problem resulting from the actual location of the resisting forces in the frit below the plate rather than at the center-plane of the plate. This is done on the cross-sections in Fig.2, in which the actual state on the left is decomposed into the extensional state on the top right and the flexural state on the bottom right. As noted in the Introduction, clearly the extensional problem is treated first and thus defines the associated flexural problem. As a consequence of the geometry of the supporting frit, we now take the displacement w_S in the tangential direction of the supporting frit to vanish.

On account of the stubbiness of the cross-section of a frit, we model its resistance to the normal displacement w_n of the center-plane of the plate as a simple variable shear strip, which yields

$$K_{(n)(n)} = \mp k w_n, \tag{3.1}$$

where[3]

$$k = 2\mu t_f / 3h^2, \tag{3.2}$$

in which μ is the appropriate shear modulus depending on the direction in which the frit runs, and we have ignored the deformation in the frit to this approximation because $h_f \ll h$. Furthermore, in accordance with the discussion in the previous paragraph concerning the tangential displacement along the frit, we have

$$w_s = 0, \tag{3.3}$$

at the frit.

As already noted, in view of the location of the resisting forces in the frit below the center-plane of the plate, we have the left cross-section shown in Fig.2, which results in the extensional problem shown at the top right that has already been discussed and the associated flexural problem is shown at the bottom right. In the flexural problem we make the conservative assumption that the bottom support plate carries none of the applied moment since we do not consider the manner in which the bottom plate is supported, which determines the flexibility. Clearly the couple associated with the flexural problem is given by

$$M = F(h + h_f). \tag{3.4}$$

4. Extensional Biasing Deformation

The extensional problem has been defined in Sec.3 with the aid of Fig.2. Clearly the equations of anisotropic plane stress, i.e., Eqs.(2.3) and (2.4), apply. From Eqs.(3.1) and (3.3) we have the boundary conditions

$$w_3 = 0, \quad K_{11} \pm k_1 w_1 = 0 \quad \text{at } X_1 = \pm a, \quad |X_3| \leq b,$$

$$w_1 = 0,\ K_{33} \pm k_3 w_3 = 0 \quad \text{at } X_3 = \pm b\ ,\ |X_3| \le \hat{a}\ . \tag{4.1}$$

We now transform the inhomogeneities from the differential equations, which consist of (2.3) with (2.4), into the boundary conditions (4.1) by writing

$$w_A = \hat{w}_A + \overset{*}{w}_A\ , \tag{4.2}$$

where

$$\overset{*}{w}_1 = A_1(X_3^2 - b^2),\ \overset{*}{w}_3 = A_3(X_3^2 - b^2) - \rho a_3 b/k_3. \tag{4.3}$$

The substitution of (4.2) with (4.3) into (2.3) with (2.4) yields

$$\begin{aligned}
&\gamma_{11}\hat{w}_{1,11} + 2\gamma_{15}\hat{w}_{1,31} + \gamma_{55}\hat{w}_{1,33} + \gamma_{35}\hat{w}_{3,33} \\
&\quad + (\gamma_{13} + \gamma_{55})\hat{w}_{3,31} + \gamma_{15}\hat{w}_{3,11} = 0\ , \\
&\gamma_{15}\hat{w}_{1,11} + (\gamma_{13} + \gamma_{55})\hat{w}_{1,31} + \gamma_{35}\hat{w}_{1,33} \\
&\quad + \gamma_{33}\hat{w}_{3,33} + 2\gamma_{35}\hat{w}_{3,31} + \gamma_{55}\hat{w}_{3,11} = 0,
\end{aligned} \tag{4.4}$$

since A_1 and A_3 are selected as

$$A_1 = \frac{\rho}{2}\ \frac{(\gamma_{33}a_1 - \gamma_{35}a_3)}{(\gamma_{33}\gamma_{55} - \gamma_{35}^2)}\ ,$$

$$A_3 = -\frac{\rho}{2}\ \frac{(\gamma_{35}a_1 - \gamma_{55}a_3)}{(\gamma_{33}\gamma_{55} - \gamma_{35}^2)}\ . \tag{4.5}$$

The further substitution of (4.2) with (4.3) and (2.4) into (4.1) yields the transformed edge conditions, which for brevity we do not bother to write here[8].

As a solution of (4.4) we write

$$\hat{w}_A = \tilde{w}_A + \sum_m \beta_{Am}\, e^{i\eta_m X_3} e^{i\nu_m X_1}\ , \tag{4.6}$$

where the $\tilde{w}_A$ are polynomials in X_1 and X_3, which satisfy (4.4) and certain boundary conditions and are discussed a little later on for convenience, and $\nu_m = m\pi/2\sigma a$ and σ is chosen to be an irrational number (here 2π) in order that neither $\cos \nu_m a$ or $\sin \nu_m a$ vanish. The substitution of (4.6) into (4.4) yields

$$\begin{aligned}
&[\gamma_{11}\nu_m^2 + 2\gamma_{15}\nu_m\eta_m + \gamma_{55}\eta_m^2]\beta_{1m} + [\gamma_{35}\eta_m^2 + \\
&\quad + (\gamma_{13} + \gamma_{55})\eta_m\nu_m + \gamma_{15}\nu_m^2]\ \beta_{3m} = 0, \\
&[\gamma_{15}\nu_m^2 + (\gamma_{13} + \gamma_{55})\eta_m\nu_m + \gamma_{35}\eta_m^2]\ \beta_{1m} + \\
&\quad + [\gamma_{33}\eta_m^2 + 2\gamma_{35}\eta_m\nu_m + \gamma_{55}\nu_m^2]\ \beta_{3m} = 0\ ,
\end{aligned} \tag{4.7}$$

for each m. Equations (4.7) 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 $\hat{\eta}_m^{(p)}$ (p = 1, 2, 3, 4) for each ν_m with four sets of amplitude ratios $\beta_{1m}^{(p)} : \beta_{3m}^{(p)}$. By scaling this is independent of m. Hence, as a solution of the boundary value problem we may write

$$\hat{w}_A = \hat{\tilde{w}}_A + \sum_m \sum_{p=1}^{4} D_p^{(m)} \beta_A^{(p)} e^{i\eta_m^{(p)} X_3}\, e^{i\nu_m X_1}\ , \tag{4.8}$$

where the $D_p^{(m)}$ are amplitude coefficients still to be determined. By satisfying the three transformed homogeneous edge conditions $\hat{w}_1 = 0$ at $X_3 = b$ and $\hat{K}_{33} \pm k_3\hat{w}_3 = 0$ at $X_3 = \pm b$ for each m, we obtain

$$D_p^{(m)} = g_p^{(m)} D^{(m)}, \tag{4.9}$$

where the $g_p^{(m)}$ are obtained by satisfying the three edge conditions[8]. The polynomials $\hat{w}_A$ are taken in the form[9]

$$\begin{aligned}
\hat{w}_1 &= D_o(X_3 - b) - \bar{D}_o(X_3 - b)[C_o C_1 X_1 + C_2(X_3 + b)]\ , \\
\hat{w}_3 &= -\frac{\gamma_{35}}{\gamma_{33} + k_3 b}\ D_o X_3 + \bar{D}_o\Big[C_o\Big(C_3 + \\
&\quad X_3\Big(\frac{\gamma_{35}}{\gamma_{55}}\ X_1 - \frac{\gamma_{13}}{\gamma_{55}}\ b\Big)\Big) + C_4(X_3^2 - b^2)\Big]\ ,
\end{aligned} \tag{4.10}$$

which satisfy the same three homogeneous edge conditions as well as (4.4). The C_r in (4.10) depend on the material coefficients and the geometry and are too lengthy to present here[8], and D_o and $\bar{D}_o$ are amplitude coefficients still to be determined. The substitution of (4.9) into (4.8) enables us to write

$$\hat{w}_A = \hat{\tilde{w}}_A + \sum_m \Big(\sum_{p=1}^{4} \beta_A^{(p)} g_p^{(m)} e^{i\eta_m^{(p)} X_3}\Big)\, D^{(m)} e^{i\nu_m X_1}\ . \tag{4.11}$$

The solution functions in (4.11) satisfy the differential equations (4.4) and the above mentioned three transformed homogeneous edge conditions, but they do not satisfy the remaining boundary conditions. We satisfy these latter conditions by means of a variational approximation procedure. Since the solution functions satisfy the differential equations exactly and one edge condition is of constraint type and the other is of natural type, all that remains of the variational principle in which all conditions are unconstrained[10] is

$$\int_{S_C} w_{(s)}\delta K_{(n)(s)}\, dS - \int_{S_N} (k_{(n)} w_{(n)} \pm K_{(n)(n)})\ \delta w_{(n)}\, dS = 0\ . \tag{4.12}$$

Substituting from (4.2) into (4.12) and utilizing the fact that certain edge conditions have been satisfied exactly, we obtain

$$-\int_{-a}^{a} \hat{w}_1 \delta\hat{K}_{13}\Big|_{X_3=-b} dX_1 + \int_{-b}^{b} (\hat{w}_3 + \overset{*}{w}_3)\delta\hat{K}_{13}\Big]_{X_1=-a}^{X_1=a} dX_3 - \int_{-b}^{b} (k_1\hat{w}_1 + \hat{K}_{11} + k_1\overset{*}{w}_1 + K^*_{11})\delta\hat{w}_1\Big|_{X_1=a} dX_3 - \int_{-b}^{b} (k_1\hat{w}_1 - \hat{K}_{11} + k_1\overset{*}{w}_1 - K^*_{11})\delta\hat{w}_1\Big|_{X_1=-a} dX_3 = 0, \quad (4.13)$$

where

$$K^*_{11} = 2(\gamma_{13}A_3 + \gamma_{15}A_1)X_3, \quad (4.14)$$

and the notation $\Big[\int_{-1}^{X_1=a} \Big]_{X_1=-a}$ means the quantity evaluated at a minus the quantity evaluated at -a and the $\delta\hat{K}_{AB}$ are obtained from (2.4) with $\hat{w}_A$ instead of w_A since the A_B are fixed and the variation of prescribed quantities vanishes. Substituting from (4.2), (4.3), (4.10) and (4.11) into (4.13) and performing the integrations, we obtain

$$\sum_m \sum_n a_{mn} D^{(m)} \delta D^{(n)} + \sum_n b_n \delta D^{(n)} = 0,$$
$$m,n = 0, \bar{0}, \pm 1, \pm 2, \ldots \pm N, \quad (4.15)$$

where 0 and $\bar{0}$ refer to the amplitudes D_o and $\bar{D}_o$, respectively. The expressions for a_{mn} and b_n are too lengthy to present and provide no additional understanding. Since the variations $\delta D^{(n)}$ are arbitrary, we obtain

$$\sum_n a_{mn} D^{(m)} = -b_n, \quad m,n = 0, \bar{0}, \pm 1, \pm 2, \ldots \pm N, \quad (4.16)$$

which constitute 2N + 2 inhomogeneous linear algebraic equations in the 2N + 2 unknown $D^{(m)}$, the inversion of which gives the approximate solution. We consistently obtain good convergence for N < 14.

5. Flexural Biasing Deformation

When the plate is supported on one side only, as shown in Figs.1 and 2 and discussed in Sec.3, flexure is induced in the resonator. From Eqs.(3.1), (3.3), (3.4) and $(3.7)_3$ of the previous paper (Ref.7), we obtain the flexural differential equation for the problem here in the form

$$\frac{2}{3} h^3 \gamma_{ABCD} w^{(0)}_{2,CDAB} = 0. \quad (5.1)$$

From the discussion in Sec.3 it is clear that the boundary conditions are given by

$$w^{(0)}_2 = 0, \; K^{(1)}_{11} = \pm \bar{m}_1 \text{ at } X_1 = \pm a,$$
$$w^{(0)}_2 = 0, \; K^{(1)}_{33} = \pm \bar{m}_3 \text{ at } X_3 = \pm b, \quad (5.2)$$

where

$$\bar{m}_1 = 2h^2 K_{11}, \; \bar{m}_3 = 2h^2 K_{33}, \quad (5.3)$$

and K_{11} and K_{33} are obtained from the extensional biasing state at $X_1 = a$ and $X_3 = b$, respectively

As in Sec.4 of Ref.7, the problem defined by (5.1) and (5.2) cannot be solved exactly. Consequently, a variational approximation procedure is employed, as in Ref.7. To this end we first transform the inhomogeneities at $X_1 = \pm a$ into the edge conditions at $X_3 = \pm b$ by writing

$$w^{(0)}_2 = \overset{*}{w}{}^{(0)}_2 + \hat{w}^{(0)}_2, \quad \overset{*}{w}{}^{(0)}_2 = -\frac{\bar{m}_1 X_1}{4h^3 \gamma_{11} a}(X_1^2 - a^2), \quad (5.4)$$

which when substituted into (5.1) yields Eq.(4.4) of Ref.7. Since as in Sec.4 of Ref.7 there are four roots of the homogeneous differential equations, by satisfying $\hat{w}^{(2)}_2 = 0$ at $X_1 = +a$ and $\hat{K}^{(1)}_{11} = 0$ at $X_1 = \pm a$, the solution function can be written in the form

$$\hat{w}^{(0)}_2 = \sum_n \Big(\sum_{p=1}^{4} g^{(n)}_p e^{i\eta^{(p)}_n X_1} \Big) D^{(n)} e^{i\nu_n X_3}, \quad (5.5)$$

where ν_n is chosen and the $g^{(n)}_p$ are found in the same way as in Sec.4 of Ref.7. For general anisotropy this problem must be solved variationally using the variational condition in (4.11) of Ref.7, which results in the linear algebra

$$\sum_n a_{mn} D^{(n)} = -b_n, \quad m,n = \pm 1, \pm 2, \ldots \pm N, \quad (5.6)$$

where a_{mn} and b_n are too lengthy to present and provide no additional understanding. Equations (5.6) constitute 2N equations in the 2N unknowns $D^{(n)}$, the inversion of which gives the approximate solution. We consistently obtain good convergence for N < 14.

It should be clear that this in-plane flexural biasing state can be completely eliminated simply by supporting the plate symmetrically on both sides.

6. 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.3. 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[6]

$$u_{1n} = \sin \frac{n\pi X_2}{2h} u_{noo} e^{i\omega_{noo} t}, \quad n = 1,3,5, \ldots, \tag{6.1}$$

where

$$u_{noo} = e^{-\alpha_n \frac{X_1'^2}{2}} e^{-\beta_n \frac{X_3'^2}{2}}, \tag{6.2}$$

in which

$$\alpha_n^2 = \frac{n^2\pi^2 \hat{c}(1)}{8Rh_o^3 M_n'}, \quad \beta_n^2 = \frac{n^2\pi^2 \hat{c}(1)}{8Rh_o^3 P_n'}, \tag{6.3}$$

and M_n' and P_n' are involved lengthy expressions that are defined in Sec.II of Ref.6 along with X_1' and 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 R denotes the radius of the spherical contour. The eigenfrequencies corresponding to the eigensolution for the harmonic modes are given by

$$\omega_{noo}^2 = \frac{n^2\pi^2 \hat{c}(1)}{4h_o^2 \rho}\left[1 + \frac{1}{n\pi}\left(\frac{2h_o}{R}\right)^{1/2} \cdot \left(\sqrt{\frac{M_n'}{\hat{c}(1)}} + \sqrt{\frac{P_n'}{\hat{c}(1)}}\right)\right], \tag{6.4}$$

where $\hat{c}^{(1)}$ is defined in Eq.(78) of Ref.6.

In the in-plane case being considered in this work only the dominant displacement field in (6.1) is needed. Consequently, the equivalent trapped energy mode employed in Ref.7 is not required here.

7. Acceleration Sensitivity for Symmetric Support Structure

In this section it is shown analytically that the in-plane acceleration sensitivity vanishes when the support configuration is symmetric with respect to the center of the mode shape. To this end recourse is had to the plan view shown in Fig.4, in which the solid coordinate system from Fig.1 is reproduced along with the dotted coordinate system obtained by means of a rotation of π about X_2. The transformation from the solid to the dotted coordinate system is given by

$$S_{KL} = \begin{Vmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{Vmatrix}. \tag{7.1}$$

The argument presented holds for the case of rectangular supports and, in fact, for any symmetric shape with symmetric boundary conditions.

When the solid coordinate system is employed from (2.3) and (2.4) the differential equations are given by

$$\gamma_{ABCD} w_{C,DA} - \rho a_B = 0, \tag{7.2}$$

and the boundary conditions are given by (4.1) When the dotted coordinate system is employed, the differential equations are given by

$$\hat{\gamma}_{ABCD} \hat{w}_{C,DA} - \rho \hat{a}_B = 0, \tag{7.3}$$

and the differentiation is with respect to the dotted coordinate system. The boundary conditions are still given by (4.1), but now the + and - locations refer to the opposite edges and that has no influence because the boundary conditions are symmetric[11]. Since A, B, C, D take the values 1 and 3 only and the transformation is given by (7.1), we have

$$\hat{\gamma}_{ABCD} = \gamma_{ABCD}. \tag{7.4}$$

If we define $\bar{w}_C = -\hat{w}_C$ and note that $\hat{a}_B = -a_B$, it is clear that (7.3) can be written in the form

$$\gamma_{ABCD} \bar{w}_{C,DA} - \rho a_B = 0, \tag{7.5}$$

which is the same as Eq.(7.2). Thus we have shown that the differential equations and boundary conditions are exactly the same in either coordinate system shown in Fig.3. Hence the w_A and $\hat{w}_A$ can contain terms only of the forms $E(X_1)\,E(X_3)$ and $O(X_1)\,O(X_3)$ or $E(\hat{X}_1)\,E(\hat{X}_3)$ and $O(\hat{X}_1)\,O(\hat{X}_3)$, respectively, because the physical result must be independent of the coordinate system[12]. Furthermore, an examination of the perturbation integral (2.2) shows that the in-plane acceleration sensitivity vanishes for all terms of the form EE and OO because of the symmetry of the mode shape. Hence, we have shown that the in-plane acceleration sensitivity of the contoured resonator vanishes when the support configuration is symmetric.

8. Calculated Results

From Sec.6 we now know g_γ^μ and from Secs.2,4 and 5 and Secs.2 and 3 of Ref.7 we know $\hat{c}_{L\gamma M\alpha}$ for in-plane acceleration when the plate is supported on one side only. Hence, we can now evaluate H_μ in Eq.(2.2). When the plate is supported symmetrically on both sides, the analysis in Sec.7 shows that H_μ always vanishes. Consequently, calculations have been performed using the known values of the second order[13] and third order[14] elastic constants of quartz when the plate is supported on one side only. The calculated components of the acceleration sensitivity for in-plane flexure of an SC-cut square plate are plotted in Fig.4 as a function of support orientation. The corresponding magnitude of the resultant in-plane acceleration sensitivity is plotted in Fig.6. The figure shows

that the in-plane sensitivity remains in the vicinity of a few parts in 10^{10} per g. However, as indicated earlier, the in-plane acceleration sensitivity can readily be made to vanish simply by supporting the plate symmetrically on both sides. This dramatic increase in the in-plane acceleration sensitivity with an asymmetric support system underscores the importance of a symmetric support system along with care in fabrication to achieve consistently low acceleration sensitivities[15].

Acknowledgements

This work was supported in part by the Army Research Office under Contract No.DAAL03-88-K-0065 and the National Science Foundation under Grant No. MSS-8908188.

References

1. D.V. Shick and H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," Proceedings of the 43rd 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.89CH2690-6, 405 (1989).
2. D.V. Shick, Y.S. Zhou and H.F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Quartz Surface-Wave Resonators Rigidly Supported Along the Edges," J. Appl. Phys., 65, 35 (1988).
3. H.F. Tiersten and D.V. Shick, "On the In-Plane Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators with Interior Rectangular Supports," J. Appl. Phys., 67, 2554 (1990).
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, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
6. 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).
7. 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," Proceedings of the 44th Annual Symposium on Frequency Control (these proceedings).
8. For more detail see H.F. Tiersten and Y.S. Zhou, "On the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Supported Along Rectangular Edges," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.
9. The inclusion of the polynomials in the solution function increases the rate of convergence by a factor of 2.
10. H.F. Tiersten, Linear Piezoelectric Plate Vibrations (Plenum, New York, 1969), Chap.6, Sec.4, Eq.(6.44).
11. The demonstration presented in this section holds for an arbitrary symmetric shape to the resonator provided only that the boundary conditions are symmetric with respect to the center of the mode shape.
12. This means that the amplitude D_o in the solution functions in (4.10) and (4.11) serves to cancel the influence of the $E(X_1)$ $O(X_3)$ and $O(X_1)$ $E(X_3)$ terms that arise from the complex exponentials employed in the sum in (4.11).
13. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 100, 1060 (1958).
14. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).
15. R.J. Besson and U.R. Peier, "Further Advances on B.V.A. Quartz Resonator," Proceedings of the 34th Annual Symposium on Frequency Control, U.S. Army Electronics Research and Development Command, Fort Monmouth, New Jersey, 175 (1980).

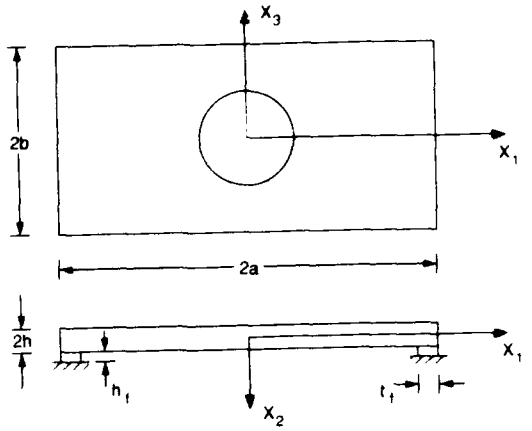


Figure 1 Plan View and Cross-Section of Rectangular Quartz Plate Supported Along Rectangular Edges

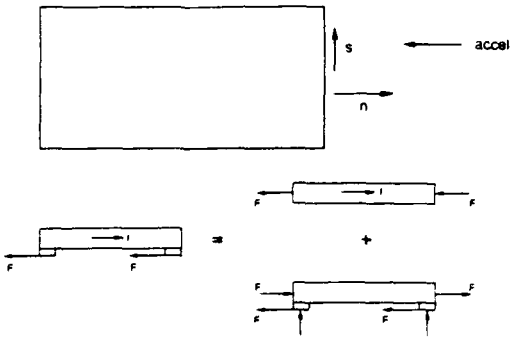


Figure 2 Plan View and Cross-Sections of Plate for Decomposition into Extensional and Flexural Biasing States

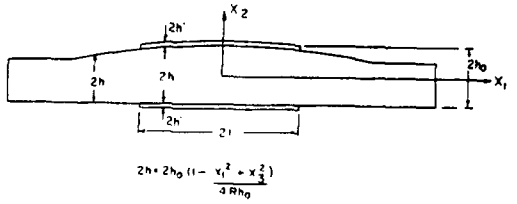


Figure 3 Cross-Section of the Plano-Convex Resonator

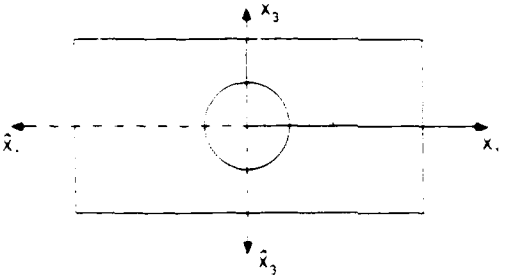


Figure 4 Plan View of Rectangular Quartz Plate with Opposing Planar Coordinate Systems

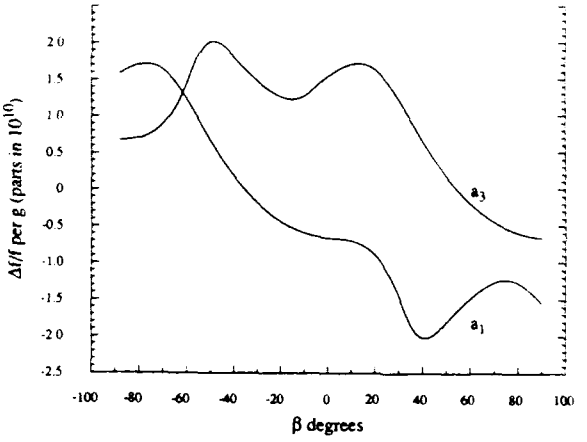


Figure 5 Calculated In-Plane Acceleration Sensitivity Versus Orientation of Support Rectangle for SC-Cut Plate Supported on One Side Only

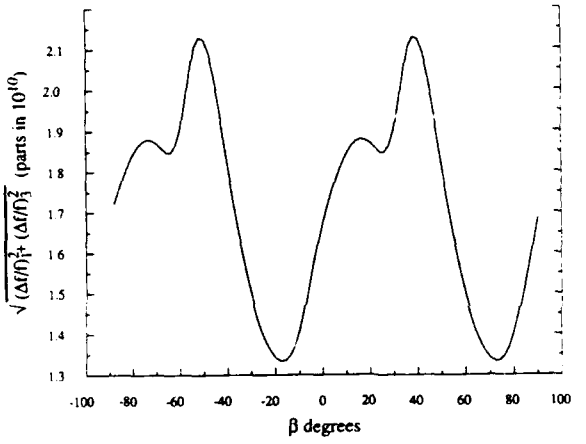


Figure 6 Resultant In-Plane Acceleration Sensitivity Obtained from Figure 4

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

Acceleration Sensitivity of Crystal Resonators Affected by the Mass and Location of Electrodes

by

P.C.Y. Lee and X. Guo

Department of Civil Engineering & Operations Research
Princeton University
Princeton, N.J. 08544

Summary

Predominant thickness-shear frequencies and modes of a crystal plate with electrodes of arbitrary shape and mass distribution are obtained by a finite element method based on Mindlin's first-order equations with platings. These frequencies and modes are, in turn, employed in a purturbation method for computing the acceleration sensitivity of crystal resonators with electrodes.

Computations are made for a square AT-cut quartz plate which is supported by a four-point mount and coated with identified square and uniform electrodes on the upper and lower faces of the plate. To study the effect of uneven distribution of electrode mass, acceleration sensitivities are calculated when a small mass is added at various locations near the edges of the square electrodes. It is found that the percent increase of the acceleration sensitivity of the resonator with a small added mass to that of the resonator without added mass ranges from 3.8% to 541.7%, depending on the location of the small mass placed at the edges of the electrodes.

1. Introduction

Acceleration sensitivity of crystal resonators depends on a number of resonator parameters. In our previous papers, effects of support numbers, orientation of the support configurations, support structures, plate thickness, and plate orientations on the acceleration sensitivity of unelectroded quartz resonators were studied.[1-3]

When the resonators which are coated with electrodes are subjected to accelerations, the additional stresses developed in the crystal plate due to the body force exerted on the electrodes were accommodated in a set of six plate equations for crystal plates with electrodes. These equations were employed for calculating the initial stresses and then the acceleration sensitivity.[4] It is found that the percent increase of the acceleration sensitivity due to the additional stresses is proportional, essentially, to the ratio of the mass of the electrodes to the mass of the plate. For usual applications in resonators, the mass ratio is about 0.5% or less. Hence, this effect on the increase of the acceleration sensitivity is very small.

Electrode mass also has the mechanical effects in (1) reducing the values of the thickness-shear resonances and (2) changing the mode shapes and trapping the vibrational energy in the plated area when the plate is excited between the cut-off frequency of an infinite plate with electrodes and that of an infinite plate without electrodes.[5] In the present paper, the frequencies and mode shapes at the predominant thickness-shear resonances are calculated for crystal plates with electrodes of arbitrary shape and mass distribution by a finite element method based on Mindlin's first-order plate equations. The thickness-shear frequencies and mode shapes are employed, in turn, in a perturbation method to compute the acceleration sensitivity of resonators coated with electrodes.

For systematic study of uneven electrode-mass distribution on the acceleration sensitivity, we first consider a square AT-cut plate which is coated with identical upper and lower electrodes of square shape and uniform mass ratio and is supported with a four-point mount. Then a small additional electrode-mass is added to a small area at the edges of the uniform and square electrodes. From the computed results it is found that the percent increase of the acceleration sensitivity of the resonator with a small added mass to that of the resonator without added mass ranges from 3.8% to 541% depending on the location of the small mass placed at the edges of the electrodes.

Based on the result of present investigation we conclude that the uneven distribution of electrode mass may be one of the dominating causes of the wide variations of acceleration sensitivity in measured values[6,7].

2. Equations of Incremental Vibrations of Plated Crystal Plates under Initial Stresses

Consider a crystal plate with thickness $2b$ and mass density ρ. It is plated with identical upper and lower platings of thickness $2b'$ and density ρ'.

CH2818-3/90/0000-468 $1.00 © 1990 IEEE

Governing equations of incremental vibrations of unplated crystal plate under initial stresses are given in previous papers[8,9]. These equations are modified slightly for the plated crystal plates as follows.

Stress Equations of Motion

$$[\beta_{ik}^{(0)} + t_{jk}^{(0)} + \beta_{ik}^{(1)} t_{jk}^{(1)}]_{,j} + q_i^{(0)} + \rho b_i^{(0)} = 2b\rho(1+R)\ddot{u}_j^{(0)},$$

$$[\beta_{ik}^{(0)} t_{jk}^{(1)}]_{,j} - (\beta_{ik}^{(0)} t_{2k}^{(0)} + \beta_{ik}^{(1)} t_{2k}^{(1)}) + q_i^{(1)} + \rho b_i^{(1)} = \frac{2b^3}{3}\rho(1+2R)\ddot{u}_j^{(1)}, \tag{1}$$

where

$$R = 2b'\rho'/b\rho \tag{2}$$

is the ratio of mass of electrodes to the mass of the plate per unit area,

$$\beta_{ik}^{(0)} = \delta_{ik} + \cup_{i,k}^{(0)} + \delta_{2k}\cup_i^{(1)},$$
$$\beta_{ik}^{(1)} = \cup_{i,k}^{(1)} + 2\delta_{2i}\delta_{2k}\cup_2^{(2)}, \tag{3}$$

are the zero-and first-order initial deformation gradients, $u_j^{(n)}, t_{ij}^{(n)}, q_i^{(n)}$, and $b_i^{(n)}$, n=0,1, are the nth-order incremental displacements, stress tensors, face-tractions, and body forces, respectively, and $\cup_i^{(n)}$ are the nth-order initial displacements.

Stress-Strain Relations

$$t_p^{(0)} = 2bK_{(p)}K_{(q)}C_{pq}^{(0)}\epsilon_q^{(0)} + \frac{2b^3}{3}K_{(p)}C_{pq}^{(1)}\epsilon_q^{(1)},$$
$$t_p^{(1)} = \frac{2b^3}{3}K_{(p)}\hat{C}_{pq}^{(1)}\epsilon_q^{(0)} + \frac{2b^3}{3}\hat{C}_{pq}^{(0)}\epsilon_q^{(1)}, \tag{4}$$

where

$$K_{(p)} = \pi/\sqrt{12} \text{ for } p = 1,3,5 \text{ and } = 0 \text{ for } p = 2,4,6,$$
$$C_{pq}^{(0)} = C_{pq} + C_{pqr}E_r^{(0)}, C_{pqr}^{(1)} = C_{pqr}E_r^{(1)}, \tag{5}$$

and $\epsilon_q^{(0)}, \epsilon_q^{(1)}$ are the incremental strains and $E_r^{(0)}$ and $E_r^{(1)}$ are the initial strains.

Strain-displacement Relations

$$\epsilon_{ij}^{(0)} = \frac{1}{2}[\beta_{kj}^{(0)}(u_{k,j}^{(0)} + \delta_{2i}u_k^{(1)}) + \beta_{ki}^{(0)}(u_{k,j}^{(0)} + \delta_{2j}u_k^{(1)})],$$
$$\epsilon_{ij}^{(1)} = \frac{1}{2}[\beta_{kj}^{(0)}u_{k,i}^{(0)} + \beta_{ki}^{(0)}u_{k,j}^{(0)} + \beta_{kj}^{(1)}(u_{k,i}^{(0)} + \delta_{2i}u_k^{(1)}) + \beta_{ki}^{(1)}(u_{k,j}^{(0)} + \delta_{2j}u_k^{(1)})]. \tag{6}$$

We note that in case of no initial stresses (1) reduce to the Mindlin's six first-order equations for plates with electrodes.

The factors (1+R) and (1+2R) associated to $\ddot{u}_j^{(0)}$ and $\ddot{u}_j^{(1)}$ in (1) are obtained in a manner similar to that given in Ref. 5, and they are identical to those in the two-dimensional equations derived by the trigonomitrical series expansion[11].

Closed form solutions of thickness-shear and flexural vibrations of an infinite AT-cut quartz plate with strip electrodes were obtained by Mindlin and Lee.[5] Thickness-shear and flexural vibrations of rectangular AT-cut plates with uniform and patched electrodes were studied by Lee, Zee and Brebbia by the finite element method for numerical solutions.[11]

In the present paper, Mindlin's first-order equations for plates with electrodes (which are obtained from (1) by letting initial displacements equal to zero) are solved numerically for crystal plates with triclinic symmetry and with electrodes of arbitrary shape and mass distribution by extending the finite element code developed previously by Lee and Tang.[12] The thickness-shear resonces and modes so obtained shall be employed in a perturbation method to compute the acceleration sensitivity of crystal resonators coated with electrodes.

3. Acceleration Sensitivity of Resonators with Electrodes

When a resonator is subject to a steady acceleration of magnitude 1g in the x_i direction (i = 1,2,3), the initial strains and initial deformation gradients due to the body force are calculated by a finite element method based on the six coupled equations for crystal plates with electrodes.[4] By inserting these initial fields into the frequency equation of the fundamental thickness vibrations of the crystal plate and employing the resonance frequencies and modes of the free incremental vibrations as unperturbed solutions, frequency changes due to the accelerations are calculated by a perturbation method and acceleration sensitivity Γ are computed according to the defintion

$$\Gamma = (\Gamma_1^2 + \Gamma_2^2 + \Gamma_3^2)^{1/2} \qquad (per \quad g) \tag{7}$$

where

$$\Gamma_i = (f_i - f_o)/f_o. \tag{8}$$

We note that f_o is the thickness-shear resonance frequency of the electroded plate without being subjected to acceleration, and f_i is the thickness-shear resonance frequency when the plate is subjected to a steady acceleration of magnitude $1g$ in the x_i direction.

For computations, we choose a square AT-cut quartz

plate of length 2ℓ=14.00mm and thickness 2b = 0.28mm. It is coated with identical square electrodes of length 2a=5.6mm on the upper and lower faces and is supported by four metal ribbons at the intersections of the plate edges with the x_1 and x_3 axes. The ribbon is made of molybdenum (E = 27.579 x 10^9dynes/mm^2, ν=0.32) with length h=1.143mm and rectangular cross-section (0.0177mm × 1.524mm). A finite element mesh with 400 elements to represent the square plate is shown in Fig. 1 in which R denotes the mass ratio in the square electroded region and R' indicates the mass ratio in a small hatched area formed by four plate elements.

The thickness-shear frequency Ω, the mode shapes $u_1^{(1)}$ along the x_1 and x_3 axes, and the acceleration sensitivity Γ are calculated for the AT-cut plate without electrodes and they are given in Fig. 2, where

$$\Omega = \omega/\omega_1, \qquad \omega_1 = \frac{\pi}{2b}\left(\frac{c_{66}}{\rho}\right)^{1/2},$$
$$n = \text{number of plate elements}, \tag{9}$$
$$\ell_e = \text{element length}.$$

Computational results for the AT-cut plate with uniform square electrodes (R=0.005) are shown in Fig. 3. By comparing Fig. 3 with Fig. 2, we see that the value of Ω is decreased, $u_1^{(1)}$ is more trapped in the electroded region, and Γ is slightly decreased when the plate is coated with electrodes.

In order to investigate the effect of the uneven distribution of electrode mass on the acceleration sensitivity, additional masses are added to the small regions (R'=0.01) denoted by the numerals 1 to 5 in Fig. 4. Computational results corresponding to the small masses placed at locations 1 to 5 are given in Figs. 5 to 9, respectively. It may be seen from these figures that the unbalanced small mass destroys the symmetry of the mode shapes and hence increases the values of the acceleration sensitivity which ranges from $\Gamma = 1.022 \times 10^{-10}/g$ at locations 3 to $\Gamma = 6.315 \times 10^{-10}/g$ at location 1, depending on the elastic symmetry of the crystal. For AT-cut plate, the effect on the acceleration sensitivity from the unbalanced mass is symmetrical with respect to the x_3-axes.

Computations are also made for the case of two identical masses being placed at locations 1 and 5 and results are shown in Fig. 10. It is seen that when two small masses are placed symmetrically with respect to the center of the plate the mode shape becomes symmetric again and hence the value of Γ reduces back almost to $\Gamma_o = 0.981 \times 10^{10}/g$ which is the acceleration sensitivity of AT-cut plate with uniform square electrodes (without added mass).

Values of Γ_1, Γ, and $\Delta\Gamma/\Gamma_o$, where $\Delta\Gamma = \Gamma - \Gamma_o$, of above discussed cases are listed in Table 1.

From the present computational results we conclude that unevenly distributed electrode masses may be one of the dominating causes for wide variations of measured values of acceleration sensitivities for crystal resonators and placing electrode mass symmetrically with respect to the center of the plate may reduce Γ practically to its lower bound value Γ_o which is the acceleration sensitivity of the plate with perfectly centered and uniform electrodes.

Acknowledgement

This work was supported by the U.S. Army Research Office, Contract No. DAAL 03-87-K-0125.

References

1. P.C.Y. Lee and M.S.H. Tang, *Proc. 41st Ann. Freq. Cont. Symp.*, pp. 277-281, 1987.

2. P.C.Y. Lee and M.S.H. Tang, *Proc. 42nd Ann. Freq. Cont. Symp.*, pp. 14-18, 1988.

3. P.C.Y. Lee and M.S.H. Tang, *Proc. 43rd Ann. Freq. Cont. Symp.*, pp. 413-416, 1989.

4. P.C.Y. Lee and X. Guo, *Proc. 43rd Ann. Freq. Cont. Symp.*, pp. 416-418, 1989.

5. R.D. Mindlin and P.C.Y. Lee, *Int. J. Solids Structures*, vol. 2, pp. 125-139, 1966.

6. R.L. Filler, J.A. Kosinski and J.R. Vig, *Proc. 37th Ann. Freq. Cont. Symp.*, pp. 265-271, 1983.

7. K.K. Tuladhar, IERE(Brit.) *Proc. Int'l Conf. on Freq. Control and Synthesis*, U. Surrey, pp. 67-73, 1987.

8. P.C.Y. Lee, Y.S. Wang, and X. Markenscoff, *J. Acoust. Soc. Am.*, vol. 57, no. 1, pp. 95-105, 1971.

9. P.C.Y. Lee and M.S.H. Tang, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. UFFC-34, no. 6, pp. 659-666, 1987.

10. P.C.Y. Lee, S. Syngellakis, and J.P. Hou, *J. Appl. Phys.* **61** (4), pp. 1249-1262, 1987.

11. P.C.Y. Lee, C. Zee and C.A. Brebbia, *J. Acoust Soc. Am.* **72** (6), pp. 1855-1862, 1982.

12. P.C.Y. Lee and M.S.H. Tang, *Proc. 40th Ann. Freq. Cont. Symp.*, pp. 152-160, 1986.

Table I: Acceleration sensitivities of a square AT-cut quartz plate with various distributions of electrode mass

Mass Ratio R	Locations of added Mass ($R' = 0.01$)	Acceleration Sensitivity (per g)				Percent Inc. $\Delta\Gamma/\Gamma_0$ (%)
		Γ_1	Γ_2	Γ_3	Γ	
0.0	—	0.0	1.06×10^{-11}	1.14×10^{-10}	1.14×10^{-10}	16.3
0.005	—	0.0	-5.51×10^{-12}	9.82×10^{-11}	0.98×10^{-10}	0.0
0.005	Location 1	0.0	3.19×10^{-10}	5.45×10^{-10}	6.31×10^{-10}	541.7
0.005	Location 2	6.5×10^{-12}	1.96×10^{-10}	4.04×10^{-10}	4.49×10^{-10}	356.4
0.005	Location 3	2.9×10^{-11}	-5.68×10^{-12}	9.77×10^{-11}	1.02×10^{-10}	3.8
0.005	Location 4	1.4×10^{-13}	-2.07×10^{-10}	-2.09×10^{-10}	2.95×10^{-10}	199.2
0.005	Location 5	0.0	-3.30×10^{-10}	-3.45×10^{-10}	4.78×10^{-10}	385.5
0.005	Locations 1 & 5	0.0	-5.44×10^{-12}	1.01×10^{-10}	1.01×10^{-10}	2.6

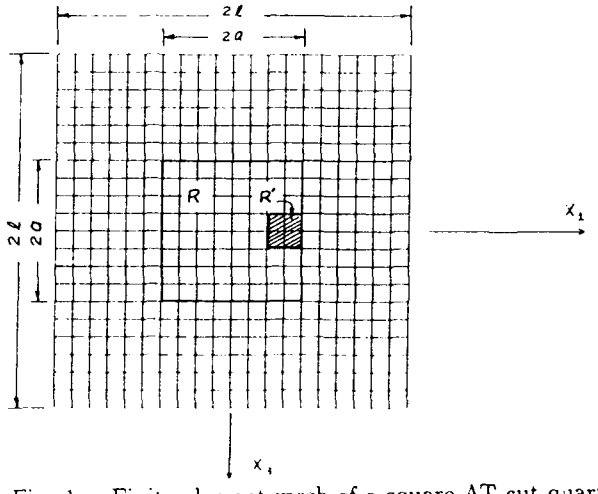


Fig. 1 Finite element mesh of a square AT-cut quartz with square electrodes.

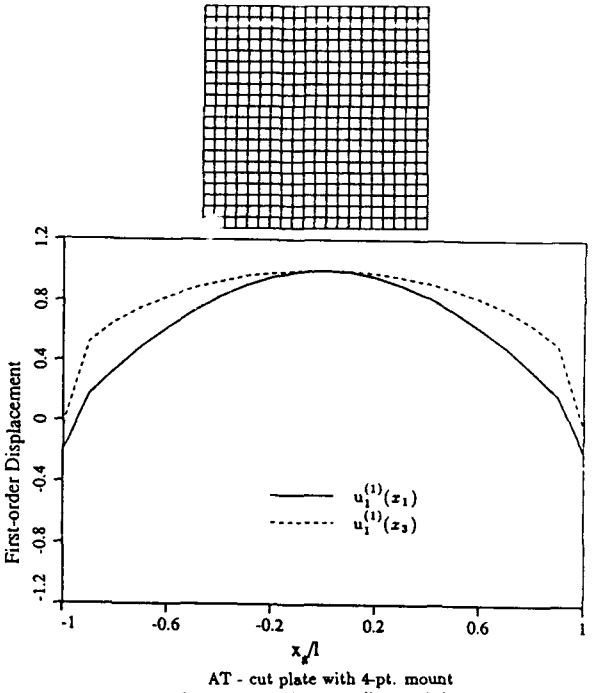


Fig. 2 $u_1^{(1)}$ vs. x_a/ℓ, Ω, and Γ for AT-cut plate without electrodes.

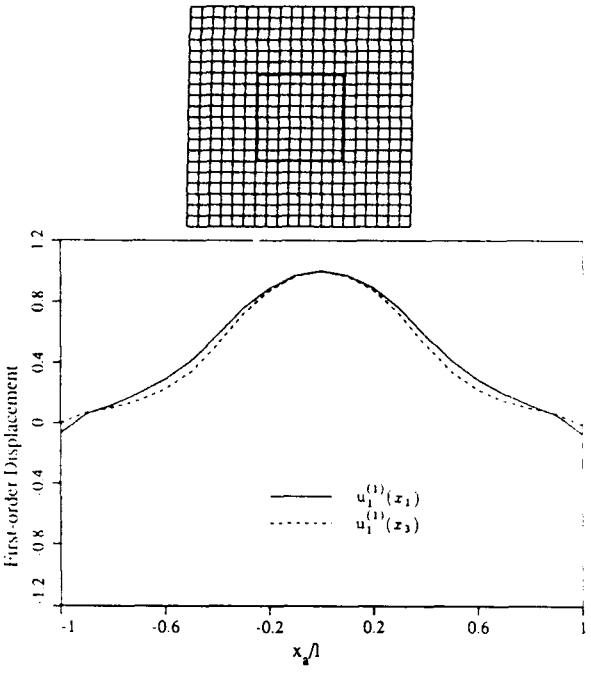


Fig. 3 $u_1^{(1)}$ vs. x_a/ℓ, Ω, and Γ for AT-cut plate plate with square electrodes.

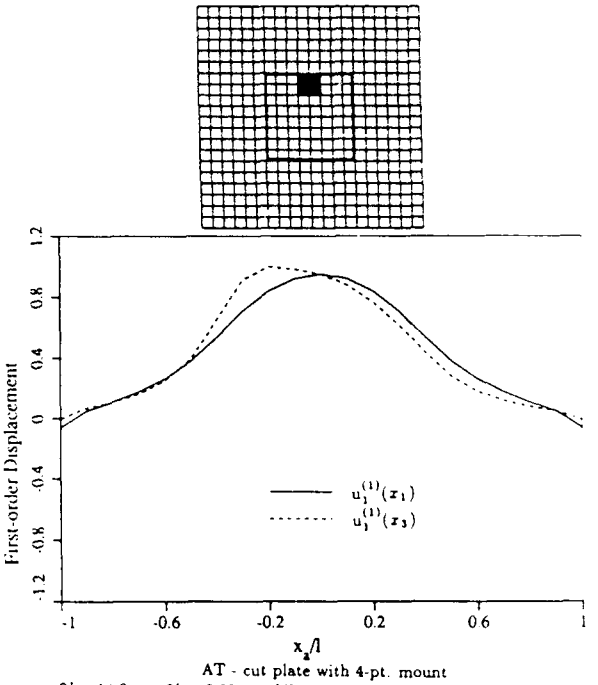


Fig. 5 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Location (1).

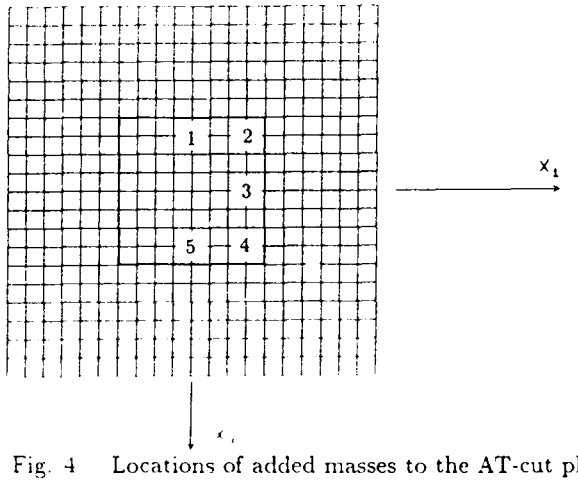


Fig. 4 Locations of added masses to the AT-cut plate with square electrodes.

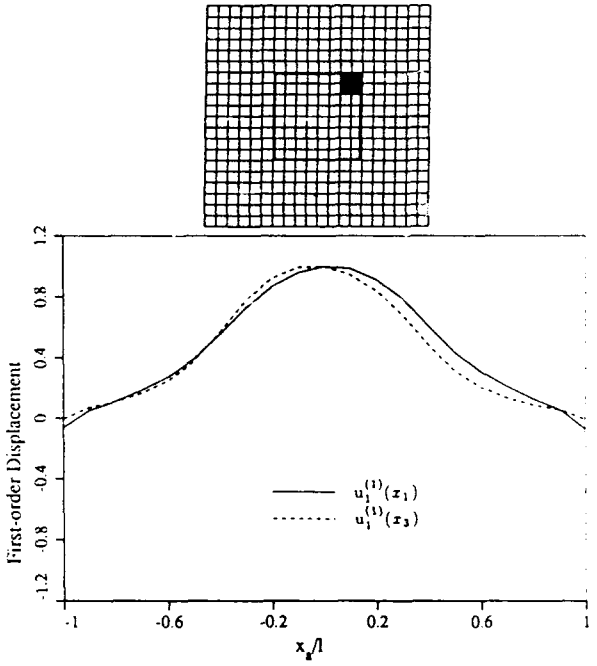
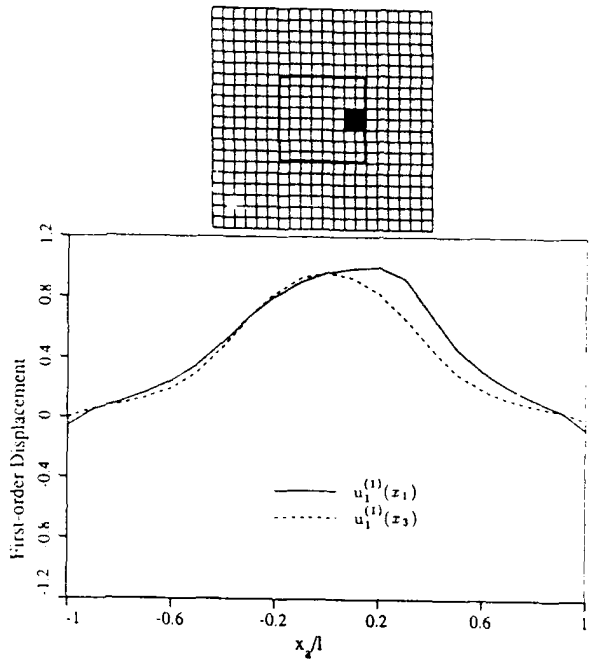
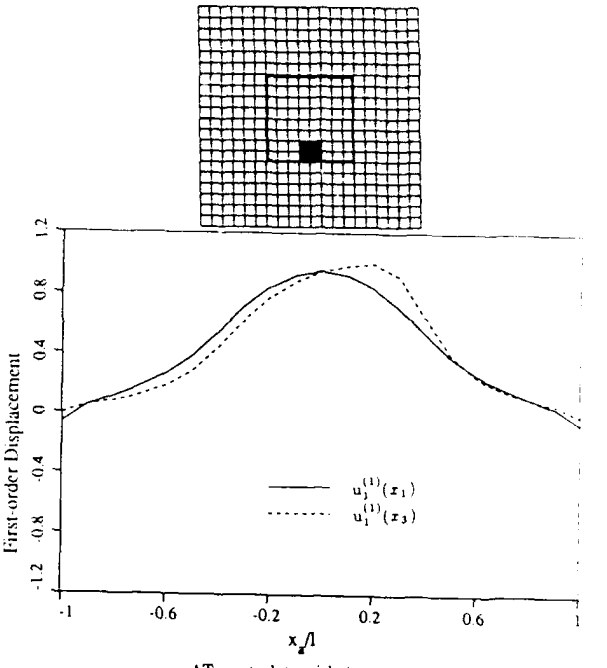


Fig. 6 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Location (2).



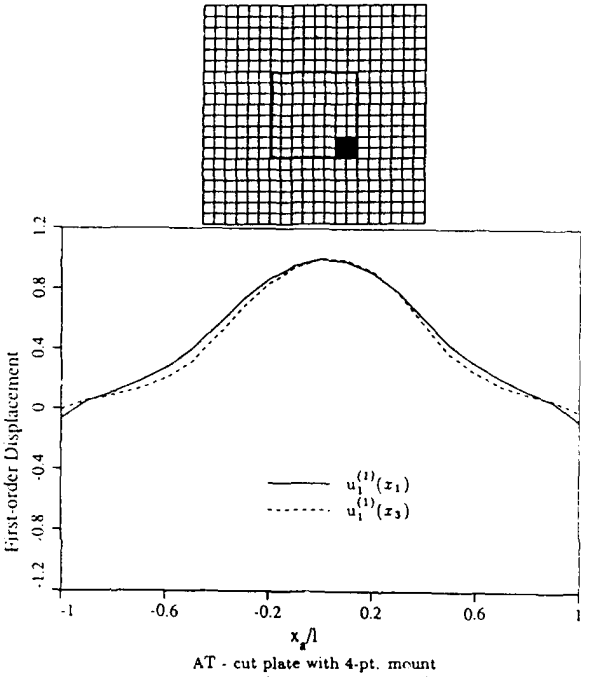
AT - cut plate with 4-pt. mount
$2l = 14.0mm, 2b = 0.28mm, l/b = 50, a/b = 20, le/b = 2.5, n = 20 \times 20$
$R = 0.005, R' = 0.01, \Omega = 0.99769, \Gamma = 1.0215 \times 10^{-10}/g$

Fig. 7 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Location (3).



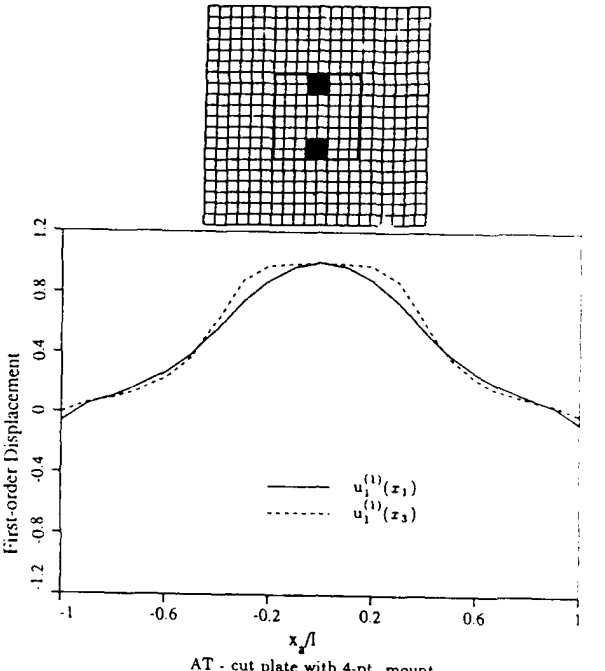
AT - cut plate with 4-pt. mount
$2l = 14.0mm, 2b = 0.28mm, l/b = 50, a/b = 20, le/b = 2.5, n = 20 \times 20$
$R = 0.005, R' = 0.01, \Omega = 0.99769, \Gamma = 4.7774 \times 10^{-10}/g$

Fig. 9 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Location (5).



AT - cut plate with 4-pt. mount
$2l = 14.0mm, 2b = 0.28mm, l/b = 50, a/b = 20, le/b = 2.5, n = 20 \times 20$
$R = 0.005, R' = 0.01, \Omega = 0.99780, \Gamma = 2.9451 \times 10^{-10}/g$

Fig. 8 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Location (4).



AT - cut plate with 4-pt. mount
$2l = 14.0mm, 2b = 0.28mm, l/b = 50, a/b = 20, le/b = 2.5, n = 20 \times 20$
$R = 0.005, R' = 0.01, \Omega = 0.99750, \Gamma = 1.0085 \times 10^{-10}/g$

Fig. 10 $u_1^{(1)}$ vs. x_a/ℓ, Ω, Γ for AT-cut plate with added mass at Locations (1) and (5).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PERFORMANCE CHARACTERISTICS OF A 10 MHz VCO WITH A DUAL BVA RESONATOR*

MALCOLM CALHOUN, PAUL KUHNLE, AND RICHARD SYDNOR

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

ABSTRACT

This paper reports the results of tests performed on two experimental 10 MHz VCOs utilizing dual BVA resonators specifically designed for low acceleration sensitivity. The goal of this experiment is to achieve very low phase noise under quiescent as well as vibration conditions. Test results indicate that the experimental VCOs have acceleration sensitivities which are an order of magnitude lower than those of VCOs presently in use.

INTRODUCTION

The Frequency Standards Test Laboratory at the Jet Propulsion Laboratory (JPL) is responsible for the test and implementation of frequency generation equipment in NASA's Deep Space Network (DSN). Certain assemblies and components of the Frequency and Timing Subsystem (FTS) are located in the cones of tracking antennas where they are subjected to uncontrolled environmental conditions — vibration, temperature extremes, stray magnetic fields, RF radiation, etc. The very stringent requirements placed on FTS equipment challenge the performance of state-of-the-art frequency sources as well as the associated distribution system.

A pair of dual BVA type resonators has been designed and fabricated for JPL by Dr. Raymond Besson of Laboratoire de Chronomètrie Electronique Piézoélectricité, Besançon Cedex, France. These quartz resonators have been oriented specifically for minimizing acceleration sensitivity. The two dual resonators have been incorporated into two 10 MHz VCOs built by Frequency and Time Systems (FTS), Inc. of Beverly, Massachusetts, under contract to JPL. This paper reports the Frequency Standards Test Laboratory's results of testing the VCOs under environmental conditions simulating those at the DSN antennas.

PERFORMANCE TESTS

Following delivery in late 1989, the two 10MHz VCOs underwent extensive testing at JPL's Frequency Standards Test Laboratory (FSTL). Phase noise measurements were made utilizing FSTL's standard test procedure [1].

Dynamic acceleration testing was performed with the VCOs in the phase noise test configuration with the exception that the test oscillator was subjected to controlled vibration to measure acceleration sensitivity [2]. Magnetic field susceptibility tests also were made with the test device in the same configuration. The test oscillator was placed inside an ac Helmholtz coil where the magnetic field magnitude and frequency were varied.

Short term stability as well as oscillator aging tests were performed with the VCOs under controlled temperature, relative humidity, and pressure in the FSTL's environmental test chamber. Finally, temperature, humidity, and pressure were cycled one at a time with the other two parameters fixed to observe the effects on the VCO frequency.

RESULTS

The first of the two VCOs (FTS 001) was resonant 70 Hz below the specified 10 MHz frequency. This offset in oscillator frequency necessitated the use of a scheme wherein the output of the VCO is multiplied up in frequency so that a valid phase noise test could be made against a known reference source. Figure 1 shows the test setup for phase noise, acceleration sensitivity, and magnetic field susceptibility tests.

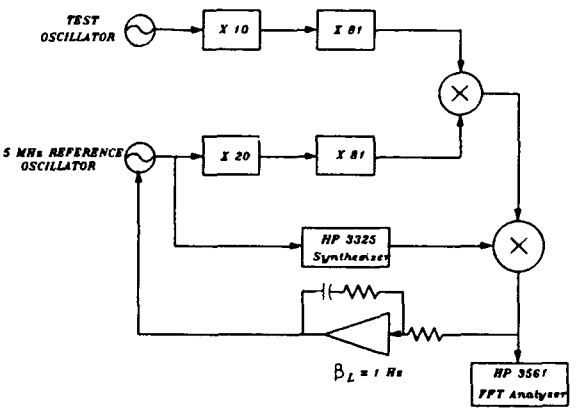


Figure 1. Phase Noise Measurement Test Setup, FTS 001

The test frequency for this configuration is approximately 8100 MHz. Due to the extremely low acceleration sensitivity, γ, of the resonator, the small argument Bessel function is valid for acceleration sensitivity calculations. Phase noise measurements, shown in

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

CH2818-3/90/0000-474 $1.00 © 1990 IEEE

Figure 2, indicate a single-sided phase noise of approximately -114 dBc at 1 Hz from the carrier and a phase noise floor of approximately -155 dBc (phase noise referred to 10 MHz). The second VCO (FTS 002), although precisely on frequency proved to have higher phase noise than FTS 001 (see Figure 3).

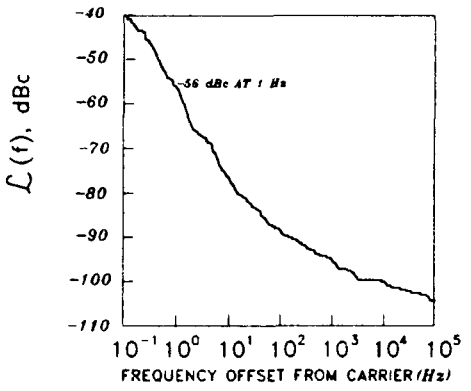


Figure 2. Phase Noise Test Results, FTS 001 at 8100 MHz

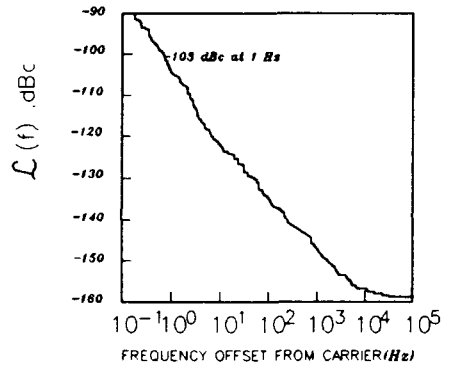


Figure 3. Phase Noise Test Results, FTS 002 at 10 MHz

Allan deviation results for the two test VCOs are shown in Figure 4 and Figure 5, respectively. The short term stability for FTS 001 is 5×10^{-13} at 1 second and at ten seconds. The aging rate for each oscillator is approximately 1×10^{-10}/Day. The short term stability for FTS 002 is 1.5×10^{-12} at one second. FTS 002 later was found to have frequency jumps of 1×10^{-10} at six to ten hour intervals.

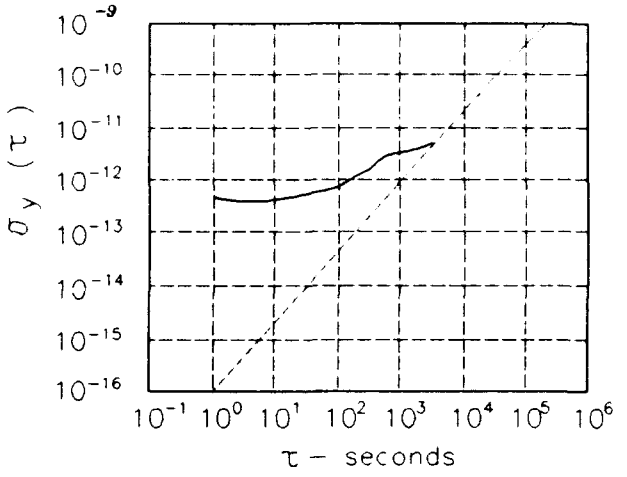


Figure 4. Allan Deviation Results, FTS 001

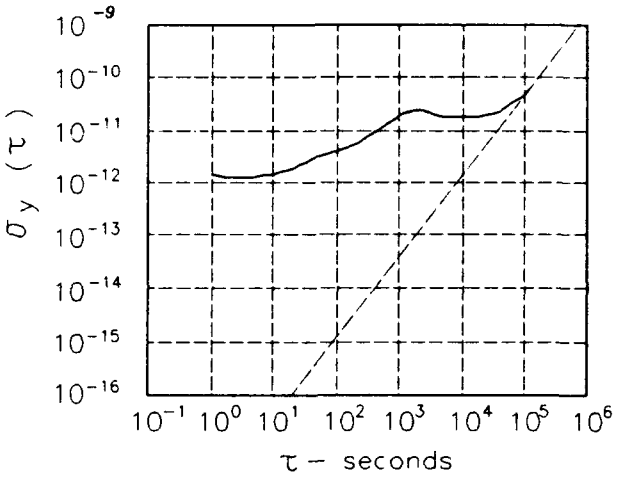


Figure 5. Allan Deviation Results, FTS 002

Dynamic acceleration test results are shown in Figures 6 through 9. The VCOs were vibrated at several different fixed frequencies as well as being subjected to random vibration tests. The phase noise floor virtually remained unchanged for both VCOs under acceleration test conditions. The acceleration sensitivity, Γ, was determined after measuring the acceleration sensitivity, γ, for each resonator axis. The calculations for γ are based on Filler's results [3].

Acceleration sensitivities for FTS 001:

$$\Gamma = \frac{2 f_v \times 10^{\frac{\mathcal{L}_v(f_v)}{20}}}{A_m f_o}$$

$f_o = 10\ MHz$

$A_m = 70.7\ mG$

$\Gamma_x = 1.77 \times 10^{-11}/G \qquad \mathcal{L}_{v_x} = -103.8\ dBc$

$\Gamma_y = 1.25 \times 10^{-11}/G \qquad \mathcal{L}_{v_y} = -107\ dBc$

$\Gamma_z = 1.64 \times 10^{-11}/G \qquad \mathcal{L}_{v_z} = -104.7\ dBc$

$$|\Gamma| = \sqrt{\Gamma_x^2 + \Gamma_y^2 + \Gamma_z^2}$$

$|\Gamma| = 2.7 \times 10^{-11}/G \qquad FTS\ 001$

Acceleration sensitivities for FTS 002:

$\Gamma_x = 1.28 \times 10^{-11}/G \qquad \mathcal{L}_{v_x} = -110.36\ dBc$

$\Gamma_y = 2.7 \times 10^{-11}/G \qquad \mathcal{L}_{v_y} = -103.75\ dBc$

$\Gamma_z = 3.3 \times 10^{-11}/G \qquad \mathcal{L}_{v_z} = -102\ dBc$

$|\Gamma| = 4.4 \times 10^{-11}/G \qquad FTS\ 002$

Acceleration sensitivities for a high quality VCO:

$\Gamma_x = 1.98 \times 10^{-10}/G \qquad \mathcal{L}_{v_x} = -86.1\ dBc$

$\Gamma_y = 2.33 \times 10^{-10}/G \qquad \mathcal{L}_{v_y} = -104.7\ dBc$

$\Gamma_z = 2.71 \times 10^{-10}/G \qquad \mathcal{L}_{v_z} = -83.37\ dBc$

$|\Gamma| = 3.36 \times 10^{-10}/G \qquad HIGH\ QUALITY\ VCO$

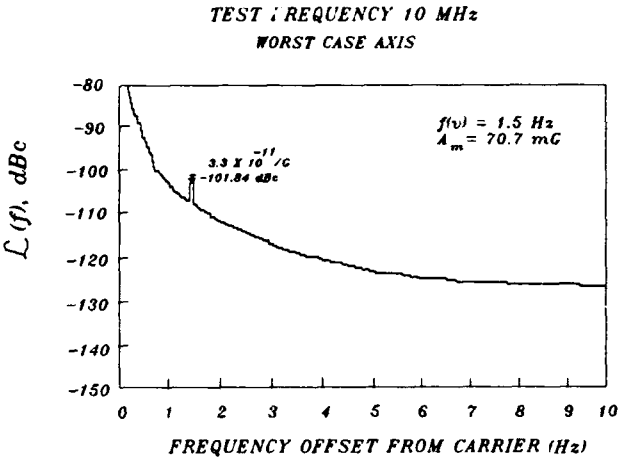


Figure 6. Acceleration Sensitivity Test, FTS 002

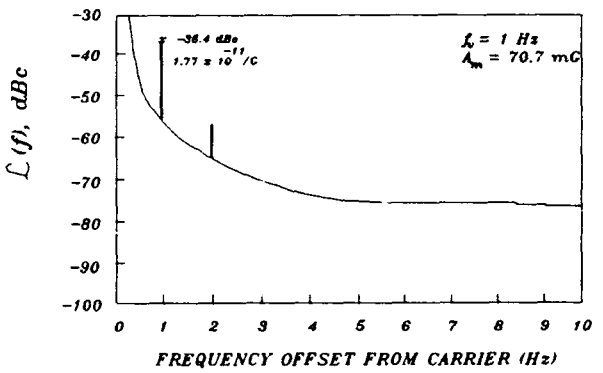


Figure 7. Dynamic Acceleration Test, FTS 001 at 8100 MHz

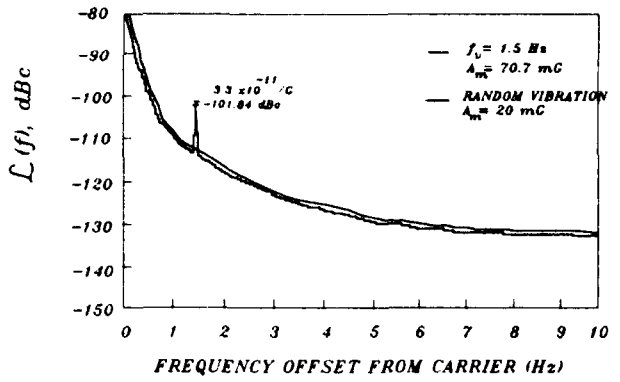


Figure 8. Dynamic Acceleration Test, FTS 002

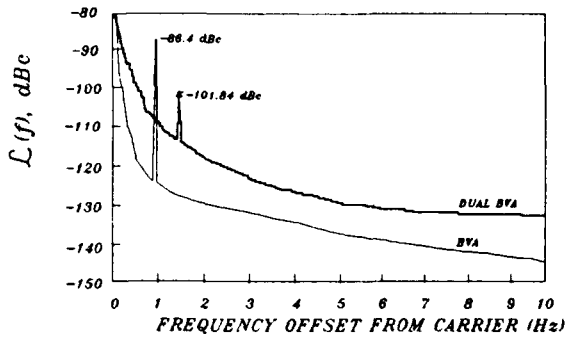


Figure 9. Acceleration Sensitivity Comparison of Dual BVA VCO (FTS 002) with High Quality VCO

Temperature dependence of the VCO frequency is shown in Figure 10. Results indicate $\Delta \frac{f}{f} = 3.6 \times 10^{-10}$/ °C. The test chamber was cycled between + 15 °C and + 35 °C with the relative humidity constant at 20% and the pressure constant at 1 atmosphere. Tests were run with the humidity and pressure cycled individually. Pressure and humidity test results are not shown in this paper.

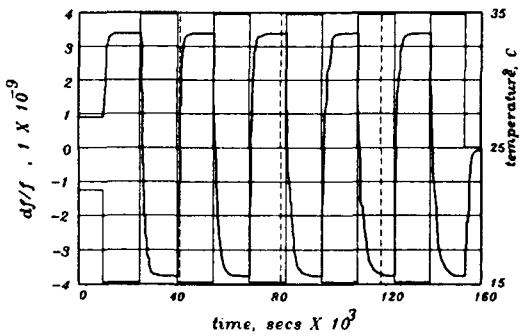


Figure 10. Temperature Dependence of Frequency Shift, FTS 001

Figure 11 is the result of exposing the VCO to a 1 gauss (rms) magnetic field. Apparently additional magnetic shielding must be added to reduce the magnetic field susceptibility.

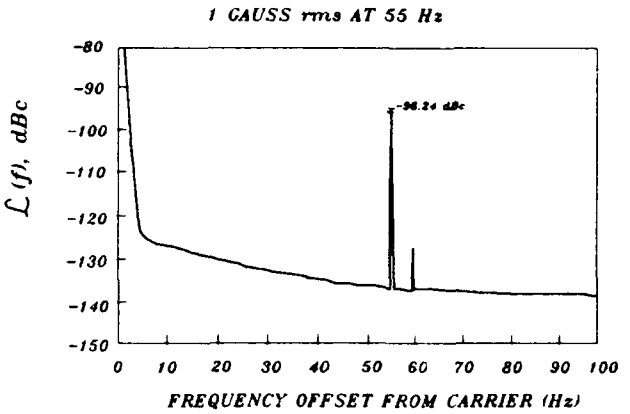


Figure 11. Magnetic Susceptibility Test, FTS 002

CONCLUSIONS

The primary goal of the experiment was realized in that the acceleration sensitivity of the two experimental VCOs is an order of magnitude lower than that of VCOs presently available. It is believed that with more stringent specifications on the VCO packaging and peripheral control circuitry that the other VCO problems can be overcome. Careful screening of the resonators is a necessity.

REFERENCES

[1] Hewlett-Packard Application Note 207, "Understanding and Measuring Phase Noise in the Frequency Domain", Hewlett-Packard, Palo Alto, CA October 1976.

[2] Instruction Manual Model 129 Shaker, Acoustic Power Systems, Inc. Carlsbad, CA.

[3] Filler, R.L. , "The Effect of Vibration on Quartz Crystal Resonator and Oscillator Production", DELET TR-80-10, May, 1980.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A New Factor Affecting the Acceleration Sensitivity of the Resonance Frequency of Quartz Crystal Resonators

W. P. Hanson, Piezo Crystal Company
T. R. Meeker, Consultant
L. C. Heishman, Piezo Crystal Company

Introduction

Progress in reducing the acceleration sensitivity of resonators to a consistent value below 10^{-10} per g has been very slow. A possible reason for the slow progress is that an important factor controlling the sensitivity of resonators has been overlooked.

When the crystal is under acceleration the electrical leads bend and stretch causing changes in the series inductance and shunt capacitance. The distributed characteristics of the other electrical parasitics may also cause changes in resonance frequency.

This suggests a coupling of elastic and electrical properties in the crystal mounting structure and associated electrical paths. A theory is presented which shows the acceleration sensitivity degradation due to this new factor can have magnitudes of 10^{-10} per g. Experimental data is also presented which tends to support the theory.

Theory

In this section the following subjects are discussed:

1. The exact electrical impedance of a lossless single thickness mode piezoelectric resonator with a series electrical inductance and a shunt electrical capacitance. All overtones are included in the derived impedance equation.

2. The computer algorithm for calculating the exact electrical impedance of a lossy single thickness mode piezoelectric resonator with a series electrical inductance and a shunt electrical capacitance. The effects of loss are included in the calculation empirically. The loss can be made frequency dependent to allow matching experimental values of motional resistance of all of the overtone modes of interest.

3. Some calculations of resonance frequency (reactance = 0) of the single thickness mode piezoelectric resonator with series electrical inductance and shunt electrical capacitance.

4. Some calculations of the movement of crystal blanks inside a package and the resonance frequency shifts associated with the resultant changes in the series electrical inductance and electrical shunt capacitance.

Now for the discussions of the above subjects.

1. The electrical impedance of the lossless single mode thickness mode piezoelectric resonator.

The electrical impedance of the lossless single thickness mode piezoelectric resonator may be written as:

$$Z_c = \frac{k^2}{j*\omega*C0} * \left[\frac{1}{k^2} - \frac{\tan(x)}{x} \right] \quad (1)$$

$$\text{where } x = \left(\frac{\pi}{2}\right) * \left(\frac{f}{fa}\right) \quad (2)$$

fa is the parallel resonance frequency, C0 is the clamped capacitance, k is the piezoelectric coupling coefficient, and ω is the angular frequency[1,2].

Two reviews of the published work on the derivation of Equation (1) give references to

CH2818-3/90/0000-478 $1.00 © 1990 IEEE

most of the earlier work [1,2].

Equation (1) expresses the electrical impedance of the resonator for all frequencies so that all overtone series resonance frequencies can be obtained.

It is assumed that Equation (1) represents the electrical impedance of a multimode resonator well enough for the purposes of this paper. Although the effect of the other modes on the mode of interest can be included in the calculations[2], these other mode effects are beyond the scope of the present paper.

The addition of a shunt electrical capacitance (Cs) and a series electrical inductance (Ls) to the resonator gives an electrical impedance that can be expressed as [3]:

$$Z_c = \frac{k^2}{j*\omega*(C0+Cs)} * \left[\frac{1}{k_e^2} - \frac{\tan(x)}{x} \right] \quad (3)$$

$$\text{where } x = \left(\frac{\pi}{2}\right) * \left(\frac{f}{fa}\right) \quad (4)$$

fa is the parallel resonance frequency, C0 is the clamped capacitance and ω is the angular frequency. Ke is an effective coupling coefficient defined in equation (5).

Equation (3) shows that the series inductance has two primary effects. The first effect is to lower the series resonance frequency as if the piezoelectric coupling coefficient (k) had increased. In fact an effective piezoelectric coupling coefficient (ke) can be defined for series resonance frequency calculations as:

$$k_e^2 = \frac{k^2}{1-\omega^2*L_s*(C0+Cs)} \quad (5)$$

The second effect is to make the series resonance frequency depend on the value of the shunt electrical capacitance. In the absence of the series inductance the series resonance frequency of the resonator does not depend on the value of the shunt capacitance (C0) of the resonator or on the added shunt capacitance (Cs).

When the packaged resonator is accelerated, the mass of the crystal blank causes the electrical leads to stretch and bend (and twist). These motions of the electrical leads produce changes in the series inductance and in the shunt capacitance, leading to changes in the resonance frequency of the resonator.

A thesis of this paper is that these changes in the electrical characteristics of the mounting leads must be controlled by proper design in order to fabricate devices with consistent acceleration sensitivities below about 1×10^{-10} per g. Some order of magnitude estimates of these effects will be discussed in Part 3 and Part 4 of this section of this paper.

2. The computer algorithm for adding loss to the calculation of the electrical impedance of the resonator.

Loss needs to be added to Equation (3) to improve the accuracy of calculated results. Electrical loss can be included in the calculations rather easily by adding a resistance in series with Ls and a second resistor in shunt with C0 and Cs. Elastic or motional loss is more difficult to model because there are three main sources, i.e., boundary condition losses, mounting losses, and material losses. For the purposes of this paper the elastic loss is added empirically to make the calculated and measured motional resistances agree. The empirical electrical loss can depend on frequency so that calculated and measured motional resistances agree for all of the overtones of interest.

The value of the parallel resonance frequency (fp - crystal reactance = 0) is also needed for the calculation. fp can be estimated from a direct measurement, or from measurements of series resonance frequency (fs), motional

inductance (L1), or motional capacitance (C1) from:

$$\omega_p * L1 - \frac{1}{\omega_p * C1} - \frac{1}{\omega_p * C0} = 0 \quad (6)$$

$$\omega_s * L1 - \frac{1}{\omega_s * C1} = 0 \quad (7)$$

$$\text{where: } \omega_p = 2 * \pi * f_p \qquad \omega_s = 2 * \pi * f_s \quad (8)$$

For the calculations reported in this paper the electrical losses are assumed to be small.

The Algorithm

a. Z = Zc; impedance of resonator from Equation (1)

b. Y = 1/Z; admittance of resonator

c. Ye = Y - j·ω·C0; elastic or motional admittance of resonator

d. Zx = (1/Ye) + Rm; motional impedance + motional resistance

e. Yx = (1/Zx) + j·ω·C0 + j·ω·Cs; shunt elements added

f. Zz = (1/Yx) + j·ω·Ls + Rs; total impedance of the resonator

g. Zz = Rz + j·Xz; definition of resistive and reactive parts of Zz

Series resonance frequencies of fundamental and overtones are the frequencies at which Xz = 0. These zeroes are found by a conventional root search.

This algorithm was used to obtain the dependence of series resonance frequency (of a particular overtone for a particular resonator) on the values of series inductance and shunt capacitance. These results are discussed in Part 3 of this section of this paper.

3. Calculation of the dependence of resonance frequency of a third overtone SC cut quartz resonator on series inductance and shunt capacitance.

The results of some of the calculations are given in Table 1.

Table 1

C0 + Cs (nominal) = 2.91 pF
Fs = 10.0045 MHz
C3 = 2.3182×10^{-4} pF
R3 = 211 ohms

	Ls (nH)	Cs(pF)	Fs3-1E7 Hz
1)	1E-4	1E-16	4405.12500
2)	20	1E-16	4404.94922
3)	40	1E-16	4404.77344
4)	1E-4	1.01E-2	4405.12500
5)	20	1.01E-2	4404.95313
6)	40	1.01E-2	4404.77734

In Part 4 of this section these results will be used with the estimates of acceleration induced crystal mount elastic motions and associated changes in electrical inductance and capacitance to estimate that part of the acceleration sensitivity due to the motions of the mounting leads.

4. Lead Stretching and Bending

In this part the stretching and bending of the electrical leads of the packaged crystal resonator are estimated. Twisting will not be considered in this paper.

The changes in series inductance and in shunt capacitance associated with the motion of the leads are also estimated in this part.

Table 2 lists the relevant material constants and structural parameters for the calculation of the motion of the parts of a particular mounting structure.

Table 2

Acceleration	= 10 g
Weight of quartz blank at 10 g	= 5.678E-3 lbs
Young's modulus for Ni lead	= 30E6 lbs/in^2
Radius of lead	= .009 inches
Length of lead from package base	= .5 inches
Number of leads	= 2

Lead stretching

The fractional length change due to lead stretching [4] is:

$$\frac{\Delta l}{l} = \frac{Fc}{E*A} \qquad (9)$$

where Fc is the force or weight of the quartz blank, E is Young's modulus, and A is the cross-sectional area of the lead.
For two leads and 10g the result is:

$$\frac{\Delta l}{l} = 3.74E-7 \qquad (10)$$

The inductance of the unstretched wire[5,6] is defined in equation (11).

$$L_0 = 5.08*Bl*[\ \ln\frac{2*Bl}{Rd} - 1 - \frac{MU}{4} + \frac{Rd}{Bl}\ ]\ E-9 \qquad (11)$$

where BL is the length of the lead, Rd is the radius of the lead, MU is the relative permeability of the material of the lead, and ln is the natural logarithm to the base e.

Table 3 presents some values of lead inductance (L0) calculated from Equation (11).

Table 3

Material	Relative Permeability (reference)	LO (x10^{-9} H)
Copper	1 (5)	17.2
Nickel	50 (6)	65.1
Iron	8000 (6)	7839.4

From Table 3 at a frequency of 10 MHz

$$.17578\ Hz\ per\ 20\ nH = 176 \times 10^{-11}\ per\ 20\ nH \qquad (12)$$

$$\frac{\Delta Ls}{Ls} = \frac{\Delta L}{L} = 3.74e^{-8}\ per\ g \qquad (13)$$

$$\frac{\Delta Fs}{Fs}(copper) = 566.1E^{-18} \qquad (14)$$

$$\frac{\Delta Fs}{Fs}(iron) = 258010x10^{-18} \qquad (15)$$

These results show that for the mounting structure considered the resonance frequency shift due to wire stretching during acceleration is at most (for a high permeability wire) $2.6x10^{-13}$ per g. At the 10^{-11} per g level wire stretching for this stiff mount plays no

significant role.

Lead Bending

When the acceleration direction is in the plane of the quartz blank, the motion of the unsupported end of the lead is toward the wall of the package. For the common case in which the package is metal and grounded to one of the crystal leads, this motion changes the shunt capacity across the crystal. A small series inductance change associated with the bending will be ignored in this paper. In the presence of the series inductance of the lead this shunt capacitance produces a resonance frequency change due to the acceleration.

Since the lead is mostly clamped at the package base and free at the quartz blank, the displacement of the lead toward the wall of the package is not the same at every point on the lead.

The equation for the displacement of the bent lead for the case in which a force (due to the acceleration of the quartz blank) is applied to the free end of the lead with no bending or twisting [7] is:

$$h(z) = \frac{Fc * z^2 * (3 * Bl - z)}{6 * E * I} \qquad (16)$$

Where z is the distance from the base, Fc is the force (weight), Bl is the length of the lead, E is Young's modulus, and I is the moment of inertia of a cross-section of the lead

$$I = \frac{\pi * Rd^4}{4} \qquad (17)$$

The capacitance per unit length of a straight wire with round cross-section to a ground plane[8] is given by:
where Ep is the permittivity (= 8.85e-12 fds/meter for vacuum), h is the distance of the wire center to the ground plane, and a is the radius of the wire.

To estimate the capacitance of the bent wire,

$$Cs = \frac{2 * \pi * Ep}{\cosh^{-1}\left(\frac{h}{a}\right)} \qquad (18)$$

the wire is divided into 50 segments. Each segment has a distance to the ground as given in equation (16). The 50 segment capacitances are added to give the capacitance of the bent lead. The capacitance of the unbent lead is subtracted to give the capacitance change associated with the bending. Table 4 gives the results of a calculation for a particular structure.

Table 4

Distance of lead to ground plane= .025 inch	
Diameter of lead	= .010 inch
Length of lead	= .5 inch
Cap. change/g for 1 lead	= 3.6×10^{-15} fds/g
Cap. change/g for 2 leads	= $1.8E \times 10^{-15}$ fds/g

From Table 1 at a series lead inductance of 20 nH:

$$\frac{\Delta Fs}{Fs} = 3.91 \times 10^{-10} \text{ per } 10^{-14} \text{ fd} \qquad (19)$$

of shunt capactance change

$$\frac{\Delta Fs}{Fs} = 3.91 \times 10^{-10} * \left(\frac{1.8 \times 10^{-15}}{1 \times 10^{-14}}\right)$$

$$= .7 \times 10^{-10} \qquad (20)$$

This result of $.7 \times 10^{-10}$ per g for the rather stiff selected mounting configuration supports the thesis of this paper that the electrical effects associated with the deformation of the mounting structure must be considered in the design of quartz resonators with expected acceleration sensitivities much less than 1×10^{-10} per g.

Experimental Methods

The effects we are looking for on acceleration sensitivity are small. Good experimental techniques are needed to achieve the measurement resolution required. Two effects were of primary concern in this experiment; cable effects and crystal orientation. Both of these effects can degrade the acceleration sensitivity measurements by an order of magnitude. When these effects have been eliminated a high degree of repeatability can be achieved.

Cable Effects

Watts, et. al.[9] and Driscoll[10] show the effect of cables on the measurement of acceleration sensitivity. Cable effects must be minimized by measuring the crystal in all six axes. Each axis must be measured with the cable up and the cable down. The magnitude of r_x is then given by:

$$|\Gamma_x| = \frac{|\Gamma_{up}| + |\Gamma_{down}|}{2} \quad (21)$$

Figure 1 shows how the cable effect adds vectorially to the r_{up} and r_{down} measurements along the x-axis. The cable must be kept in the same position if the acceleration effects of the cable are to be corrected. Changing the position of the cable may change the direction and magnitude of the cable acceleration sensitivity contribution. All three axes must be corrected in this same manner.

The cable effect was the spark that started the thinking into the effect of vibration on the acceleration sensitivity. What causes the cable effect? Could it be the change in shunt capacitance or series inductance as the cable vibrates.

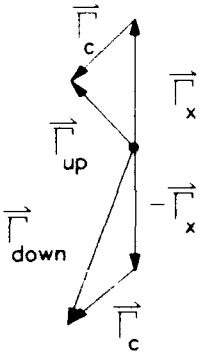


Figure 1 Cable effects on the measurement of r

Resonator Orientation

The orientation of the resonator in the mounting structure is known from theory to affect the acceleration sensitivity. The number of mounts is also known to affect the acceleration sensitivity. Shick and Lee describe these effects in recently published works[11,12].

For the experiments reported on in this paper a two point crystal mount was selected because of its simple symmetry and ease of implementation. The optimum orientation of the blank in a two point mount had to be measured. A fixture was designed to find this orientation as simply as possible.

The fixture had to have several important features. The fixture must mount the crystal blank semirigidly and be capable of reorienting the crystal easily. The fixture must also be easily flipped in all three axis to measure cable effects.

A semirigid mount was needed to minimize the coupled elastic and electrical effects in the fixture. This allows a measurement of the effects of selected wires added in series with the crystal. The fixture was designed in such a way that the crystal could be easily

reoriented with respect to the mounts. The crystal was plated so that it could be measured at any angle from zero degrees to greater than 90 degrees. This allowed the crystal to be optimally oriented for minimum acceleration sensitivity.

The fixture is also capable of incorporating a conventional mounting structure. This allows for correlation to standard resonators. Figure 2 is a diagram of the crystal as it was plated around the periphery.

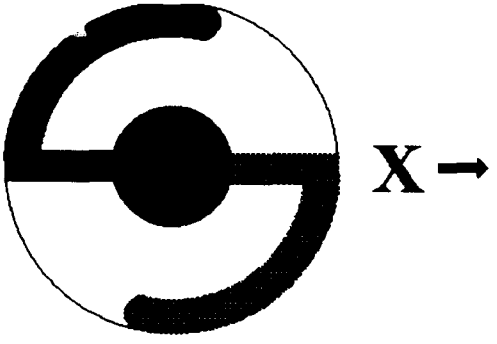


Figure 2 Crystal and plating pattern for easy acceleration sensitivity measurements

The fixture was designed to easily incorporate a series wire or ribbon. Figure 3 is a diagram of the fixture in the rigid configuration.

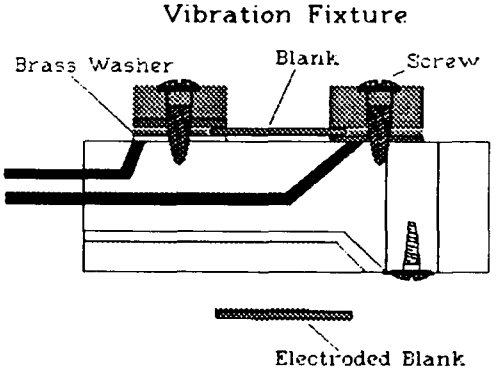


Figure 3 Fixture for Semi-rigid Crystal Setup

Figure 4 is a block diagram of the fixture as it is setup for measurement with a series wire added.

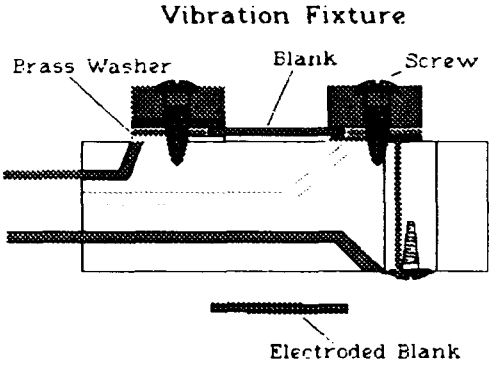


Figure 4 Fixture Setup for Series Wire

Experimental Results

The first experiment was to find the optimum orientation for the mounting. The plated resonator was mounted at two points. At each mount the resonator was clamped between two .040 inch wide .035 inch thick steel plates. The resonator had a .050 wide bevel where the crystal was clamped. The resonator was carefully placed in the fixture and measured at many orientations. Repeatability was checked frequently and found to be $2\text{-}3\text{x}10^{-11}$/g in each axis for the semirigid configuration. Figure 5 shows the measured acceleration sensitivity as a function of orientation.

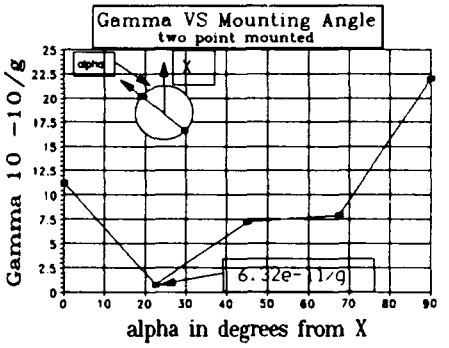


Figure 5 Measured acceleration sensitivity as a function of plate orientation.

After the orientation for minimum acceleration sensitivity was found the resonator remained at that orientation during acceleration sensitivity measurements with a lead in series with one crystal lead.

A Wire in Series

After good repeatability was found in the minimum r orientation an 18 mil wire was placed in series with one lead. The gamma was then measured in each axis. The total gamma at 10g's and 50 Hz was measured as $3.8 \times 10^{-9}/g$. This measurement is larger than what was expected from a series inductance change in the wire. Dielectric effects from the surrounding Delrin may have caused greater changes. Some measurement problems may have occurred that were not noticed. Further investigations will have to be made on this configuration.

A Ribbon in Series

A ribbon was placed in series with one lead and the acceleration sensitivity measured in all three axes at a range of frequencies. Mounting resonances were found easily. Figure 6 is a graph of the acceleration sensitivity in all three axes. Figure 7 is a graph of the total gamma for the ribbon. In this configuration the fixture was wired in the semirigid configuration with both wires at the top of the fixture. Only the ribbons were added into the fixture. The ribbons were cemented to the crystal at the points were the minimum acceleration sensitivity was found. The resulting acceleration sensitivity was very poor.

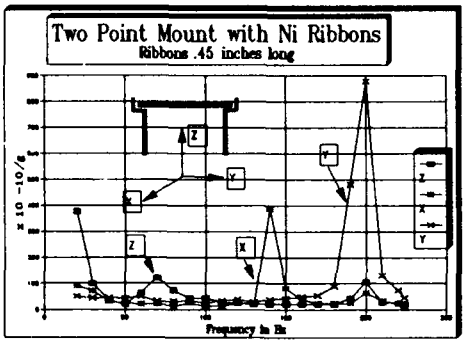


Figure 6

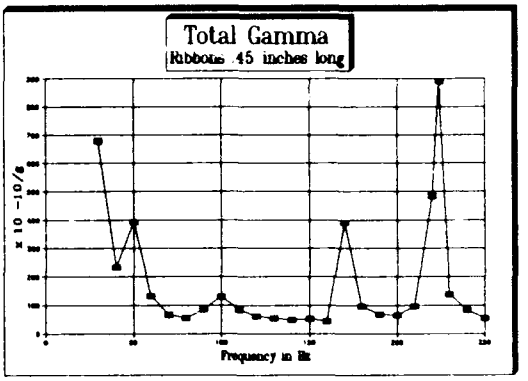


Figure 7

Figure 8 is a summary of all acceleration sensitivity measurements taken with the fixture. Reasonable repeatability can be seen at the minimum acceleration sensitivity orientation. The measurements in Figure 8 are temporal covering about 35 hours on the vibration machine. At the end of the test the fixture was becoming worn. Some degradation in the measurements is probably due to fixture wear.

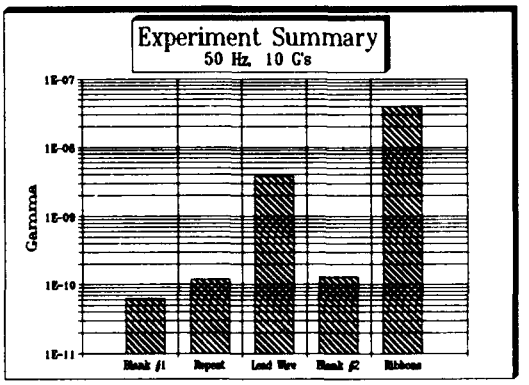


Figure 8

Conclusions

1. In a simple fixture a quartz SC blank can be made to have 10^{-11}/g acceleration sensitivity reproducible.

2. Adding a loose series wire appears to have produced a large increase in acceleration sensitivity.

   a. fixture problems included contact and solder joint problems, as well as cable motion problems.

3. Adding a loose series ribbon produced even larger increases in acceleration sensitivity at certain shake frequencies.

4. The slopes of the series resonant frequency vs series inductance curve are:

   a. 1.37e-9/nH for a 10 MHz 3rd SC

   b. 1.37e-7/nH for a 100 MHz 3rd SC

Acknowledgements

The authors would like to thank Mike Driscoll for helpful discussions.

References

1. T. R. Meeker, Thickness Mode Piezoelectric Transducers, Ultrasonics, Vol. 10, No. 1, January 1972, pps. 26-36.

2. T.R. Meeker, Bulk Acoustic Waves and Resonators, Precision Frequency Control, Editors E. A. Gerber and A. Ballato, Vol. 1, Chapter 2.1 and Chapter 2.2, pps. 47-118.

3. S. N. Wickstrom and T. R. Meeker, Electrical Measurement of High Frequency ceramic Resonators, Proceedings of the 1989 IEEE Ultrasonics Symposium, October 3-9, 1989, Montreal, Canada, pps. 717-722.

4. L. D. Landau and E. M. Lifshitz. Theory of Elasticity (Translation), Pergamon Press, London, and Addison-Wesley, Reading, Massachusetts, 1959, p. 13.

5. F. E. Terman, Radio Engineers' Handbook, McGraw Hill, New York, 1943, p. 48.

6. Reference Data for Radio Engineers - Fifth Edition, ITT - H. W. Sams, Indianapolis, 1968 (Sixth Printing - 1974), p, 4-32.

7. L. D. Landau and E. M. Lifshitz. Theory of Elasticity (Translation), Pergamon Press, London, and Addison-Wesley, Reading, Massachusetts, 1959, p. 13., p. 91, Problem 3, and p. 75.

8. American Institute of Pnysics Handbook, Second Edition, McGraw Hill, New York, 1957, p. 5-15.

9. M. H. Watts and E. P. EerNisse, R. W. Ward, R. B. Wiggins, "Technique for Measuring the Acceleration Sensitivity of SC-Cut Quartz Resonators," Proceedings 42nd Annual Symposium on Frequency Control, pp 442-446, 1988.

10. M. M. Driscoll, "Quartz Crystal Resonator G-Sensitivity Measurement Methods and Recent Results," 43rd Annual Symposium on Frequency Control, pp 419-426, 1989.

11. P.C.Y. Lee and M. S. H. Tang, "Acceleration Insensitivity of Thickness Frequencies of Doubly Rotated Quartz Crystal Disks," 42nd Annual Symposium on Frequency Control, pp 14-18 1988.

12. D. V. Shick and H. F. Tiersten, "An Analysis of the In-Plane Acceleration Sensitivity of Contoured Quartz Resonators Rigidly Supported Along the Edges," 43rd Annual Symposium on Frequency Control, pp 405-412, 1989.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

MEASUREMENTS OF ACCELERATION-INDUCED PHASE NOISE IN SURFACE ACOUSTIC WAVE DEVICES

J. Kosinski, A. Ballato, and T. Lukaszek

U.S. Army Electronics Technology and Devices Laboratory
LABCOM, Fort Monmouth, NJ 07703-5000

ABSTRACT

Stable local oscillators based on surface acoustic wave (SAW) resonators have been proposed for future use in Army electronic systems designed for airborne and vehicular applications. The vibration spectra of these systems in conjunction with the vibration sensitivity of the SAW resonators leads to a degradation in oscillator phase noise which may then exceed the total allowable system phase noise. This report details a study of the vibration induced phase noise in a group of SAW oscillators developed for a low-noise radar application.

INTRODUCTION

Stable local oscillators based on surface acoustic wave (SAW) resonators have been proposed for future use in Army electronic systems designed for airborne and vehicular applications. The advantages of the SAW oscillator in such systems are reductions in size, weight, and power consumption due to the high frequencies (1 GHz or better) obtainable without the use of frequency multiplier chains. The small size and simplicity of SAW resonators leads to low-cost UHF oscillators with spectral purity superior to that obtainable using any other UHF frequency control device.

As with bulk acoustic wave (BAW) quartz resonators, SAW resonators are sensitive to external accelerations. The resonant frequency f_o of the SAW device experiences a perturbation δf given by

$$\delta f = f_o \, (\Gamma \cdot \mathbf{A}), \qquad (1)$$

where $\Gamma \cdot \mathbf{A}$ is the scalar product of the SAW device acceleration sensitivity Γ (a vector property of the SAW device) and the external acceleration field $\mathbf{A}$. Reported values for $|\Gamma|$ for SAW devices range from 10^{-8} per g (1g = earth's gravitational field) for single SAW devices to 3×10^{-10} per g for dual SAW device structures employing compensation techniques.

In this study, measurements of the acceleration sensitivities of a group of SAW resonators developed for a low-noise radar application are presented. The SAW resonator designs were systematically varied in order to obtain the best possible phase noise in a quiescent environment. In this report the relationships between the design changes and the vibration induced phase noise will be discussed.

EXPERIMENTAL SAMPLES

The SAW devices used in this study have been extensively detailed elsewhere [1]. The experimental matrix included three device types (one-port resonator, two-port resonator, delay line), two metallization types (pure Al,

TABLE 1. EXPERIMENTAL MATRIX

Device Type	Transducers Thickness	Transducers Material	Busbars Thickness
Delay Line	400Å	Pure Al	unknown
One-Port	600Å	Pure Al	2,000Å
Two-Port	600Å	Pure Al	2,000Å
Two-Port	600Å	Cu-doped Al ≈0.5%	2,000Å
Two-Port	600Å	Pure Al	6,000Å-10,000Å
Two-Port	400Å	Pure Al	2,000Å
Two-Port	800Å	Pure Al	2,000Å
Two-Port	600Å	Cr flash (≈60Å) + pure Al	2,000Å

US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT

Cu-doped Al), three transducer thicknesses (400Å, 600Å, 800Å), and three busbar thicknesses (2000Å, 6000Å, 10,000Å). Details of the experimental matrix are shown in Table 1. A total of twenty-two devices were tested, consisting of twelve each delay lines, one each one-port resonator, and nine each two-port resonators.

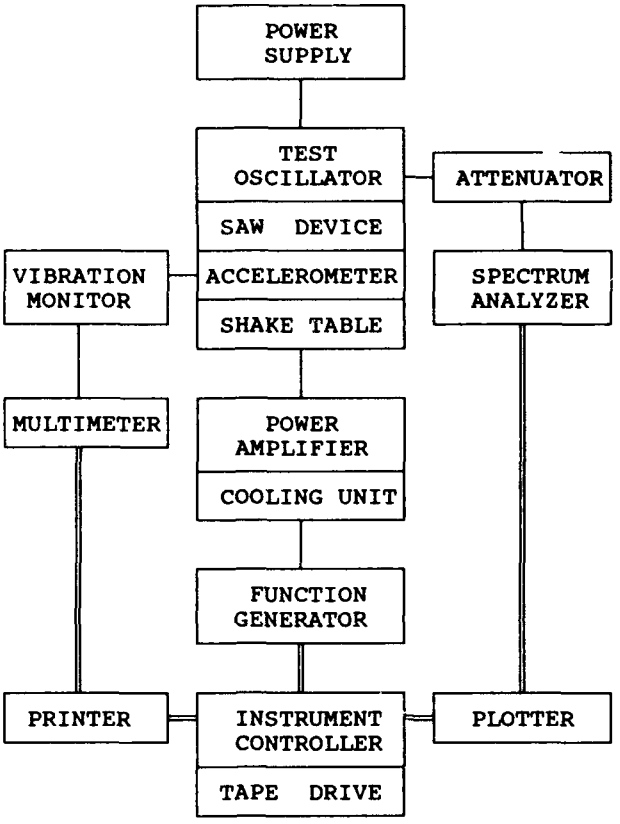


Figure 1. Test system block diagram.

TEST SYSTEM

A block diagram of the test system employed is shown in Figure 1. In the presence of a sinusoidally varying acceleration field, the output signal (carrier) of the SAW based oscillator will undergo frequency modulation at the sinusoidal frequency of the acceleration field. The effect is analogous to tone modulation and the ratio of the power in the n^{th} acceleration-induced sideband to the power in the carrier is given simply as

$$f_a^{(n)} = 20 \log [J_n(\beta)/J_o(\beta)] , \qquad (2)$$

where β is the modulation index determined from

$$\beta = \frac{\Omega_o \, (\Gamma \cdot \mathbf{A})}{\Omega_a} . \qquad (3)$$

In equations (2) and (3), $\mathbf{A}$ is the peak acceleration, Ω_o is the unperturbed output frequency, Ω_a is the acceleration frequency, and Γ is the acceleration sensitivity. If the modulation index is less than 0.1, equation (2) simplifies to

$$f_a^{(n)} \approx 20 \log \frac{f_o \, (\Gamma \cdot \mathbf{A})}{2 \, f_a} . \qquad (4)$$

We can therefore measure Γ by applying a sinusoidally varying acceleration field via a shaketable and measuring the power in the acceleration-induced sidebands using a spectrum analyzer. Inasmuch as Γ is a vector quantity, we perform the measurement along three orthogonal axes.

The limitations of the measurement system as implemented may be found by first solving equation (4) for Γ and then examining the sensitivity of the calculated Γ value to each of the measured quantities f_c, f_a, $|\mathbf{A}|$, and $f_a^{(n)}$. Errors in measuring the sideband levels contribute

$$\frac{\delta\Gamma}{\Gamma} = 10^{(\epsilon/20)} - 1 , \qquad (5)$$

errors in measuring the acceleration frequency contribute

$$\frac{\delta\Gamma}{\Gamma} = \frac{\epsilon}{f_a} , \qquad (6)$$

and errors in measuring either the peak acceleration or carrier frequency contribute

$$\frac{\delta\Gamma}{\Gamma} = \frac{-\epsilon}{\phi+\epsilon} , \qquad (7)$$

In equations (5) through (7), ϵ refers to the error in the measurand of interest, and ϕ in equation (7) refers to the measurand itself. For the measurement system as implemented here, the dominant source of error is the sideband power ratio measurement which limits the measurement of $|\Gamma|$ to +19/-16% accuracy.

EXPERIMENTAL RESULTS

Sample data obtained using the measurement system are shown in Figure 2. The maximum and minimum values obtained at each acceleration frequency tested are shown as a function of the test frequency. The labels "Normal" and "In-plane" refer to measurements normal to the plane of the

SAW resonator and in the plane of the SAW resonator respectively. In this device we observe a mechanical mounting related resonance near 1kHz which degrades net device sensitivity by up to a factor of five in the resonance region. We also observe that the mounting technique used produces greater repeatability normal to the plane of the SAW as compared to in the plane of the SAW. The results of all the delay line measurements are summarized in Table 2 and the results of the resonator measurements are summarized in Table 3.

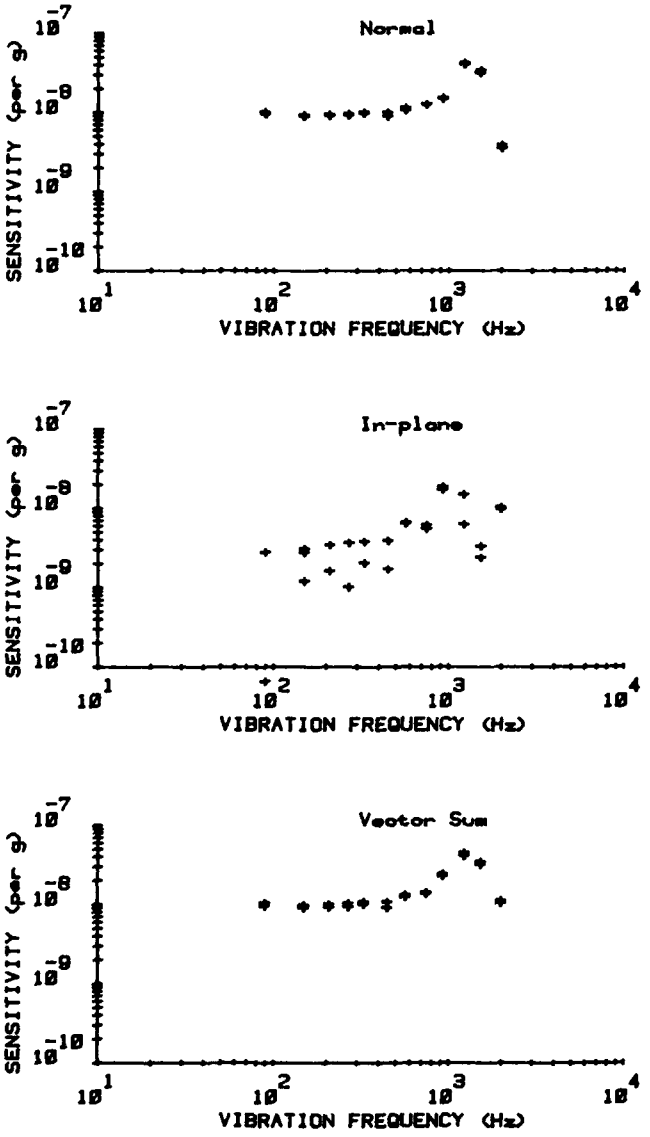


Figure 2. Sample experimental data.

TABLE 2. DELAY LINE MEASUREMENTS ($x10^{-9}/g$)

S/N	Normal	In-plane	Total
Q1359A	22	6.0	25
Q1359#2	22	8.0	25
Q1357A	14	<10	16
Q1357B	14	<20	25
Q1357C	18	<20	28
Q1398A	21	<9.0	21
Q1398B	12	<20	30
Q1402D	22	<20	30
Q1816A	10	<20	20
Q1816B	4.5	<14	14
Q1816C	9.0	<1.0	14
Q1817F	5.0	<50	50

TABLE 3. RESONATOR MEASUREMENTS ($x10^{-9}/g$)

Type (Comment)	S/N	Normal	In-plane	Total
One-Port (baseline)	Q1758-7	12	2.5	12
Two-Port (Cu-Al)	Q1773-6	25	1.2	25
Two-Port (Cu-Al)	Q1773-9	21	2.0	21
Two-Port (10,000Å busbars)	Q1823D	14	5.5	15
Two-Port (800Å transducers)	Q1824A	<2.0	5.0	5.0
Two-Port (10,000Å busbars)	Q1825A	<4.0	4.0	4.0
Two-Port (10,000Å busbars)	Q1825D	<6.0	6.0	8.0
Two-Port (10,000Å busbars)	Q1826B	9.0	2.0	9.0
Two-Port (10,000Å busbars)	Q1826D	9.0	2.5	9.5
Two-Port (400Å transducers)	Q1829A	9.5	2.2	9.5

INFLUENCE OF FABRICATION PARAMETERS

The results obtained on the two-port resonators may be used to study the influence of certain fabrication parameters on SAW device acceleration sensitivity, although caution is advised in drawing conclusions from a small sample lot.

Figures 3 through 5 show the variations in $|\Gamma|$ as a function of busbar thickness for units with 600Å thick transducers, including two units with Cu-doped Al transducers. We observe that as the busbar thickness is increased, the in-plane acceleration sensitivity is increased while the normal acceleration sensitivity is decreased, with the rate of improvement in normal sensitivity slightly larger than the rate of degradation of in-plane sensitivity. The in-plane sensitivity is typically less than the normal sensitivity for the units tested here. The trend implied by the data is for equal in-plane and normal components at a busbar thickness slightly in excess of 10,000Å.

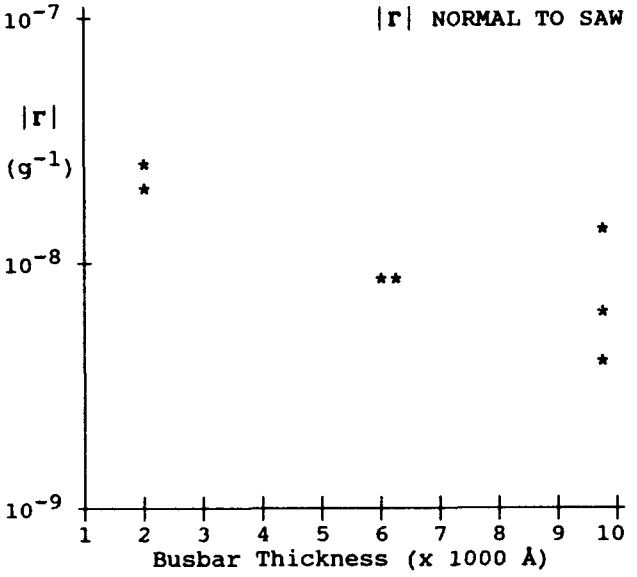


Figure 3. Two-port resonator acceleration sensitivity normal to the plane of the SAW device versus busbar thickness.

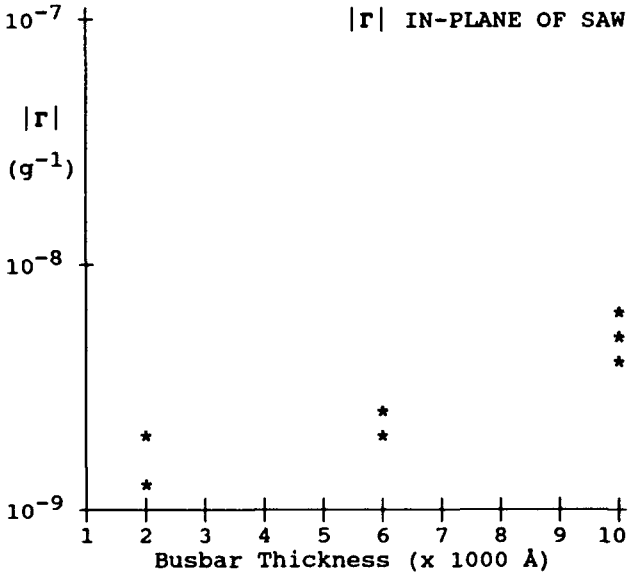


Figure 4. Two-port resonator acceleration sensitivity in the plane of the SAW device versus busbar thickness.

Figures 6 through 8 show the variations in $|\Gamma|$ as a function of transducer thickness for units with 2000Å thick busbars, including two units with Cu-doped Al transducers. Without further experimental samples, it is difficult to determine whether any trends are present or whether the large differences in $|\Gamma|$ between units with Al transducers and those with Cu-doped Al transducers are due to the metallization differences.

DISCUSSION

The devices tested here were developed for a low-noise radar application. The quiescent phase noise among the two-port devices of the various metallization types varied by 3dB, whereas the acceleration induced phase noise varied by 16dB. Unpublished data [2] indicating similar effects of similar magnitude for bulk wave plate resonators have recently been confirmed and theoretical models for the effect proposed [3-5]. Based on the theoretical and experimental work to date, it is suggested that careful control of transducer and busbar metallization may be used to optimize for acceleration induced phase noise without seriously compromising quiescent phase noise.

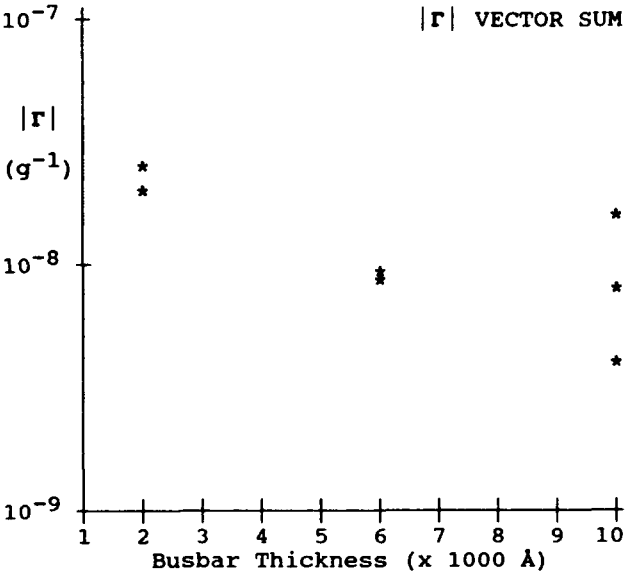


Figure 5. Two-port resonator acceleration sensitivity net vector sum versus busbar thickness.

REFERENCES

[1] T. Parker and J. Loan, "SAW Oscillators for Air Defense Systems," USA LABCOM Technical Report SLCET-TR-84-0401-F, U.S. Army Electronics Technology & Devices Laboratory, Fort Monmouth, N.J., 184 pp., Dec. 1986.

[2] J. Kosinski, USA ERADCOM Laboratory Notebook DELET-21, US Army Electronics Technology & Devices Laboratory, Fort Monmouth, N.J., 150pp., Jan. 1982-Aug. 1984.

[3] P. C. Y. Lee and X. Guo, "Effect of Electrodes on the Acceleration Sensitivity of Crystal Resonators," Proc. 43rd Annual Frequency Control Symposium, pp. 416-418, May-June 1989.

[4] 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 Annual Frequency Control Symposium, pp. 377-387, May-June 1989.

[5] E.P. EerNisse, R.W. Ward, and O.L. Wood, "Acceleration-Induced Frequency Shifts in Quartz Resonators," Proc. 43rd Annual Frequency Control Symposium, pp. 388-395, May-June 1989.

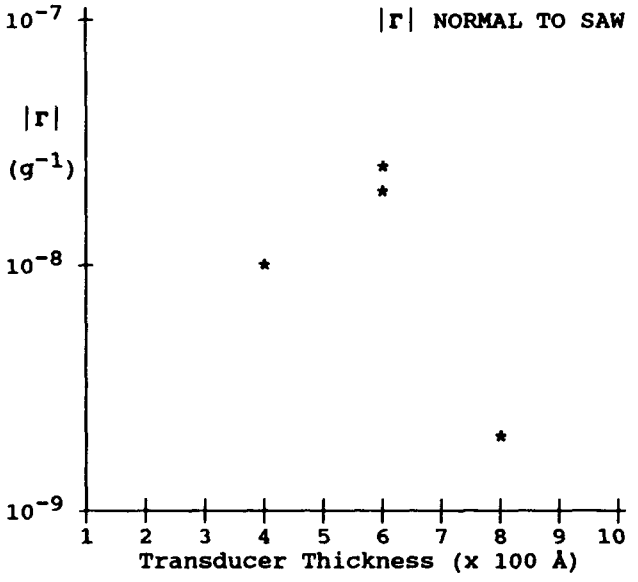


Figure 6. Two-port resonator acceleration sensitivity normal to the plane of the SAW device versus transducer thickness.

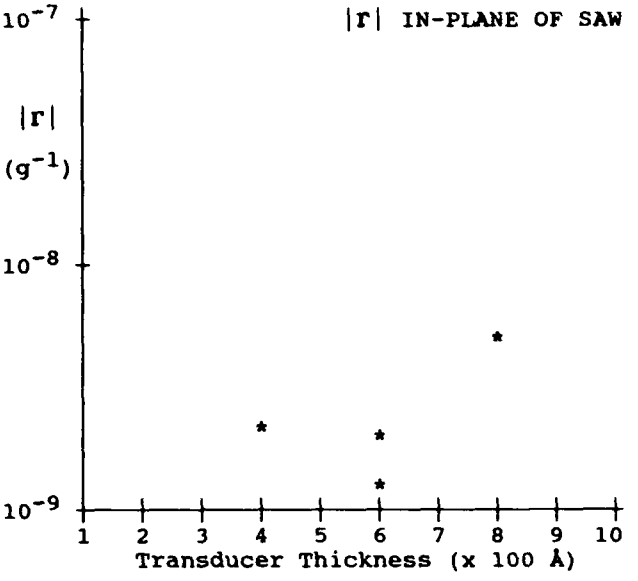


Figure 7. Two-port resonator acceleration sensitivity in the plane of the SAW device versus transducer thickness.

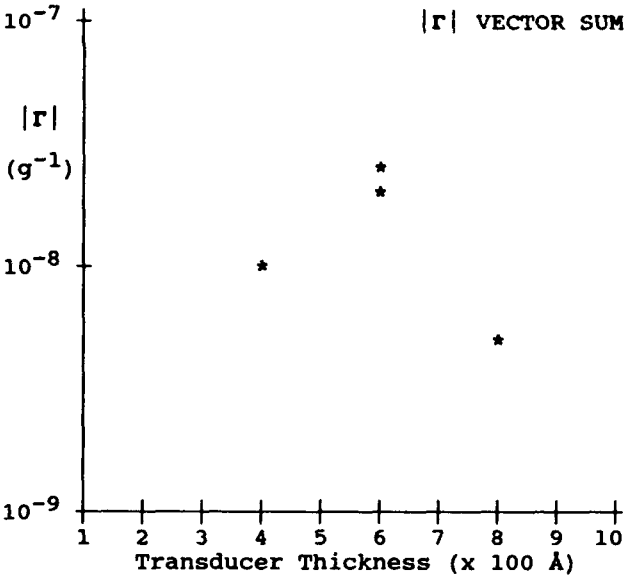


Figure 8. Two-port resonator acceleration sensitivity net vector sum versus transducer thickness.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

PHASE NOISE MEASUREMENTS ON VIBRATING CRYSTAL FILTERS

R.L. Clark

Penn State University, York, PA 17403

ABSTRACT

It is well known that the resonant frequencies of bulk and surface acoustic wave resonators can be slightly perturbed by external acceleration forces. The resulting phase modulation which occurs when the resonator is used in a filter is reviewed in terms of an effective acceleration sensitivity vector for the filter. The degradation in spectral purity is expressed as a function of this sensitivity. The phase shift through the filter is then quantified in terms of the acceleration sensitivity and the filter group delay. By considering typical values of acceleration sensitivity for bulk-wave resonators, estimates of the phase noise induced by the filter can be made. Measurement results are presented for phase noise degradation in sine and random vibration. Measured results on a 5 MHz signal passed through a crystal filter show an increase in $\mathcal{L}$ of greater than 20 dB in random vibration over static values. The degradation in Allan variance is also discussed.

I. INTRODUCTION

The resonant frequency of mechanical resonators can be slightly perturbed by external acceleration forces [1]. If the resonator is used as a circuit element, then the electrical performance becomes a function of acceleration forces. This effect is well known and has been thoroughly investigated for the case where the resonator is used to stabilize the frequency of an oscillator [2],[3]. If the resonator is used as a frequency selective filter, then the transfer function of the filter must also become a function of external acceleration forces. In the case of a vibrating filter, the transfer function becomes a function of time. Since the equivalent circuit for the resonator is varying with time, a rigorous solution of the output spectrum for this case would then involve the solution of a differential equation with varying coefficients. This often is not a practical approach. Because the resonator shift is a small percentage of the actual frequency, a numerical solution is also not trivial.

By estimating the phase shift due to acceleration, expressions for the ensuing reduction in frequency stability for the case of a sinusoidal input signal have been developed [4]. These expressions are briefly reviewed below.

II. FREQUENCY STABILITY DEGRADATION

For the case of a sinusoidal input signal to a crystal filter, a small change in the center frequency of the filter will result in a phase shift, $\Delta\phi$, of

$$\Delta\phi = 2\pi\frac{d\phi}{d\omega}\Delta F = 2\pi\tau_g\Delta F \qquad (1)$$

where ΔF is the change in the center frequency of the filter and τ_g is the filter group delay. With an estimate of phase shift, the traditional measures of frequency stability can be analyzed. In the frequency domain, the expression for sinusoidal vibration becomes

$$\mathcal{L}_{dB}(f) = 20\log\left[\pi\tau_g F_o \vec{\Gamma}_{eff}\cdot\vec{A}_p(f)\right] \qquad (2)$$

where $\vec{\Gamma}_{eff}$ denotes the effective G-sensitivity of the filter. Note

CH2818-3/90/0000-493 $1.00 © 1990 IEEE

that the 20 dB per decade rolloff which occurs in crystal oscillators does not occur in crystal filters. Since the acceleration sensitivity vector of the filter may differ from that of the individual resonators, an effective acceleration sensitivity is defined to distinguish the two quantities.

For the case of random vibration, the induced noise may be estimated from

$$\mathcal{L}_{dB}(f) = 20\log\left[\pi\tau_g F_o \vec{\Gamma}_{eff}\cdot\vec{u}\sqrt{2S_g(f)}\right] \quad (3)$$

where S_g is the spectral density of vibration in G^2/Hz and $\vec{u}$ is a unity vector in the direction of acceleration.

The reduction in frequency stability in the time domain is given by

$$\Delta\sigma_y(\tau) = \frac{2\tau_g \vec{\Gamma}_{eff}\cdot\vec{A}_p}{\tau}\sin^2(\pi f_v) \quad (4)$$

A subtle point which must be noted in these expressions is that τ_g is the group delay at the input frequency of the signal, not at the induced noise offset frequency. Thus, in the case of random vibration with white frequency distribution, one would expect that the induced noise spectrum would not vary significantly with offset frequency, (ie. "flat" vibration input causes "flat" induced noise spectrum). In practice, this is a good approximation for single crystal filters but some variation is seen for multiple resonators. This is not suprising, considering that each resonator acts, in a sense, as a filter for signals induced by other resonators. In addition, vibration resonances may cause significant variations in the noise response.

III. PHASE NOISE MEASUREMENTS

Phase noise measurements on a vibrating filter can be performed in the same manner as any two port measurement. [5] Of course, one must take care to ensure that fixtures, cabling and other elements do not contribute excessive phase modulation. Two methods are commonly used to measure the phase noise of two port devices. Figure 1 shows a setup which can be used when two filters are available, and Figure 2 can be used for one device. The mixer and baseband analysis setup used in the measurements described below was an HP 3048A phase noise test system. Path lengths were adjusted to achieve quadrature at the mixer inputs.

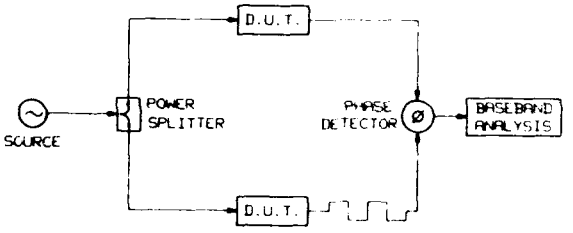


Figure 1. Typical test configuration for two filters.

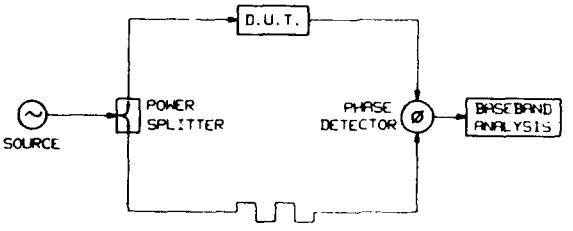


Figure 2. Typical test configuration for one filter.

The configuration of Figure 2 provides a suppression of the source phase noise of [5]

$$Att(dB) = -20\log\left|2\sin(\pi f\tau_g)\right| \quad (5)$$

Figure 3 shows that the attenuation of the

reference oscillator noise is quite effective for small offset frequencies. Note that since $\sin\theta \approx \theta$ for small angles, the source noise attenuation decreases by 20 dB per decade for offset frequencies much less than $1/2\pi\tau_g$. Since the noise of an oscillator decreases by 10 dB per decade at small offsets, the noise spectrum shows a 1/f characteristic, or a 10 dB per decade slope.

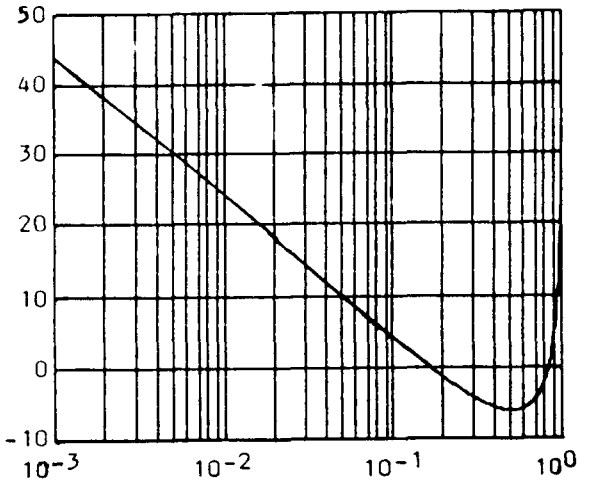


Figure 3. Source attenuation vs. product of group delay-offset frequency.

The use of two filters provides an additional advantage in that the effective group delay for source noise suppression is the difference in group delay between the two signal paths. This advantage reduces the source noise out tc considerably higher offset frequencies.

IV. MEASURED RESULTS

To verify the measurement apparatus and cabling, measurements were taken under static and vibration conditions with the filter removed. Figure 4. shows the measured results for both cases. Since the environment where the measurements were performed was quite noisy, no attempts were made to suppress the floor noise. The actual floor noise of the filters was considerably lower than the noise floor of the system used for vibration testing. It can be seen that the contribution of the cables was negligible however.

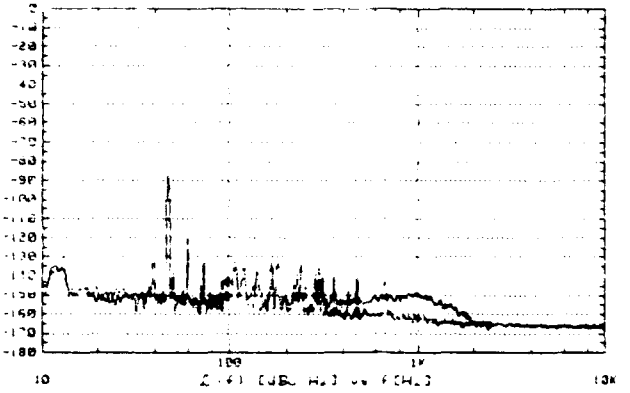


Figure 4. Static and vibrating cables only.

Figure 5. shows measured sideband level vs. vibration amplitude for a 5 MHz crystal filter with a group delay of 147 μS at the center frequency. The calculated value of effective acceleration sensitivity for this filter is $2.73 \cdot 10^{-9}$. This is not an unreasonable value considering that the resonators are AT cut and not optimized for G-sensitivity.

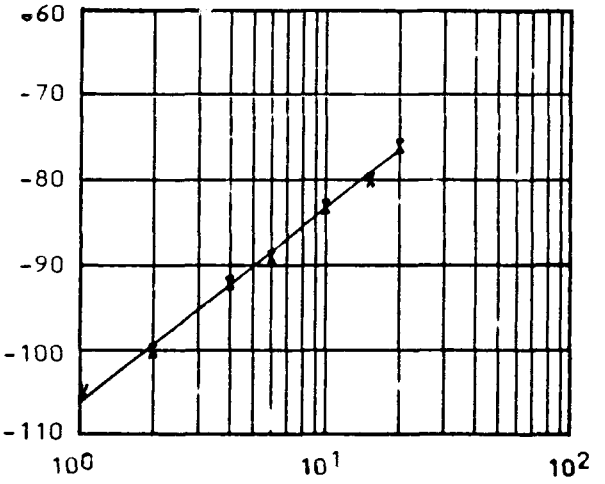


Figure 5. Measured sideband levels vs. acceleration amplitude for a 5 MHz filter.

Figure 6 is a plot of phase noise during random vibration in two different axis of a 5 MHz filter. The vector properties of acceleration sensitivity are obvious.

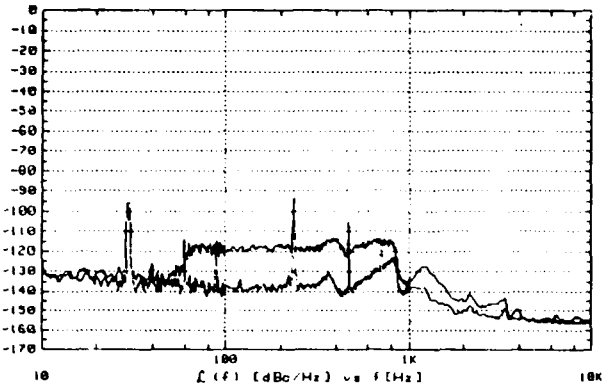


Figure 6. Phase noise in two orthogonal directions for a crystal filter.

A test was performed on a 10.25 MHz monolithic filter with the source frequency set to the edge of the filter passband. In this particular case, the input signal was 10.25045 MHz. The filter response and phase noise characteristics are shown in Figures 7. Figure 8 shows the phase noise with a random vibration input of .125 G^2/Hz over 400 to 1200 Hz. No suppression of noise outside of the passband could be observed for this filter. This result should not be generalized, however, since in multicrystal filters, resonators closer to the output may provide some suppression of noise induced by other resonators. For this case, the effective acceleration sesitivity vector will vary somewhat with offset frequency.

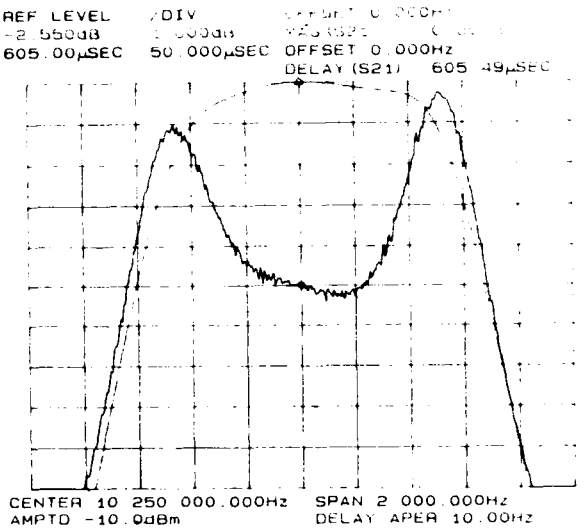


Figure 7. Filter frequency response and group delay.

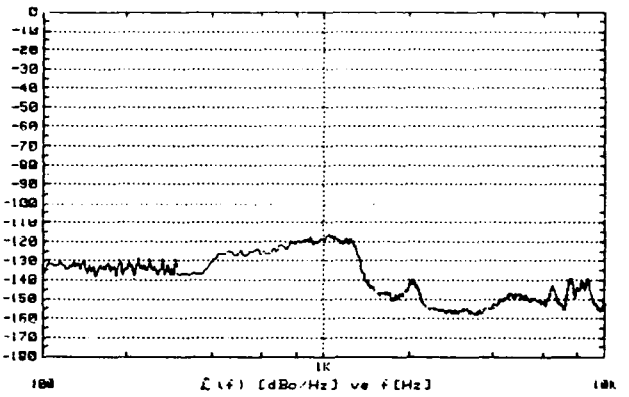


Figure 8. Noise performance during random vibration for the filter of Figure 7.

The spurious frequencies at higher offset frequecies where repeatable and not present at low vibration levels. A plausible explanation for this is that nonlinear mechanisms in the mechanical support structure excite high order, high Q, mechanical resonances during random vibration. The author has also observed this phenomenon in oscillators under high level random vibration.

V. Conclusion

It has been shown that crystal filters (BAW and SAW) may induce phase modulation during mechanical vibration. Estimates for the resulting degradation in frequency stability of a sinusoidal input in the time and frequency domain have been reviewed. Measured results of the phase modulation correlate with typical values for AT cut resonators.

Acknowledgements

The author wishes to acknowledge valuable discussions with D. Junk and E. Smith of McCoy Electronics, along with the assistance of J. Lewis in performing the measurements in the environmental laboratory at McCoy Electronics.

References

[1] M. Valdois, J. Besson, and J.J. Gagnepain, "Influence on Environment Conditions on a Quartz Resonator," Proc. of 28th Ann. Symp on Freq. Cont. 1974, pp. 19-32.

[2] R.L. Filler, "The Effect of Vibration on Frequency Standards and Clocks," Proc. of 35th Ann. Symp. on Freq. Cont., 1981, pp. 31-39.

[3] R.L. Filler, "The Acceleration Sensitivity of Quartz Oscillators: A Review," in IEEE Trans. Ultrason. Ferroelec. Preq. Cont., UFFC-35, No. 3, pp. 297-305, May, 1988.

[4] R.L. Clark, "Spurious Signals Induced By Vibration of Crystal Filters", IEEE Ultra. Symp. Proc., Vol. I., 1988, pp. 365-368.

[5] T. R. Faulkner, R. E. Temple, "Residual Phase Noise and AM Noise Measurements and Techniques", Hewlett Packard Part No. 03048-90011.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SPECTRAL PURITY OF DIRECT DIGITAL FREQUENCY SYNTHESIZERS

VENCESLAV F. KROUPA

Czechoslovak Academy of Sciences
Praha, 182 51 Czechoslovakia

Abstract.

We start with a short introduction to the problems of frequency synthesis and proceed by investigating mathematical models suitable for DDFS which provide quasi-periodic omission of clock pulses. In that case the spurious phase-time modulation does not exceed one T_i .Its simplification for easier computation of major spectral lines for pulse rate and square wave outputs is suggested. Finally, truncation of LSB's in sine wave output is considered.

I. Introduction.

In the last 25 years frequency synthesizers replaced single frequency oscillators practically in all fields of applications: in communications, in measurement instruments, in radars, in home radio sets, etc. [1-6]. In addition they are encountered throughout the whole range of electromagnetic waves from audio and RF to microwaves and optical frequencies (Fig.1).

In the present paper we are concerned with Direct Digital Frequency Synthesizers (DDFS) the output frequency of which extends from very low frequencies up to tens of MHz and lately even exceeding one hundred MHz.

Their major advantages are short switching times, very small frequency steps, use of digital IC on a large scale, compatibility with the microprocessor techniques, etc.

However, there are also disadvantages, namely, a rather limited range of maximum output frequencies and often a bad spectral purity. The last problem has been discussed by many authors [7-14] from different points of view in the past. Here we want to provide some new insights.

II. Basic mathematical approach to frequency synthesis.

An analysis of working modes reveals that all systems employ a common software algorithm i.e. a step by step approximation of normalized output frequencies.

In any case we have to start by normalizing the output frequency or frequencies with respect to the input standard frequency, i.e. to find out the set

$$x = \langle f_x/f_i \rangle \qquad (1)$$

In practice, x is generally a rational fraction either "ex decisione" or because its accuracy is limited by practical reasons.

CH2818-3/90/0000-498 $1.00 © 1990 IEEE

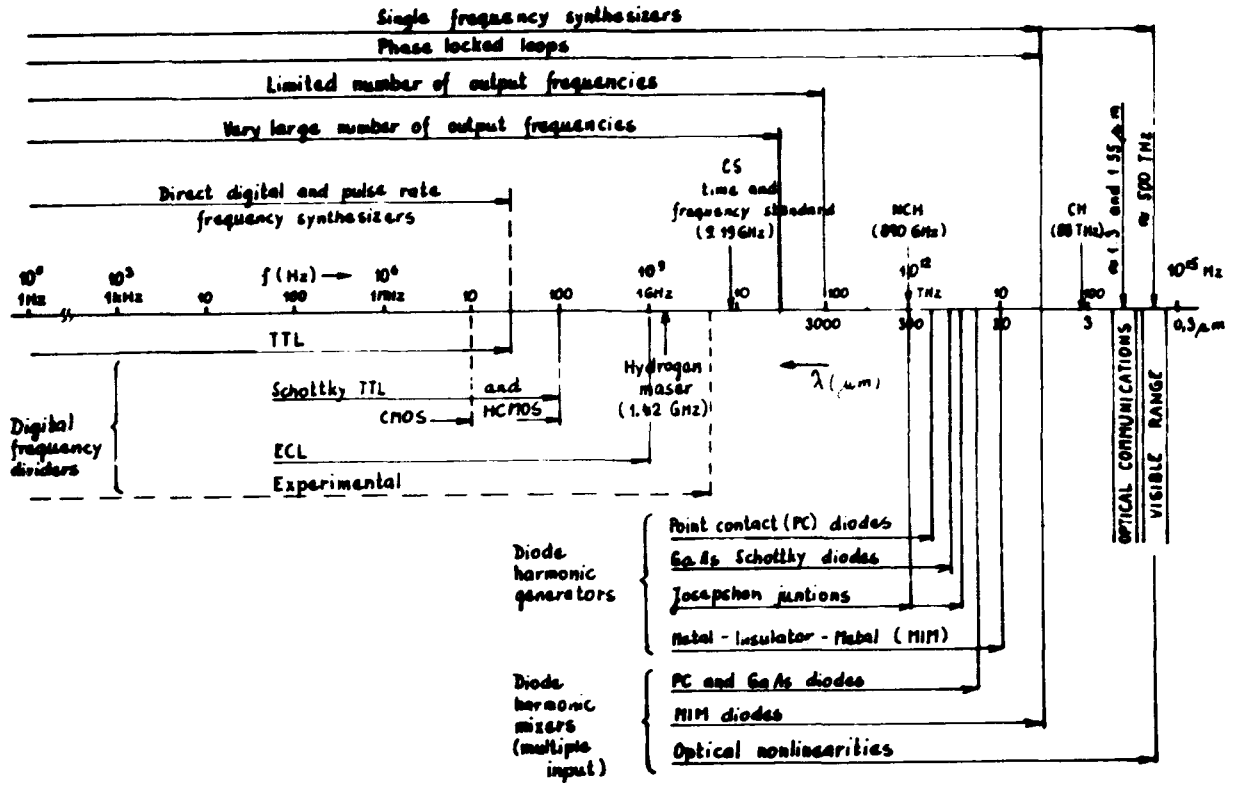


Fig.1 Present situation of frequency synthesis techniques in electromagnetic wave ranges from audio frequencies till the visible wavelengths.

In the first step we have to choose the actual approximation process which is modified by different limitations caused by the number of output frequencies, the required frequency range, the used hardware, the desired spectral purity, the acceptable cost etc.

Let us briefly recall the basic mathematical models suitable for the frequency synthesis software.

A/ Series expansions:
Cantor series,
synthesizer series [1],
systematic fractions,
modified Engel series [15];
B/ continued fraction expansions;
C/ product expansions;
D/ Modulo-N arithmetics.

III. Mathematical models suitable for DDFS's.

A/ One single output frequency.

If we need one single output frequency only it is a better choice to use a special software since in these instances we can synthesize the output frequency with no or extremely small error and in addition we get the smallest possible content of spurious side bands. The difficulty is that we are not able to use DDFS chips from the shelf.

The leading idea is to remove pulses in a quasiperiodic manner from a clock pulse train. There are two approximation softwares which make possible the

process: continued fraction and modified Engel series expansion [15].

A/ One single output frequency.

If we need one single output frequency only it is a better choice to use a special software since in these instances we can synthesize the output frequency with no or extremely small error and in addition we get the smallest possible content of spurious side bands. The difficulty is that we are not able to use DDFS chips from the shelf.

The leading idea is to remove pulses in a quasiperiodic manner from a clock pulse train. There are two approximation softwares which make possible the process: continued fraction and modified Engel series expansion [15].

1/ Continued fraction expansion.

By assuming that the normalized output frequency is given by

$$f_x = f_o = X/Y \qquad (2)$$

where both X and Y are relatively prime integers we proceed the step by step approximation as follows

$$f_o = b_o + 1/f_1 \; ; \; b_o = \text{int}\, f_o \qquad (3)$$

$$f_1 = b_1 + 1/f_2 \; ; \; b_1 = \text{int}\, f_1 \qquad (4)$$

etc.

$$f_o = b_o + \cfrac{1}{b_1 + \cfrac{1}{b_2 + \cfrac{1}{\ddots}}} \qquad (5)$$

The block diagram of the necessary hardware is shown in Fig. 2.

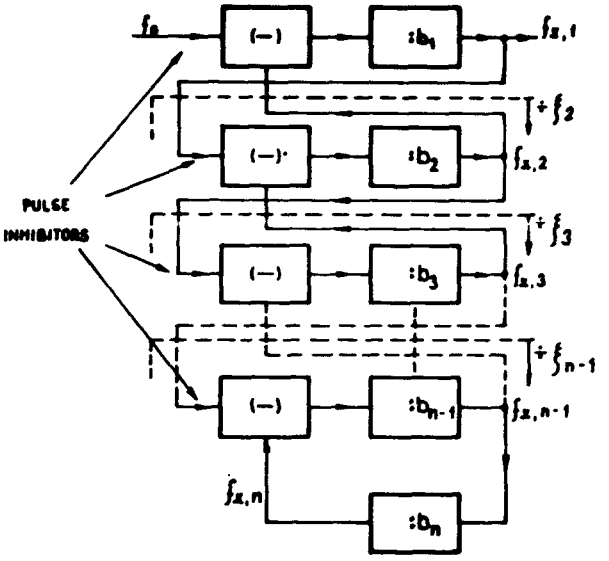


Fig.2 Block diagram of a pulse rate DDFS based on continued fraction expansion.

2/ Modified Engel series expansion.

This step by step approximation has been successfully used as a mathematical model for DDFS. The algorithm applied on the normalized frequency is as follows:

$$f_o = \gamma_o = c_o - 1/\gamma_1 \qquad (6a)$$

$$c_o = \text{int}(\gamma_o + 1) \qquad (6b)$$

$$\gamma_1 = 1/(c_o - \gamma_o) \qquad (7)$$

and for $k \geq 2$

$$\gamma_{k+1} = \gamma_k/(\gamma_k - q_k) \qquad (8)$$

where

$$q_k = \text{int}(\gamma_k) \quad k = 1,2,.. \qquad (9)$$

and the final expansion is given by

$$\gamma_o = c_o - \frac{1}{q_1}\left(1 - \frac{1}{q_2}\left(1 - \ldots\right)\ldots\right)$$

$$= c_o - \sum_{n=1} \frac{(-1)^{n+1}}{q_1 q_2 \ldots q_n} \qquad (10)$$

The application of this expansion is illustrated in Fig. 3 where a block diagram for generation of the half of color carrier in the European PAL TV system is shown.

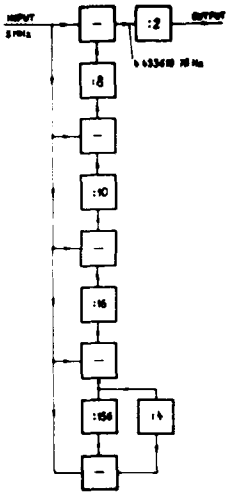


Fig.3 Block diagram of a DDFS based on Modified Engel series expansion (the numerical example shows generation of the European PAL color subcarrier).

B/ Very large number of output frequencies.

Frequency synthesizers with one output frequency are required only in rare instances. Whereas devices with variable output frequencies are encountered much more often and DDFS's based on IC may simplify the hard-ware considerably and at the same time provide very small tuning steps. Low frequency synthesizers from "DC" to some 30 to 130 MHz are based directly on Modulo-N arithmetics. However, in instances with limited or higher frequency ranges we take recourse to systems with fractional frequency dividers.

1/ Modulo-N arithmetics.

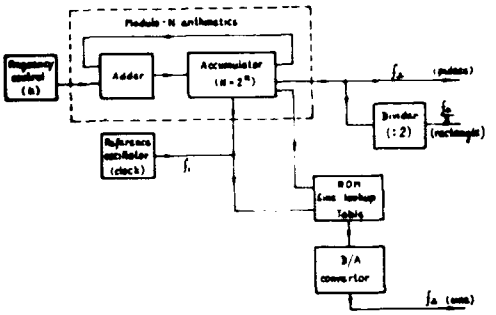


Fig.4 Block diagram of a DDFS based on modulo-N arithmetics.

The principle can be deduced from Fig.4. At each cycle or pulse of the clock signal a number , k , which represents a phase increment, is added to that stored in the accumulator-register. Every time the register overflows an output pulse is generated and a new period of accumulation is started, however, with a modulo - N remainder (smaller than k) in the memory. As a consequence the output pulse rate has the mean frequency

$$f_x = \frac{k}{N} f_i \qquad (11)$$

which can easily be changed by reprograming the "word" k in the steering memory block - see Fig.4 Further, in practical devices N is generally equal to a large

power of 2 [12, 14 etc.], i.e.

$$N = 2^R \qquad (12)$$

By starting with zero in the accumulator we arrive after m_1 clock pulses at its first overflow. As a consequence we have

$$m_1 k - N < k \qquad (13a)$$

For the second overflow we get

$$m_2 k - 2N < k \qquad (13b)$$

etc. (cf. Fig. 5a) till

$$m_n k - nN = 0 \quad (n=k) \qquad (13c)$$

Generalization reveals

$$mk - rN < k \quad (r<k) \qquad (14)$$

from which

$$m - r\frac{N}{k} = \varepsilon \qquad (15)$$

where

$$1 > \varepsilon \geq 0 \qquad (16)$$

since m is an integer both (15) and are (17) met when

$$\mathrm{int}\left(m - r\frac{N}{k}\right) = 0 \qquad (17a)$$

or

$$m = \mathrm{int}\left(r\frac{N}{k}\right) \qquad (17b)$$

After multiplying (15) by the clock period we find that the spurious phase-time modulation does not exceed one clock period T_i .As a consequence modulo-N arithmetics also provides mathematicla model for DDFS's meeting

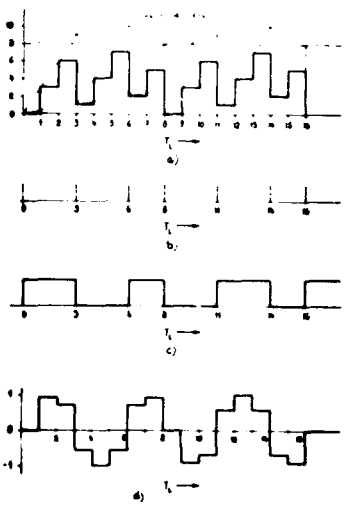


Fig.5 Waveforms of DDFS for $f_x = 3/16$
a) accumulator contents;
b) pulse rate output (dashed additional pulses are for $f_x = 3/8$);
c) square wave output;
d) sine wave output.

the requirement for quasiperiodic omission of pulses in the steering pulse train.

2/ Fractional frequency dividers.

Fractional frequency dividers make it possible to generate very small tuning frequency steps in a limited frequency range. The principle is shown in Fig. 6. The basic circuit is a phase locked loop (PLL) with a pulse inhibitor and a frequency divider in the feedback path.

Actually we subtract from the output frequency f_x its fraction generated in a DDFS from the reference frequency f_r, i.e. we apply the "modified Engel series" algorithm and get

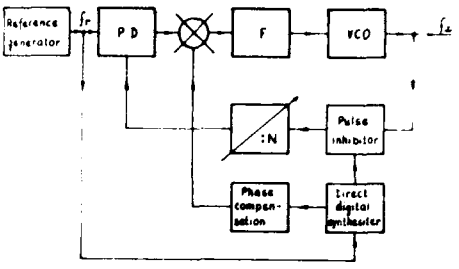


Fig.6 Phase lock loop with a fractional frequency divider.

$$(f_x - f_r\frac{k}{2^R})/N = f_r \qquad (18)$$

since

$$f_x/f_r = N_{eff} \qquad (19)$$

the effective division ratio is

$$N_{eff} = N + \frac{k}{2^R} \qquad (20)$$

The pulse inhibitor is easily realized with the assistance of dual moduls divider the division ratio of which is changed for one clock period from N and N+1 in the rate of the output pulse rate frequency supplied by the DDFS.

IV. Spurious side-bands in DDFS.

In all frequency synthesizers with quasiperiodic omission of pulses we encounter spurious phase-time modulation.

In this section we shall investigate three different output waves met in practice and shown in Fig. 5 for a simple case of the normalized output frequency

$$f_x = 3/16$$

In systems with fractional frequency dividers the quasiperiodic pulse rate output of DDFS is sufficient. However, this is not sufficient in instances where the output of DDFS is the end product. In these cases we have to enhance the fundamental output frequency by generating either a square wave or sine wave output.

A/ Pulse train output.

The quasiperiodic omission of pulses introduces a phase-time modulation the amplitude of which does not exceed one clock period T_i. This condition can be written as

$$s(t_r) = mT_i - rT_x < T_i \qquad (21)$$

After introducing results found for modulo-N arithmetics we get

$$s(t_r) = T_i[r\frac{N}{k} - \mathrm{int}(r\frac{N}{k})] \qquad (22)$$

By plotting the above equation for the simple case of $f_x = 7/8$ we find that function $s(t_r)$ can be idealized to a sawtooth wave as illustrated in Fig. 7.

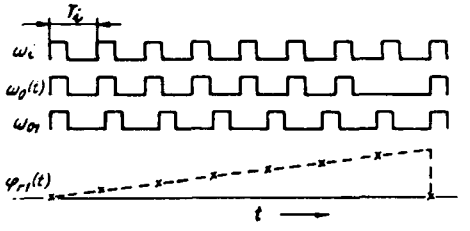


Fig.7 Idealized phase-time modulation $s(t_r)$ for the normalized frequency $f_x = 7/8$.

In a more general case where both X and Y in (2) are larger integers we get the behaviour of the modulation function as shown

in Fig. 8. Note that the plot reveals superposition of several sawtooth waves. We shall discuss the problem in depth in the next section.

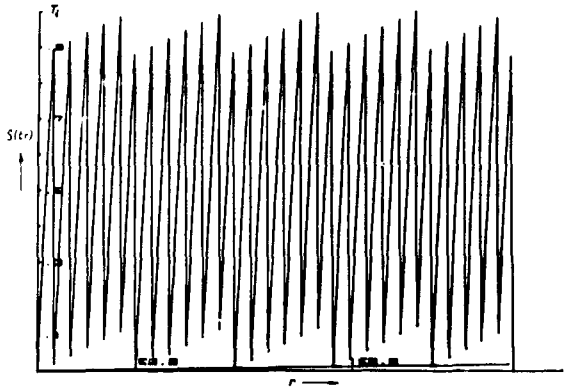


Fig.8 Plot of the phase-time modulation $s(t_r)$ for the 3rd approximation of the PAL color subcarrier i.e. for $f_x = 256/227$.

Reverting to Fig.5a and eqs. (15) and (17) we easily arrive at the conclusion that information about $s(t_r)$ is stored in the accumulator and can be used for compensation of spurious modulation in frequency synthesizers using fractional frequency dividers with the assistance of an analog phase interpolator (API - system). See e.g.[5 pp. 136-140].

Recently B. Miller and B. Conley suggested a network pushing the spectral lines close to the carrier out of the pass-band of the phase locked loop [16].

B/ Square wave output.

In order to enhance the fundamental harmonic in the output signal, the rectangular output wave should have the space-mark ratio as close to 1:1 as possible. This requirement can be met by different means, the simplest of which is to put a flip-flop or binary divider at the end of the synthesizing chain. In this instance both leading and trailing edges of the output wave are synchronized by the, say, leading edges of the original pulse train. As a consequence its frequency must be twice the desired frequency f_x i.e.

$$f_{xp} = \frac{2X}{Y} f_i \qquad (23)$$

where X and Y are integers having no common divisor.

In this case the phase time modulation is again

$$s(t_r) = k\frac{f_i}{f_x}T_i - rT_i < T_i \qquad (24)$$

and can be rewritten after application of (2) to

$$s(t_k) = T_i[k\frac{Y}{2X} - \mathrm{int}(k\frac{Y}{2X})] \qquad (25)$$

The actual spectrum can be easily calculated with the assistance of the Fourier analysis of the rectangular output wave with zero crossings in accordance with (25), cf. Fig. 5c. For the cosine coefficient of the n-th harmonic of the k-th rectangle we have

$$a_{nk} = \frac{2}{YT_i}A \int_{T_i\mathrm{int}(kY/2X)}^{T_i\mathrm{int}[(k+1)Y/2X]} \cos(n\frac{2}{YT_i}t)dt \qquad (26)$$

and for the n-th harmonic of all X rectangles

$$a_n = \frac{A}{\pi n}\sum_{k=0,1,2\ldots}^{2X-1}(-1)^{k+1}\sin\left[\frac{2\pi n}{Y}\,\mathrm{int}\left(\frac{kY}{2X}\right)\right] \qquad (27)$$

and similarly for the sine coefficient

$$b_n = \frac{A}{\pi n}\sum_{k=0,1,2\ldots}^{2X-1}(-1)^{k}\cos\left[\frac{2\pi n}{Y}\,\mathrm{int}\left(\frac{kY}{2X}\right)\right] \qquad (28)$$

Evidently, for evaluating the output spectrum, in instances where both X and Y are large, the use of a computer is inevitable. It is true that the respective software is very simple; however, the computation time can be considerable, and at the same time, one may lose orientation in the wealth of data. The remedy provides approximation of Y/2X with a ratio of smaller integers by applying the shortened version of continued fraction expansion [17] (see also Appendix). By its introduction together with eq. (23) into (25) we get

$$s(t_k) = T_i\left[k\left(A_o + \frac{a_1}{B_1} - \frac{a_1a_2}{B_1B_2} + \ldots\right) - \mathrm{int}\left(k\frac{Y}{2X}\right)\right] \qquad (29)$$

At the first order approximation of Y/2X by A_1/B_1 the fundamental $s(t_k)$ is a simple sawtooth wave with the amplitude T_i. The second order approximation reveals another sawtooth wave with the amplitude T_i/B_1 and the new modulation period A_2T_i which is also imposed on the first sawtooth wave. The same happens with every other approximation, see Fig. 8.

The above discussion of (29) leads to the following conclusions:

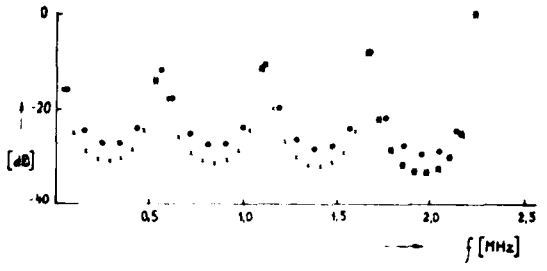
1) There are always present components caused by the sawtooth waves with large amplitudes: T_i, T_i/B_1, etc. (see Fig. 8 and 9).
2) To get the knowledge about the approximate distribution of major spectral lines it is sufficient to choose an appropriate approximation A_r/B_r of Y/2X with B_r small enough to keep the computation time in reasonable limits. See spurious signal levels in DDFS in Fig. 9 for color subcarrier from Fig. 3.
3) Due to the above mentioned intermodulation between individual sawtooth waves there are peaks of spurious signal around the major ones (see Fig. 9).
4) The power level of the spurious components closest to the carrier supplies the last sawtooth wave with the amplitude T_i/B_{n-1}.
5) In instance where Y is even all even spurious harmonics are missing.

From (29) we get for the spurious signal levels "near the carrier"

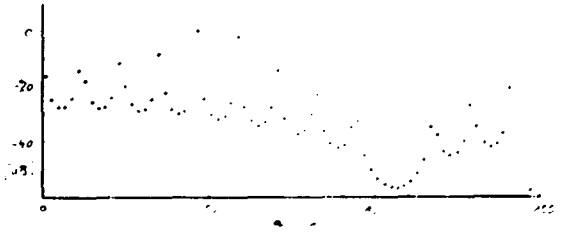
1-st order approximation

$$20\log\frac{X}{n_1Y} \quad (n_1 = 1,2,..) \qquad (30)$$

with the modulation frequency



a)



b)

Fig.9 Spurious spectral lines of DDFS for the PAL color subcarrier as found by the first (o) (A_1/B_1=8/9) and second (o) (A_2/B_2=47/53) order approximations and theoretical (X) and experimental (o) results of a truncated Modified Engel series expansion.
a) In the frequency range from 0 to f_x;
b)In the frequency range from 0 to $3f_x$

$$f_{m1} = f_i/A_1 \qquad (31)$$

2-nd order approximation

$$20\log\frac{X}{n_2 Y B_1} \quad (n_2 = 1,2,..) \qquad (32)$$

with

$$f_{m2} = f_i/A_2 \qquad (33)$$

etc. till the smallest spurious signals with amplitudes

$$20\log\frac{X}{Y}\cdot\frac{1}{n_{n-1}B_{n-1}} \quad (n_{n-1} = 1,2..) \qquad (34)$$

and

$$f_{m,n-1} = f_i/A \qquad (35)$$

(Note: the above approximate relation for the level of spurious signals are valid only for small values of n_1, n_2, etc. They supply the better results the smaller is the ratio X/Y or X/YB_k. Since in these instances we take into account the first harmonic of the respective sawtooth wave and ap-proximate the Bessel function $J_1(z)$ by $z/2$).

These, sometimes large spurious signals present the major inconvenience of DDFS systems. However, there are several means for alleviation of the problem.

In single frequency synthesizers the application of an PLL output filter provides the best solution (cf. Fig.3). However,in these instances we need to know the level of spurious signals very close to the carrier. For the above mentioned example of the PAL color carrier with

$$f_x = 2\ 216\ 809.375/5*10^5$$

we get for the last but one (i.e.

the 8th) approximation for

f_{m8} = 40 Hz the sp. level -99dB

and for the 9th approximation

f_{m9} = 6.25 Hz the sp.lev. -108dB

The preceding example is intended to call attention of designers to check spurious signals close to the carrier withe assistance of relations (34), (35) etc. and not to be content only with experimental results fund by analyzer.

C.E. Wheatley and D.E. Phillips [10] proposed to add into the accumulator, at each clock pulse a random number, smaller than k, and in this way to smear out large coherent spurious signals.

B/ Sine wave output.

An efficient suppression of spurious side-bands provides the sine wave output.

Since the number stored in the accumulator of the modulo-N device contains information about the instantaneous phase it can be changed into a sine wave with the assistance of ROM look-up table and a digital to analogue converter (cf. Fig. 4). By having recourse to the Fourier analysis of each rectangle in the output staircase curve (cf. Fig. 5d) we get

$$a_{nm} = \frac{2}{YT_i}\int_{mT_i}^{(m+1)T_i} \sin\left(2\pi m\frac{X}{Y}\right)\cos\left(2\pi n\frac{t}{YT_i}\right)dt \qquad (36)$$

where

$$\sin\left(2\pi m\frac{X}{Y}\right) \qquad (37)$$

are amplitudes of individual rectangles.

After performing the integration and applying some trigonometric rearrangements we arrive at

$$a_{nm} = \frac{2}{\pi n}\sin\left(\pi\frac{n}{Y}\right)\sum_{m=1,2,..}^{Y-1} \cos\left(\pi n\frac{2m+1}{Y}\right)\sin\left(2\pi m\frac{X}{Y}\right) \qquad (38)$$

and finally for the n-th harmonic of all Y rectangles at

$$a_n = \frac{2}{\pi n}\sin\left(\pi\frac{n}{Y}\right)\sum_{m=0,1,..}^{Y-1} \cos\left(\pi n\frac{2m+1}{Y}\right)\sin\left(2\pi m\frac{X}{Y}\right) \qquad (39)$$

and similarly for the sine coefficients

$$b_n = \frac{2}{\pi n}\sin\left(\pi\frac{X}{Y}\right)\sum_{m=0,1,..}^{Y-1} \sin\left(\pi n\frac{2m+1}{Y}\right)\sin\left(2\pi m\frac{X}{Y}\right) \qquad (40)$$

Investigation of both equations

(39) and (40) reveals that ideally all spurious signals disappear. Unfortunately the real life is not such easy. The one difficulty is that ROM look-up tables refer only to several most significant bits and disregard a lot of least significant bits. The other source of errors is the digital to analog converter which has a limited bit size and precision.

Let us first investigate the former problem when from R bits in eq. (12) we disregard B LSB's. As a consequence we check the output sine wave in effectively

$$Y/2^B \tag{41}$$

points only with the resolution of individual amplitudes reduced to

$$\sin(2\pi \frac{mX/2^B}{Y/2^B}) \tag{42}$$

Since we are referring only to integers of the factor $mX/2^B$ in the numerator the argument of the m dependent sine factors in relations (39) and (40) changes to

$$\sin[2\pi \frac{2^B}{Y} \mathrm{int}(\frac{mX}{2^B})] = \tag{43a}$$

$$\sin[2\pi \frac{2^B}{Y}(\frac{mX}{2^B} - \varepsilon_m)] \tag{43b}$$

However, since

$$\varepsilon_m < 1 \tag{44}$$

we can assume

$$\frac{2^B}{Y}\varepsilon_m << 1 \tag{45}$$

and relation (43) can be simplified to

$$\sin\frac{2\pi mX}{Y} - 2\pi\frac{2^B}{Y}\varepsilon_m\cos\frac{2\pi mX}{Y} \tag{46}$$

As an example we show for the normalized frequency

$$f_x = 7/32$$

the outputwave forms for increasing B in Fig. 10.

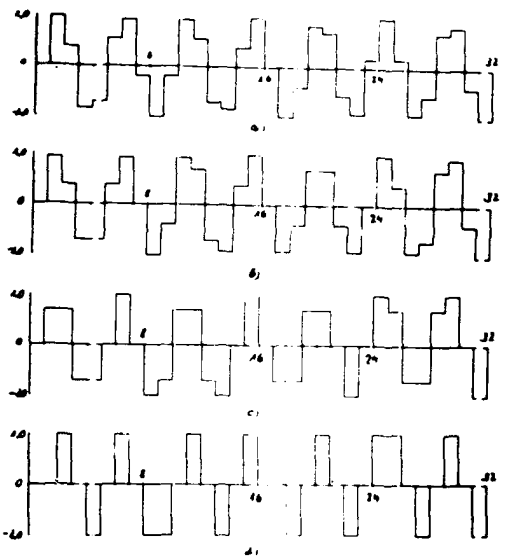


fig.10 Output waveforms of DDFS for different B of discarded LSB: a) B=0; b) B=1 c) B=2; d) B=3.

For B=R-1 we get the same results as for the rectangular wave output. See also Tab.1 for computer simulated results.

X= 7	Y= 32	B= 6	X= 7	Y= 32	B= 2
N= 7	(A<DB>= 0		N= 7	(A<DB>= 0	
N= 6	(A<DB>=-224.51		N= 6	(A<DB>=-223.64	
N= 5	(A<DB>=-230.33		N= 5	(A<DB>=-226.59	
N= 4	(A<DB>=-240.01		N= 4	(A<DB>=-251.1	
N= 3	(A<DB>=-230.86		N= 3	(A<DB>=-240.92	
N= 2	(A<DB>=-222.03		N= 2	(A<DB>=-226.34	
N= 1	(A<DB>=-213.22		N= 1	(A<DB>=-17.26	

X= 7	Y= 32	B= 1	X= 7	Y= 32	B= 3
N= 7	(A<DB>= 0		N= 7	(A<DB>= 0	
N= 6	(A<DB>=-225.67		N= 6	(A<DB>=-218.03	
N= 5	(A<DB>=-230.86		N= 5	(A<DB>=-19.45	
N= 4	(A<DB>=-244.43		N= 4	(A<DB>=-255.72	
N= 3	(A<DB>=-245.34		N= 3	(A<DB>=-13.07	
N= 2	(A<DB>=-228.56		N= 2	(A<DB>=-229.18	
N= 1	(A<DB>=-221.53		N= 1	(A<DB>=-17.26	

Tab.1 Spurious signals in a DDFS with the sine wave output for different B of discarded LSB ($\xi_x = 7/32$).

V. Conclusions.

After a short review of frequency synthesis systems in the whole range of electromagnetic waves from audio to optical frequencies we have discussed the respective mathematical models, particularly, those suitable for DDFS's: continued fraction expansion, Modified Engel series expansion, and Modulo-N arithmetics. Their advantage is the quasiperiodic omission of pulses from the input pulse train.

We have shown that in these instances the peaks of the phase-time modulation function do not exceed one clock period T_i.

With the assistance of the Fourier series algorithm we can find all spurious signals in the output wave. However, the procedure might be lengthy and the use of computer is inevitable.

For rectangular output waves approximation of the ratio Y/2X with smaller integers, arrived at by application of the shortened version of continued fraction expansion, provides information about major spectral lines.

By considering sine wave output we have paid attention to the problem of omission of some least significant bits.

Appendix.

By expanding the continued fraction (5) into an equivalent series we have found that after each $b_k = 1$ two neighbouring terms might be contracted into one term. This new series results from the shortened version of the continued fraction where the original Euklid algorithm (3), (4), etc. has been changed as follows:

$$\xi_o = b_o + \frac{1}{\xi_1} \qquad \xi_1 > 2$$

$$\xi_1 = b_1 + \frac{1}{\xi_2} \qquad \xi_2 > 2$$

$$b_k = \operatorname{int} \xi_k$$

$$\xi_o = b_o + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{\ddots}}}$$

The respective series expansion is

$$\xi_o = A_o + \sum_{r=1}^{n} (-1)^{r-1} \frac{a_1 a_2 \ldots a_r}{B_{r-1} B_r}$$

where

$a_1, a_2, \ldots$ are either +1 or -1.

References:

[1] V.F. Kroupa, "Frequency Synthesis: Theory, Design, and Applications". London, England: Griffin, 1973: New York Wiley 1973 .

[2] J. Gorski-Popiel, " Frequency Synthesis: Techniques and Applications". New York: IEEE Press, 1975.

[3] V. Manassewitsch, "Frequency Synthesizers, Theory and Design". New York Wiley, 1976, 1980.

[4] W.F. Egan, "Frequency Synthesis by Phase Lock". New York: Wiley 1981.

[5] U.L. Rohde, "Digital PLL Frequency Synthesizers, Theory and Design". Eglewood Cliffs: Prentice-Hall,1983.

[6] R.C. Stirling, "Microwave Frequency Synthesizers". Englewood Cliffs: Prentice-Hall, 1987.

[7] J.Tierney, C.M.Radar, and B. Gold, "A Digital Frequency Synthesizer" IEEE Trans.U-19, March 1971, pp.43-48.

[8] B. Bjerede and G. Fisher, "An efficient hardware implementation for high resolution frequency synthesis". Proc. 30th AFCS, pp. 318-321

[9] A. L. Bramble,"Direct Digital Frequency Synthesis". Proc. 35th AFCS May 1981, pp. 406-414.

[10] C.E. Wheatley, "Spurious Suppression in Direct Digital Synthesizers". Proc. 35th AFCS, May 1981, pp 428-435.

[11] W.C. Troxell, "Performance Analysis of the Numerically Controlled Oscillator". Proc. 40th AFCS, 1986, pp. 373-378.

[12] H.T. Nicholas and H. Samueli, "An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation". Proc. 41st AFCS, 1987, pp. 495-502.

[13] H.T. Nicholas, H. Samueli and B. Kim, "The Optimization of Direct Digital Frequency Synthesizer Performance in the Presence of Finite Word Length Effects". Proc. 42nd AFCS, 1988, pp. 357-363.

[14] E.M. Mattison and L.M. Coyle, "Phase Noise in Direct Digital Synthesizers" Proc. 42nd AFCS 1988, pp. 352-356.

[15] V.F. Kroupa, "Approximating Frequency Synthesizers". IEEE Trans. on Instrementation and Measurement, IM-23, December, 1974, pp. 521-4.

[16] B. Miller and B. Conley, "A Multiple modulator fractional divider". This issue.

[17] V.F. Kroupa, "Spectra of Pulse Rate Frequency Synthsizers". Proc. IEEE, vol. 67, December 1979, pp. 1680-1682.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN EXACT SPECTRAL ANALYSIS OF A NUMBER CONTROLLED OSCILLATOR BASED SYNTHESIZER

Joseph F. Garvey and Daniel Babitch
California Microwave Inc. Communications Technology Division
990 Almanor Ave., Sunnyvale, California, 94086

e-mail: cmic!garvey
cmic!babitch

ABSTRACT

A Direct Digital Synthesizer (DDS) based on a Number Controlled Oscillator (NCO) has advantages in many applications due to its wide frequency range, quick frequency changes, and fine resolution. However, because of the digital nature of the NCO, conventional wisdom has deemed the DDS spectrally unclean. In addition, because of the digital nature of the NCO, it has been difficult to determine what the ultimate performance levels might be. This paper presents an exact analysis of an NCO-based synthesizer. We believe this is the first exact analysis, and will be extremely important to those who design DDS', NCOs, and DACs.

This paper describes the architecture of a DDS, presents an exact spectral analysis and discusses some actual performance observations.

1.0 INTRODUCTION

This paper examines the ultimate performance of an NCO-based frequency synthesizer by going through a number of steps:

- A description of an NCO-based synthesizer is provided for those unfamiliar with the architecture of these units;
- A continuous Fourier Transform is developed for an arbitrary stepped waveform like the DDS output;
- A method for generating the mapping PROM used by a DDS is described, so that the output waveform distortion is minimized;
- Computer-based calculations are presented so that the reader may gain insight to the workings of a DDS;
- A final discussion is presented to summarize what the authors have learned and observed from their efforts.

2.0 DESCRIPTION OF AN NCO-BASED SYNTHESIZER

2.1 Description Of A DDS

A DDS based on an NCO takes a fixed frequency reference clock and generates an analog waveform of variable frequency, based on digital frequency commands.

An NCO-based frequency synthesizer, as shown in Figure 1, consists of the following pieces:

NCO: Really just a big accumulator, which we describe as holding the "phase" of the output signal. As the NCO accumulator changes, the values represent a linear phase ramp. The NCO accumulator has **P** bits, and is incremented by **S** every time the NCO is clocked.

Mapping PROM: This Programmable Read Only Memory maps the most significant bits (MSB) of the NCO into a sinusoid. In other words, the linear phase from the NCO is converted to a series of values representing a sinusoidal waveform. In our case a cosine is used. It has **A** address bits, and A <= P, always. All data lines from the mapping PROM go to the DAC.

DAC: The Digital to Analog Converter is used to generate the analog output voltage that is the output of the synthesizer. It has **D** bits.

2.2 Further Description Of The NCO

As previously mentioned, the NCO is simply a large accumulator (typically 32 bits, or more). The value of the accumulator at any given time corresponds to a phase. It is useful to picture the accumulator value as an angle, which, like the second hand of some watches, steps from second to second with each "tick" of the fixed frequency reference clock. Unlike the second hand of a watch, however, the size of each step can be varied by changing the phase step size. For the second hand of a watch the phase step size is 1 sec, with a corresponding output frequency of 1/60th of a Hz. If we could change the step size to 2, then the output frequency would be 1/30 of a Hz. Changing the step size to 10 would result in an output frequency of 1/6 Hz.

CH2818-3/90/0000-511 $1.00 © 1990 IEEE

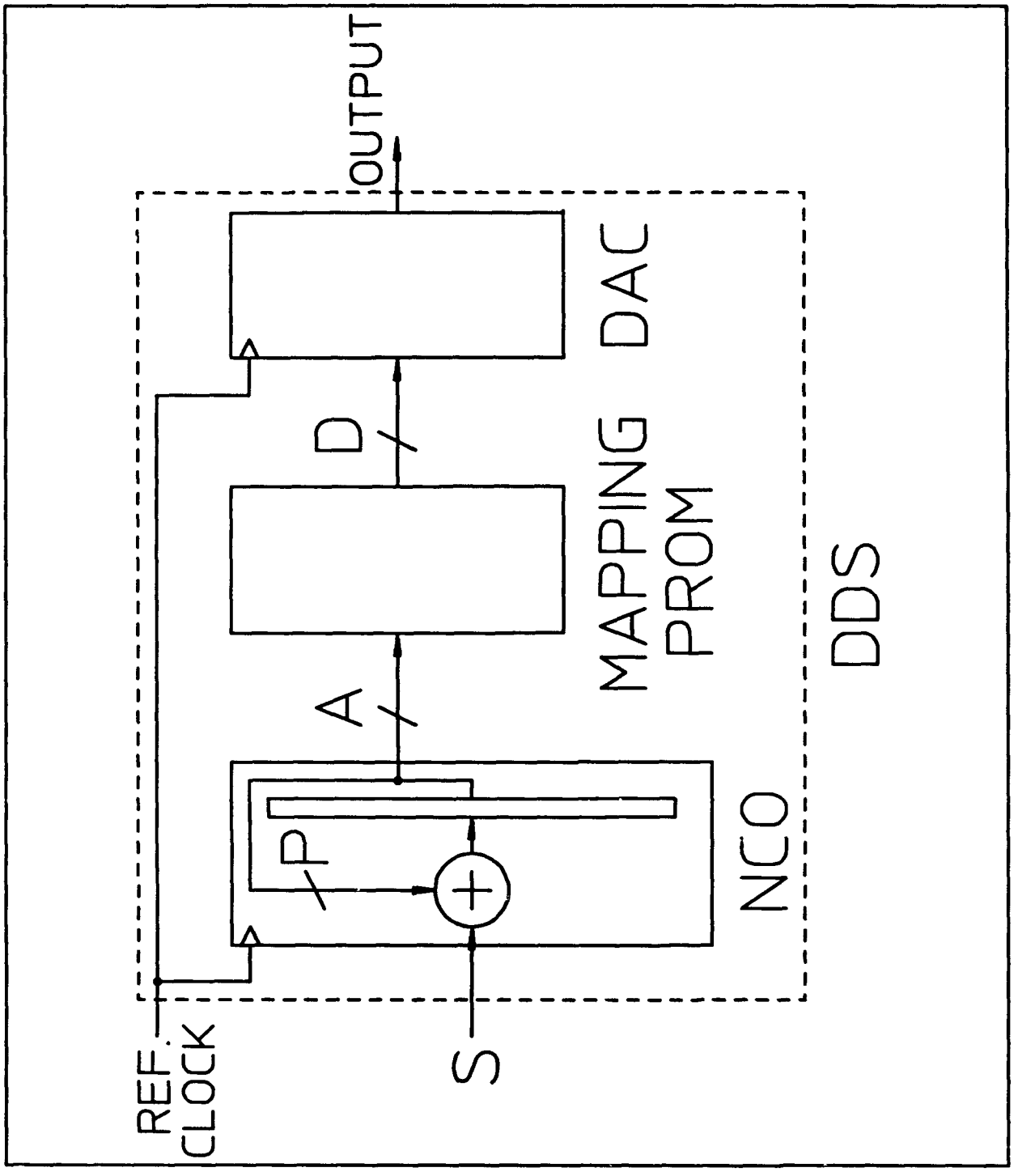


Figure 1.
DDS Block Diagram

Using this analogy, we can readily deduce the simple ratios that relate the phase step size (S), size of the accumulator (P), reference clock frequency (F_c), and output frequency (F_o).

$$\frac{\text{Step Size}}{\text{Accumulator Size}} = \frac{S}{2^P} = \frac{F_o}{F_c} = \frac{\text{Output Frequency}}{\text{Clock Frequency}} \quad (1)$$

From this we can see that frequency resolution is of the NCO synthesizer is simply:

$$\text{Resolution} = \frac{\text{Clock Frequency}}{\text{Accumulator Size}} = \frac{F_c}{2^P} \quad (2)$$

3.0 FOURIER TRANSFORM OF A REPEATING PULSE TRAIN

3.1 Grand Repetition Period

Though it may not be immediately obvious, the output of the synthesizer is a stepped waveform that at some point in time repeats. This occurs when the NCO contains exactly the same phase value as it started with. During the time it takes for this to occur, many sinusoidal cycles may occur, each slightly different from the preceding cycle. The minimum time for the NCO accumulator value to repeat is known as the Grand Repetition Period.

First, let's verify a repeating pattern occurs, and put an upper bound on how long this takes. To start, assume an initial NCO accumulator value of zero. Then take advantage of the fact that addition of a value to itself is equivalent to shifting the value "left" by one bit, and filling the least significant bit (LSB) with a zero.

To load the step size into the accumulator would take one tick. The next tick would add this value to itself. The value in the accumulator is the step size value shifted one bit to the left with a 0 filling in the LSB. To add this value to itself would take two more ticks for a total of four ticks. If we repeat the process, of adding the current accumulator value to itself, we keep doubling the number of ticks required for each iteration, and shifting left, and filling in with zeros. After P iterations of this procedure, we will have filled the accumulator with zeros, which is where we started. Counting up the number of ticks used to do this gives us an upper limit to the grand repetition period. So, the number of ticks in the worst-case grand repetition period is 2^P.

In the worst case grand repetition period, we may have crossed zero one or more times during these 2^P ticks. As we find out later, this has a direct effect on computational complexity, so let's pursue this issue a little further and find the grand repetition period exactly.

To find the grand repetition period, continue to think about what happens in the accumulator for different step sizes. Consider the case of P=4, for each possible step size.

Table 1

	Tick	P	P-1	...	4	3	2	1	0
Start, 0 phase	0	0	0	...	0	0	0	0	0
Add 3	1	0	0		0	0	0	1	1
Add 3	2	0	0		0	0	1	1	0 [1]
Add 3	3	0	0		0	1	0	0	1
Add 3	4	0	0		0	1	1	0 [2]	0 [2]
Add 3	5	0	0		0	1	1	1	1
Add 3	6	0	0		1	0	1	0	1
Add 3	7	0	0		1	0	1	0	1
Add 3	8	0	0		1	1	0 [3]	0 [3]	0 [3]

Note 1: After 2^1 ticks, the first LSB is 0.
Note 2: After 2^2 ticks, the first and second LSBs are 0.
Note 3: After 2^3 ticks, the first, second and third LSBs are 0.

Table 2

Ticks	Step Size (P=4) 1	2	3	4	5	6	7	8
	0	0	0	0	0	0	0	0
1	1	2	3	4	5	6	7	8
2	2	4	6	8	10	12	14	0
3	3	6	9	12	15	2	5	
4	4	8	12	0	4	8	12	
5	5	10	15		9	14	3	
6	6	12	2		14	4	10	
7	7	14	5		3	10	1	
8	8	0	8		8	0	8	
9	9		11		13		15	
10	10		14		2		6	
11	11		1		7		13	
12	12		4		12		4	
13	13		7		1		11	
14	14		10		6		2	
15	15		13		11		9	
16	0		0		0		0	
Ticks to Repeat	16	8	16	4	16	8	16	2

If the step size is a multiple of two, then the waveform repeats multiple times in 2^P ticks. In fact for each factor of two in the step size the wave form repeats.

Thus, if we eliminate all the factors of two from the step size, we can determine the grand repetition period exactly. One simple way to find the number of factors of two in the step size is to find the Greatest Common Divisor between the number needed to roll the accumulator over, and the step size. Then the grand repetition period can be directly calculated.

$$\text{Min. Number of Ticks in Grand Repetition Period} = \frac{2^P}{GCD\left(2^P, S\right)} \equiv \mu \qquad (3)$$

μ ≡ min. number of "ticks" in Grand Repetiton Period

GCD = greatest common divisor

3.2 Fourier Transform Of A Pulse

Now that we know the interval on which the NCO is periodic, we can perform a classical Fourier Transform.

To do this, let's consider each step of the output waveform as a separate pulse. Adding all these pulses up with proper phase and amplitude will create the original stepped waveform. Furthermore, let's just consider one of those pulses at this time:

As noted in Figure 2, the amplitude of the pulse is a function of its position (n). Notice that this will make the

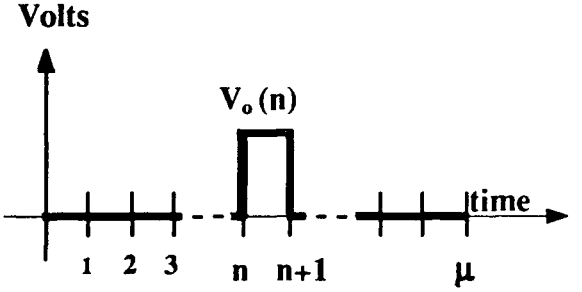


Figure 2.

analysis general – that is, it will work for any set of repeating pulses. Later, we'll substitute in the sinusoidal terms of the mapping PROM for the output voltage $V_o(n)$.

Now for the Fourier Transform itself:

Assume the pulse can be made of complex sinusoids:

$$V_o(t) = \sum_{i=-\infty}^{\infty} C_i\, e^{j\omega_i t} \quad \text{where } \omega_i = \frac{2\pi i}{\mu T_c} \tag{4}$$

$$T_c \equiv \frac{1}{F_c}$$

C_i are complex coefficients

Multiplying both sides by $e^{-j\omega_b t}$ and integrating over a period of $V_o(t)$:

$$\int_0^{\mu T_c} V_o(t)\, e^{-j\omega_b t}\, dt = \int_0^{\mu T_c} C_i \sum_{i=-\infty}^{\infty} e^{j(\omega_i-\omega_b)t}\, dt \tag{5}$$

Since this is an integration of a bounded well-behaved function, over a finite interval, we can fearlessly interchange the order of integration and summation:

$$\int_0^{\mu T_c} V_o(t)\, e^{-j\omega_b t}\, dt = \sum_{i=-\infty}^{\infty} C_i \int_0^{\mu T_c} e^{j(\omega_i-\omega_b)t}\, dt \tag{6}$$

The integral over a full period of two sinusoidal functions $e^{j\omega_i t}$ and $e^{-j\omega_b t}$ is always zero unless i = b. That's the definition of orthogonal functions. This eliminates all terms of the summation, except when i = b. When i = b, the complex exponential cancels to 1, $e^{j(\omega_i-\omega_i)t}$.

$$\int_0^{\mu T_c} V_o(t)\, e^{-j\omega_i t}\, dt = C_i\, \mu\, T_c \tag{7}$$

Rearranging terms we get:

$$C_i = \frac{1}{\mu T_c} \int_0^{\mu T_c} V_o(t)\, e^{-j\omega_i t}\, dt \tag{8}$$

For the pulse this becomes:

$$C_i = \frac{V_o(n)}{\mu T_c} \int_{nT_c}^{(n+1)T_c} e^{-j\omega_i t}\, dt \tag{9}$$

After integration, factoring out the exponential in the result, and a little algebra, we get:

$$C_i = \frac{V_o(n)}{\mu}\, \text{sinc}\, \frac{\pi i}{\mu}\, e^{-j\omega_i\left(n+\frac{1}{2}\right)T_c} \tag{10}$$

Fourier Transform Of A Pulse Train

Knowing the Fourier Transform of a single pulse with a period of μ allows us to use the simple concept of superposition to find out what the entire spectrum would look like. For a series of μ pulses, phased to create the stepped output waveform:

$$C_i = \sum_{n=0}^{\mu-1} \frac{V_o(n)}{\mu}\, \text{sinc}\, \frac{\pi i}{\mu}\, e^{-j\omega_i\left(n+\frac{1}{2}\right)T_c} \tag{11}$$

Rearranging terms, and factoring out part of the exponential, we get:

$$C_i = \frac{e^{-j\frac{\pi i}{\mu}}}{\mu}\, \text{sinc}\, \frac{\pi i}{\mu} \sum_{n=0}^{\mu-1} V_o(n) W_\mu^{ni} \tag{12}$$

$$W_\mu \equiv e^{-j\frac{2\pi}{\mu}}$$

Notice when stated in this form that the summation is just the Discrete Fourier Transform of the synthesizer output over the grand repetition period, and since μ is composed of factors of two only, the Radix Two Fast Fourier Transform can be used. This makes computation of the coefficients quick and easy. After performing a Fast Fourier Transform, the coefficients are scaled by the term in front of the summation to get the Continuous Time Fourier Transform. An exact spectrum!

3.4 Verification

3.4.1 By Summation: A computer program was created to generate the output spectrum of the NCO based on equation 12. The program generated two primary lines, the upper and lower sidebands, which are centered on half the clock frequency, as expected. In order to verify the spectrum generated by the Fourier Transform of the NCO output was correct, a program that did a Fourier Summation was used. This allowed the spectrum to be converted back into a time domain waveform, and thereby verifying that the proper coefficients had been calculated.

One of the important features observed is that harmonics above the NCO clock frequency are required for accurate reproduction of the time waveform. In order to get a reasonably clean waveform harmonics up to 64 times the NCO clock rate were required.

3.4.2 Known Spectrum: A second test was made by checking what happens to the equation for C_l when half the output frequency was requested. In this case the formula must and does simplify to that of a square wave.

In addition, proper behavior is observed when a very low or very high frequency is requested. The lower sideband amplitude approaches unity (0 dB) at low frequency, and the upper sideband goes to 0 ($-\infty$ dB) as it approaches the clock frequency. Near half the clock frequency, each sideband is 3 dB down, as it should be.

3.4.3 By Measurement: DAC clock leakage in our experimental DDS prevented us from measuring the individual spurious frequencies. However, the DAC clock leakage is too small to significantly affect the amplitudes of the upper and lower sidebands. These were checked and were in excellent agreement with the predictions of the model (hundredths of a dB), over a wide range of frequencies. In addition, the measured spurious level of -65 dBc was consistent with theoretical predictions of -67.8 dBc cleanliness.

4.0 GENERATING THE MAPPING PROM

4.1 Need For Symmetry

To generate the cleanest possible output signal, the digitized waveform should be quantized as symmetrically as possible. This includes symmetry about the time (phase) and amplitude axes. Figure 3 shows a sinusoid, and it's corresponding quantization.

Notice, by selecting the quantization levels in the middle of each time period (pulse), the amplitude error for half the pulse is in one direction, and the error in the second half of the pulse is in the opposite direction.

Also notice that by drawing a line through the 90, and 270 degree marks, the quantized waveform is symmetric about these points as well. This again minimizes distortion, and preserves the symmetry one expects of a sinusoid.

4.2 Quantized Sinusoid - Mathematical Description

If we quantize a cosine wave form, then we want to put one cycle of the cosine into the Mapping PROM. The phase resolution of this PROM is determined by the number of address bits it has:

$$\text{PROM Phase Resolution} = \frac{\text{One Cycle}}{\#\text{ of Points}} = \frac{2\pi}{2^{A}} \qquad (13)$$

In order to maintain symmetry, we need to sample the cosine wave at the mid-point between two pulses. This is obtained simply by averaging the two times.

$$\begin{aligned}\text{Midpoint between two pulses} &= \frac{\text{phase of point 1} + \text{phase of point 2}}{2} \\ &= \frac{m + (m+1)}{2}\end{aligned} \qquad (14)$$

$m \equiv$ address applied to PROM at time n

Finally, examining the amplitude of the sinusoid, we notice one bit can be considered a sign bit, and the remaindering bits are the magnitude of the amplitude. Also notice, the 0 value out of the DAC is never used.

$$\text{Amplitude of PROM} = 2^{(D-1)} \qquad (15)$$

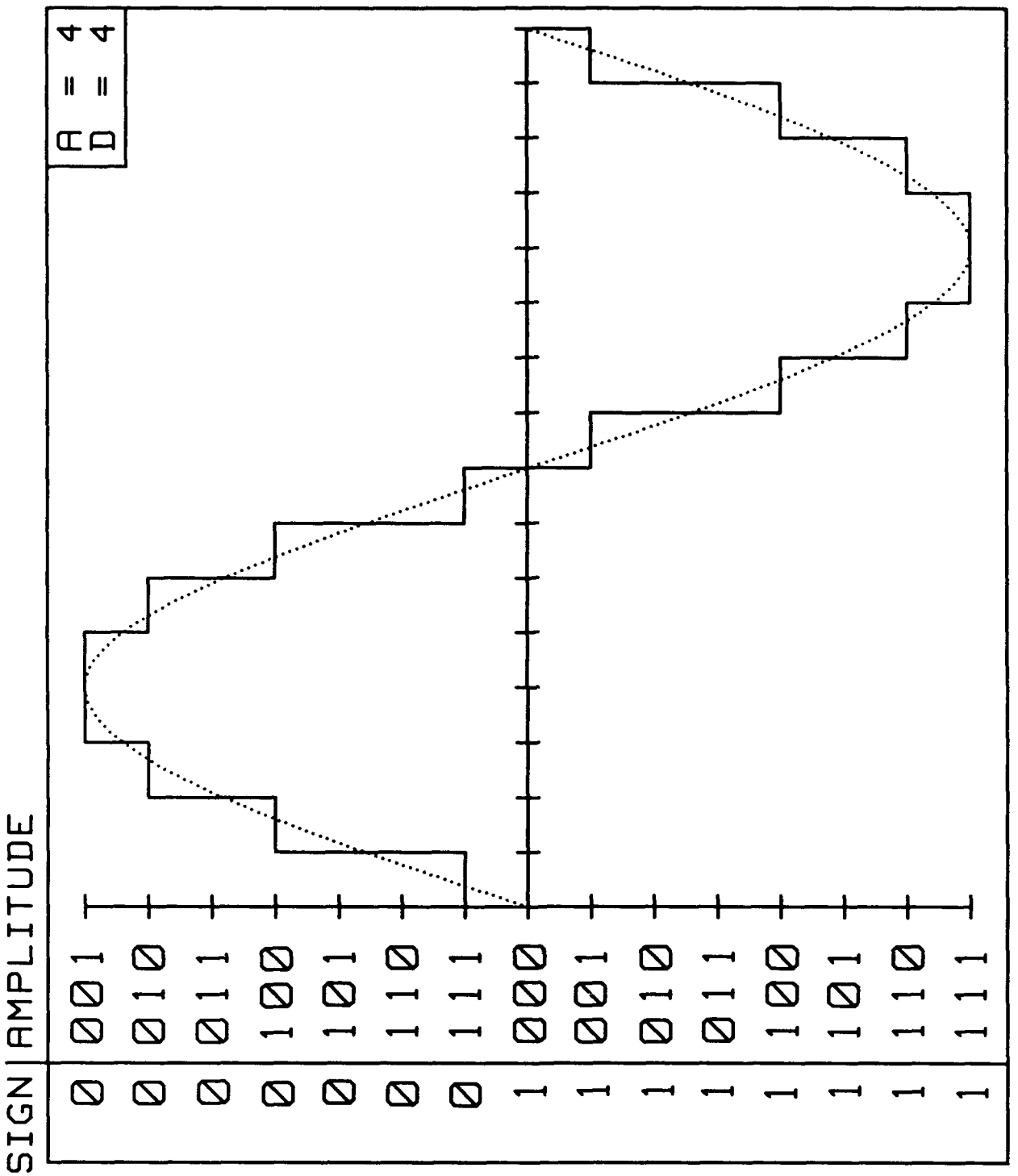


Figure 3.
Symmetrically Quantized Sinusoid

Putting this all together we have the mathematical description of the contents of the Mapping PROM.

$$V_o(n) = 2^{(D-1)} \cos\left(\frac{n+(n+1)}{2}\frac{2\pi}{2^A}\right) \qquad (16)$$

A word of caution: the value actually used in the PROM is a rounded integer value of $V_o(n)$. To avoid inducing any extra distortions, these values should be rounded symmetrically. This can be achieved by rounding the absolute value of $V_o(n)$, and then attaching the proper sign.

5.0 PRESENTATION OF RESULTS

5.1 Spectral Cleanliness

Using simulation software based on the equations in this paper, we modeled various DDS configurations. Figures 4 and 5 graphically present the results of these analyses.

All analyses presented here were performed with the output frequency set to 1/3 of the reference clock frequency. This is where the worst spurious occurs based on empirical evidence.

Figure 4 shows that the output cleanliness is <u>not</u> a function of the number of phase bits in the NCO. Rather, the phase bits in the NCO simply represent the resolution with which frequencies may be selected. There is a section of Figure 4 where the cleanliness actually goes up as the number of phase bits goes down. However, this range is extremely impractical to use, since it is characterized by the condition P=A=D. As a result of this condition, there is limited frequency resolution for real world values, and output slew rates tend to be too great to implement cleanly.

The big surprise is Figure 5. It shows that the cleanliness is a 6 dB/bit function of the number of mapping PROM address bits A, and an 8.5 dB/bit function of the number of DAC bits D.

Figure 5 clearly indicates there is a range of values where having extra DAC bits results in NO performance gain. This means that designers who push to incorporate large mapping PROMs into NCOs may besurprised that performance doesn't improve, if they haven't selected the right Address-Data bit combinations (A=8,D=6), (A=9,D=7), (A=10,D=7), (A=11,D=8), (A=12,D=9), (A=13,D=10), etc.

6.0 ADDITIONAL THOUGHTS

6.1 DAC Clock Leakage

One of the reasons an actual spectrum is not compared against a predicted spectrum is DAC clock leakage causes a severe distortion of the actual spectrum at the DAC output. When observed through filtering and limiting circuits the clock leakage sidebands were -65 dBc.

We used a 25 MHz NCO with a high-speed 8-bit DAC in our experimental DDS. This DAC had clock leakage 45 dB below the output signal at the clock frequency. This clock leakage signal mixed with the switched current sources in the DAC, so that mixing signals appeared at multiple points across the output spectrum. This is consistent with the concept of the DAC acting as a switching mixer, where the DAC MSB is acting as the switch control and the clock leakage (main spectral line) acting as the input.

The original analysis program was modified to verify this behavior. It correctly indicated where the DAC spurious would occur, but was not good at estimating spectral magnitude. To accurately model the magnitude of DAC spurious, the internal layout, parasitics, and process parameters would have to be included. Nevertheless, we believe the basic mechanism of clock leakage mixing sidebands is correctly modeled because of the match between predicted frequencies and observed frequencies.

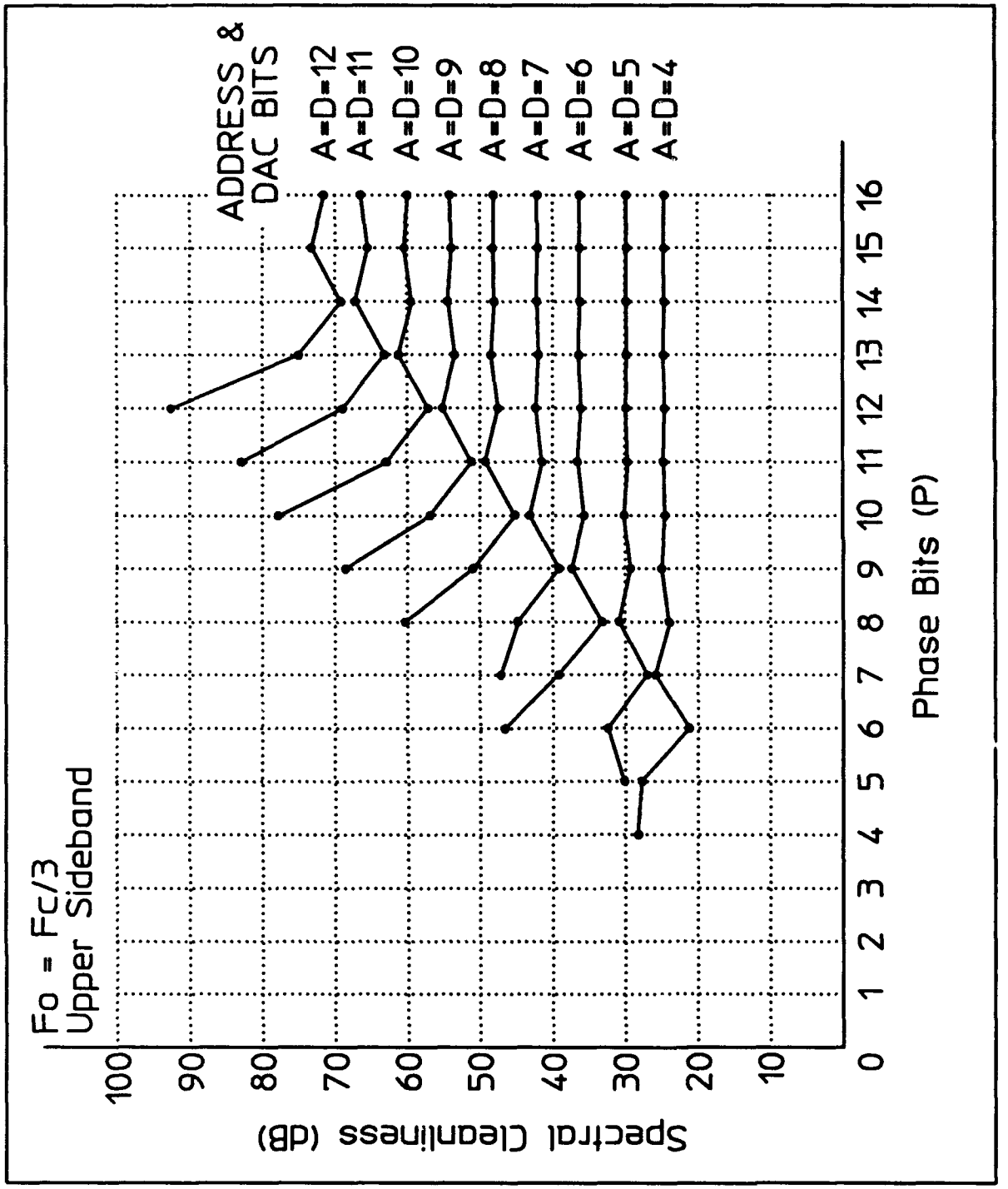


Figure 4.
Number of Phase Bits vs. Spectral Cleanliness

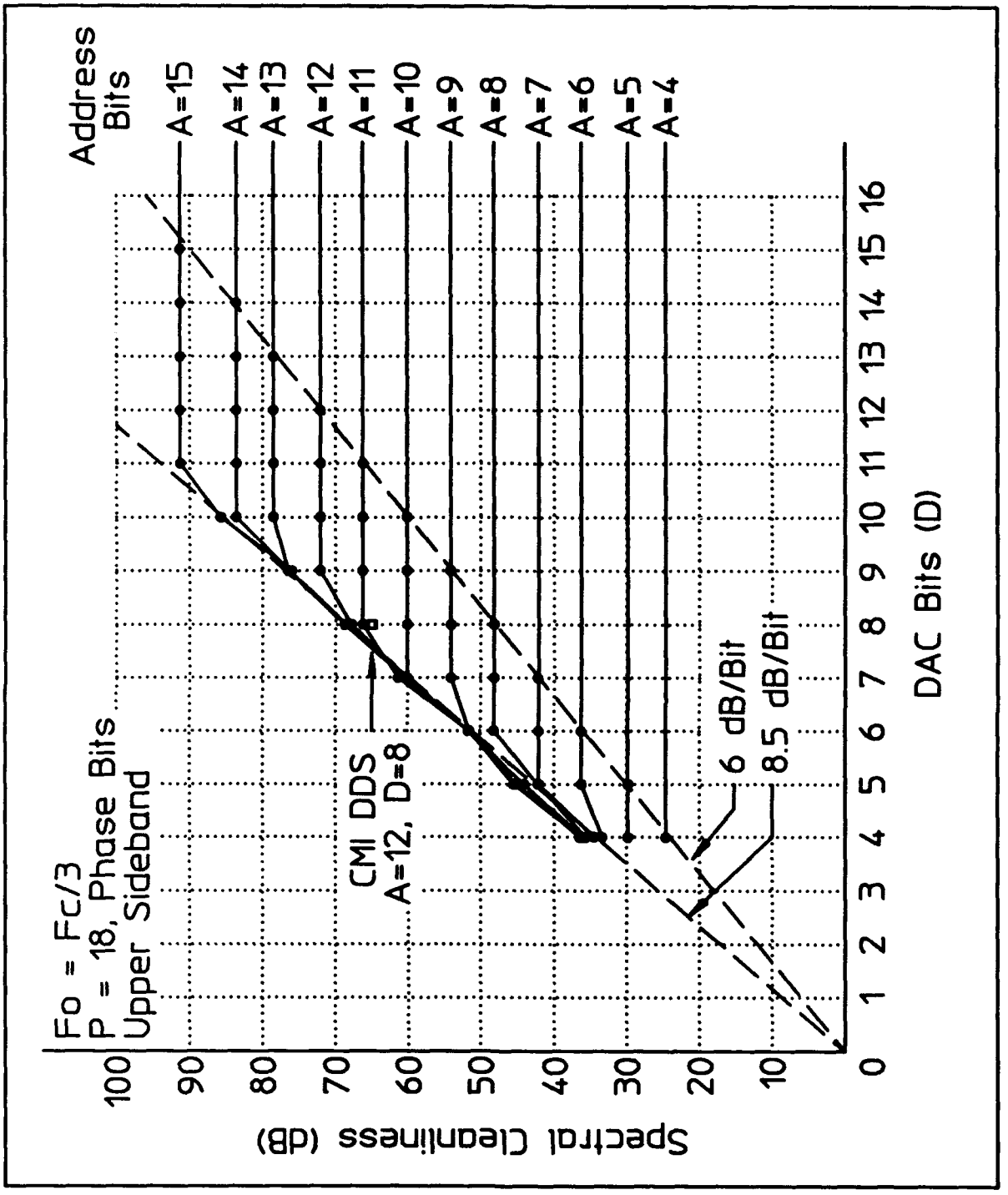


Figure 5.
Number of DAC Bits vs. Spectral Cleanliness

7.0 CONCLUSION

A DDS based on an NCO can be spectrally clean, which is contrary to conventional wisdom. Furthermore, its spectrum can be calculated exactly using a conventional radix-2 DFT, and applying a frequency dependent correction factor. Examining the results of these calculations, one discovers there are very specific configurations of the DDS architecture that will achieve maximum spectral purity with the least circuitry.

Currently, much work remains to be done in the actual implementation of an NCO-based DDS. Specifically, modulation of the DAC output by the clock is a significant problem. This is especially so with available high-speed DDS' having clock rates of hundreds of megahertz, where performance is still somewhat limited.

ACKNOWLEDGEMENTS

The authors would like to thank Kathy Stanley for her diligent efforts in typesetting this paper and her discerning editorial comments. A great deal of the paper's organization is due to the efforts of Rod Kronschnabel. He prodded us to distill our unnecessarily complex ideas down to the simple, straightforward form presented here.

ANNOTATED BIBLIOGRAPHY

An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation by Nicholas and Samueli, Proceedings of the 41st Annual Frequency Control Symposium, 1987, presents what appears to be a less computationally intensive method for estimating spurious due solely to phase accumulator truncation. Because of the approximations used, the reader is also left without detailed knowledge of the trade-offs between DAC resolution and size of the mapping PROM. Of course with modern PC's and the widespread availability of DFT software an exact analysis is relatively straightforward to accomplish, now that one knows how to "correct" the DFT values.

The Optimization of Direct Digital Frequency Synthesizer Performance in the Presence of Finite Word Length Effects by Nicholas, Samueli, and Kim, Proceedings of the 42nd Annual Frequency Control Symposium, 1988. This is a follow up paper by Nicholas and Samueli. It describes and analyzes a hardware implementation of the sinusoid lookup table for those interested in alternatives to the conventional mapping PROM.

Performance Analysis of the Numerically Controlled Oscillator by Kisenwether, Proceedings of the 40th Annual Frequency Control Symposium, 1986, presents an analysis of a DDS using the NCO's most significant bit as the output. For many applications this is impractical since the required high slew rates cannot be achieved.

Phase Noise in Direct Digital Synthesizers, by Mattison and Coyle, Proceedings of the 42nd Annual Frequency Control Symposium, 1988, presents an interesting set of experimental results. Here the author compares the results of a DFT with actual measurements. He also observes that Kisenwether's approach of using the MSB has slew rate problems.

Number Controlled Modulated Oscillator Theory and Applications, by McCune of Digital RF Solutions in Santa Clara, Ca presents an interesting discussion on modulating the NCO output digitally.

DESIGN TECHNIQUES FOR ACHIEVING STATE-OF-THE-ART OSCILLATOR PERFORMANCE

G. K. MONTRESS and T. E. PARKER

Raytheon Research Division
131 Spring Street
Lexington, Massachusetts 02173

ABSTRACT

Recently, state-of-the-art phase noise performance has been demonstrated for both surface acoustic wave (SAW) and dielectric resonator (DR) stabilized oscillators [1]-[4]. The same basic feedback-loop oscillator [5] design philosophy was applied in each case in order to achieve these results. This paper describes the design, fabrication, component selection, and performance of extremely low noise SAW resonator and delay line oscillators, as well as an extremely low noise L-band DR oscillator. It will be shown that a basic 50-Ω modular component, feedback-loop oscillator architecture incorporating a two-port frequency stabilizing device is easy to set-up, and perhaps much more importantly, is easy to trouble shoot when something does go wrong.

I. INTRODUCTION

The residual phase noise properties of an oscillator's electronic components, for example RF amplifiers and electronic phase shifters, are extremely important if reproducibly low oscillator phase noise levels are to be achieved. For SAW oscillators, it goes without saying that the residual flicker noise of the SAW resonator or SAW delay line must also be minimized if truly state-of-the-art phase noise performance is to be achieved. For DR stabilized oscillators, we have not as yet seen any evidence indicating that the DR is a significant source of residual flicker noise. A considerable portion of the following discussions describes state-of-the-art residual noise measurement techniques which may be used to accurately and reproducibly screen component residual noise properties.

To illustrate our design procedure, several specific examples are presented. First, prototype SAW resonator voltage controlled oscillators operating at 500 MHz have demonstrated white PM noise floors of -184 dBc/Hz for carrier offset frequencies greater than 100 kHz, while flicker FM noise levels of -83 dBc/Hz at 10 Hz carrier offset have been achieved. Analogous designs using a 1 GHz dielectric resonator (DR) have demonstrated similar phase noise performance, except now at approximately twice the operating frequency of the SAW resonator oscillators. Engineering prototype 400 MHz SAW delay line stabilized oscillators with electronic frequency tuning ranges of ±150 kHz (±375 ppm) have demonstrated white PM phase noise floors of -170 dBc/Hz, along with flicker FM noise levels of -70 dBc/Hz at 10 Hz carrier offset. Finally, other aspects of an oscillator's overall performance, such as vibration sensitivity, long-term stability, etc., are not discussed in detail since they have been adequately covered elsewhere [1]-[4]. Quite often trade-offs must be made between the various performance criteria that might otherwise be desired. These trade-offs are mentioned, when appropriate, in the context of the design examples presented.

II. FEEDBACK-LOOP OSCILLATOR BASICS

Figure 1 shows a simple block diagram for a general purpose SAW device stabilized feedback-loop oscillator design of the type first analyzed by Leeson [6]. It should be noted that the SAW device can, in principle, be replaced with any two-port

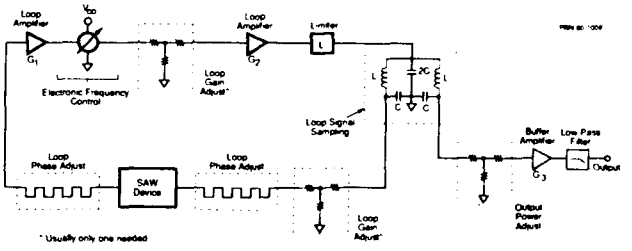


Fig. 1 Circuit Diagram for a General Purpose Feedback-Loop Oscillator.

CH2818-3/90/0000-522 $1.00 © 1990 IEEE

frequency stabilizing element, for example, a dielectric resonator, without in any way whatsoever altering the discussion to follow. Although individual oscillators will always differ in specifics of their particular implementation, they all will share the following common features: 1) one or more loop amplifiers (G_1, G_2) of sufficient gain to overcome total feedback-loop losses; 2) some means for gain limiting (compression) within the feedback loop to insure stable oscillation (in many instances this gain-limiting action will actually occur in the second stage (G_2) or high level feedback-loop amplifier; 3) provisions for both gain and phase adjustment within the feedback loop in order to establish the proper loop conditions for oscillation to occur, namely approximately 3 dB (2 to 4 dB) of excess small-signal gain (nominally 3 dB of gain compression when equilibrium is reached) and $2\pi N$ radians of net transmission phase shift through the loop, where N is an integer; 4) feedback-loop signal sampling, which may be either capacitive, resistive, etc.; and 5) a buffer amplifier (G_3) to isolate the feedback loop from external load variations. The need for an electronic phase shifter in the feedback loop to provide frequency modulation and a low-pass filter at the output to suppress undesired harmonic signals may or may not be necessary depending upon a specific oscillator's performance requirements.

The basic feedback-loop oscillator circuit of Fig. 1 is particularly well suited for two-port SAW devices and is generally the most commonly used configuration for SAW oscillators, both resonator and delay line based. It is also particulary convenient to choose components that have been designed to work well in a 50-Ω environment. By using amplifiers, power splitters or directional couplers, and other electronic components that nominally have been designed with 50-Ω input and output impedances, the proper conditions for oscillation can be quickly established using a vector network analyzer. By breaking the oscillator loop open at any convenient location the open-loop gain and transmission phase of the combined oscillator components can then be readily established. The equilibrium oscillator loop power can be determined by increasing the network analyzer's test signal level until gain compression gives 0 dB of open-loop gain. By observing the gain and phase conditions at frequencies other than the desired oscillation frequency, the potential for spurious oscillations can also be evaluated. Finally, if a CW RF test signal is applied to the input port of the opened-loop at the power level corresponding to 0 dB of open-loop gain, then the oscillator's ouput power level may be verified. Clearly, other measurements may be performed to establish the oscillator's approximate electronic tuning sensitivity (in the case of a VCO), power supply pushing sensitivity, load pulling sensitivity, etc.

The simplest SAW oscillator may contain only an amplifier, a SAW device, an output coupler, and means for setting the loop's excess gain level and transmission phase shift, e.g., an attenuator and a short length of coaxial cable. The saturation of the loop amplifier provides the necessary gain compression. More capable oscillators may contain a buffer amplifier, an electronic phase shifter, amplitude limiter, or any of the other components shown in Fig. 1.

As indicated in Fig. 1 there is some discretion as to where the loop gain adjust attenuator is placed. For a feedback-loop oscillator with a *single loop amplifier* it doesn't really matter at all where the attenuator is positioned insofar as its basic contribution to the oscillator's white phase noise floor goes, although it may end up affecting the oscillator's overall performance in secondary ways, e.g., if the attenuator were positioned between the output of the high level loop amplifier and the coupling circuitry. However, for a feedback-loop oscillator with *two loop amplifiers* there is a distinct performance advantage to be realized if the attenuator is positioned between the output of the first stage loop amplifier (G_1) and the input of the second stage loop amplifier (G_2). If the attenuator is not positioned inter-amplifier stage the oscillator's phase noise floor will degrade dB for dB according to the value of attenuator used. This will not happen if the attenuator is introduced inter-amplifier stage, as just noted. In this case the attenuator's contribution to the oscillator's phase noise floor is effectively reduced ("shielded") by the gain of the first loop amplifier stage. The net effect is to reduce the noise floor degradation factor to only about a tenth of a dB per dB according to the value of the attenuator used. The approximate amount of noise floor degradation may be estimated using the following equation

$$\Delta(\text{dB}) = 10\times\log_{10}[(F_1G_1 + F_2L - 1)/(F_1G_1 + F_2 - 1)] \quad (1)$$

where F_1 (= $10^{[(NF_1)/10]}$) and F_2 (= $10^{[(NF_2)/10]}$) are the *noise factors* for the first and second stage loop amplifiers, respectively, G_1 is the gain of the first stage loop amplifier (expressed numerically), and L is the loss of the inter-stage attenuator (expressed numerically). Let's examine a specific example, as follows: $F_1 = F_2 = 3.16$ (i.e., $NF_1 = NF_2 = 5$ dB), $L = 3.16$ (i.e., a 5 dB attenuator), and $G_1 = 25.12$ (i.e., the small-signal gain of the first stage loop amplifier is 14 dB). If the attenuator in positioned inter-stage then the oscillator's white phase noise floor degradation will only be about 0.35 dB, whereas if the attenuator is placed anywhere else within the feedback loop the oscillator's white phase noise floor will be

degraded by the full 5 dB value of the attenuator. This represents a significant noise floor improvement with :t incurring any performance penalty whatsoever. Clearly, it would be foolhardy to place the attenuator anywhere except inter-stage when two amplifiers are required in the feedback loop!

Perhaps the greatest advantage which accrues from using the feedback-loop oscillator architecture is the ability to readily isolate problem components, including the ability to pre-screen critical loop components such as the amplifier, SAW device, and electronic phase shifter for their residual phase noise levels before the oscillator is even assembled. This ability is particulary useful since these rather basic measurements may be used directly to predict the oscillator's full phase noise spectrum. As has been shown previously, for most SAW resonator oscillators, the following approximate expression may be used to estimate the oscillator's expected phase noise performance for carrier offset frequencies greater than 0.1 Hz [7]

$$S_\phi(f_m) = [\alpha_R F_o^4/f_m^3] + [\alpha_E/\{(2\pi\tau_g)^2 f_m^3\}] + [(2\alpha_R Q_L F_o^3)/f_m^2] + [(2GFkT/P_o)/\{(2\pi\tau_g)^2 f_m^2\}] + \alpha_E/f_m + (2GFkT)/P_o \quad (2)$$

where G is the compressed power gain of the loop amplifier, F is the noise factor of the loop amplifier(s), k is Boltzmann's constant, T is the temperature in °K, P_o is the carrier power level (in watts) at the output of the loop amplifier, F_o is the carrier frequency in Hz, f_m is the carrier offset frequency in Hz, τ_g is the SAW resonator's group delay in seconds, Q_L $(=\pi F_o \tau_g)$ is the loaded Q of the SAW resonator in the feedback loop, and α_R and α_E are the flicker noise constants for the SAW resonator and loop amplifier, respectively. The flicker noise constants may be evaluated from residual noise measurements using the following expressions

$$\alpha_R = [1/(2Q_L F_o)^2][2f_m x 10^{\{\mathcal{L}'(f_m)/10\}}] \quad 3(a)$$

and

$$\alpha_E = 2f_m x[10^{\{\mathcal{L}'(f_m)/10\}}] \quad 3(b)$$

where Q_L, F_o, and f_m are as previously defined and $\mathcal{L}'(f_m)$ is taken to represent a measurement of either the resonator's or amplifier's single sideband residual flicker PM noise. Note that the "prime" symbol is intended to denote a residual rather than oscillator single sideband noise spectrum.

A similar, although somewhat simpler approximate expression may be used to estimate a SAW delay line oscillator's phase noise spectrum for carrier offset frequencies greater than 1 Hz, namely

$$S_\phi(f_m) = [(\alpha_D + \alpha_E)\{(2\pi\tau_g)^2 f_m^3\}] + [(2GFkT/P_o)/\{(2\pi\tau_g)^2 f_m^2\}] + \alpha_E/f_m + (2GFkT)/P_o \quad (4)$$

where G, F, k, T, P_o, F_o, f_m, and α_E are as previously defined, τ_g is now the SAW delay line's group delay in seconds, and α_D is the flicker noise constant for the SAW delay line. The amplifier's flicker constant in this case is again evaluated using Equation 3(b), while the SAW delay line's flicker constant is calculated in a similar manner, i.e., using

$$\alpha_D = 2f_m x[10^{\{\mathcal{L}'(f_m)/10\}}] \quad (5)$$

where $\mathcal{L}'(f_m)$ now represents the result of a single sideband residual flicker noise measurement on the SAW delay line device. On occasion, instead of the symbols α_E and α_D, the symbols a_E and a_D have been used to denote the amplifier's and delay line's flicker constants, respectively. This was done in order to indicate that they are calculated in a different manner from their respective component's measured residual flicker noise, when compared to the flicker constant for a SAW resonator device as defined in Equation (2). As a direct consequence α_E and α_D have the units of radians^2, while α_R has the units of radians^2/Hz^2. Finally, the magnitude of α_R is markedly different from the magnitudes of either α_E or α_D. A typical value for α_R might be $2x10^{-40}$, while typical values for α_E and α_D might range anywhere from $2x10^{-15}$ to $2x10^{-13}$, depending upon the particular amplifier or delay line under consideration.

Since state-of-the-art phase noise performance is the central theme of this paper, Section III which follows briefly describes the typical SAW resonator and delay line designs which we routinely use.

III. SAW DEVICE DESIGN

All of our recent SAW oscillator work is based upon the

"all quartz package" (AQP) which has been described previously [8]-[10]. AQP SAW resonator or delay line devices are packaged in a similar manner. Typically, our SAW resonator devices are designed as shown in Fig. 2. The active acoustic aperture ranges from 150 to 300 λ, where λ denotes an acoustic wavelength at the device's nominal resonant frequency. An important feature common to all of our SAW resonator designs is the rather large free surface region between the two transducers, typically 200 λ. This results in an effective cavity length [11] of approximately 365 λ. As a consequency of this large cavity length the resonator will support three distinct cavity modes. This is illustrated in Fig. 3 for a 500 MHz two-port SAW resonator device design with an active acoustic aperture of 300 λ. The three cavity mode resonances are clearly evident, but note that the transmission phase shifts of the two outer resonances are offset by 180° with respect to the transmission phase shift of the central resonance peak. For most filter applications the existence of these satellite resonances would be totally unacceptable, but in an oscillator application the phase reversals exhibited by the satellite peaks totally eliminate them from consideration if the feedback-loop's transmission phase shift has been properly set-up for oscillator operation on the central resonance peak. The large cavity length reduces the fraction of the effective cavity length covered by metal to only about 7 percent, and in most instances eliminates the need to apodize the transducers [12], even for the relatively wide apertures (150 to 300 λ) used in our designs. Another advantage of a large cavity size (both length and width) is that it increases the power-handling capability of the resonator [13]. SAW resonator devices incorporating this basic design have typically exhibited 4 to 10 dB of insertion loss, depending upon the operating frequency, number of fingers in each transducer, and the actual acoustic aperture used.

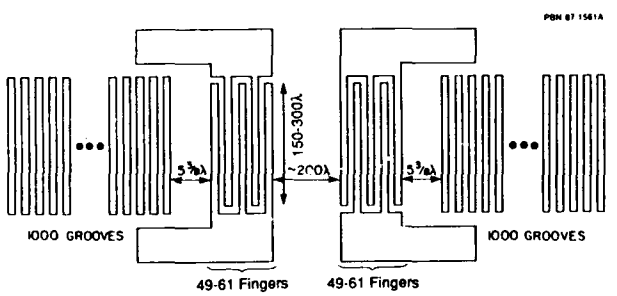


Fig. 2 Typical SAW Resonator Design.

The key features of the transducers for a SAW delay line designed for an oscillator application are shown in Fig. 4. For good frequency stability the delay line should have as large a delay time (group delay) as possible. However, the delay line

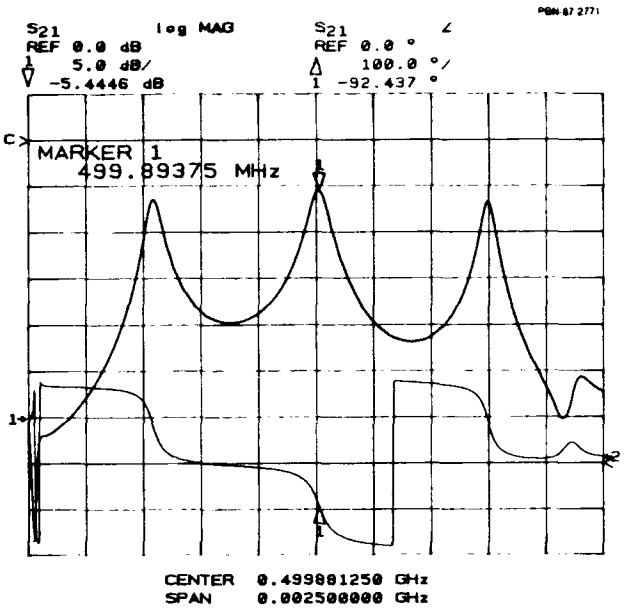


Fig. 3 Insertion Gain and Transmission Phase Responses for a 500 MHz SAW Resonator Device.

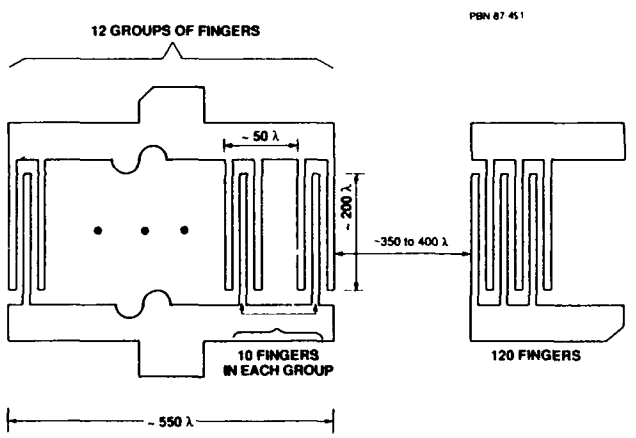


Fig. 4 Typical SAW Delay Line Design.

must also provide a frequency filtering function to limit the amplitude response of the delay line to a relatively narrow frequency range. Generally, the value of group delay selected will represent a compromise since most SAW delay line oscillators are intended to operate as frequency agile (voltage tunable) sources. The long "thinned" transducer on the left in Fig. 4 serves this purpose of frequency filtering, while the center-to-center distance between the two transducers determines the group delay, τ_g. To ensure that only one frequency can satisfy the oscillation conditions at any given time, the combined length of the two transducers should be no less than approximately 90 percent of the center-to-center distance between the two transducers. It has been found

generally desirable to limit the total number of fingers in each transducer to about 120, thus the long transducer is usually thinned into a ladder type structure [14], as illustrated in Fig. 4. This results in a device with only about 5 to 6 percent of the active acoustic area covered with aluminum in the transducers. Additional fingers can be used to achieve lower insertion loss, but this increases the undesirable influence of the metalization on turn-over temperature, triple transit reflections, and frequency accuracy. If desired, electrically inactive fingers (so called "dummy" electrodes) may be used in order to reduce ripple in the insertion loss and transmission phase responses. This might be done, for example, in order to achieve a smoother frequency versus tuning voltage transfer characteristic, as well as to achieve a smoother frequency versus temperature characteristic for the oscillator design.

Additional factors influencing SAW resonator and delay line designs for oscillator applications are discussed in Ref. [5].

IV. COMPONENT RESIDUAL NOISE MEASUREMENTS

Residual phase noise measurements have been used for some time to characterize devices in the transmission [15]-[18], as well as reflection [19] modes. The realization of extremely low phase noise VHF, UHF and microwave oscillators is critically dependent upon the use of components, e.g., amplifiers, SAW devices, electronic phase shifters, etc., which possess verifiably low residual phase noise levels. The utility of residual phase noise measurements is due to the fact that there is a one-to-one relationship between an oscillator's closed-loop (oscillator) flicker FM noise ($1/f_m^3$) and open-loop (residual) flicker PM noise ($1/f_m$) levels [4], namely

$$\mathcal{L}_O(f_m) = \mathcal{L}_R(f_m) + 10\log_{10}\{1/(2\pi\tau_g)\}^2 + 10\log_{10}\{1/f_m\}^2 \quad (6)$$

where $\mathcal{L}_O(f_m)$ is the oscillator's single sideband flicker FM noise spectrum expressed in dBc/Hz, $\mathcal{L}_R(f_m)$ is the residual flicker PM noise spectrum for the oscillator when measured looking through the feedback loop (same test configuration as described previously for oscillator set-up) expressed in dBc/Hz, f_m is the carrier offset frequency in Hz, and τ_g is the SAW device's group delay, in seconds, when the oscillator loop is closed. For a SAW resonator its group delay may also be found using the relationship $\tau_g = Q_L/(\pi F_o)$, where F_o is the oscillator's carrier frequency in Hz. It is important to realize that $\mathcal{L}_R(f_m)$ represents the algebraic sum of *all* the potential flicker phase noise contributions in the feedback loop due to the loop amplifier(s), SAW device, electronic phase shifter, etc. Therefore, if a given level for the oscillator's flicker FM noise level ($1/f_m^3$) is desired, Equation (6) may be used to set an upper limit on the acceptable residual flicker PM noise for each component in the loop which is found to significantly contribute to the oscillator loop's residual flicker phase noise level. We have found in our own work that for SAW resonator and delay line oscillators such as those described in Section V that it is often quite difficult, if not impossible, to actually uniquely identify the particular component which is dominating the oscillator's flicker FM noise level. In reality, one or more components (typically the limiting loop amplifier and the SAW device) may both be contributing on a more or less equal basis.

The basic residual phase noise measurement set-up is shown in Fig. 5. The choice of a test source depends upon a number of factors. Generally, a frequency synthesizer based test source configuration is most useful when evaluating the residual flicker phase noise level of a SAW resonator or delay line device, since it may readily be adjusted in frequency in order to test a wide variety of relatively narrow bandwidth devices. We have previously shown [20] that a state-of-the-art, extremely low noise (both PM *and* AM) SAW resonator oscillator is quite useful when evaluating the residual phase noise properties of components which may exhibit appreciable AM-to-PM conversion sensitivities, e.g., amplifiers, electronic phase shifters, etc. Figure 6 shows the measured residual flicker phase noise level for a 500 MHz SAW resonator device with an active acoustic aperture of 300 λ and an effective cavity length of approximately 365 λ, with $\tau_g \sim 2.5$ μsec. The solid line ($1/f_m$ slope) shown in Fig. 6 represents the limiting residual phase noise system sensitivity floor for measurements on SAW resonator devices in this frequency range (typical insertion loss less than 7 to 9 dB). A Hewlett-Packard 8662A frequency

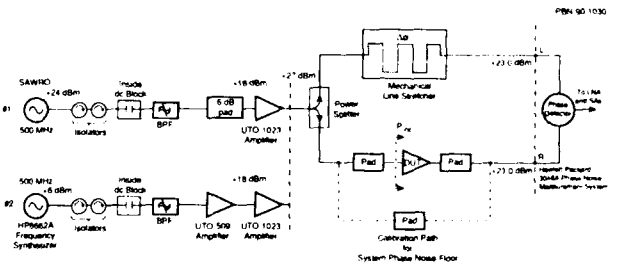


Fig. 5 Residual Phase Noise Measurement Set-up at 500 MHz.

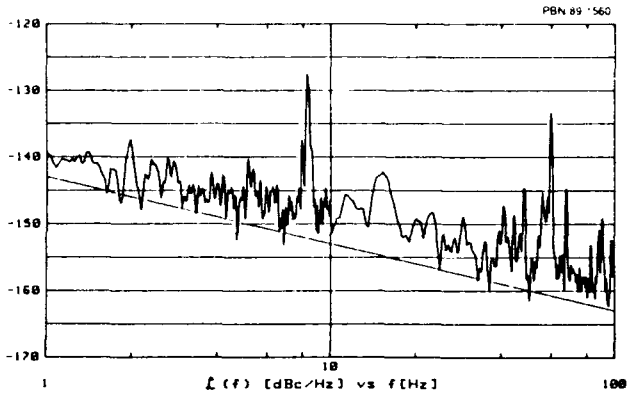


Fig. 6 Measured Residual Flicker Phase Noise Level for a Typical 500 MHz AQP SAW Resonator Device After a High Power RF Burn-in.

synthesizer was used as the test source for the measurement. Figure 7 illustrates a typical residual flicker phase noise level measured on a 675 MHz SAW delay line device ($\tau_g \sim 0.45$ μsec). The solid line ($1/f_m$ slope) shown in Fig. 7 represents the limiting residual flicker phase noise system sensitivity floor for measurements on SAW delay line devices in this frequency range (typical insertion loss less than 20 to 22 dB). Once again a Hewlett-Packard 8662A Frequency Synthesizer was used as the test source for the measurements. It should be noted that the measured flicker noise level for both cases just illustrated are each within 3 dB of the limiting system sensitivity noise floor, and therefore the actual residual flicker phase noise levels for the SAW devices are substantially (at least 3 dB) lower than the measurements would otherwise suggest. Generally, residual flicker noise measurements on SAW resonators are limited to carrier offset frequencies less than about 100 Hz, while for SAW

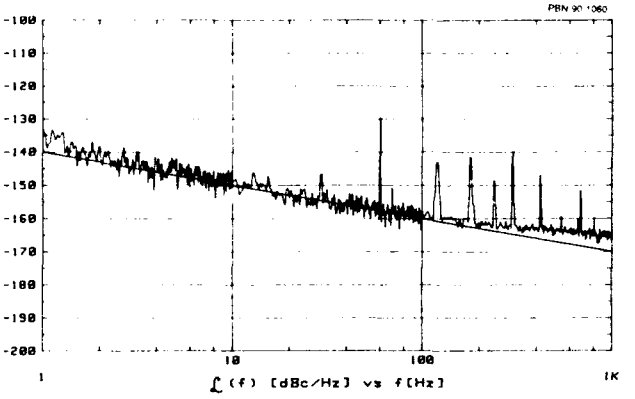


Fig. 7 Measured Residual Flicker Phase Noise Level for a Typical 675 MHz AQP SAW Delay Line Device Without a High Power RF Burn-in.

delay line devices they are limited to carrier offset frequencies less than approximately 1 kHz, since de-correlation of the test source's phase noise will occur due to the rather substantial unmatched delay introduced in one test leg by the device under test.

It is interesting to note that while the SAW resonator device whose residual flicker FM noise was shown in Fig. 6 received a burn-in using a high power incident RF signal level [21], the delay line device illustrated in Fig. 7 was not burned-in. In general, we have found that the high power RF burn-in technique may be used to lower the flicker phase noise levels of both SAW resonator as well as SAW delay line devices. However, many of our devices, both resonators and delay lines, are sufficiently close to the limiting system sensitivity noise floor even prior to burn-in that we are really unable to tell directly whether the burn-in has actually accomplished its objective or not, let alone to what extent the burn-in may have improved the device's flicker PM noise level. In these circumstances, the only way to really tell whether the high power RF burn-in has effected a decrease in a device's flicker noise level is to evaluate the device's performance in a "test bed" oscillator circuit of demonstrated low residual flicker noise. Clearly, this approach will ultimately be limited by the residual flicker noise level of the "test bed" oscillator circuit itself!

If a suitably low noise test source is used, then the basic residual phase noise test set-up shown in Fig. 5 is capable of evaluating the residual phase noise properties of electronic components, e.g., amplifiers and electronic phase shifters, for carrier offset from 1 Hz to 40 MHz with unprecedented sensitivity and reproducibility [20]. Figure 8 illustrates the measurement technique for three similar amplifier designs from three different vendors. The specific amplifiers are: 1) Cougar

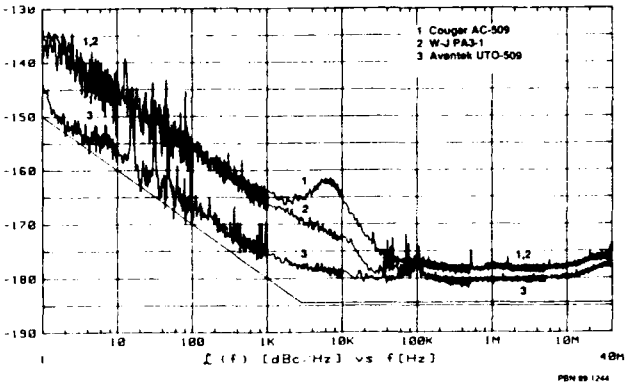


Fig. 8 Residual Phase Noise Measured for Three Similar Amplifiers Using SAWRO Test Source.

AC-509, 2) Watkins-Johnson PA3-1, and 3) Avantek UTO-509. For all three amplifiers the test conditions were identical, namely an incident RF test power level of +12.0 dBm at the amplifier's input. These three amplifiers are all remarkably similar in their nominal electrical performance parameters, which are summarized in Table I, along with the results of our own electrical measurements on the devices which were tested. The UTO-509's residual phase noise performance shown in Fig. 8 is typical of that observed on more than one hundred devices, while the specific PA3-1 and AC-509 data shown in Fig. 8 was essentially identical for samples of three and four devices, respectively. It is interesting to note that both the PA3-1 and the AC-509 have consistently higher levels of residual flicker noise, when compared to the UTO-509. Also, all four AC-509s which were tested exhibited an anomalous residual phase noise "bump" near 6 kHz carrier offset frequency. Finally, the higher residual white phase noise floor measured for the PA3-1 is consistent with the fact that its noise figure was found to be approximately 2 dB higher than that of the UTO-509. However, the comparatively high white phase noise floor for the AC-509 cannot be explained in a similar manner. Separate residual phase noise measurements were performed on each amplifier using a frequency synthesizer (Hewlett-Packard 8662A) based test source, primarily for its quantitatively higher AM noise level, thereby confirming that the AC-509 does indeed have a considerably higher AM-to-PM conversion factor than either of the other amplifiers. These results are shown in Fig. 9. It is evident that even with the extremely low noise SAW resonator oscillator based test source configuration, it simply isn't possible to accurately measure this particular amplifier's (the AC-509) residual white phase noise floor properties, at least when only a single device is tested at one time.

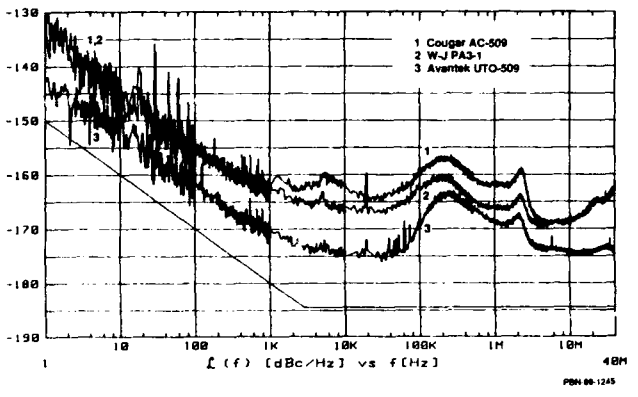


Fig. 9 Residual Phase Noise Measured for Three Similar Amplifiers Using Synthesizer Test Source.

Table I

Summary of Typical RF Amplifier Characteristics at T = 25°C

Amplifier	Small Signal Gain	Noise Figure	Pout @ 1 dB Compression	Pout @ 3 dB Compression	I_{DC} @ V_{DC}	Residual Phase Noise[2] @ (f ≥ 1 Hz)
Avantek/UTO-509						
Vendor Typical	14.3 dB	4.5 dB	+22.3 dBm	N.A.	90 mA @ +15V	N.A.
Measured	(14.2 dB)[1]	(4.2 dB)[1]	(+22.2 dBm)[1]	(+23.3 dBm)[1]	(100 mA @ +15V)	(- - 140 dBc/Hz)[1]
Watkins-Johnson/PA3-1						
Vendor Typical	14.5 dB	7.0 dB	+22.5 dBm	N.A.	91 mA @+15V	N.A.
Measured	(14.2 dB)[1]	(6.8 dB)[1]	(+22.6 dBm)[1]	(+23.6 dBm)[1]	(92 mA @+15V)	(- - 135 dBc/Hz)[1]
Cougar/AC509						
Vendor Typical	13.5 dB	5.0 dB	+22.0 dBm	N.A.	86 mA @ +15V	N.A.
Measured	(14.0 dB)[1]	(5.3 dB)[1]	(+22.0 dBm)[1]	(+23.1 dBm)[1]	(87 mA @ +15V)	(- - 136 dBc/Hz)[1]

[1] Measured at 500 MHz [2] Measured at 3 dB of Gain Compression N.A. = Not Applicable

PBN 88-1274

Recently, we have found that testing two amplifiers at a time (one in each test leg shown in Fig. 5) improves the situation dramatically for DUTs with large AM-to-PM conversion factors. This is because the resulting PM in each test leg due to the test source's AM noise will be correlated and therefore largely cancel, in a manner similar to the measurement configuration's suppression of the test source's PM noise. One disadvantage of this approach is that three DUTs would have to be tested (#1 vs. #2, #1 vs. #3, and #2 vs. #3) in order to uniquely extract information about each individual amplifier's residual phase noise properties.

It is perhaps interesting to note that neither an amplifier's residual flicker noise level, nor the possible presence of "bumps" in its residual phase noise spectrum, are predictable based upon any other electrical characteristic that we are aware of, including a small-signal noise figure measurement. Furthermore, as has been shown, even very similar amplifier designs from different vendors can differ markedly in their residual phase noise characteristics, even though their other nominal electrical parameters are nearly identical. It has been our experience that every single amplifier must be "screened" before use in an oscillator design where truly state-of-the-art phase noise performance is desired. Otherwise, you will constantly be attempting to diagnose which component (or components if Murphy's Law is found to prevail!) is the culprit when a particular oscillator's phase noise spectrum simply doesn't meet its intended specification.

The basic residual phase noise measurement technique may also be applied to other linear electronic components as well, e.g., an electronic phase shifter. Since an electronic phase shifter is quite often used in the feedback loop of a voltage controlled oscillator, its residual phase noise performance is very important. Figure 10 shows the circuit schematic for a low noise, electronic phase shifter with an extremely linear

transmission phase shift versus tuning voltage characteristic. This basic phase shifter design has been used with considerable success in a number of low noise, 500 MHz SAWROs [1]-[3]. Figure 11 shows a typical residual phase noise measurement on an electronic phase shifter. The test was performed with an RF power level of +23.5 dBm incident on the device. This is almost 10 dB higher than the actual RF power level which the phase shifter sees in the oscillator circuits it has been used in. Once again, the measured residual phase noise level is within 3 dB of the limiting system sensitivity noise floor for the measurement, as denoted by the solid line ($1/f_m$ slope) in Fig. 11. Since this result is so close to the measurement system's flicker phase noise floor, we can only infer that the electronic phase shifter's 1 Hz intercept for its residual flicker phase noise is approximately -150 dBc/Hz, or possibly even lower.

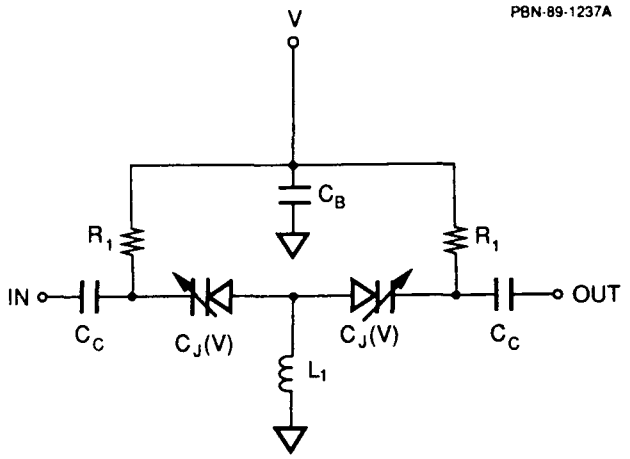


Fig. 10 Electronic Phase Shifter Circuit.

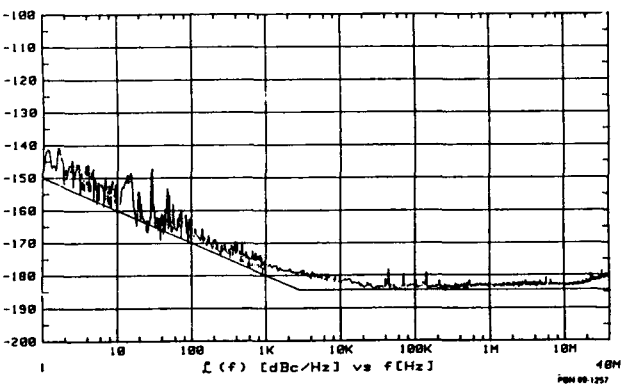


Fig. 11 Residual Phase Noise Measured on a Typical Electronic Phase Shifter.

It is perhaps worth commenting here that the limiting measurement system sensitivity noise floors shown in Figs. 6, 7, and 8, for residual phase noise measurements on a SAW resonator, a SAW delay line, and an RF amplifier, respectively, are due to the varying loss or gain exhibited by these devices. While an amplifier such as the UTO-509 may be tested as shown explicitly in Fig. 5, one (for a SAW resonator) or more (for a SAW delay line) RF amplifiers must be used when testing devices with significant insertion losses (greater than 2 or 3 dB) in order to achieve the best limiting system sensitivity noise floor for the measurement. Thus the amplifiers which must be used to test the residual flicker noise levels of SAW resonators and delay lines are part of the "basic" residual phase noise set-up, and their own flicker noise levels are therefore part of the limiting system sensitivity flicker noise level.

While there are some obvious limitations upon ones ability to accurately test devices for their residual flicker noise levels when these levels are less than about -145 dBc/Hz, the residual noise measurement technique just described is still extremely useful as a means of "weeding out" components which are clearly too noisy for the intended application. Finally, as even lower residual noise phase detectors (mixers), RF amplifiers and video (baseband) low noise amplifiers (LNAs) are developed our ability to characterize component residual noise properties will correspondingly improve. We are slowly approaching the point where a significant improvement in our limiting residual phase noise system floors will be necessary in order to successfully pursue a further order of magnitude reduction (to -60 to -65 dBc/Hz at 1 Hz carrier offset for a 500 MHz SAW resonator based oscillator and to -50 dBc/Hz to -55 dBc/Hz at 1 Hz carrier offset for a 400 MHz SAW delay line {τ_g = 1.55 μsec} based oscillator) in a SAW oscillator's flicker FM noise level.

V. OSCILLATOR DESIGN EXAMPLES

A. Introduction

While many factors enter into the design of an oscillator in response to a desired set of performance parameters, we have generally found that the specified phase noise spectrum is given considerable weight when evaluating whether a particular design is truly capable of addressing a specific oscillator requirement. Other performance parameters, e.g., vibration sensitivity, harmonic and spurious output signal levels, load pulling sensitivity, dc power supply pushing sensitivity, etc., may take a back seat in comparison to phase noise although an increased emphasis is now evident for low phase noise oscillators with minimal vibration sensitivities (~ $1x10^{-10}$/g per axis) for airborne and missile operating environments. Reference [7] presents

detailed discussions of the various factors which may enter into the design of an oscillator in response to a particular set of requirements. In the examples which follow, the oscillators' phase noise spectra were considered of paramount importance, and every attempt was made to achieve the best overall phase noise performance in each case.

B. Low Phase Noise SAW Resonator VCOs

We have previously described our efforts to develop extremely low phase noise SAW resonator VCOs [1]-[3], based upon the feedback-loop oscillator architecture illustrated in Fig. 12. Typically, for more than twenty engineering prototype oscillators, their phase noise performance was as illustrated in Fig. 13. A white phase noise floor of approximately -177 dBc/Hz for carrier offset frequencies greater than 100 kHz, and a flicker FM noise level of -75 dBc/Hz at 10 Hz carrier offset were routinely achieved. While this is not quite as low as described in Ref. [1], some compromises were necessary in order that the oscillators could be manufactured in production with acceptable yields. The various electronic components used in the oscillator design were described in Ref. [3].

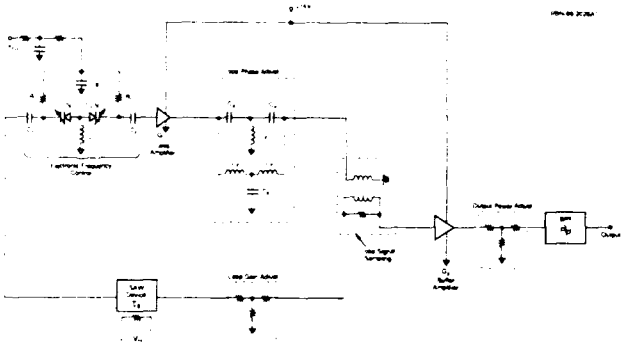


Fig. 12 Circuit Diagram for a SAW Resonator Based UHF Voltage Controlled Oscillator.

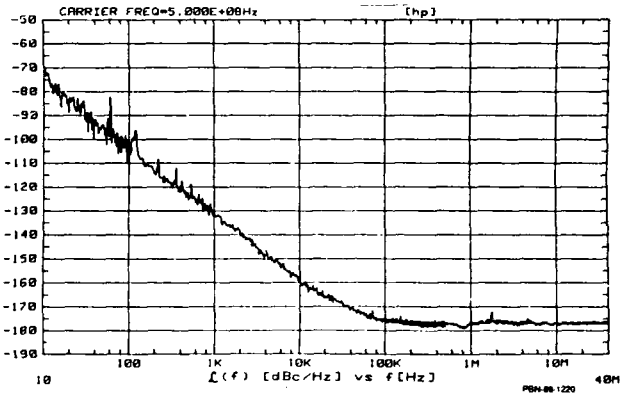


Fig. 13 Typical Phase Noise Spectrum Measured for One 500 MHz SAW Resonator VCO.

More recently, we have realized noticeable (> 3 dB) improvements in both the close-to-carrier flicker FM phase noise level, as well as the white PM noise level far-from-the-carrier, for laboratory prototype 500 MHz SAW resonator oscillators. This has been achieved by incorporating a new dual aperture SAW resonator design, as shown schematically in Fig. 14, as well as by using an even higher power, yet still suitably low residual flicker noise TO-8 RF amplifier, namely the UTO-1023 (Avantek). The result of these enhancements is shown in Fig. 15 for a laboratory prototype oscillator. The small "dip" (4 or 5 dB) in measured phase noise near 750 kHz carrier offset is attributable to noise degeneration in the loop due to the two satellite SAW device resonances, since the net transmission phase shift through the loop is almost exactly 180° at this carrier offset frequency! This is the only observable consequence of our basic multi-cavity mode SAW resonator design described earlier. It should be noted that even for two SAW resonators fabricated at the same time in reasonably close proximity on a single quartz substrate, they will quite often not have identical resonant frequencies. This is illustrated by the frequency response characteristic shown in Fig. 16(a). It is readily apparent that the overall frequency response of the composite

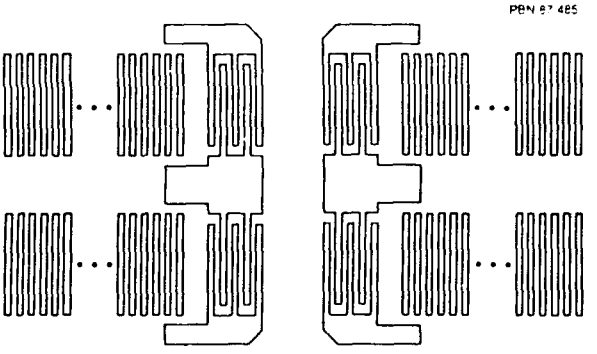


Fig. 14 Dual Aperture SAW Resonator Design.

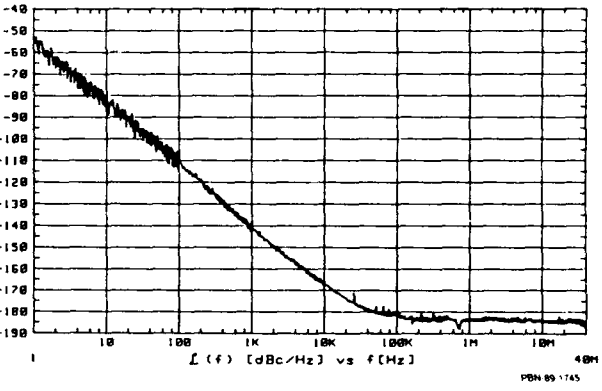


Fig. 15 Phase Noise Spectrum for a Laboratory Prototype 500 MHz SAW Resonator VCO.

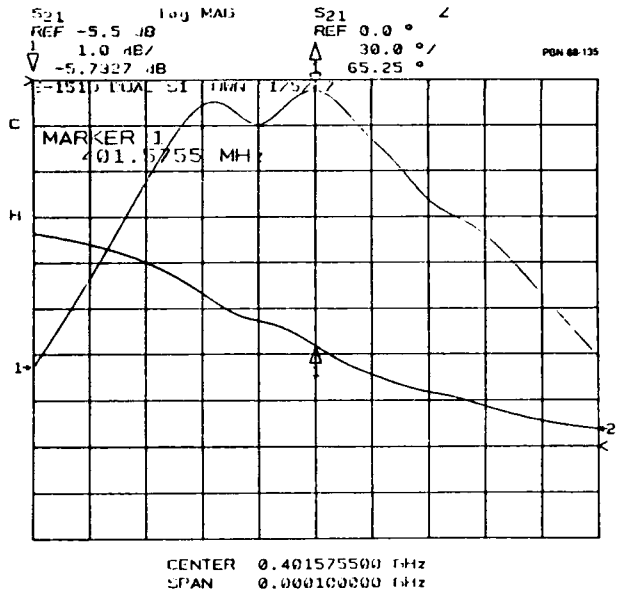


Fig. 16 Insertion Gain and Transmission Phase Plots for a Dual Channel SAW Resonator Device. (a) Before Laser-Trimming. (b) After Laser-Trimming.

dual channel device is substantially degraded due to this splitting of the individual device resonant frequencies. As described previously, laser-trimming of the individual resonant frequencies was used in order to eliminate this problem [10]. The resulting improvement in the device's frequency response is shown in Fig. 16(b), after laser-trimming. Each resonator device had an active acoustic aperture of 150 λ, thus the dual resonator device was essentially equivalent to the single 300 λ resonator devices which we have described previously [1]-[3], except that the unloaded Qs for the dual aperture devices were substantially better than those measured on single wide aperture resonators. This is attributable to the significant reduction in finger resistance (and thereby ohmic losses) for the dual aperture device design.

While the phase noise spectrum shown in Fig. 14 represents the current state-of-the-art for a 500 MHz SAW resonator oscillator, by no means does it represent a limit to what may ultimately be achieved with further improvements in both SAW resonator design and fabrication techniques, as well as anticipated improvements in RF amplifier residual phase noise levels.

C. *Low Noise SAW Delay Line VCOs*

While SAW delay line oscillators have generally been considered too noisy for most low noise applications, we have recently found that AQP SAW delay line devices can possess flicker phase noise levels which are substantially better than previously expected. This was indicated in Fig. 7 for a 675 MHz AQP SAW delay line device. In general, some of this *improvement undoubtedly* results from the fact that a high power RF burn-in has also been found to be effective in reducing the flicker phase noise of SAW delay line devices. However, as noted the delay line devices whose residual flicker noise levels were shown in Fig. 7 *did not* receive a high power burned-in. Careful AQP design for SAW delay line devices helps to minimize reflected signals, while the recessed transducer metallization (and "dummy" electrodes if necessary) will help to eliminate spurious triple transit reflections. It has been shown that these spurious signal may contribute to a device's measured flicker phase noise level [22]. It goes without saying that the transducer metallization process (Cu-doped aluminum in all of our SAW resonator and delay line devices, deposited using standard E-beam evaporation techniques) should obviously be well controlled and reproducible, since we have seen isolated instances where a device's flicker noise has been significantly degraded due to improper film deposition conditions.

Figure 17 shows the basic circuit schematic for an engineering prototype 400 MHz SAW delay line oscillator. The SAW device's group delay was approximately 1.55 μsec. All three RF amplifiers were Avantek UTO-509s. This particular oscillator has a usable tuning range of ±150 kHz (sufficient to

maintain phase lock to a stable reference signal over temperature as well as all other environmentally induced frequency changes), yet it still has a white phase noise floor of approximately -170 dBc/Hz and a flicker FM noise level of -70 dBc/Hz at 10 Hz carrier offset, as illustrated by the measured phase noise spectrum shown in Fig. 18. The broad "dip" in the oscillator's phase noise floor which is evident near 350 kHz carrier offset is due to noise degeneration in the feedback loop at frequencies $\pm 1/(2\tau_g)$ (~ ±325 kHz) where the net transmission phase through the feedback loop is 180° and the SAW delay line device's insertion loss is not too large. In general, the smaller group delay and higher insertion loss of a SAW delay line in comparison to a SAW resonator will preclude delay line based oscillators from achieving phase noise levels comparable to those demonstrated by SAW resonator oscillators, but in many applications, e.g., very wideband VCOs, the considerably larger tuning range of a SAW delay line oscillator is quite important. The phase noise performance shown in Fig. 18 serves to illustrate that good phase noise performance may still be achieved even for a relatively wideband SAW delay line oscillator.

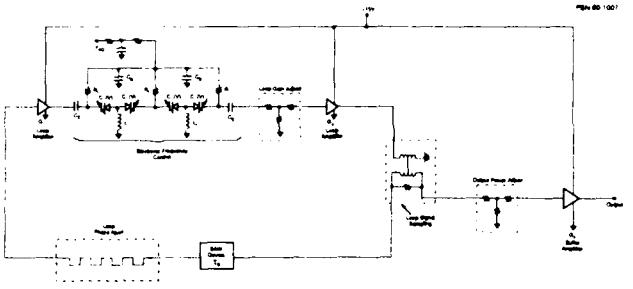


Fig. 17 Circuit Diagram for a SAW Delay Line Based UHF VCO.

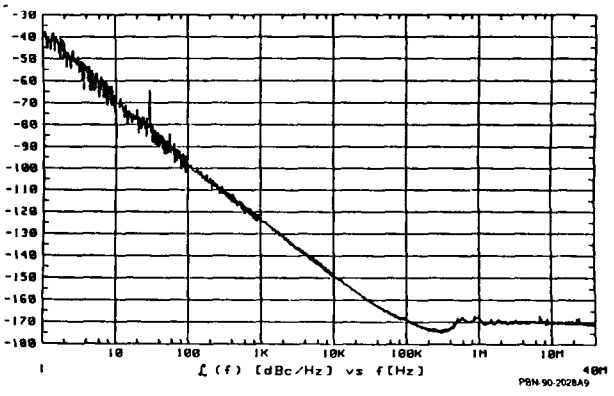


Fig. 18 Measured Phase Noise Spectrum for One Engineering Prototype 400 MHz SAW Delay Line VCO.

More recently, a laboratory prototype 675 MHz SAW VCO was evaluated based upon devices similar to those whose residual flicker noise was shown in Fig. 7. The same basic circuit design as shown in Fig. 17 was once again used, but now the three amplifiers were each UTO-1005s. This particular amplifier is essentially "identical" to the UTO-509, except that reactive broadbanding has been used to achieve comparable electrical performance to 1 GHz, rather than to the 500 MHz upper limit of the UTO-509. The measured phase noise spectrum for one such oscillator is shown in Fig. 19. Since the group delay of this device is only 0.45 μsec, an oscillator properly configured with this delay line device would be capable of electronic frequency tuning over a ±550 kHz range, with negligible degradation in its phase noise spectrum. In spite of this extremely wide tuning range potential, the oscillator still possesses a white phase noise floor of approximately -167 dBc/Hz, and a flicker FM noise level of -60 dBc/Hz at 10 Hz carrier offset. The broad "dip" in the oscillator's phase noise floor near 1 MHz carrier offset is once again attributable to noise degeneration in the feedback loop for frequencies $\pm 1/(2\tau_g)$ (~ 1.1 MHz) removed from the carrier, where the net loop transmission phase shift is 180°. The approximately 10 dB degradation in flicker FM noise level for this delay line oscillator, in comparison to the result for the 400 MHz SAW delay line oscillator's flicker FM noise level shown in Fig. 16, is due almost entirely to the difference in group delay between the two devices (i.e., $10\text{x}\log_{10}\{1.55/0.45\}^2 = 10.7$ dB) [5]. The difference in white phase noise floor levels between the two delay line oscillators is primarily due to the somewhat inferior noise figure of the UTO-1005 in this frequency range in comparison to the UTO-509, as well as to the fact that the output power at 3 dB of gain compression for the UTO-1005 is also somewhat reduced compared to that for the UTO-509.

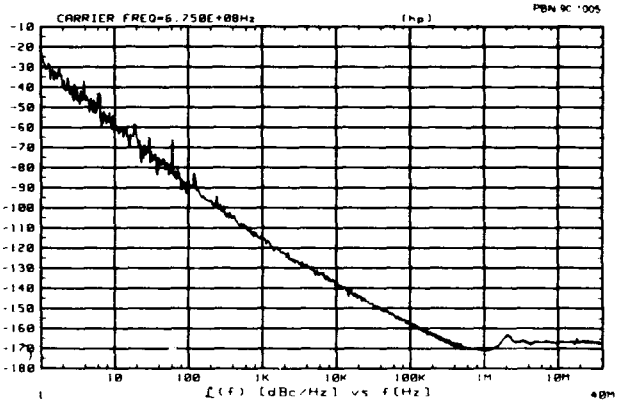


Fig. 19 Measured Phase Noise Spectrum for One Laboratory Prototype 675 MHz SAW Delay Line VCO.

In either case, it is evident that careful SAW device design, as well as close attention to processing parameters, along with close monitoring of electronic component residual phase noise levels can result in SAW delay line VCOs with extremely wide tuning ranges, while simultaneously maintaining very low phase noise levels.

D. An L-band Dielectric Resonator Oscillator

While an L-band dielectric resonator oscillator may be quite large (and heavy to boot), if properly designed it is capable of phase noise performance which cannot be equaled using any other currently practical oscillator technology. However, it has also been noted that the vibration sensitivity likely to be achievable with dielectric resonator oscillators will be inferior to that readily achievable with SAW devices ($1x10^{-9}$/g per axis) [4]. If phase noise is the *primary* motivating performance requirement, then a dielectric resonator is worthy of consideration. Figure 20 shows the very basic feedback-loop oscillator design which we have implemented with a conventional two-port, transmission mode dielectric resonator cavity design. The resulting phase noise spectrum for the 982 MHz dielectric resonator oscillator is shown in Fig. 21. Note that it is superior (by anywhere from 3 to 6 dB) to the SAW resonator result shown in Fig. 15, after that result has been referenced to L-band (~ 6 dB phase noise degradation). It is worth noting, in view of our earlier comment about the relatively high vibration sensitivities reported for DR oscillators, that the noticeable perturbations to the DRO's phase noise spectrum in the 10 Hz to 40 Hz offset range are likely due to mechanical vibrations coupling into the dielectric resonator oscillator assembly, in spite of reasonable efforts to eliminate the problem by vibration isolating the oscillator with foam, as well as other materials. The basic design approach and component evaluation techniques previously described were applied here to the dielectric resonator oscillator as well.

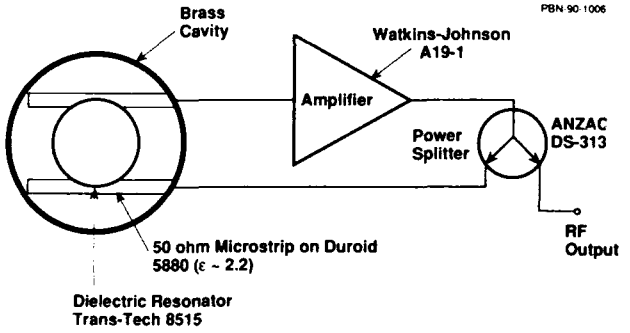


Fig. 20 Circuit Diagram for a Basic Dielectric Resonator Stabilized Feedback-loop Oscillator.

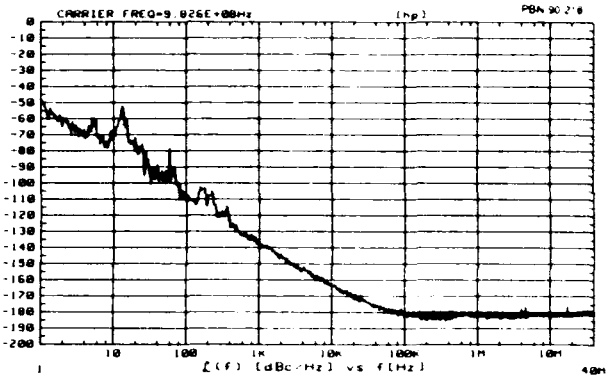


Fig. 21 Measured Phase Noise Spectrum for One L-band (982 MHz) Dielectric Resonator Oscillator.

E. Summary

It is indeed gratifying that the basic feedback loop oscillator design, which was first analyzed by Leeson, may be successfully applied under such a wide variety of circumstances, including the potential use of many different classes of frequency stabilizing element with equally satisfactory results. In fact, it is perhaps appropriate to conclude with the observation that when properly designed and implemented, the feedback-loop oscillator configuration is well suited to the realization of truly state-of-the-art oscillator performance. Based upon our own experience, this statement applies not only to the exceptionally low phase noise levels which have been achieved, but also to any of the other measures of an oscillator's performance which might be applied, as we have previously demonstrated [1]-[3].

VI. SUMMARY & CONCLUSIONS

We have shown, using several diverse design examples, that the basic feedback-loop oscillator configuration is well suited to high performance oscillator design in the VHF, UHF and microwave frequency ranges, based upon the incorporation of a two-port frequency stabilizing element. State-of-the-art performance was realized for SAW resonator and delay line based oscillators, as well as for a very simple L-band dielectric resonator oscillator design.

The design and evaluation techniques just discussed have proven to be invaluable in selecting the proper types of components for obtaining truly low noise oscillator performance. As shown in Fig. 8 not all amplifiers are created equal, at least with regard to their residual phase noise levels. Residual phase

noise measurements are also very valuable in screening out obviously defective components that would otherwise contribute unacceptably high levels of residual phase noise. The proper evaluation of amplifier characteristics has allowed us to predict an oscillator's white phase noise floor to within about ±1 to ±2 dB. However, the ability to predict flicker FM noise levels on individual oscillators (based upon component residual noise measurements) has been somewhat limited due to the fact that in many cases the residual flicker PM noise levels of the components are at or below measurable levels. Experience with more than thirty fully characterized PC board oscillators (SAW resonator, SAW delay line and dielectric resonator based) has given us considerable confidence in the feedback-loop oscillator based design procedures which we've described.

However, there are always exceptions, indicating that there is still more to be learned. A small fraction of some recently fabricated L-band hybrid circuit SAW resonator oscillators (using Avantek MSA-0770 RF amplifiers) have shown unusually high flicker FM noise levels, even after the components had all been carefully screened for residual noise prior to the oscillators' assembly. In one case, we even measured the open-loop residual flicker PM noise of an oscillator and found it to be consistent with its high level of closed-loop flicker FM noise. Yet, repeated measurements of the *individual component*'s residual flicker PM noise levels failed to reveal the culprit. In spite of our best efforts, we have not been able to determine why that particular oscillator had such a high flicker noise level. Clearly, there are factors present that are not as yet fully understood.

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] 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.

[3] 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.

[4] M. J. Loboda, T. E. Parker, and G. K. Montress, "Frequency Stability of L-Band, Two-Port Dielectric Resonator Oscillators", *IEEE Transactions on Microwave Theory and Techniques*, Vol. MTT-35, No. 12, pp. 1334-1339, December 1987.

[5] T. E. Parker and G. K. Montress, "Precision Surface Acoustic Wave (SAW) Oscillators", *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 3, pp. 342-364, May 1988.

[6] D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proceedings of the IEEE*, Vol. 54, No. 2, pp. 329-330, February 1966.

[7] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators," in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 99-110.

[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 Wave Resonator," in *Proceedings of the 42nd Annual Symposium on Frequency Control*, 1988, pp. 239-251.

[10] J. A. Greer, G. K. Montress, and T. E. Parker, "Applications of Laser-Trimming for All Quartz Package, Surface Acoustic Wave Devices," in *Proceedings of the IEEE Ultrasonics Symposium*, 1989, Vol. 1, pp. 179-184.

[11] W. J. Tanski and H. van de Vaart, "The Design of SAW Resonators on Quartz with Emphasis on Two Ports," in *Proceedings of the IEEE Ultrasonics Symposium*, 1976, pp. 260-265.

[12] W. R. Shreve and P. S. Cross, "Surface Acoustic Wave Resonators," in *Precision Frequency Control: Acoustic Resonators and Filters*, Vol. 1, F. A. Gerber and A. Ballato, Eds. New York: Academic Press, 1985, pp. 118-145.

[13] W. R. Shreve, R. C. Bray, S. Elliott, and Y. C. Chu, "Power Dependence of Aging in SAW Resonators," in *Proceedings of the IEEE Ultrasonics Symposium*, Vol. 1, 1981, pp. 94-99.

[14] J. Crabb, M. F. Lewis, and J. D. Maines, "Surface-Acoustic-Wave Oscillators: Mode Selection and Frequency Modulation," *Electronics Letters*, Vol. 9, No. 10, pp. 197-199, May 17, 1973.

[15] 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 Transactions on Instrumentation and Measurement*, Vol. IM-24, No. 1, pp. 15-20, March 1975.

[16] F. L. Walls and C. M. Felton, "Low Noise Frequency Synthesis," in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 512-518.

[17] T. R. Faulkner and R. E. Temple, "Residual Phase Noise and AM Noise Measurements and Techniques," Hewlett-Packard Application Note, H-P Part Number 03048-90011.

[18] G. S. Curtis, "The Relationship Between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques," in *Proceedings of the 41st Annual Symposium on Frequency Control*, 1987, pp. 420-428.

[19] T. F. O'Shea, M. R. Lewis, and B. R. Horine, "Low-Noise Voltage Controlled SAW Oscillator for Phase Lock Loop Application," in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 53-59.

[20] G. K. Montress, T. E. Parker, and M. J. Loboda, "Residual Phase Noise Measurements of VHF, UHF, and Microwave Components," in *Proceedings of the 43rd Annual Symposium on Frequency Control*, 1989, pp. 349-359.

[21] M. J. Loboda, T. E. Parker, J. A. Greer, and G. K. Montress, "Reduction of Close-to-Carrier Phase Noise in Surface Acoustic Wave Resonators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1987, Vol. 1, pp. 43-46.

[22] R. L. Baer, "Phase Noise in Surface-Acoustic-Wave Filters and Resonators," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, Vol. UFFC-35, No. 3, pp. 421-425, May 1988.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

UNIVERSAL, COMPUTER FACILITATED, STEADY STATE OSCILLATOR ANALYSIS THEORY

Benjamin Parzen
consulting engineer
San Diego, CA 92103

ABSTRACT

The theory of oscillator analysis in the immittance domain is presented. This theory enables the computer simulation of the steady state oscillator. The simulation makes practical the calculation of the oscillator total steady state performance, including noise. Some oscillator applications of PC program, BPT, created for the simulation, are listed.

1. INTRODUCTION

1.1 During the past 12 years, the writer has been developing theory for oscillator analysis in the immittance domain.

"IMMITTANCE DOMAIN" means that the basic relationships are expressed in terms of immittance, $Z = R + jX$ or $Y = G + jB$.

Effective and facile application of the theory to real problems also necessitated the creation of a PC computer circuit analysis program. The results of this effort are

- A remarkably simple oscillator theory which fully describes the oscillator operation and which is readily translatable into the real world.
- A very user friendly multipurpose and powerful circuit analysis program, BPT.

The theory and program are universal in that they apply to all oscillators, past, present, and future.

The theory and program are symbiotic, in that

- To apply the program to oscillators, the theory must be used.
- The program is very helpful and almost indispensable in applying the theory to real oscillator problems including that of improving the theory.

It should be noted that, as BPT does not incorporate any basic information peculiar to oscillators, it follows that the theory may be used, in conjuction with other circuit analysis programs, to analyze oscillators, but with much greater, and perhaps prohibitive, difficulty.

1.2 The theory is based upon the combination of 2 mutually compatible oscillator models

- The negative resistance model [2] as the primary model.
- The noise source, amplifier, filter, model [3] in which, positive feedback produces the necessary extremely high effective amplifier gain, as the secondary model.

2. THE NOISELESS OSCILLATOR
(See Ref 2)

2.1 Fundamental relationships

At steady state, in any and every mesh of a hypothetically noiseless oscillator,

$$Zt = \sum Z = 0 \qquad (1)$$

From which,

$$\sum R = 0 \qquad (1a)$$

$$\sum X = 0 \qquad (1b)$$

Defining,

$$RN = \sum Rnegative, \quad RT = \sum Rpositive \qquad (2)$$

Then,

$$RT = -RN \qquad (3)$$

The dual statement of Eq 1 is,

At steady state, across any and every 2 nodes of a hypothetically noiseless oscillator,

$$Yt = \sum Y = 0 \qquad (4)$$

It is noteworthy that Eqs 1 and 4 are completely analagous to Kirchoff's first and second network laws. It may, therefore, be in order to name Eqs 1 and 4 as the oscillator first and second laws.

It is interesting to speculate that similar sets of laws may exist in other fields and that efforts should be made to discover and apply them.

2.2 Non-linearity considerations

If the linear and non-linear elements are grouped together, then, from Eqs 1

$$\sum Rlin + \sum Rnlin = 0 \qquad (5)$$

CH2818-3/90/0000-536 $1.00 © 1990 IEEE

from which,

$$\Sigma\, Rnlin = - \Sigma\, Rlin \qquad (5a)$$

Similarly,

$$\Sigma\, Xnlin = - \Sigma\, Xlin \qquad (5b)$$

where the non-linear values are the effective values.

Eqs 5 are very important since they state that it is only necessary to study the linear R and X elements to compute the totals of the non-linear R and X elements. This fact drastically reduces the labor involved in the analysis and, very often, it eliminates the need for considering the detailed behavior of non-linear elements.

2.3 Noiseless oscillator model

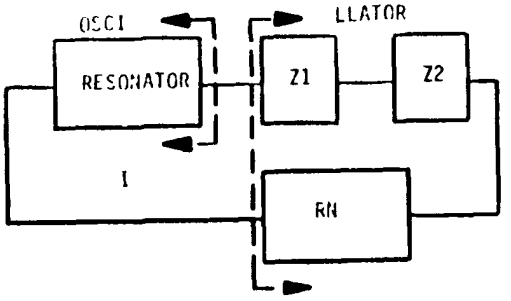


Fig. 1 Noiseless Oscillator Model

Fig 1 is the simplefied diagram of the oscillator resonator mesh. (The mesh containing the resonator is usually chosen for analysis as it has the most information.)

As shown in Fig 1, that part of the oscillator containing the resonator circuitry is called the osci, symbol *os*. The remainder is called the llator, symbol *LL*.

From Eq 1,

$$Z_{os} = - Z_{LL} \qquad (6)$$

Also,

$$\Delta F = - \Delta X_{LL} / (\delta X_{os} / \delta f) \qquad (7)$$

Eq 7 enables the easy determination of oscillator frequency shifts caused by llator changes.

3. THE REAL OSCILLATOR

3.1 Introduction

The real oscillator always includes one or more llator noise sources, in addition to the resonator frequency noise. The calculation of the contribution of the resonator noise to the oscillator noise is very simple (See Section 3.3) Howver, the calulation of the contributions of the llator noises is extremely difficult and constitutes a major topic of this paper.

3.2 Frequency relationships

The total frequency, F, is described by

$$F = fo + f \qquad (8)$$

where fo is the carrier frequencuy
f is the offset or Fourier frequency

3.3 Contribution of the resonator frequency noise to the oscillator noise.

Let the resonator frequency noise be described as

$$[S_f\ (f)]_{os} \qquad (9)$$

then the resonator frequency noise contribution to the oscillator frequency noise at all oscillator locations is identical to that in Eq 9.

The resonator frequency noise contribution to the oscillator phase noise at all oscillator locations is

$$[S_f\ (f)]_{os} / f^2 \qquad (10)$$

The total phase noise at any oscillator location is given by the sum of that of Eq 10 and the sum of the contributions of the llator noise sources to the phase noise at that location.

3.4 Contributions of the llator noise sources to the oscillator noise at location, m.

Let the noise of noise source, n, be given by power spectrum,

$$PS_n\ (f) \qquad (11)$$

then the contribution of this source to the oscillator noise power spectrum at location m, is

$$PS_{mn}\ (f) \qquad (12)$$

and

$$PS_{mn}\ (f) = PS_n\ (f) * CF_{mn}\ (f) \qquad (12a)$$

where $CF_{mn}\ (f)$ is defined as the contribution function of noise source, n, to the oscillator noise at location m. Note that CF_{mn} may have dimensions.

The total oscillator noise, PSt, at location m, obviously is

$$PSt_m\ (f) = \sum_{1}^{n} PS_{mn}\ (f) \qquad (13)$$

3.5 Real oscillator, special case, N configuration

Fig 2 is the complete diagram for the N configuration (N meaning noise) of a special case real oscillator. The fact that makes this oscillator a special case is that the noise source, V_n, represents the total equivalent noise contributions of all the noise sources in the llator.

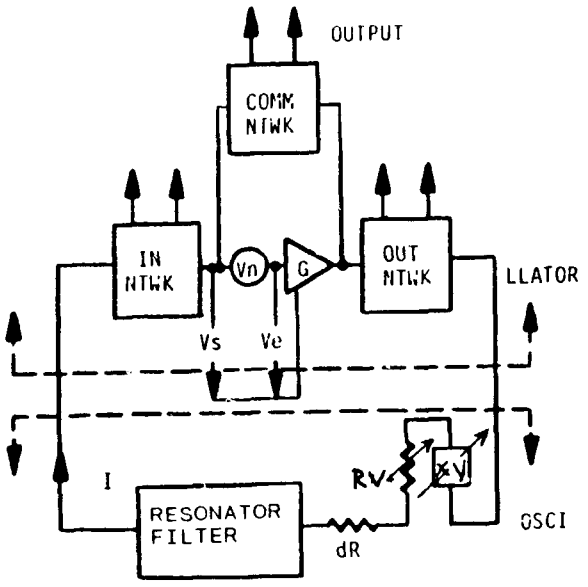


Fig 2. Real Oscillator, Special Case, N configuration

All real oscillators always contain a non-physical very small value resistance, *dR*. The value of *dR* is determined by the oscillator limiting, or ALC, system which sets the level of the oscillator output.

Fig 2 includes the input, output, and common networks and the multiple potential output points, available in every real oscillator. The noise in the output depends upon the output point location.

Fig 2 also shows *RV* and *XV* which do not exist in the real oscillator. Their function is described in Sect 3.7.

The oscillator, at steady state, at f=0, (the carrier frequency) satisfies Eqs 14 to 16, similar to and derived from Eqs 1 to 3,

$$Zt = \Sigma Z = dR \qquad (14)$$

from which,

$$\Sigma R = dR \qquad \Sigma X = 0 \qquad (15)$$

$$RT = - RN + dR \qquad (16)$$

Also, it has been proven that

$$\left|\frac{RN}{dR}\right|^2 = \left|\frac{Vs(0)}{Vn(0)}\right|^2 \qquad (17)$$

This equation is used in calculating the oscillator operating Q.

Let $(RN / dR)^2 = (Ar)^2$ (18)

3.6 Real oscillator operation

a. The oscillator is constructed and power is applied.
b. The operator sets the output level by adjusting the limiting circuitry, and thus setting *dR*.
c. The oscillator determines *fo*, V_n, *RT*, and all other operating conditions.
d. The oscillator determines the phase noise as a function of *f*.

3.7 The real oscillator, special case, Z configuration

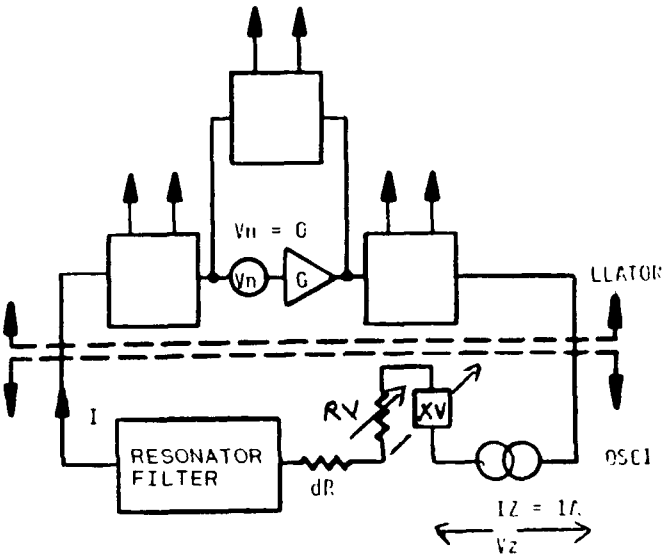


Fig. 3 Real Oscilator, Special Case, Z Configuration

Fig 3 is the complete diagram for the Z configuration (Z meaning impedance) of the special case real oscillator.

The purpose of this configuration is to enable the precise setup of the necessary and sufficient oscillatory condition of Eq 14. Once this is done, we are certain that this configuration when converted to the N configuration represents a true oscillator at the desired frequency.

As will be noted, the Z configuration differs from the N configuration in the following:

a. The value, V_n, of the noise is 0.
b. The circuit has been broken and a 1 A current source inserted. The value of the voltage, V_z, across this source is the value of the impedance *Zt* of Eq 14. When *Zt* has been precisely adjusted to *dR* at the desired frequency, *fo*, then this Z configuration is ready for conversion to a N configuration. The adjustment procedure is called " zeroing ".

RV and *XV* are very small value trimmable resistor and reactor, respectively, provided to facilitate zeroing. Their values are identical in the Z and N configurations.

The zeroing procedure has been automatized in the computer simulated laboratory and thus simplefies the computer oscillator simulation.

3.8 Computer simulated oscillator operation

In the program oscillator, Z config. (See Fig 3)

a. The oscillator is constructed by entering the circuit into the computer.
b. The operator sets fo, and $V_n = 0$.
c. The program sets the effective gain of the active circuitry, so that Eq 14 is almost satisfied.
d. The program adjusts *RV* and *XV* so that the mesh impedance, $Zt = V_z$ in Fig.3, satisfies Eq 14 to a very high degree of precision. The Z configuration is then converted into the N configuration.

In the program oscillator, N config. (see Fig 2)

e. The operator sets V_n to V_n real.
f. The operator sets *dR* to the value where the desired oscillator output is obtained as measured by the ac current in any component or ac voltage between any 2 points.
g. The program determines the voltages and currents at all other points.
h. The program determines the phase noise at and between all points.
i. The program determines $Zt(f)$, (see Fig 3) when *dR* is at the value of Step f, and V_n is set to 0.

3.9 Computation of the contribution functions, CF_{mn} in the real oscillator

This is best done by using the program. The procedure is the following

a. Start with the oscillator configured as in Sect 3.8, Step d. Include all the noise sources, which are assumed to be uncorrelated.
b. Create the equivalent N configuration.
c. To compute CF_{mn}, set the magnitudes of all noise sources to 0, excepting the source at location n.
d. Make the source, n, a unit white noise source.
e. Determine the noise response at location m,

$$PS_{mn}\ (f)$$

f. Then from Eq 12a,

$$CF_{mn}\ (f) = PS_{mn}\ (f) \qquad (19)$$

g. Repeat steps c to f for all noise sources.

3.10 Oscillator noise in all real oscillators, due to V_n (See Fig 2)

3.10.1 The resonator current, I_x, phase noise, due to V_n, is given by

$$\mathcal{L}_{I_x}(f) = \mathcal{L}_{v_n}(f) * \left|\frac{RT}{Zt(f)}\right|^2 \qquad (20)$$

$Zt(f)$ is obtained with the Z configuration and

$$\mathcal{L}_{v_n}(f) = PS_{vn}(f) / (V_s(0))^2 \qquad (21)$$

3.10.2 Noise at other locations

The phase noise in V_s is almost the same as in I_x, except at the higher values of f.

Now that the phase noise in I_x is known, the determination of the phase noise, at all other locations, is straightfoward but, depending upon the circuit complexity, can be very difficult and tedious, and is best done with the program.

However those, desiring to perform the calculations, should keep in mind the following rules when combining noises at any location:

a. Noise voltages and currents, due to the same noise source, should be combined as phasors.

b. Noise powers, due to different uncorrelated noise sources, should be combined as scalars.

3.11 The general relationship between $PS(f)$ and the phase noise, $\mathcal{L}(f)$ at the same location, is

$$\mathcal{L}(f) = PS(f) / \text{Carrier Power} \qquad (22)$$

$$= PS(f) / PS(0) \text{ for all values of } f \qquad (23)$$

when PS (0) does not approach infinity.

Eq 22 is also useful for the case where $V_n(f)$ has a flicker or other noise component which theoretically approaches infinity as f approaches 0 and thus theoretically also makes $PS(f)$ approach infinity as f approaches 0. In this case, the oscillator noise power spectrum, at f when $|Xt(f)| >> dR$, is independent of *dR*, which is set by the oscillator limiting system at the desired carrier. Therefore, at these values of f,

$$\mathcal{L}(f) = PS(f) / PC \qquad (24)$$

where PC is the arbitrary desired carrier reference power.

3.12 The phase noise contribution function $PCF_{mn}(f)$

$PCF_{mn}(f)$, the phase noise contribution function, is defined as

$$\frac{CF_{mn}(f)}{CF_{mn}(0)} \qquad (25)$$

and is dimensionless.

$PCF_{mn}(f)$ is usefully interpreted as the phase noise contribution, at location m, of a white noise source, n, for the same circuit configuration. It may be considered as a figure of merit for the oscillator phase performance of the circuitry for equal noise sources.

For example, Eq 20 may now be rewritten as

$$\mathcal{L}_{Ix}(f) = \mathcal{L}_{vn}(f) * (Ar)^2 * PCF_{Ix\,vn}(f) \qquad (26)$$

where

$$PCF_{Ix\,vn}(f) = \left|\frac{dR}{Zt(f)}\right|^2 \qquad (27)$$

4. PROGRAM APPLICATIONS IN OSCILLATORS

Some applications to oscillators are cited below to demonstrate the utility and power of program BPT when used in conjunction with the theory. It shoukd be remembered that BPT's high resolution and accuracy and virtual ground node facility are of prime importance as they enable the precise zeroing of the Z configuration required to simulate the oscillator.

4.1 Automatic zeroing of the Z oscillator configuration (coarse, active device gain setting and fine, impedance trimming) for the derivation of the $Zt(f)$ relationship.

4.2 Automatic calculation of the oscillator operating frequency.

Both 4.1 and 4.2 functions are high speed and high resolution executions of their respective functions.

4.3 Calculation of the DC operating point of oscillators.

4.4 Includes a procedure for determining the AC operating point of a self-limiting oscillator.

4.5 Includes a procedure for setting the AC and DC operating points in ALC type oscillators.

4.6 The program is a linear one but is designed to be capable of being interfaced with non linear operating conditions. Items 4.3 to 4.5 are examples.

4.7 Investigation of the $Zt(f)$ or $Zt(?)$ function of an oscillator or circuit such as a llator.

The following 5 examples illustrate the very useful information obtained from noiseless llator studies.

4.7.1 The effect of component changes upon frequency.

4.7.2 Overtone and mode selector gain margins and the effect of the resonator overtone and mode selector curcuitry upon the oscillator stability..

4.7.3 Starting gain margin (loop gain, ALo)

4.7.4 Effect of power dissipating components upon the loop gain and the operating Q.

4.7.5 Effect of component tolerance and environment upon the above 4 items

4.8 Determination of bypass and coupling capacitor adequacy.

4.9 The performance of subcircuits such as tuning networks.

4.10 The determination of circuit isolation properties.

4.11 Calculation of operating Q, from Eq 17.

4.12 Studies of impedance properties of resonators and other devices requiring high resolution.

4.13 Investigation of the effect of component and subcircuit noise upon oscillator noise.

4.14 Determination of the cicrcuit configuration for optimum noise performance.

4.15 Calculating and plotting oscillator phase noise for white noise and white plus flicker noise sources. It is not necessary to assume symmetrical noise sidebands.

4.16 Calculating and plotting the resonator phase noise for white noise sources from the $Zt(f)$ relationship.

4.17 In setting up an oscillator, the oscillator Z configuration is first created and then zeroed. (See Item 4.1). A facility is provided to automatically convert the zeroed Z configuration into an N configuration and vice versa.

5. THE PROGRAM AS A RESEARCH TOOL

The following suggested applications illustrate the power and usefulness of the program in research activities:

5.1 The formulation, checking, and confirmation of new theory. and new oscillator designs.

5.2 Determination of component aging from experimentally derived llator aging.

5.3 Determination of component temperature characteristics from experimentally derived llator temperature performance.

5.4 Determination of llator noise by experimentally deriving the oscillator phase noise and then program calculating the $V_n(f)$ function required to produce that oscillator noise.

6. CONCLUSIONS

Extremely simple and powerful oscillator theory has been presented. This theory has been used in the crestion of a cumputer program for the universal analysis of oscillator steady state performance.

The importance of $Zt(f)$, RT, and dR and the tremendous power of analysis in the immittance domain complemented by computer aided analysis have been demonstrated. An important additional advantage is that this anlysis method provides considerably greater understanding of the operation of the real oscillator.

7. REFERENCES

1. Leeson, D. B., "A Simple Model of Feedback Oscillator Noise Spectrum.", Proc. I.E.E.E. 54, 329-330 (Feb. 1966)

2. Parzen, B. "Design of Crystal and Other Harmonic Oscillators", Wiley, New York, 1983

3. Robins, W.P. " Phase Noise in Signal Sources", Peter Peregrinus Ltd., London, 1982

4. Parzen, B. "Clarification and a Generalized Restatement of Leeson's Oscillator Noise Model", Proc. 42nd Annual Frequency Control Symposium, pp.348-351, June 1988

5. Parzen, B. "Oscillator and Stability Analysis in the Immittance Domain ", in preparation.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH SPECTRAL PURITY X-BAND SOURCE

F.L. Walls and C.M. Felton
Time and Frequency Division
National Institute of Standards and Technolgoy
Boulder, CO 80303

and

T.D. Martin
Gravity Research Institute
Boulder, CO

ABSTRACT

We have developed a X-band frequency source that has very high spectral purity and is suitable for frequency synthesis and many kinds of high-resolution spectroscopy. A commercial dielectric resonator oscillator (DRO) is frequency locked to a high-Q cavity used as a frequency discriminator. Many systems have been developed in the past for locking a source to a reference cavity. The distinguishing features of our approach are (1) the system is relatively simple and inexpensive, (2) our approach does not use any modulation techniques, and (3) the phase noise in the 1 to 100 kHz region is comparable to the best that has been reported-S_ϕ(10 kHz) approximately -135 dBc- for a free-running, room temperature, X-band source. The lack of modulation on the source means that it can be used for very high order frequency synthesis (in principle up to 250 THz without loss of the carrier). Both the transmitted and reflected signal from the cavity are used to obtain a very steep discriminator curve. Phase compensation of the amplified discriminator signal is used to extend the unity gain point well beyond the half bandwidth of the discriminator cavity. This allows the servo system to reduce the phase noise of 1 kHz from the carrier by about 60 dB relative to the free-running performance. Phase noise inside of 1 kHz can be controlled by a signal multiplied up from a low frequency crystal oscillator for even lower phase noise near the carrier. We describe the design of the discriminator cavity, the phase compensated frequency-locked loop, and the phase and amplitude noise performance. We also discuss vibration sensitivity, techniques for automatically locking to the low frequency reference, and the reduction of 60 Hz sidebands.

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

INTRODUCTION

A perfect frequency multiplier increases the phase noise of the output signal by a factor of N^2 over the phase noise of the input source. As the phase noise increases, the power in the carrier is distributed over a wider and wider frequency interval about the carrier. At high enough multiplication factor, which depends on the phase noise characteristics of the source, the carrier signal can no longer be separated from the noise [1, 2]. This condition is called "carrier collapse" and is one of the two primary limitations to high order frequency multiplication. The other limitation is the degree of nonlinearity in the multiplier [2, 3]. The quadratic growth of the phase noise and eventual carrier collapse impose very severe requirements on the phase noise of sources that are used for high order frequency multiplication.

We have kept these considerations in mind in the development of a low phase noise source at X-band. Traditional methods for locking a noisy oscillator to a stable passive reference frequency usually impose some form of modulation on the source signal [4-7]. The modulated carrier signal then interacts with the frequency reference to yield an ac error signal that is compared with the imposed modulation (usually in a lock-in detector) to derive a dc error signal to correct the frequency of the noisy oscillator. In such an approach it is often difficult to remove the modulation signals from the carrier. The presence of the modulation sidebands on the carrier could then limit the range of multiplication factors that can be realized before carrier collapse occurs. In contrast to the traditional approach, our source uses a DRO that is frequency-locked to a high-Q cavity without the use of any modulation of the carrier signal. This avoids the necessity of removing modulation sidebands from the carrier. The long-term frequency stability of systems using a dc frequency lock is susceptible to a number of systematic effects. We avoid most of these problems by locking the

entire system to a signal derived from a low frequency quartz crystal controlled oscillator. This controls the phase noise close to the carrier without significantly degrading the wideband phase noise. The entire assembly is enclosed in a magnetic shield to reduce power line generated modulation sidebands. The combination of these techniques leads to a source that in principle can be multiplied to approximately 250 THz before carrier collapse occurs [1, 2].

Frequency Discriminator

Figure 1 shows a block diagram of the signal path around the frequency discriminator. The X-band cavity has a unloaded Q-factor of approximately 55 000 and operates on the TE 023 mode. The inside dimensions are approximately 7.48 cm diameter and 7.93 cm long. The cavity is made from Invar plated with copper and then gold. The copper is polished to yield a near optical finish before it is plated with gold. The purpose of the gold is to reduce changes in surface properties with time. Each end cap has a $\lambda/4$ choke section and is threaded with 2.36 threads /mm. The cavity frequency can be adjusted about 20 MHz without significant change in Q-factor. The signal is coupled in or out of the cavity using a small iris and a stub to adjust the coupling coefficient. To obtain an unloaded Q-factor above 50 000 we found it necessary to use mica windows over the irises to suppress some of the spurious modes. The temperature coefficient of the discriminator system is approximately 20 kHz/K. The transmitted signal through the resonator is phase shifted by an amount

$$d\phi = 2\,Q\,\frac{\nu_0 - \nu_R}{\nu_0} \qquad (1)$$

which depends on the difference between the frequency of the input oscillator, ν_0, and the resonance frequency of the cavity, ν_R. For small frequencies differences the reflected signal is phase shifted by a factor of 2 more than the transmitted signal. The two phase shifted error signals are combined in the double-balanced mixer used as a phase detector. The phase shifter is used to maximize the slope of the error signal versus the frequency difference and to adjust the symmetry of the error curve.

Figure 2 shows a typical error signal versus frequency difference between the input oscillator and the resonance frequency of the cavity. The width of the linear central portion of the error curve is limited by the bandwidth of

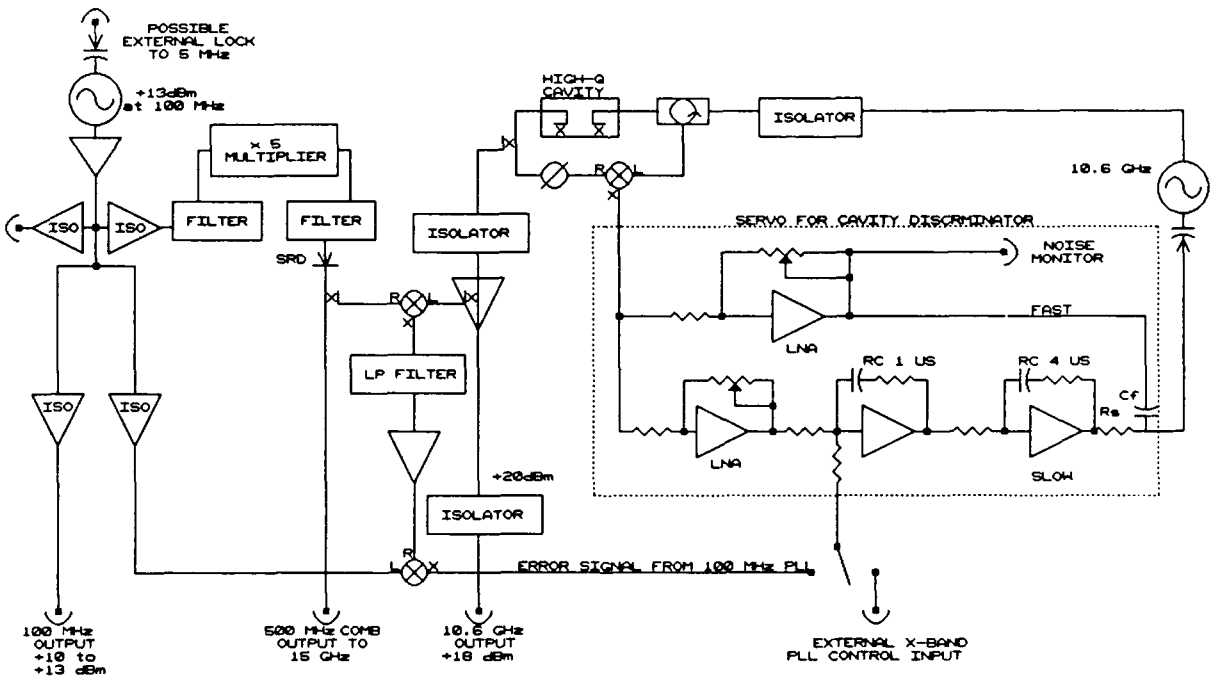


Figure 1 Block diagram of the cavity discriminator and the key elements of the servo system. The dotted portion contains the fast servo electronics to lock the DRO to the cavity discriminator. The phase noise near the carrier can be reduced by phase locking the 10.6 GHz source to a harmonic of a 100 MHz oscillator.

the cavity to approximately ± 40 kHz. The phase delay in the error signal is zero at line center, reaches approximately 90° at 100 kHz and eventually reaches approximately 120° at very high frequency differences.

The slope of the output voltage for such a frequency discriminator can be estimated from

$$d\phi = 6\, Q_L \frac{\nu_0 - \nu_R}{\nu_0}, \qquad (2)$$

$$V_{mixer} = K_d\, d\phi, \qquad (3)$$

where the extra factor of 3 in equation 2 comes from using both the transmitted and reflected signal from the cavity, Q_L is the loaded Q-factor, and K_d is the phase to voltage conversion factor for the mixer at 0 output voltage. For a loaded Q-factor of 20 000, and a mixer sensitivity K_d = 0.5 V/rad, the expected mixer output is 6×10^{-6} V/Hz frequency difference. The measured value is 5×10^{-6} V/Hz.

The noise floor of the discriminator (in terms of spectral density of fractional frequency fluctuations, $S_y(f)$) can now be estimated by dividing the expected noise in the double-balanced mixer by ν_0^2 and the square of the discriminator slope in volts^2/Hz^2. For our mixer

$$V_n(f)_{mixer} = 10^{-14}/f + 10^{-17} \qquad [V^2/Hz],$$

which yields

$$S_y(f) = \frac{V_n}{\nu_0^2 (3.6 \times 10^{-11})} = 2.8\times10^{-24}/f + 2.8\times10^{-27}. \qquad (4)$$

The equivalent phase noise is then given by

$$S_\phi(f) = \frac{\nu_0^2}{f^2} S_y(f) = \frac{2.8 \times 10^{-4}}{f^3} + 2.8 \times \frac{10^{-7}}{f^2} \quad f < 40\ kHz. \qquad (5)$$

SERVO SYSTEM

The dotted box in Fig. 1 shows the functional block diagram of the servo system that is used to provide very fast control of the frequency of the 10.6 GHz oscillator. If a simple proportional control loop is used, the bandwidth is limited to approximately 100 kHz by the phase shift of the error signal. This phase shift reaches more than 90^0 for error signals occurring at rates beyond approximately 100 kHz. The upper servo path labeled "fast" drives resistor R_s through capacitor C_f. This provides a phase advance of about 90^0 for correction rates slower than about 7 MHz which is used compensate for the phase retardation by the frequency discriminator circuit. At correction rates of approximately 1 MHz the proportional part of the gain (lower amplifiers) dominate the loop gain. At correction rates of approximately 160 kHz, the first integrator begins to increase the loop gain by 6 dB/octave, and at frequency below approximately 40 kHz the gain is increased by another 6 dB/octave by the second integrator. The relative gains of the stages are adjusted to minimize the noise detected at the noise monitor port in the upper servo path. If the gain is set too high, there are regions in the closed loop noise spectrum that are significantly increased over the open-loop noise.

The phase noise of the X-band source in the region below about 1 kHz can be controlled by a quartz crystal controlled oscillator. A small portion of the 10.6 GHz signal from the output amplifier is beat against the 21st harmonic of 500 MHz derived from a low noise 100 MHz crystal controlled oscillator. The beat frequency at 100 MHz is amplified and phase compared in another double balanced mixer. The dc error signal is then injected into the first integrator of the fast servo amplifier controlling the frequency of the 10.6 GHz oscillator. This is equivalent to moving up and down on the frequency discriminator curve shown in Fig. 2. The dc error signal is limited to approximately ∓0.1 V or ±20 kHz so that it does not greatly perturb the response of the fast loop. The long term frequency of the discriminator is controlled by controlling the temperature of the cavity. Even lower phase noise at Fourier frequencies below about 300 Hz can be obtained by locking the 100 MHz oscillator to the multiplied signal derived from a low noise 5 MHz oscillator. Improvements of 30 dB in phase noise for Fourier frequencies below about 30 Hz appear possible [8].

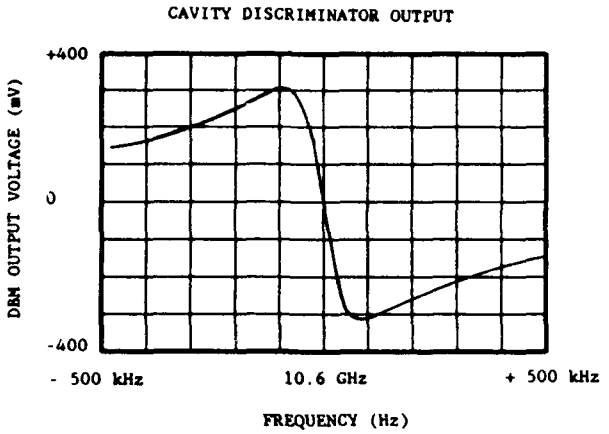


Figure 2 Typical error signal from the double-balanced mixer in the cavity discriminator of Fig. 1.

RESULTS

Figure 3 shows the comparison between the open and closed loop phase noise of the signal transmitted through the cavity. A post-amplifier with a gain of about 14 dB has been used to boost the output power to +18 dBm. We have measured the noise on 4 different systems which use this approach. The improvements in phase noise are generally from 50 to 70 dB for Fourier frequencies of 1 kHz and depend on the open-loop noise in the source. The noise floor in the region from 1 Hz to 1 kHz is generally about 6 to 10 dB above that calculated in equation 5 above. We have, however, measured one system that is within about 2 dB of the calculated noise floor in this region. We suspect that the mixer performance and possibly amplitude noise in the oscillator play a role in determining the noise floor. In the region from about 10 kHz to 100 kHz, the noise floor is determined by the loop gain and the open-loop phase noise of the oscillator. With enough loop gain the noise floor should be approximately -165 dB(rad^2/Hz) at a Fourier frequency offset of 100 kHz. Loop gain in this region can be increased only by improving the match between the phase shifted signal and the phase compensation provided by the differentiation stage.

Table 1 gives the phase noise of one source determined using a frequency discriminator system similar to that shown in Fig. 1. The raw data from this measurement is shown in Figs. 4-6. Several low-frequency spurs, not harmonically related to the power lines, are noticeable in Fig. 6. These were determined to be due to mechanical resonances in the cable connecting the source to the cavity discriminator used to measure the phase noise. All the lines in the source are glued to the support structure to

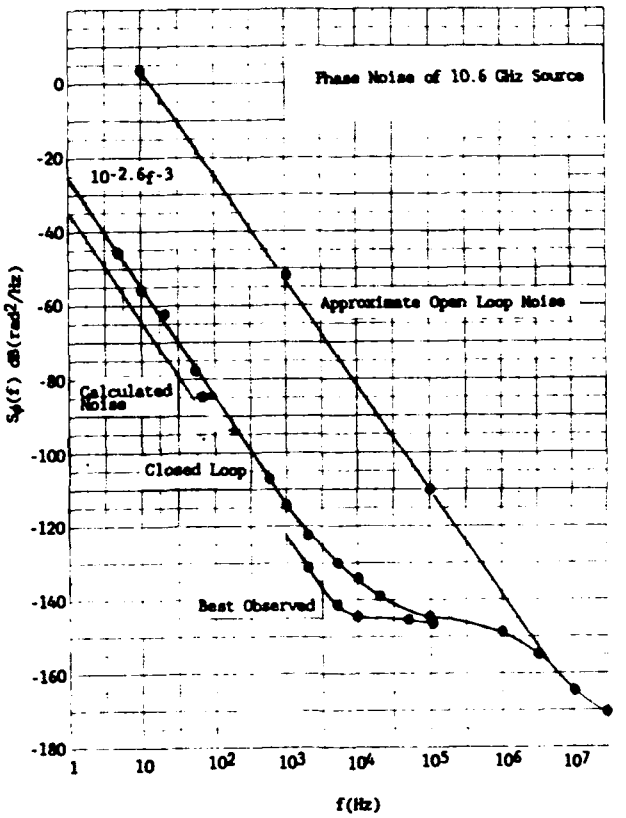


Figure 3 Comparison between the open and closed loop phase noise performance of a 10.6 GHz DRO locked to a cavity discriminator. The system is not locked to a crystal oscillator. Also indicated is the best performance that has been measured and the performance expected from equation 5.

minimize these effects. Also given is the phase noise at high Fourier frequencies determined by using the two-oscillator method [9,10]. The amplitude noise of the output signal is also indicated. For Fourier frequencies above approximately 10 kHz, the amplitude noise is similar to the phase noise. For lower Fourier frequencies the amplitude noise is much less than the phase noise. The lowest three lines in the amplitude noise column give the amplitude of the first three power line modulation sidebands. The amplitude noise measurement system consisted of a diode detector, battery powered low-noise amplifier and spectrum analyzer. The amplitude measurement system can be calibrated using the scheme described in [10]. (The recommended notation for specifying amplitude noise has recently been changed and is described in NIST Technical Note 1337 [11].) The simplicity of the setup made it possible to reduce the spurs associated with the power line below that possible in the phase noise measurement system.

Because of the difficulty in eliminating 60 Hz pickup, we believe that the actual phase noise at 60 Hz and harmonics is much closer to the AM values than that measured by the cavity discriminator method. The sidebands associated with the power line are extremely low because the entire system is enclosed in a magnetically shielded box of thickness 0.32 cm.

CALCULATION OF THE COLLAPSE FREQUENCY

The limit to which a source can be multiplied before the power density in the carrier drops below that in the noise pedestal (carrier collapse) can be determined from Eqs. 9-22 of [1]. The collapse frequency is roughly given by N × 10.6 GHz where

$$\Phi_p = N^2 \int_p S_\phi(f)df \approx 10, \qquad (6)$$

Φ_p is the mean squared phase modulation due to the wide-band noise pedestal. This is calculated from the closed loop phase noise of 10.6 GHz source given in Fig. 3 and Table 1 assuming a bandpass filter with a width of ± 300 MHz. This is easily achieved using passive passband filters. The collapse frequency is calculated to be approximately 250 THz. At 250 Thz the carrier signal would be about 2.8 kHz wide. At lower frequencies, for example, 30 THz, the linewidth could be reduced to approximately 10 Hz if the phase noise inside of 300 Hz was controlled by a low noise 5 MHz quartz controlled oscillator. To reach a collapse frequency of 250 THz all the discrete modulation sidebands outside the carrier linewidth of ±1400 Hz must be more than 90 dB below the carrier. The data of Figs. 2-6 and Table 1 show that we have accomplished this. At lower carrier frequencies the primary question is to what extent the vibrational sidebands observed in the few Hz range are due to the measurement system and not the X-band source since they might preclude achieving a linewidth of a few hertz.

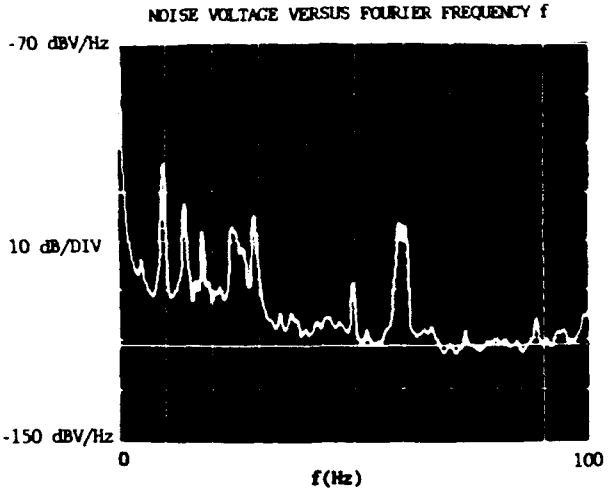


Figure 4 Raw data from the output of the cavity discriminator used to measure the phase noise of a 10.6 GHz source for Fourier frequencies from 6 to 100 Hz. The calibration is $S_\phi(f) = 10 \log (V_n^2/Hz) - 20 \log f - 86$ dB in dB(rad^2/Hz).

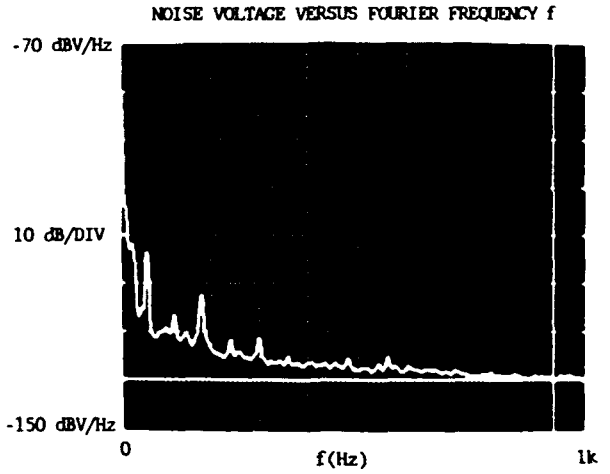


Figure 5 Raw data for the output of cavity discriminator used to measure the phase noise of a 10.6 GHz source for Fourier frequencies from 100 Hz to 1 kHz. The calibration is $S_\phi(f) = 10 \log (V_n^2/Hz) - 20 \log f - 86$ dB in dB(rad^2/Hz).

CONCLUSION

We have described an X-band source that is specifically designed for high order frequency multiplication and precision spectroscopy. The wideband phase noise is controlled by frequency- locking a DRO source to a high-Q cavity with a dc loop. This avoids the need for modulation on the source signal that might interfere with high order multiplication. The phase noise close to the carrier can be controlled by phase-locking the 10.6 GHz signal to a harmonic of a low-noise quartz crystal controlled oscillator. Modulation sideband due to the power line and harmonics are suppressed far below the random noise by enclosing the entire source in a magnetic shield. The phase noise of the completed source is the lowest that has been reported for a free-running, room temperature, X-band source. In principle the 10.6 GHz signal could be multiplied to approximately 250 THz before carrier collapse would occur. At 250 THz the free-running linewidth would be approximately 2.8 kHz. At a frequency of 30 THz we

would expect a linewidth of order 10 Hz if the phase noise near the carrier was controlled by a harmonic of a low-noise 5 MHz oscillator.

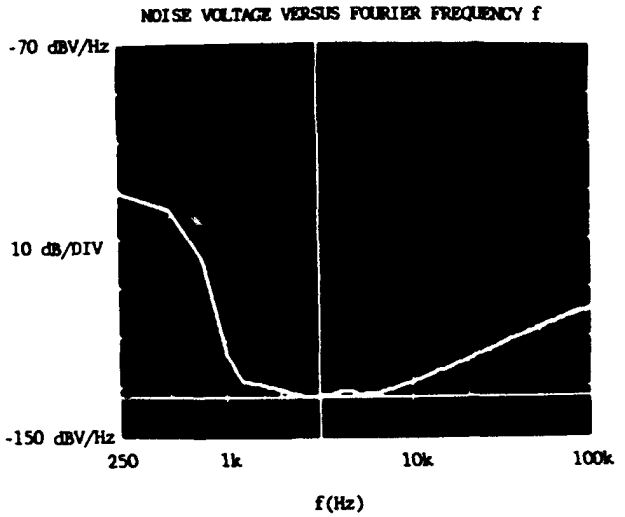


Figure 6 Raw data for the output of cavity discriminator used to measure the phase noise of a 10.6 GHz source for Fourier frequencies from 2 to 50 kHz. The calibration is $S_\phi(f) = 10 \log (V_n^2/Hz) - 20 \log f - 86$ dB in $dB(rad^2/Hz)$.

ACKNOWLEDGEMENTS

The authors are grateful to M. Vanek for help in the early stages of this project and to A. DeMarchi for many fruitful discussions.

REFERENCES

[1] F.L. Walls and A. DeMarchi, "RF Spectrum of a Signal after Frequency Multiplication Measurement and Comparison with a Simple Calculation," IEEE Trans. Instrum. Meas. IM-24, pp. 210-217 (1975).

[2] F.L. Walls and S.P. Beaton, "High-Order Harmonic Mixing with GaAs Schottky Diodes," these proceedings.

[3] K.M. Evenson, F.R. Petersen, and J.S. Wells, Laser Frequency Measurements: A Review, Limitations, Extension to 197 THz (1.5 μm) in Laser Spectroscopy III, (J.L. Hall and J.L. Carlsten, Eds. Springer Verlag, 1977) pp. 56-68.

[4] R. V. Pound, Electronic Stabilization of Microwave Oscillators, Rev. Sci. Instrum. 17, 490-505 (1946).

[5] F. W. Walls, "Errors in determining the center of a resonance line using Sinusoidal Frequency (Phase Modulation)," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, UFFC-34, 592-597 (1987).

[6] S. R. Stein and J. R. Turneure, "The Development of the Suyperconducting Cavity Stabilized Oscillator," Proc. of the 27th Annual Frequency Control Symposium, 414-418 (1973).

[7] J.J. Jimenez and A. Septier, "S- and X-Band Superconducting Cavity Stabilized Oscillators," Proc. of the 27th Annual Frequency Control Symposium, 406-413 (1973).

[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," Proc. of the 43rd Annual Symposium on Frequency Control, 80-84 (1989).

[9] 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 the 42nd Ann. SFC, 432-441 (1988.)

[10] F.L. Walls, C.M. Felton, A.J.D. Clements, and T.D. Martin, "Accuracy Model ofr Phase Noise Measurements," Proc. of the 21st Annual Precision Time and Time Interval (PTTI) Meeting, 295-310 (1989).

[11] "Characterization of Clocks and Oscillators," D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls, Eds. Nat. Inst. Stand. Tech (U.S) Tech Note 1337 1990.

[12] J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," Proc. of the 43rd Symposium on Frequency Control, 107-114 (1989). See also J. Dick, J. Saunders, and T. Tucker, "Ultra-Low Noise Microwave Phase Stabilizer Using a Ring Cavity," these Proceedings.

Table I

Fourier Frequency f	$S_\phi(f)$ of 10.6001 Fig. 1 dB rel to 1 rad^2/Hz	$S_\phi(f)$ of 10.6001 vs Ref Fig. 2 dB rel to 1 rad^2/Hz	$S_a(f)$ of 10.6001 dB V/Hz	Best Estimates $S_\phi(f)$ for 10.6001 dB rel to 1 rad^2/Hz
1 GHz		-170		-170
100 MHz		-170		-170
10 MHz		-165		-165
1 MHz		-149		-149
500 kHz		-147		-147
200 kHz		-145		-145
100 kHz		-143	-138	-143
50 kHz	-136	-143	-136	-140
20 kHz	-134	-138	-135	-136
10 kHz	-132	-131	-135	-132
5 kHz	-130		-134	-130
2 kHz	-122		-134	-122
1 kHz	-114		-133	-114
500 Hz	-107	-107	-132	-107
200 Hz	-94		-130	-94
100 Hz	-84	-84	-128	-84
50 Hz	-78			-78
20 Hz	-62		-121	-62
12 Hz	-56		-119	-56
6 Hz	-47			-47
	dB below carrier	Power Line Harmonics	dB below carrier	
60 Hz	-52		-136	
120 Hz	-78		-141	
240 Hz	-72		-138	

The accuracy of column 2 is ± 3 dB from 20 Hz to 20 kHz and ± 5 dB elsewhere. Column 3 was measured by the two oscillator method. The accuracy of the measurements is ± 3 dB.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

LOW PHASE NOISE DIELECTRIC RESONATOR OSCILLATOR

Rommel Jones and Vaughn Estrick
Hughes Aircraft Company
1901 W. Malvern Ave
Fullerton, CA 92503

Abstract

An extremely low phase noise 9.0 GHz Dielectric Resonator Oscillator (DRO) has been developed for use as a clean fundamental microwave source. In order to achieve minimum phase noise, a bipolar junction transistor (BJT) was selected as the active device for the oscillator. Given the limited number of unconditionally stable, high frequency BJT's, the conventional feedback oscillator was undesirable. Instead, a negative resistance topology was selected. The dielectric resonator material was $(ZrSn)TiO_4$ which has a dielectric constant of 38 and an unloaded Q of 7000 at 7 GHz. A unique bipolar voltage tuning design provided +/-500 kHz of linear voltage tuning. The phase noise of this circuit measured -114 dBc/Hz at 10 kHz from the carrier with a wideband floor for offsets greater than 1 MHz of -150 dBc/Hz [1].

Dielectric Resonator Background

The dielectric puck acts as a microwave resonator due to the boundary established at the interface between the air and the high dielectric device. This boundary confines approximately 80-90% of the energy within the device. The fact that as much as 20% of the energy radiates from the dielectric puck makes it very easy to couple RF power to this device providing the resonance of the oscillator circuit. The resonant frequency of the puck is primarily determined by its dimensions. Therefore, the frequency fluctuates very little producing a more stable frequency source with a corresponding improvement in phase noise. As with any resonant cavity, an infinite number of modes can exist within the puck and care must be taken that the appropriate one is selected without exciting other modes. The most commonly used mode for the dielectric resonator is the $TE_{01\delta}$. This is considered the fundamental mode for these devices since it exhibits the lowest resonant frequency. The primary method of establishing the mode of operation is by proper setting of the diameter (D) to thickness (t) ratio of the resonator. A computer aided design (CAD) program, Resomics, produced by Murata Erie, was used to aid in the design. This particular CAD program varies such parameters as, the desired frequency of operation, the quality factor, and the thermal coefficient of expansion. The resulting puck dimensions are a function of the cavity in which it will be placed and materials such as the substrate, epoxy, and spacer with which it must interface. Through this design process, the excitation of spurious modes were minimized and the Q of the device was maximized. Pucks are presently available which operate from 900 MHz to 30 GHz. The first is about the size of a half dollar while the latter is approximately 85 mils in diameter.

Quality Factor

An important parameter of the dielectric resonator which influences the oscillator's phase noise is the unloaded quality factor (Q_0). This parameter is a ratio of the stored energy within the puck to the amount of energy which is dissipated by the resonator and serves as a figure of merit for the device. Unloaded Q's of 20,000 are now possible to 10 GHz. The quantity which is of ultimate importance for an oscillator is, however; the loaded quality factor (Q_L) since it will determine the oscillator's phase noise response. As shown by Leeson's equation [2], a higher Q_L implies a lower value for the FM noise component. This parameter is a ratio of the energy stored within the puck per cycle to that delivered to the load per cycle.

The Q_L of the circuit can be degraded by the materials used to fabricate the oscillator. Depending upon the mechanical design of the enclosing cavity or on the selection of substrate material, the resultant value of the loaded Q can be reduced regardless of the resonator's unloaded Q. It is thus very important for the designer to appreciate the impact each circuit component has on the phase noise of the circuit. The effect of the ceramic substrate on Q_L can be minimized by assuring that the thickness of the substrate is less than one-quarter of the puck height and that the dielectric constant of the low loss substrate is less than half that of the resonator. Possible materials include alumina, beryllia, quartz, or fosterite with the first being the most commonly utilized.

CH2818-3/90/0000-549 $1.00 © 1990 IEEE

The degradation which the enclosing cavity has upon the loaded quality factor can be reduced by providing sufficient distance between the puck and the enclosure walls. Typically, all walls should be at least one resonator diameter away from the puck and the lid should be at least twice the puck thickness above it. There have been however, compromises to these cavity "rules" to minimize the circuit size while still providing good performance [3]. The more conservative approach is recommended so that the cavity does not control the DRO performance. An additional consideration on the cavity design is to use a good RF seal so that the loaded Q is not degraded by leakage.

Another factor which can significantly affect the loaded Q of the final circuit is the amount of coupling of the dielectric resonator to the active circuit. A detailed procedure to measure the loaded and unloaded quality factors of the resonator as a function of coupling to a microstrip transmission line has been presented by Khanna [4]. This test procedure allows the designer to select the optimum height of low dielectric constant material used as the spacer between the resonator and the substrate surface. This value is critical in maximizing the Q_L of the overall circuit since it extends the height of the puck above the ground layer to a value greater than that of the substrate thickness. If the distance from the ground plane is not sufficient, the field lines are constrained by electromagnetic boundary conditions to terminate in a manner which provides less than optimum performance. Insufficient distance from the ground plane can also cause moding (generation of spurious signals). A low dielectric constant material such as glass, fused quartz, or fosterite can be used for the spacer so that it does not become an electrical extension of the dielectric resonator.

The dielectric resonator material that was selected for our design was $(ZrSn)TiO_4$ which has a dielectric constant of 38 and an unloaded Q of 7000 at 7 GHz. From Khanna's test procedure, it was determined that a 30 mil fosterite spacer resulted in a coupling coefficient to a microstrip test line of 44 which provided a loaded Q of 70. In contrast, a 60 mil spacer gives a loaded Q of nearly 1500 with a coupling factor of 1.5. The optimum position of the resonator to the microstrip line to which it couples was located at about 1/4 of an electrical wavelength away and at a height of 60 mils above. The calculation of the parallel equivalent or tank circuit of the DR is then determined from the relationship between the coupling and Q_0 of the device.

Temperature Stability

The resonator puck is typically selected with an initial thermal coefficient (TC) of approximately 0 ppm/^oC. This TC is obtained by varying the chemical composition of the resonator material with a resultant increase in dielectric constant for positive values and a decrease for negative TC's. The composite oscillator TC is also effected by the substrate, enclosure, and spacer materials. Ideally, all of these other effects would produce a frequency variation which is as close to zero as practicable. Since it is usually not possible to achieve this, the next alternative is to select materials with either complementary characteristics or ones that have minimal effects. For example, the spacer beneath the resonator must be selected so that it does not become a significant contributor to the overall TC. Both fosterite and quartz materials have been used in our circuit with good results.

The temperature characteristics of the adhesive material that bonds the puck to the spacer and the spacer to the substrate must also be evaluated to assure that it will not compromise circuit performance. It should also have a low dielectric constant so that it does not become an electrical extension of the dielectric resonator. Acrylics have a low dielectric but are brittle and do not perform well over temperature. Low dielectric epoxies function quite well as bonding materials and were successfully used in this design. Care must be taken, however; in using a minimal amount of epoxy since the material does interact and affect the electromagnetic boundary conditions.

A secondary and equally important contributor to the oscillator TC is variations of transistor parameters as a function of temperature. Virtually all parameters of the active device vary as a function of temperature. Since many of the variations are unpredictable, the designer's only recourse is to build the complete unit with the dielectric resonator and measure overall circuit TC. A second dielectric resonator can then be selected having a temperature coefficient which is the negative of that of the combined circuit. When used to replace the initial puck, a composite TC as near to zero as practicable will be measured. Care must be taken with this substitution since the dielectric constant of the puck will change and alter the resonant frequency of the circuit slightly.

Oscillator Design

The next critical facet in the development is the active device and the topology used to establish oscillations. Gallium arsenide field effect transistors (GaAs FETs) are often used for an efficient oscillator at higher microwave frequencies (above 4 GHz). These devices are capable of operation at 10 GHz and beyond producing an output power in excess of +10 dBm. In the past, the silicon bipolar junction transistor (BJT) was only capable of providing sufficient power at frequencies below about 4 GHz but recent developments in processing methods have allowed device capability into X-band. BJT's have

Table 1
Summary of DRO Performance Characteristics
(Reference 5)

Device	HBT Rockwell	FET NE72084	Bipolar NE68135
Oscillation frequency (MHz)	3990.4	4006.8	3968.0
Output power (dBm)	10.2	16.0	14.9
Power of second harmonic frequency (dBm)	-6.0	-1.0	-8.0
Efficiency (%)	29.4	44.2	12.9
Frequency stability (ppm/^{o}C) -30 to +70^{o}C	2.802	1.361	5.29
Tuning range of 1 dBm power var (mech) (%)	2.8	2.1	3.0
FM noise (dBc/Hz) at 1 kHz off-carrier frequency	-73.0	-61.0	-78.0
at 10 kHz off-carrier frequency	-95.0	-95.0	-108.0

significant advantages over their GaAs FET counterparts since they have approximately 10 dB better flicker and shot noise characteristics. The disadvantage of BJT's at X-band is their lower output power. Another possible active device useful for DRO applications is the heterojunction bipolar transistor (HBT). The HBT appears to have the high frequency gain of FET's with the low noise characteristics of BJT's.

Capabilities of an oscillator developed from an NPN grounded emitter HBT with 1.2 to 1.5 micron emitter width were demonstrated in 1986 and are shown in Table 1 [5]. The performance of a GaAs FET and BJT oscillator are also compared in the table. It is apparent from this data that the lowest source of flicker noise is still the BJT and should be selected in designs seeking minimum noise levels until lower noise HBT's become available.

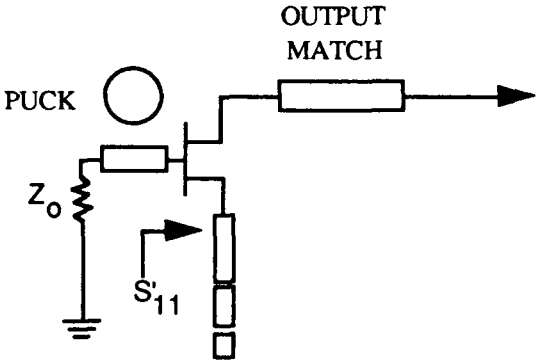


Figure 1. A series feedback approach to the negative resistance oscillator topology in which matching at the source is used to create the desired instability[8].

The available low phase noise oscillator topologies using microwave active devices are either a feedback or negative resistance configuration. The first uses an active device which is unconditionally stable, or will not oscillate when either the input or output is presented with any value of passive load. The GaAs FET normally meets this criterion for X-band applications. This oscillator circuit can be implemented easily with a dielectric resonator completing the feedback path. The puck can be coupled to the microstrip lines of both the gate and drain of a FET thus allowing the output power of the device to be coupled into its input where it is amplified to generate or sustain oscillations.

The selected transistor can not always be guaranteed of unconditional stability as is frequently the case with BJT's operating at X-band. Therefore, the alternative architecture of a negative resistance oscillator can be utilized. This form of oscillator was demonstrated by Lakshminarayana [6] to be a very effective means of realizing a low phase noise oscillator with a BJT. The design work of this author was extended to determine further areas of phase noise improvement. The potential instability of an active device provides a convenient method of establishing and sustaining oscillations. For this configuration, the region of instability is first carefully characterized and defined on the Smith Chart. A load value from within this region is then extracted from the Smith Chart and used to yield an input reflection coefficient (S'_{11}) which is greater than one. A value of S'_{11} greater than one implies that oscillations will be initiated. Several methods that can be used to make the device unstable are series and shunt feedback of lumped elements or microstrip open and short circuit stubs at the device ports. With the attainment of transistor reflection gain, the puck can then be positioned to resonate the input port. This second condition is satisfied by the equation[7]:

$$(S'_{11})(\Gamma_r)=1.0$$

where Γ_r is the refection coefficient, seen by the input port, caused by the dielectric resonator. The important relationship here is that Γ_r is the inverse of S'_{11}. The most critical aspect of this relationship is that the phase quantities of the two parameters cancel. This phase relationship is controlled by the position of the puck relative to the microstrip transistor, typically 1/2 the wavelength of the operating frequency. At resonance, the puck will appear as a resistive load, determined by the amount of coupling, which creates the reflection coefficient. Figure 1 illustrates an approach to the negative resistance oscillator configuration [8].

Several BJT's offered by California Eastern Labs were reviewed for this application including the NE64587 and the NE243287. Both devices were found to be potentially unstable and therefore candidates for the negative resistance approach. Microstrip matching networks which would induce oscillations were calculated for both ports and then optimized using Touchstone, a trademark of EEsof Inc. The open stub matching required to attain a maximum value for the input reflection coefficient (S_{11}') was calculated. Based on an optimization of this matching scheme, the NE64587 was found to be capable of achieving more gain and power than the other device and was selected for this application. It was found that an open stub match at the emitter with the tank circuit of the dielectric resonator coupled to the base as illustrated in Figure 2 would sustain oscillations. A picture of the completed unit is shown in Figure 3.

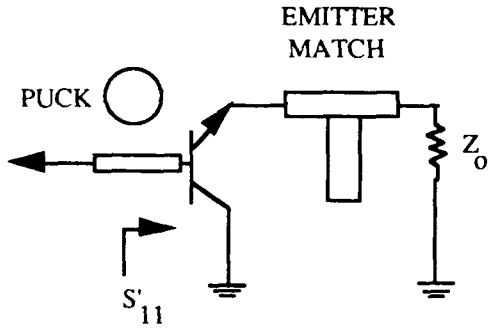


Figure 2. This negative resistance circuit topology was used with the NE64587 to yield an extremely low phase noise oscillator. The matching at the emitter created the instability while the resonator at the base stabilized the design.

Tuning

One method of achieving the desired frequency of the DRO is through lapping of the puck. Very fine, wet sandpaper was used to polish the resonator and reduce its thickness. This increases the frequency of the circuit but care must be taken that the puck's dimensions are not changed by more than about 8% or mode problems will result. Once the puck is polished, the DRO can be mechanically tuned to provide coarse adjustment of its fundamental by 1-5%. This amount will depend upon the frequency of operation and the amount of stability that can be sacrificed. A reduction of both the unloaded Q and the temperature stability of the resonator can result from this tuning circuit. For this design, tuning was achieved by varying the distance of the ground plane directly above the resonator with a tuning screw. Moving the plane toward the resonator will vary the boundary conditions resulting in an increase of the fundamental. The tuning screw was used as the variable ground plane and contained a metal shaft set upon extremely fine threads.

Voltage tuning is normally achieved through the use of varactor diodes which are coupled to the dielectric resonator via a microstrip line. The capacitance of the varactor then becomes a parallel extension of the DR equivalent circuit. The capacitance of the diode varies as a function of bias voltage and is used to vary the fundamental of the resonator by a very small percentage, typically 0.1%. A change in the diode capacitance thus induces a change in the resonator capacitance and its effect is a function of the coupling coefficient established. Increased tuning can be achieved by tight coupling at the expense of the loaded Q and stability of the circuit.

An alternate voltage tuning topology is outlined in a paper by Znojkiewicz [9] for use with bipolar devices. Linear tuning is achieved by varying the equivalent parameters of the transistor as shown in Figure 4. The emitter current is used to control the emitter junction resistance while the collector-base voltage is used to control the collector junction capacitance. The fundamental can then be lowered by raising the emitter current and increased by raising the collector to base voltage. Since this architecture offers a unique, simplified alternative to the varactor tuning scheme, it was selected for our design.

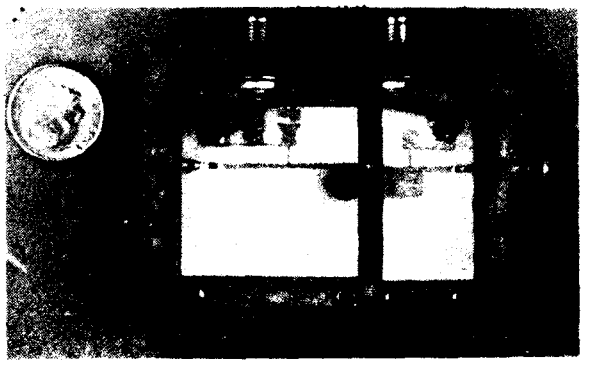


Figure 3. 9.0GHz DRO with extremely low phase noise.

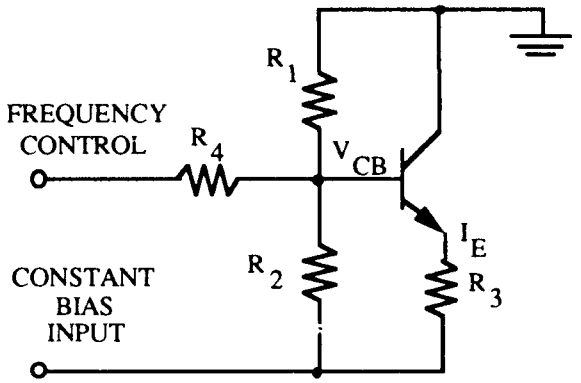


Figure 4. A unique biasing technique which provides linear frequency tuning of a bipolar oscillator by varying the equivalent parameters of the transistor[9].

Oscillator Test Data

The voltage tuning characteristics of this oscillator are shown in Figure 5 for an operating point of Vce=8 V and Ic=10 mA. Other results of the DRO test data are summarized in Table 2. The dielectric resonator selected for the first iteration of this design had a resonant frequency of 8.73 GHz and as shown in Table 2, set the oscillator's frequency. The dielectric resonator was later lapped in order to realize the resonant frequency of 9.0 GHz with equivalent phase noise performance. The output power of this iterated version was -20 dBm as opposed to that of -2.5 dBm measured at 8.73 GHz. The reason for this power difference was that the matching network had been designed for optimum power at the resonance of the initial puck and had not been altered for the increased frequency. Harmonic and spurious suppression matched well to those values documented in similar DRO circuits. The phase noise performance of

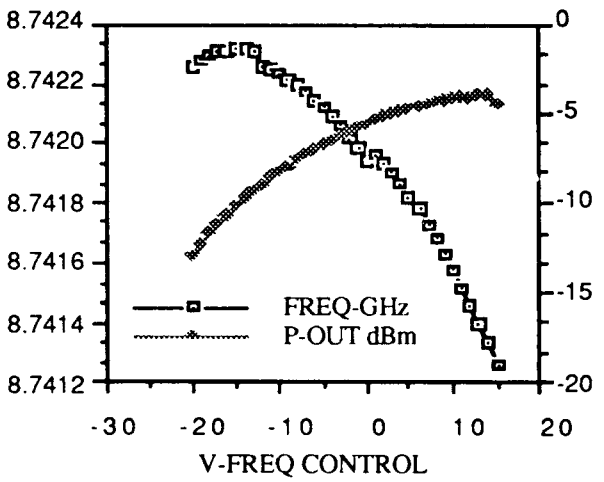


Figure 5. Measured voltage tuning characteristic of the BJT DRO.

the two oscillators were measured on the Hewlett Packard 3047A Phase Noise Measurement System and is shown in Figure 6. In comparison to this data is that from a low noise GaAs FET oscillator designed by Bianchini, et al[10]. As you can see, there is approximately a 20 dB advantage in close-in phase noise with the BJT topology. Bianchini et al, continued with a frequency discriminator method of noise degeneration which improved their close-in noise by about 20 dB but the initial circuit is a closer baseline to our topology and is used here for comparison. Since two identical oscillators are being measured against each other with identical phase noise data, there will be a summation of voltage fluctuations. For this reason, a 3 dB correction factor is added to this data to demonstrate the actual characteristic of each oscillator circuit. The resultant temperature stability was measured to be -9.8 ppm/°C over a range of +85°C to -55°C with a resonator having a coefficient of 0 ppm/°C. The 3 ppm/°C desired frequency stability could be achieved by the purchase of a dielectric resonator having a positive thermal coefficient which would negate that of the circuit.

Table 2
DRO MEASURED DATA

Fundamental Frequency	8.73 GHz
Electrical Tuning	+/-500 kHz
Output Power	- 2.5 dBm
Output Power vs Tuning	+/- 3.4 dBm
Harmonic Suppression	-23.4 dBc
Spurious Suppression	-73.8 dBc
Temperature Stability	+/-9.8 ppm/°C

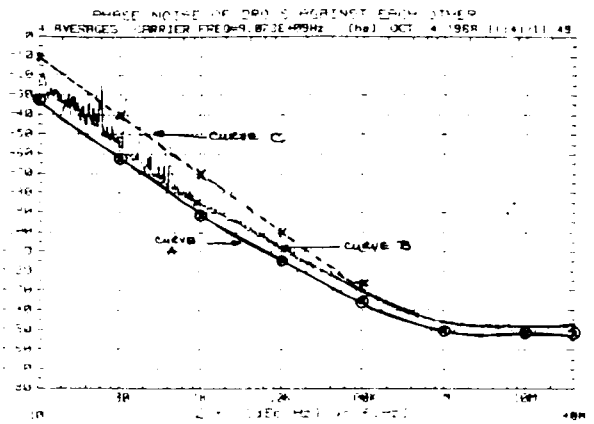


Figure 6. Curve A above provides the phase noise response of a single BJT DRO. This data is found from a 3dB correction to the measured data of curve B. The phase noise of a low noise GaAs FET DRO is shown in curve C for comparison.

Conclusions

The purpose of this paper has been twofold, to provide a tutorial on the fundamental aspects of DRO's as well as to demonstrate the usefulness of these devices in creating an extremely low noise source. The phase noise improvements gained with this topology are nearly 20 dB better than those of the more commonly used FET DRO's and provide a means of establishing a clean source without the need of multipliers and thus will greatly minimize hardware requirements. Recent developments in low noise HBT devices offer opportunities to enhance the phase noise characteristics of DRO's while providing higher output levels.

Acknowledgements

We greatly appreciate the support and direction of Dr. Mysore Lakshminarayana of Cal Poly, Pomona who helped to make this effort such a success. We also wish to thank Richard Docter, Manager RF Product Design Dept for his support and contributions to this design effort.

References

[1] Rommel Jones, "Development of a Dielectric Resonator,Oscillator Phase Locked to a Stable Reference". Accepted by the Faculty of California Polytechnic University, Pomona, 1988.

[2] D.B. Leeson, "A simple model of feedback oscillator noise spectrum," Proc. IEEE, Vol 54, pp. 329-330, 1966.

[3] Kevin J. Anderson and Anthony M. Pavio, "Reduced-cavity DROs exhibit excellent noise performance," Microwaves & RF, pp. 135-138 Oct. 1985.

[4] A.P.S. Khanna, "Q measurement of microstrip-coupled dielectric resonators," Microwaves & RF, pp. 81-86, Jan. 1984.

[5] Krishna K. Agarwal, "Applications of GaAs heterojunction bipolar transistors in microwave dielectric resonator oscillators," Microwave Journal, pp. 177-182, Nov. 1986.

[6] Mysore Lakshminarayana, "A low noise dielectric resonator stabilized x-Band bipolar oscillator," Proc. 7th Biennial University/ Government/Industry Microelectronics Symp. pp. 211-213, 1987.

[7] George Vendelin, Design of Amplifiers and Oscillator by S-Parameter Method, New York: John Wiley & Sons, 1982.

[8] Darko Kajfez and Pierre Guillon, Dieletric Resonators, Norwood, MA: Artech House, Inc., 1986.

[9] M.E. Znojkiewicz, "8 Ghz low noise bias tuned VCO," IEEE MTT-S Digest, pp. 489-491, 1984.

[10] M.J. Bianchini, Et al, "A single-resonator GaAs FET oscillator with noise degeneration," IEEE-MTT-S Digest, pp. 270-273, 1984.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

HIGH-ORDER HARMONIC MIXING WITH GaAs SCHOTTKY DIODES*

MARCO PREVEDELLI
Universita 'di Firenze
Pisa, Italy

and

F.L. WALLS AND STUART P. BEATON
Time and Frequency Division
National Institute of Standards and Technology
Boulder, CO 80303

ABSTRACT

In many areas of precision frequency metrology, it is useful to use a single nonlinear element to both multiply a source frequency and mix this harmonic with a higher frequency source. This process is commonly called harmonic mixing. When the frequency span is great, it is advantageous to use large harmonic numbers to reduce the complexity, the number of local oscillators, and ultimately the cost. Planar GaAs Schottky diodes are good candidates for high-order harmonic mixing because of their high cut-off frequency. Their robust construction should lead to excellent lifetime, reliability, and reproducibility. We have investigated the performance of one type of fast planar Schottky diode for harmonic mixing with harmonic numbers from 8 to 201. Our measurements indicate that, with optimum biasing and power, harmonic mixing with harmonic numbers up to approximately 80 can be achieved with good signal-to-noise ratios. Detectable signals have been observed for harmonic numbers of 201. In a test of both the nonlinear and the high-frequency performance, we were able to obtain a beat signal between the 34th harmonic of a 74.17 GHz klystron and a laser operating at 2522 GHz. The resulting signal-to-noise ratio was 48 dB (bandwidth of 3 kHz). This is more than sufficient to phase lock the laser to the multiplied reference signal. From the nonlinear characteristics of the diode and the signal-to-noise ratio obtained at lower frequencies, it looks as though we can multiply from 70 GHz to about 4.5 THz in one step.

INTRODUCTION

The absolute measurement of frequency and precision frequency synthesis up to approximately 10 THz often requires the use of harmonic mixers. Many types of diodes have been successfully used in this frequency range with the metal-insulator-metal (MIM) the most common [1-7]. Josephson junctions based on niobium have been used up to approximately 3.5 THz [8-10]. With the advent of the new high temperature superconductors it may eventually be possible to extend this upper limit by a factor of 10 or more. We report here the results for harmonic mixing with harmonic numbers from 8 to 201 using GaAs schottky diode constructed with planar geometry [11]. As many as 1000 separate diodes are fabricated on one die and attached to a 5 mm diameter post. A 25 μm diameter tungsten whisker of approximately 9 mm length and sharpened to point of approximately 100-1000 Å radius of curvature is used to select the particular diode, to couple in both the high frequency and the low frequency radiation into the diode and to couple the IF signal into a 50 Ω coaxial line. Alternately we have used coaxial tees and directional couplers at lower frequencies to inject the subharmonic into the diode, and to extract the IF frequency. Both schemes lead to a high signal-to-noise ratio (SNR) for harmonic mixing with harmonic numbers up to 80 with detectable signals up to harmonic numbers of 201. There is no evidence of significant roll-off with this approach up to 2.5 THz, which is consistent with the reported cut-off frequency of 11 Thz for this diode structure[11]. These diodes appear to have a long lifetime and excellent diode-to-diode reproducibility. Failures were traced to input power in excess of about 15 mW or severe shock. Their use has the potential for substantial simplification in the hardware and reduction in the number of local oscillators required for frequency synthesis in the 30 Ghz to 10 THz region.

EXPERIMENTAL TESTS

Figure 1 shows the block diagram of the setup used to compare the SNR for high-order harmonic mixing. The phase shifter is a simple line stretcher. The low-pass filter

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

is used to separate the high frequency and the subharmonic from the IF frequency. For these measurements the high-frequency signal was
fixed at 74.17 Ghz and the subharmonic signal derived from a frequency synthesizer. Fixing the high frequency eliminated uncertainties in the SNR with harmonic number associated with frequency roll-off. The low-phase noise 74.17 GHz signal was coupled into the diode by placing the WR 15 waveguide about 5 mm from the tungsten whisker. The subharmonic signal is coupled into the diode using a coaxial tee or directional coupler. The induced bias current is maximized for low subharmonic power by adjusting the phase shifter. The subharmonic power is then adjusted to create a bias current of 0.5 mA through a 1 kΩ resister. The beat frequency is coupled out

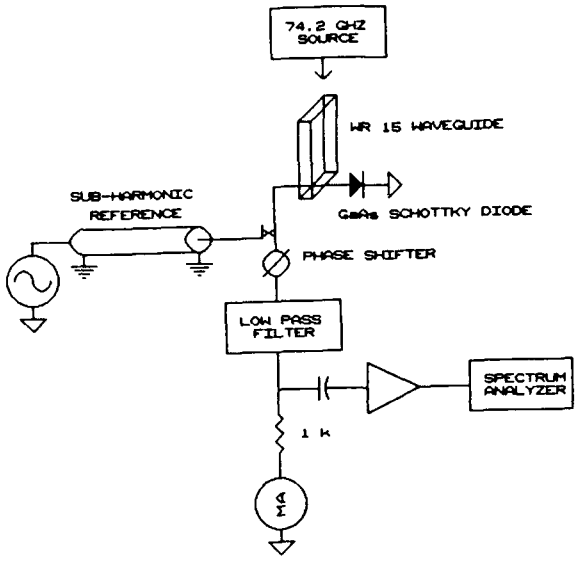


Figure 1. Schematic diagram of the test setup used to measure the high-order multiplication and mixing properties of planar GaAs Schottky diodes. The phase shifter was of the line stretcher type.

through the low pass filter and amplified in an amplifier with a noise figure of approximately 9 dB. The SNR of the beat frequency observed on a spectrum analyzer for harmonic numbers from 8 to 80 is shown in figure 2. The noise bandwidth is 3 kHz. These results were highly reproducible from day to day and from diode to diode. Figure 3 shows the spectrum analyzer trace on a linear scale at harmonic mixing of 201 with a noise bandwidth of 3 kHz.

It is notable that the difference between odd and even harmonic mixing is very small for harmonic numbers above 11. This is in considerable contrast to results with many other diodes [12]. The decrease in SNR with increasing harmonic number is also remarkably small compared to many other systems [12].

A diode from the same die as above has also been used to harmonically mix the 74.17

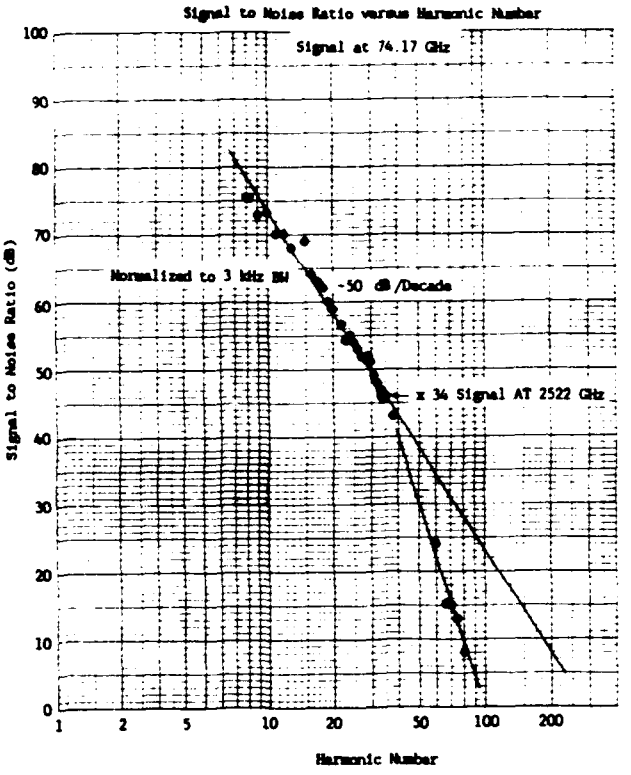


Figure 2. Measured SNR versus harmonic number for planar GaAs Schottky diodes. The upper reference frequency was held at 74.17 GHz for all except the point labeled "x 34 signal at 2522 GHz" where the upper reference frequency was 2522 GHz. Repeatability of the SNR measurements was typically +/-3 dB.

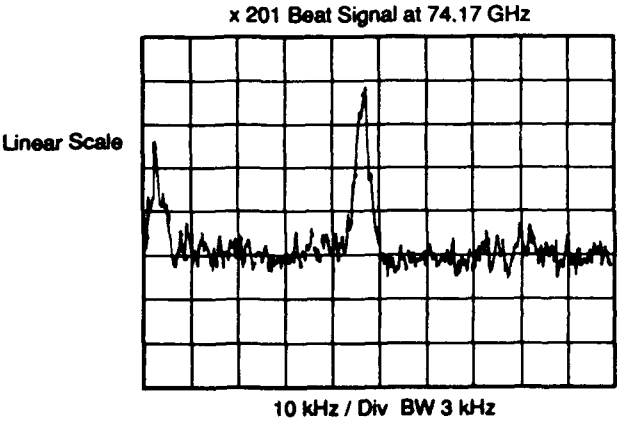


Figure 3. Spectrum analyzer plot for the beat frequency between the 74.17 GHz reference and the 201 st harmonic of the frequency synthesizer. The drive level was about 6 dB higher than for the data of figure 2.

GHz signal with the 118.8 μm (2522.7816 GHz) laser line of CH_3OH (methyl alcohol). The block diagram is shown in Figure 4. The 74.17 GHz klystron source was phase locked 30 MHz below the 7th harmonic of a very stable 10.6 GHz source[13]. The width of the phase-locked 74.17 GHz source was approximately 5 Hz. The 118.8 aim line was oscillating with approximately 1 mW of power. Figure 5 shows the spectrum analyzer trace for the beat frequency with a noise bandwidth of 10 kHz.

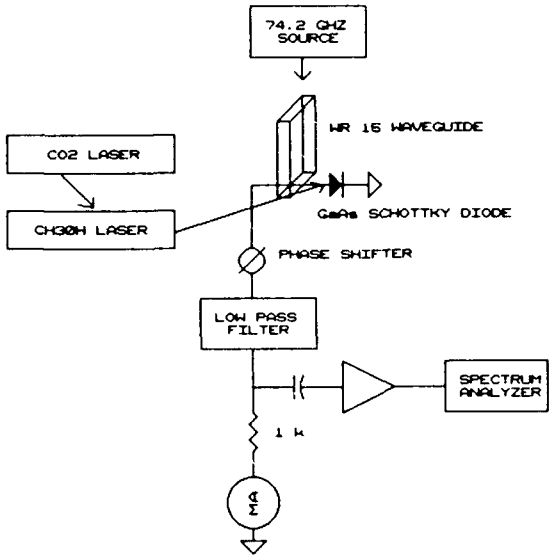


Figure 4. Schematic diagram of the test setup used to measure the SNR for the beat frequency between the 34 th harmonic of 74.17 and the 2.522 THz methyl alcohol line.

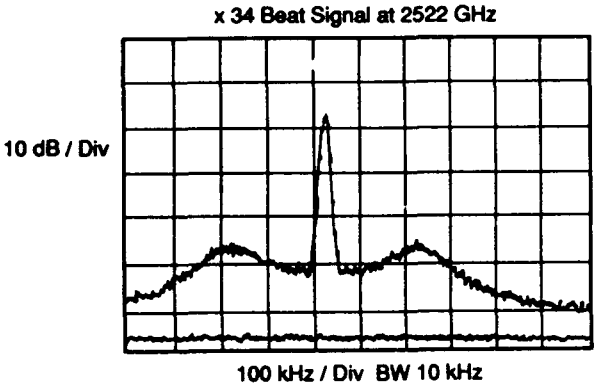


Figure 5. Spectrum analyzer plot for a 1 MHz sweep over the beat frequency between the 34th harmonic of the 74.17 GHz reference and the 2.522 THz line of methyl alcohol line. The noise bandwidth was 10 kHz.

Figure 6 shows the spectral analyzer for a noise bandwidth of 3 kHz. The increase in the apparent noise floor at about 200 kHz is due to the residual phase noise in the klystron source that was not removed by the phase-locked-loop. This increase can in principle be significantly reduced by increasing the bandwidth of the phase-locked-loop. The SNR for this signal is about 46 dB if one considers the noise in the wings which are not effected by the noise bump. This result is in close agreement with that measured at low frequencies and shown in figure 2. Evidently the roll-off effects in the diode are still small at 2.5 THz.

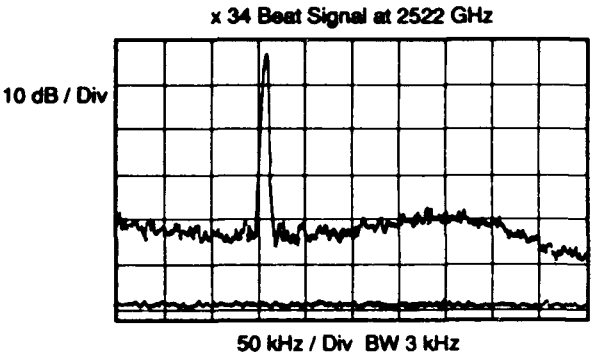


Figure 6. Spectrum analyzer Plot for a 500 kHz sweep over the beat frequency between the 34th harmonic of the 74.17 GHz reference and the 2.522 THz line of methyl alcohol. The noise bandwidth was 3 kHz.

CONCLUSION

We have shown that sample GaAs Schottky diode of the planar type are sufficiently nonlinear to yield very high SNR for harmonic mixing up to order 80. Detectable signals can even be obtained for harmonic number of 201 for stable sources where the noise bandwidth can be reduced to 100 Hz. Frequency roll-off effects in the diode were not detectible at 2.5 THz, which is consistent with the calculated value of 11 THz based on the mask geometries. The SNR is nearly identical for either odd of even order harmonic mixing for harmonic numbers above 11. The sample diodes tested were very stable in time and reproducible diode to diode on the same die.

ACKNOWLEDGEMENTS

The authors are very grateful to K.M. Evenson and D.A. Jennings for many fruitful discussions and help in aligning the CH_3OH laser.

REFERENCES

[1] T.Y. Chang and T.J. Bridges, "Laser Action at 452, 496, and 541 μm in Optically Pumped CH_3F," Opt. Commun. 1, 423 (1970).

[2] H.R. Fetterman, B.J. Clifton, P.E. Tannenwald, and C.D. Parker, "Submillimeter Detection and Mixing Using Schottky Diodes," Appl. Phys. Lett. 24, 70 (1974).

[3] C.O. Weiss and A. Godone, "Harmonic Mixing and Detection with Schottky diodes up to the 5 THz Range," IEEE J. Quantum Electron. QE-20, 97 (1984).

[4] K.M. Evenson, D.A. Jennings, and D.J.E. Knight, "Optical Frequency Measurements," IEEE J. Quantum Electron., 74, 168-179 (1986).

[5] D.J.E. Knight, G.J. Edwards, P.R. Pearce, and N.R. Cross, "Measurement of the Frequency of the 3.39 μm Methane-Stabilized Laser to ± 3 Parts in 10^{11}," IEEE Trans. Instrum. Meas. IM-29, 257 (1980).

[6] T.G. Blaney, D.J.E. Knight, and E.K.M. Lloyd, "Frequency Measurements of Some Optically-Pumped Laser Lines in CH_3OD," Opt. Commun. 25, 176 (1978).

[7] F.R. Petersen, K.M. Evenson, D.A. Jennings, J.S. Wells, K. Goto, and J. Jiménez, Correction to "Far Infrared Frequency Synthesis with Stabilized CO_2 Lasers, Accurate Measurements of the Water Vapor and Methyl Alcohol Laser Frequencies," IEEE J. Quantum Electron. QE-12, 86 (1976).

[8] D.G. McDonald, A.S. Risley, J.D. Cupp, and K.M. Evenson, "Harmonic Mixing of Microwave and Far-Infrared Laser Radiation Using a Josephson Junction," Appl. Phys. Lett. 18, 162-164 (1971).

[9] T.G. Blaney and D.J.E. Knight, "Direct 825th Harmonic Mixing of a 1 GHz Source with an HCN Laser in a Josephson Junction," J. Phys. D: Appl. Phys. 7, 1882-1887 (1974).

[10] T.G. Blaney, N.R. Cross, and D.J.E. Knight, "Harmonic Mixing and Frequency Measurement at 2.5 THz Using Josephson Junctions," J. Phys. D: Appl. Phys., 9, 2175-2181 (1976).

[11] W. C. B. Peatman and T. W. Crowe, "Design and Fabrication of 0.5 Micron Gallium Arsenide Schottky Barrier Diodes for Low-Noise Terhertz Receiver Applications", Int. Journ. Infared and MM Waves; 11, #3, 355-365, (1990).

[12] K.M. Evenson, F.R. Petersen, and J.S. Wells, "Laser Frequency Measurements: A Review, Limitations, Extension to 197 THz (1.5 μm) in Laser Spectroscopy III, (Ed. J.L. Hall and J.L. Carlsten, Springer Verlag 1977) pp. 56-68

[13] F.L. Walls, C.M. Felton, and T. Martin, "High Spectral Purity X-Band Source," these proceedings.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A MULTIPLE MODULATOR FRACTIONAL DIVIDER

Brian Miller, member IEEE, and Bob Conley

Hewlett-Packard
1620 Signal Dr. Spokane, WA 99220

Fractional-N synthesis allows a PLL to achieve arbitrarily fine frequency resolution. Because the technique modulates the instantaneous divide ratio, fractional-N synthesizers suffer from fractional spurs. Various cancellation schemes allow fractional spur reduction to about -70 dBc at the expense of hardware cost and complexity.

Recent advances in oversampling A/D conversion technology can be incorporated into fractional-N synthesis, allowing the spectrum of error energy to be shaped so that fractional synthesis error energy is pushed away from the carrier. Based on this new technology, a CMOS integrated fractional-N divider was successfully developed. A complete fractional-N PLL was constructed utilizing only the CMOS divider, a dual modulus prescaler, a simple loop filter and VCO. The resulting PLL exhibits no fractional spurs.

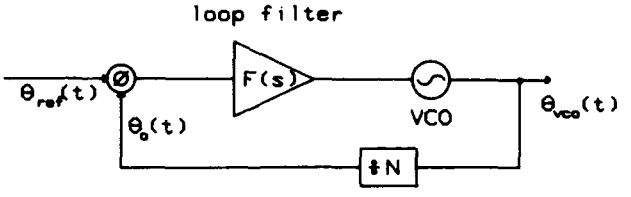


Fig. 1 Basic PLL.

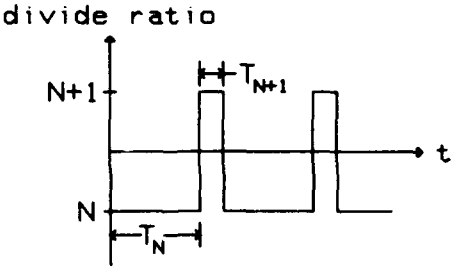


Fig. 2 Alternating divide ratio of fractional-N PLL.

I) Review of Fractional-N Synthesis

In a conventional PLL (Fig. 1) $F_{vco} = N \times F_{ref}$. The divide ratio, N, must be an integer and therefore the frequency resolution of the locked loop is F_{ref}. Fine frequency resolution requires a small F_{ref} and a correspondingly small loop bandwidth. Narrow loop bandwidths are undesirable because of long switching times, inadequate suppression of VCO phase noise, and susceptability to hum and noise.

Fractional-N synthesis was developed to allow a phase-locked loop to have frequency resolution finer than F_{ref} [1], [2]. In a fractional-N divider, the integer divide ratio is periodically altered from N, to N+1 (Fig. 2). The resulting average divide ratio will be increased from N by the duty cycle of the N+1 division.

$$\text{Average } F_{vco} = \frac{1}{(T_N + T_{N+1})} \left[T_N \times N \times F_{ref} + T_{N+1} \times (N+1) \times F_{ref}\right]$$

$$= \left[N + T_{N+1}/(T_N + T_{N+1})\right] \times F_{ref} \equiv (N.f) \times F_{ref}$$

...where N denotes the integer portion of the divide ratio and .f is the fractional component of the average divide ratio.

Typically, the overflow from an accumulator is used to modulate the instantaneous divide ratio (Fig. 3) [3]. Given that;

a) VCO frequency = N.f × Fref
b) Accumulator maximum capacity = C-1
c) X = .f × C (so that X/C = .f)

CH2818-3/90/0000-559 $1.00 © 1990 IEEE

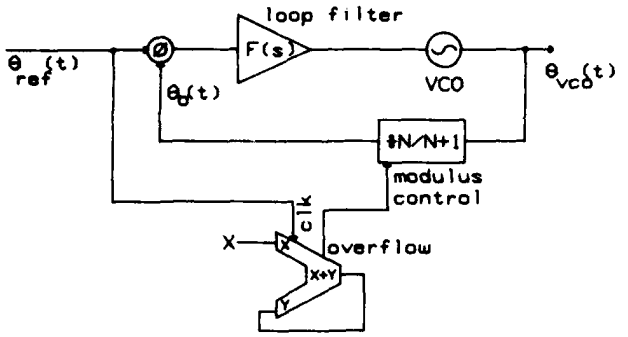


Fig. 3 Fractional-N PLL showing accumulator control of divide ratio.

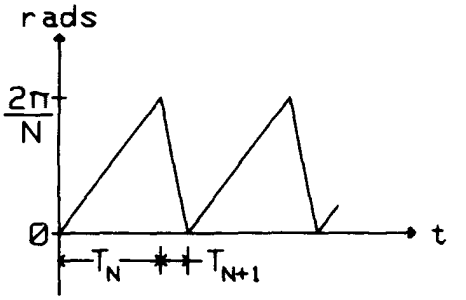


Fig. 4 Sawtooth phase error of conventional fractional-N synthesis.

...then, while the divider is programmed to divide by N, the VCO signal at the phase detector will be at a frequency of Fref + (.f/N)×F_{ref} and the loop phase error will begin to advance at a rate of 2π×(.f/N)×F_{ref} rads/sec. The phase error, referred to the VCO, advances at a rate of 2π×.f×F_{ref} rads/sec. Because the accumulator is summing the same fraction as the VCO phase-error/reference-cycle, an accumulator overflow corresponds to a VCO phase error exceeding 2π radians and indicates the need to remove 2π of phase from the VCO output; accomplished by changing the divider modulus to N+1 for a single reference cycle.

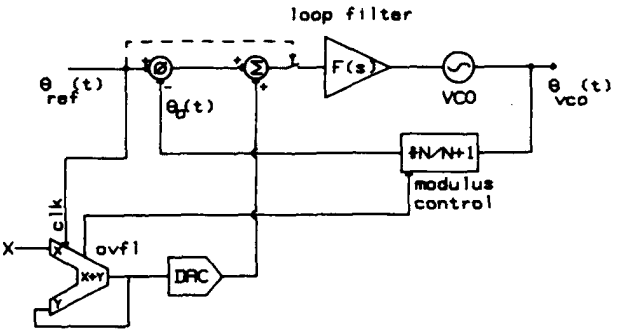


Fig. 5 Fractional-N PLL incorporating phase interpolation.

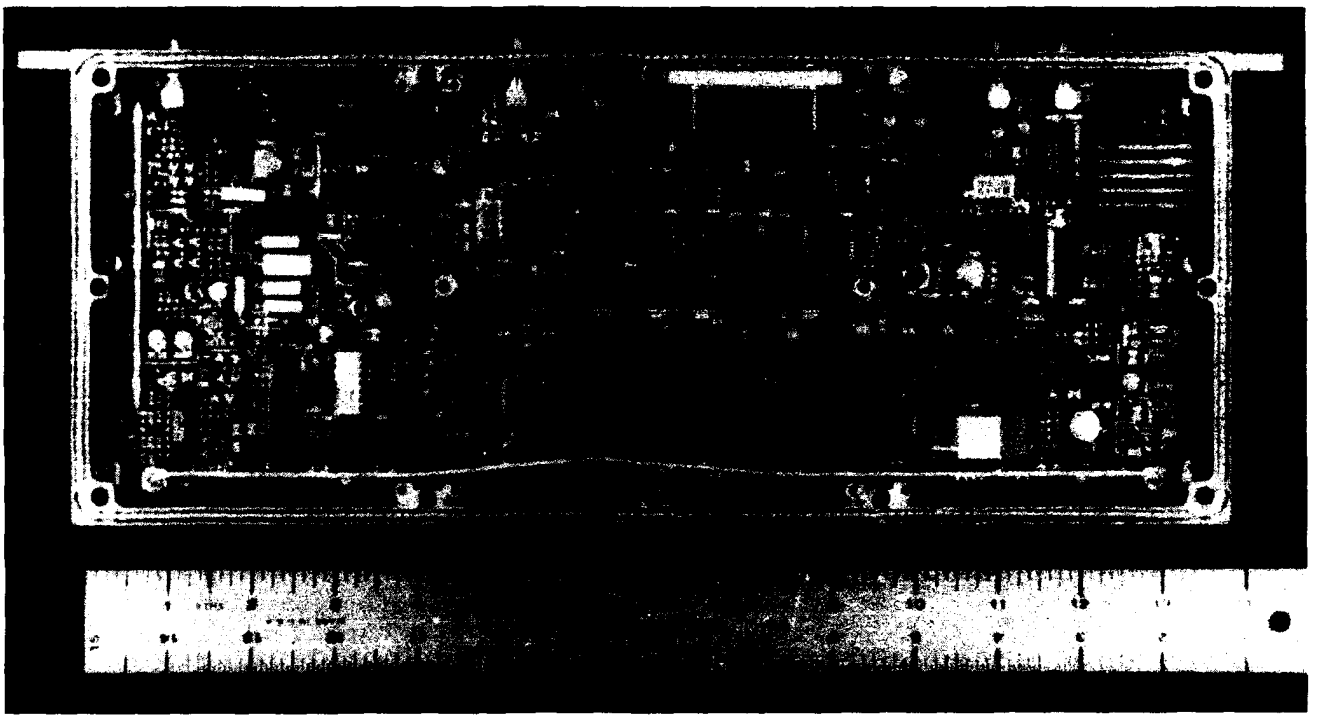


Fig. 6 High performance fractional-N PLL.

This periodic modification of the divider modulus gives rise to a sawtooth phase error (Fig. 4). If unfiltered, the phase error causes severe spurious tones (fractional spurs) at all multiples of the offset frequency ($.f \times F_{ref}$).

Fortunately, it is easy to predict what the phase error will be at any given time and a compensating signal can be summed into the PLL to cancel the error signal (Fig. 5) [4]. The phase detector output is sampled to allow settling of the interpolation DAC every reference cycle. This form of correction is called phase interpolation and fractional spurs are reduced to the extent that the phase interpolation signal exactly matches the phase error. Through the use of precision DACs and carefully designed phase detector and sampler circuitry, fractional spurs of -70 dBc can be achieved. The complexity and expense of the interpolation circuitry makes this form of fractional-N synthesis unsuitable in many applications.

Fig. 6 shows a 0.5 – 1 GHz fractional-N PLL designed for a high quality RF signal generator (HP 8645). This implementation utilizes phase interpolation, and fractional spurs are suppressed to approximately -70 dBc. The circuitry of Fig. 6 does not include the loop VCO, but does contain the complete fractional divider, phase detector, loop amplifier, circuitry to allow FM inside the PLL bandwidth, and microprocessor interface.

II) A New Approach To Fractional Division

Interpolative A/D converters primarily based on sigma-delta modulators, have recently been developed into a viable technology for low frequency measurement and audio systems [5-7]. Interpolative A-D converters operate by greatly oversampling the input with a coarse (usually one-bit) converter and then digitally filtering the one-bit output stream to eliminate out-of-band quantization noise. S/N is further enhanced by embedding the one-bit converter in a recursive filter structure which shapes the quantization noise present at the converter output so that most of the noise energy lies outside the band of interest and will be removed during filtering.

The same concept may be applied to fractional-N synthesis. Fractional-N synthesis attempts to achieve fine frequency resolution thru manipulation of a coarse, integer divider. The desired fractional frequency is analogous to the analog input of an A-D. The integer-restricted divider is analogous to the one-bit converter utilized in interpolative A-D converters. Consequently, most of the recent developments in interpolative A-D converter technology are also applicable to fractional-N synthesis.

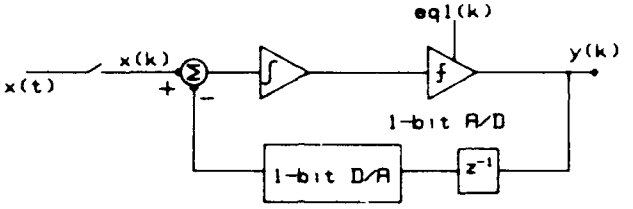


Fig. 7 Sigma-delta modulator.

A multi-stage, sigma-delta modulator architecture based on the work of Matsuya, et al. [5] was chosen to implement the new fractional-N divider. Fig. 7 shows the basic modulator used in sigma-delta A-D converters. x(k) is the modulator input, y(k) is the modulator output, and $e_{q1}(k)$ is the quantization error added by the one-bit A/D. In fractional-N synthesis applications, the input to the sigma-delta modulator is the desired fractional offset, which is a digital word. Consequently, the integrator may be digitally implemented and the one-bit D-A is not required. Fig. 8 shows a sigma-delta modulator suitable for fractional-N synthesis.

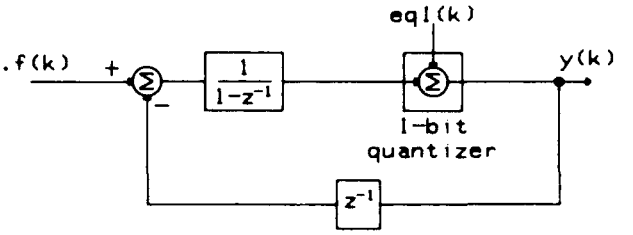


Fig. 8 Sigma-delta modulator suitable for fractional-N synthesis.

$$Y(z) = \frac{1/(1 - z^{-1})}{1 + z^{-1}/(1 - z^{-1})} \times (.F(z)) + \frac{1}{1 + z^{-1}/(1 - z^{-1})} \times e_q(z)$$

$$= .F(z) + (1 - z^{-1}) \times e_q(z) \qquad (1)$$

A block diagram of a 3-modulator fractional divider is shown in Fig. 9. System behavior is assumed sufficiently random to justify modeling each one-bit quantizer as a unity gain element with added quantization noise [8]. Under these assumptions the quantization noise will be uniformly distributed with

$\sigma^2 = (1)^2/12$. N.f is the desired rational divide ratio and $N_{div}(k)$ is the actual sequence presented to the integer restricted divider. Using Eq. 1;

$$N_1(z) = (1-z^{-1})e_{q1}(z) + .f(z) \quad (2)$$

$$N_2(z) = -e_{q1}(z) + (1-z^{-1})e_{q2}(z) \quad (3)$$

$$N_2'(z) = -(1-z^{-1})e_{q1}(z) + (1-z^{-1})e_{q2}(z) \quad (4)$$

$$N_3(z) = -e_{q2}(z) + (1-z^{-1})e_{q3}(z) \quad (5)$$

$$N_3'(z) = -(1-z^{-1})^2e_{q2}(z) + (1-z^{-1})^3e_{q3}(z) \quad (6)$$

$$N_{div}(z) = N(z) + N_1(z) + N_2'(z) + N_3'(z) = N.f(z) + (1-z^{-1})^3e_{q3}(z) \quad (7)$$

In a locked PLL, $f_{out}(k) = N_{div}(k) \times \vec{r}_{ref}$

Using Eq. 7 we can then write;

$$f_{out}(z) = N.f(z) \times F_{ref} + (1-z^{-1})^3F_{ref}e_{q3}(z) \quad (8)$$

Where the 1st term of Eq. 8 is the desired frequency, and the second term represents frequency noise due to fractional division. This form is not useful for frequency synthesis applications and needs to be converted into single-sideband phase noise, £(f).

e_{q3} has variance = 1/12 over a bandwidth of F_{ref}

Consequently, the power spectral density (PSD) of $e_{q3} = 1/(12F_{ref})$.

Defining, $\nu(z) \equiv$ frequency fluctuations of $f_{out}(z)$

$$S_\nu(z) = |(1-z^{-1})^3F_{ref}|^2 \times (1/12F_{ref}) \quad (9)$$

$$= |1-z^{-1}|^6 \times F_{ref}/12 \quad (10)$$

We want phase fluctuations, not frequency fluctuations.

$$\phi(t) = \int w(t)\,dt = 2\pi\int \nu(t)\,dt$$

Employing a simple rectangular integration to represent $\int dt$ in the z-domain;

$$\phi(z) \quad \frac{T_sW(z)}{1 - z^{-1}} = \frac{2\pi T_s\nu(z)}{1 - z^{-1}} \quad (11)$$

Using $T_s = 1/F_{ref}$ and Eq. 10 in Eq. 11, we obtain;

$$S_\phi(z) = \frac{(2\pi)^2}{|1-z^{-1}|^2F_{ref}^2} \times \frac{|1-z^{-1}|^6F_{ref}}{12}$$

$$= \frac{(2\pi)^2}{12F_{ref}} |1-z^{-1}|^4 \text{ rad}^2/\text{Hz} \quad (12)$$

If $S_\phi(f)$ is a two-sided PSD, then; £(f) = $S_\phi(f)$. Therefore;

$$£(z) = \frac{(2\pi)^2}{12F_{ref}} |1-z^{-1}|^4 \text{ rad}^2/\text{Hz} \quad (13)$$

Converting to the frequency domain and generalizing to any number of modulator sections;

$$£(f) = \frac{(2\pi)^2}{12F_{ref}} \left[2\sin(\pi f/F_{ref})\right]^{2(m-1)} \text{ rad}^2/\text{Hz} \quad (14)$$

...where m is the number of modulator sections.

Typically we are concerned with offset ranges small compared to the reference frequency allowing;

$$£(f) \approx \frac{(2\pi)^2}{12F_{ref}} \left[\frac{f}{F_{ref}/2\pi}\right]^{2(m-1)} \text{ rad}^2/\text{Hz} \quad (15)$$

Eq. 15 gives the colored quantization noise produced by the multiple modulator synthesis technique. Instead of discrete spurs, error energy produced by the fractional division will be manifested as noise. This noise must be filtered prior to the VCO to prevent unacceptable degradation of spectral purity. Interpolative A-D converters utilize digital filters to remove out-of-band quantization noise. In a fractional-N synthesis application, the PLL lowpass characteristic may be utilized to filter the quantization noise. A circuit example is given in Section IV.

The system of Fig. 9 can be simplified when it is recognized that an accumulator is a compact realization of the sigma-delta modulator. Fig. 10 shows an accumulator based sigma-delta modulator. The feedback of Fig. 8 occurs implicitly in the internal logic of the accumulator. Incorporating accumulators into the three-stage sigma-delta modulator yields Fig. 11. The topology of Fig. 11 is readily amenable to integration. The multiple modulator fractional-N control system reduces to a forward path of accumulators and a reverse path of differentiators.

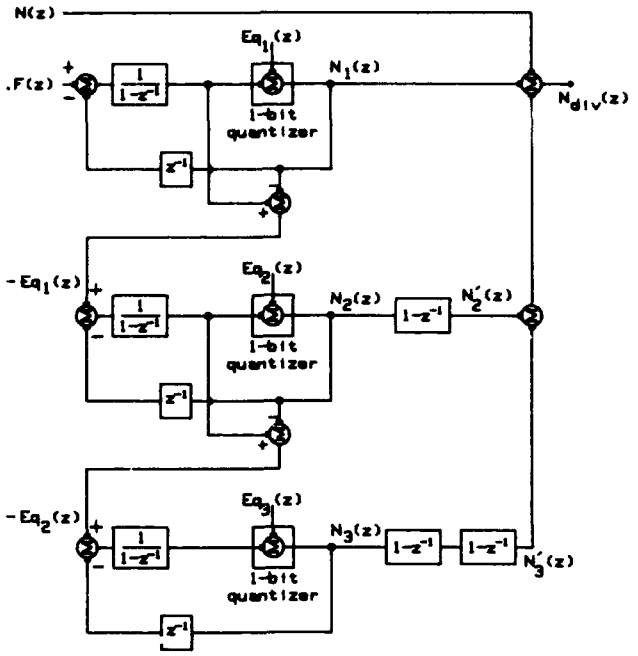


Fig. 9 Three-modulator fractional-N divider.

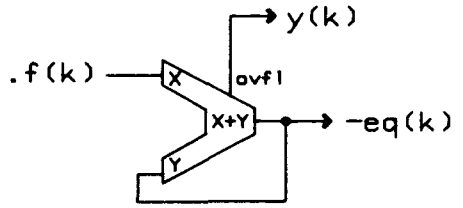


Fig. 10 An accumulator regarded as a sigma-delta modulator.

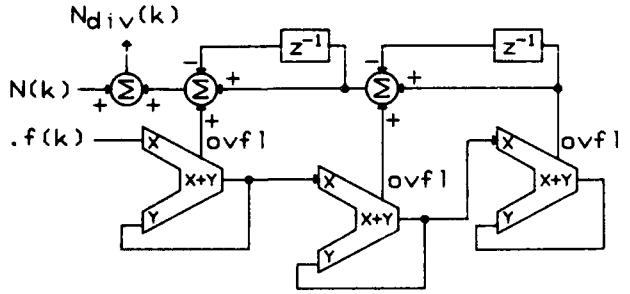


Fig. 11 Multiple modulator divider implemented with accumulators.

III) Fractional Divider Implementation

An IC was conceived to implement all the digital functions of a PLL incorporating a multiple modulator fractional divider. The part is fabricated in a 1.5 micron, 5 Volt CMOS process and packaged in a 44 pin plastic leaded chip carrier. The part dissipates 75mW when clocked at 15 MHz.

Fig. 12 illustrates the overall block diagram of the IC. This diagram can be broken down into two distinct sections:

1) **Basic Synthesizer**
   - a) N-divider
   - b) Phase Frequency Detector
   - c) Reference Divider
   - d) Out of Lock Detector
   - e) Serial Interface

2) **Computation and Control**
   - a) Sequential Math Section
   - b) Dedicated Math Section
   - c) Data Flow Control

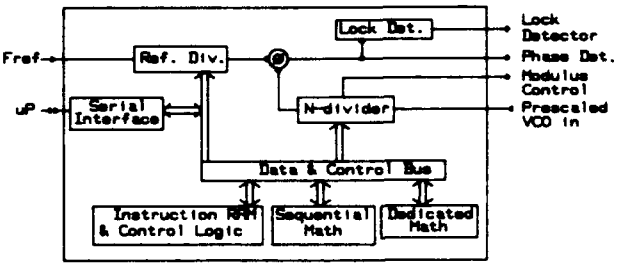


Fig. 12 Block diagram of the fractional-N IC.

Not shown in Fig. 12 but included on the IC is circuitry to support FM inside the PLL bandwidth, synthesized sweep, and a general purpose 12-bit parallel data bus for control of external perhipheral devices.

Basic Synthesizer

Although the basic synthesizer looks very much like the parts available from Motorola [11], Fujitsu [12] and others, significant differences exist, particularly in the N-divider. Those who have built and tested fractional-N PLLs know the importance of avoiding spurious coupling from the N-divider, reference divider, and the fraction computation logic onto the edges presented to the phase detector. With this in mind, the following basic premise governed the architecture of the N-divider and Math sections of the IC.

"Avoid spurious coupling of divider and computation events to the phase detector edges by restricting both to separate times within the PLL cycle."

This, coupled with the following two observations, resulted in the general PLL Cycle requirements illustrated in Fig. 13.

a) To reduce the delay of FM, sweep and other external requests, the requests should be sampled and their associated calculations executed at the last possible moment before the N-divider is reloaded.

b) The VCO/N edge placement must represent exactly the number of VCO cycles,calculated by the sigma-delta modulators. No incidental PM can be allowed to modulate the VCO/N edge. The start of the reference cycle is "quiet" and reserved for the phase detector edges. The last portion of the cycle is reserved for the noisy calculations and prescaler modulus control changes.

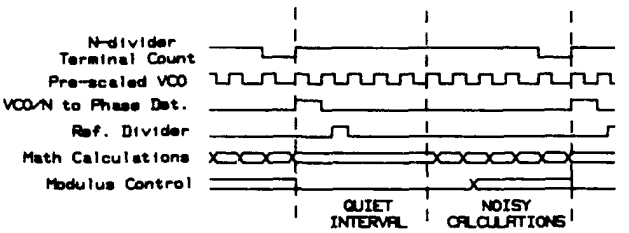


Fig. 13 Timing diagram of the basic PLL cycle.

The N-divider controller is based on the well known dual modulus prescaler [13] technique. In a typical prescaler based divider, prescaler modulus changes are referred to the beginning of the reference cycle as illustrated in Fig. 14. However, unlike most dividers found in single chip synthesizers, the new IC references changes of the prescaler modulus to the end of the PLL reference cycle (Fig. 15) to facilitate time separation of phase detector edges and noisy computations. Instead of using two counters to generate prescaler modulus control, a counter and comparator are used. The Q outputs of the down counter are compared to the value stored in the modulus control register. Although other dual modulus dividers have been built based on a single counter and comparators [14], they do not reference the modulus control changes to the end of the PLL cycle.

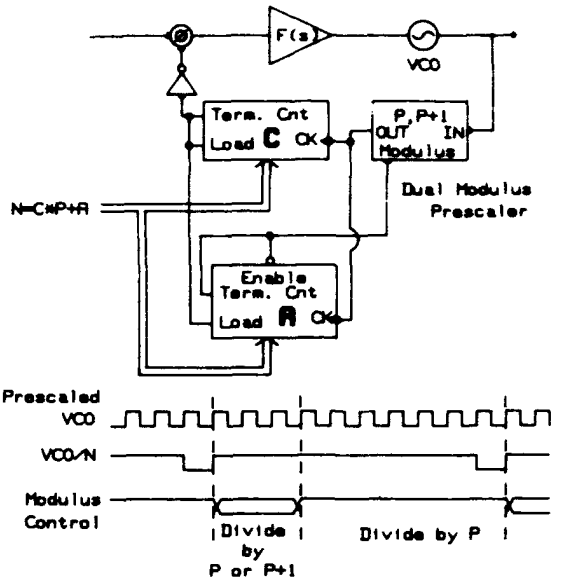


Fig. 14 Prescaler controller with modulus changes referenced to the beginning of the reference cycle.

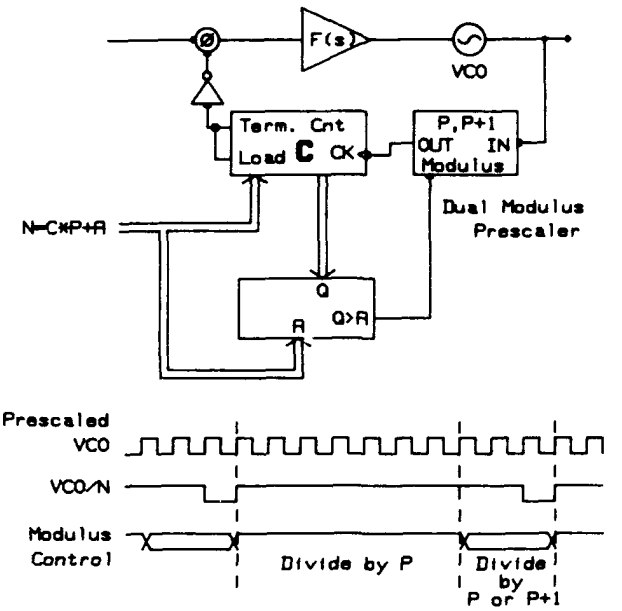


Fig. 15 Prescaler controller with modulus changes referenced to the end of the reference cycle.

Computation and Control

Consider the implementation of the three modulator divider illustrated in Fig. 11. In addition, let the desired fractional resolution require a 24 bit word width in each modulator and an N value less than 4096. The major logic required to implement this is cataloged as follows:

1) One 24 bit adders
2) Four 24 bit registers
3) Six 24 bit data paths
4) Two $1-z^{-1}$ operators
5) One 12 bit adder
6) Three 12 bit data paths

This could be a large integration problem. Since multiple prescaler output cycles are available every PLL cycle, a sequential approach to the accumulations was chosen. This approach complicates IC control but saves considerable die area. The number of prescaler output cycles per PLL cycle are limited, so the small valued computations such as the $1-z^{-1}$ operators are implemented in dedicated logic. The $1-z^{-1}$ operators can be scaled since their peak numeric value is bounded (e.g. +1, -1 for the third modulator section). This further reduces die requirements.

The control section, in addition to scheduling accumulations, is required to fetch dedicated computations into the sequential section and pass new values off to the N-divider and peripherials used in FM [15] and sweep functions. The state machine required to implement the sequential math and control could have been implemented in PLA, ROM or RAM. Cell based RAM was selected for its obvious versatility as well as overall speed when compared to available cell based ROM. The IC has a 16 instruction control RAM which must be programmed upon application of power to the part. RAM based control hastened development of the part and its intended application because it was easy to reconfigure the IC for experimentation. The RAM based control also provides flexibility to PLL designers who, with the IC, can program unique combinations of correction, specialized sweep and FM or PM.

IV) Complete Fractional-N Synthesis PLL

The new fractional-N divider was used to construct a 0.5-1 GHz PLL (Fig. 16). The divider IC was programmed to implement a three modulator interpolative divider. An external 32/33 prescaler was utilized in conjunction with the fractional-divider IC to implement the N-divider function. The prescaler is preceeded by a divide-by-two because the 32/33 prescaler is not able to toggle at 1 GHz.

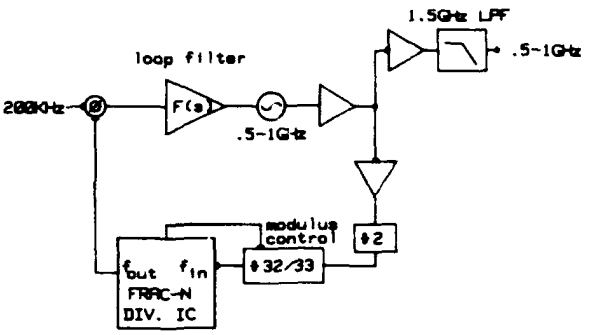


Fig. 16 0.5–1 GHz fractional-N PLL.

The VCO is a single band, 0.5–1 GHz, varactor tuned transistor oscillator. The loop F_{ref} is 200 KHz. In addition to the normal phase control objectives, the loop filter must also provide adequate rejection to the synthesis noise. The predicted synthesis noise is overlaid on a plot of the free running VCO phase noise (Fig. 17) to determine the loop filter design parameters. The synthesis noise plotted in Fig. 17 is increased 6 dB over Eq. 14 because of the divide-by-two prescaler between the VCO and the fractional divider circuit. Note how the fractional division error energy has been pushed away from the carrier. In this case, the error energy intersects the VCO phase noise at an offset greater than where the phase detector noise intersects the VCO phase noise. Notice also that the error energy rises at 40 dB/decade. For this application, the servo requirements were satisfied with a type II, second order loop of approximately 750 Hz bandwidth. The filtering requirements were met by adding two additional real-axis poles (1.8 KHz, 3 KHz) to the loop filter.

The results of a phase noise measurement on the locked loop are given in Fig. 18. The fractional divider was programmed for a frequency offset of 2 KHz. None of the traditional 2 KHz (and multiples thereof) spurs are present.

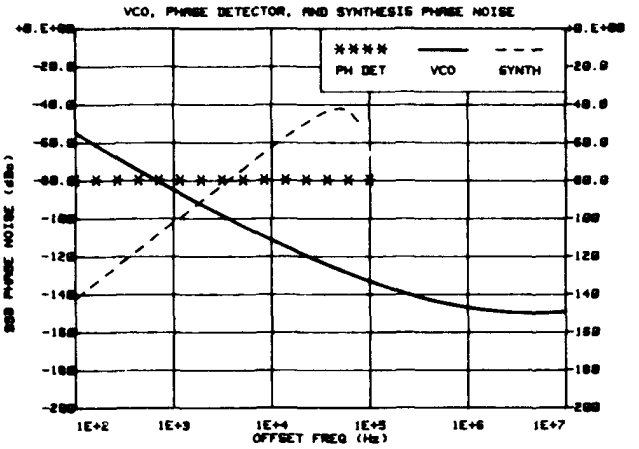


Fig. 17 Phase noise sources in the 0.5--1 GHz PLL.

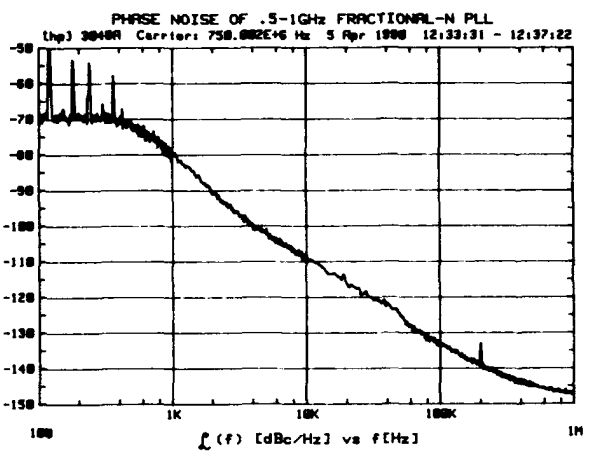


Fig. 18 Measured phase noise of the PLL.

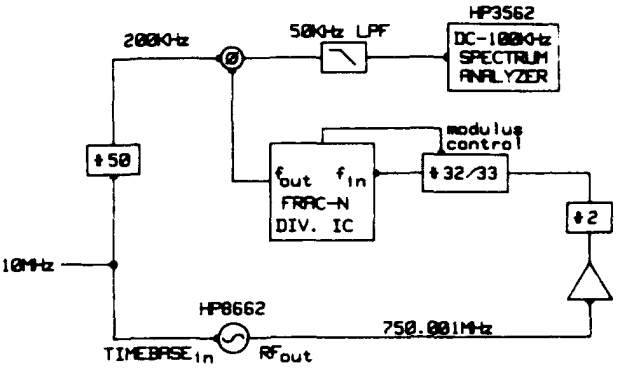


Fig. 19 Apparatus to measure synthesis noise of new fractional-N technique.

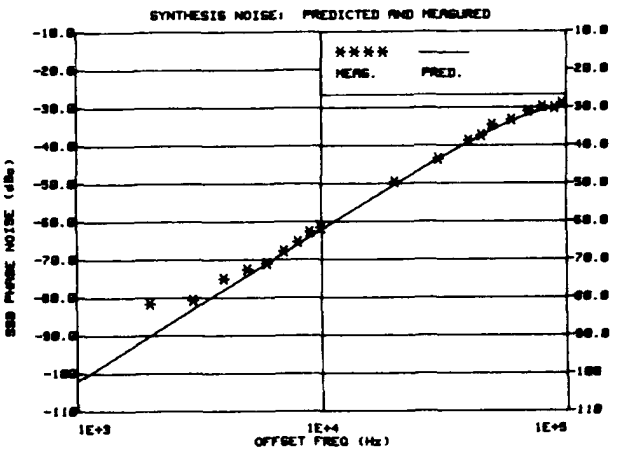


Fig. 20 Theoretical and measured synthesis noise.

Eq. 14 was verified with the apparatus of Fig. 19. N.f was chosen to divide the applied RF (750.001 MHz) down to the reference frequency (200 KHz). The baseband noise recorded on the spectrum analyzer was adjusted by the reciprocal of the low pass filter gain; which was measured separately. Results are given in Fig. 20. The noise floor of the measurement system was approximately -80 dBc.

The mathematical prediction of synthesis error energy (Eq. 14) relies on a uniform quantization noise model. The reader may rightly wonder if this is justifiable. Several authors [9,10] have addressed the exact output spectra of a sigma-delta modulator but have limited their analyses to single stage modulators. For single stage modulators, the output spectra is strongly dependant on the DC input to the modulator and a quantization noise model is not appropriate [9]. Heuristically, higher order coders decrease the correlation between the input signal and the quantization error [10], permitting a quantization noise model.

We found the three-stage ΣDM to comply with the uniform noise model if the first accumulator experiences activity in, or near, the LSB bit position. Inputs which excite only bits near the MSB position, such as $.f = .5, .75, .25$ etc, result in a limit cycle of short duration and insufficient randomness to decorrelate the quantization error. Fortunately, it is easy to always add one LSB to the desired frequency offset if the desired .f does not already have the LSB set. Because the accumulator length is so large, a one LSB frequency error is easily tolerable. In our test circuit utilizing a 24 bit accumulator, one LSB corresponds to 0.0238 Hz, or 4.77×10^{-5} ppm worst case.

V) SUMMARY

Fig. 21 shows a 0.5–1 GHz PLL incorporating the new fractional-N synthesis technique. The circuit contains the complete PLL (including VCO), circuitry to allow FM inside the PLL bandwidth and a microprocessor interface. This implementation is over 2.5 times smaller than its phase interpolation based predecessor.

A new technique for fractional-N synthesis has been developed. The technique causes error energy to be suppressed at small offsets, allowing the remaining error to be rejected by simple filtering. The technique provides a tremendous advantage in cost, size and complexity over traditional fractional-N synthesis techniques which utilize error correction.

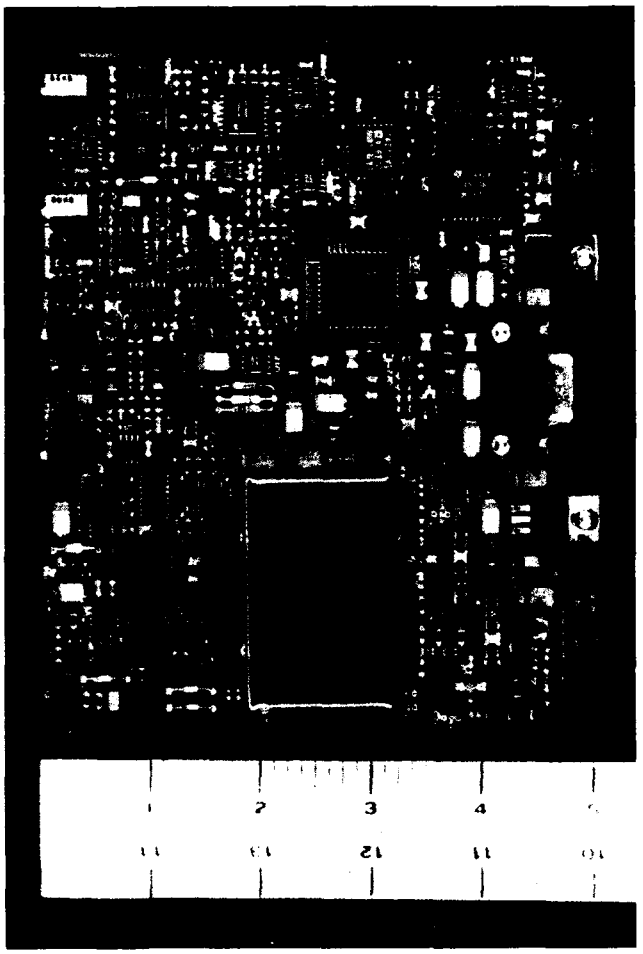


Fig. 21 Complete fractional-N PLL utilizing new divider IC.

Acknowledgement

The authors would like to thank Scott Grimmett, Marcus DaSilva, Mark Talbot, Ben Flugstad and Jim Catlin for their contribution to the successful development of the IC.

References

[1] V. Manassewitsch, Frequency Synthesizers, New York, N.Y., John Wiley & Sons, 1987, pp. 43-48

[2] W. F. Egan, Frequency Synthesis by Phase Lock, New York, N.Y., John Wiley & Sons, 1981, pp. 196-202

[3] C. A. Kingsford-Smith, Patent No. 3,928,813 (Washington, D.C.: U.S. Patent Office, December 23, 1975)

[4] R. G. Cox, Patent No. 3,976,945 (Washington, D.C.: U.S. Patent Office, August 24, 1976)

[5] Y. Matsuya, et al., "A 16-Bit Oversampling A/D Conversion Technology Using Triple Integration Noise Shaping", IEEE Journal Solid State Circuits, Vol. SC-22, pp. 921-929, Dec. 1987

[6] J. C. Candy, "A Use of Double-integration in Sigma-delta Modulation," IEEE Trans. Commun., Vol. COM-33, pp. 249-258, Mar. 1985

[7] D. R. Welland, et al., "A Stereo 16-Bit Delta-Sigma A/D Converter for Digital Audio," J. Audio Eng. Soc., Vol. 37, No. 6, pp. 476-484, June 1989

[8] B. Agrawal and K. Shenoi, "Design Methodology for ΣDM", IEEE Trans. Commun., Vol. CE-31, pp. 360-370, Mar. 1983

[9] J. C. Candy and O.J Benjamin, "The Structure of Quantization Noise from Sigma-delta Modulation," IEEE Trans. Commun., Vol. COM-29, pp. 1316-1323, Sept. 1981

[10] R. M. Gray, "Oversampled Sigma-delta Modulation," IEEE Trans. Commun., Vol. COM-35, pp. 481-489, May 1987

[11] Motorola CMOS Special Functions Data Book, Q2 1988

[12] Fujitsu Telecommunications Data Book, 1987

[13] Motorola Application Note AN-827

[14] J. A. Borras, Patent No. 4,427,820, (Washington, D.C.: U.S. Patent Office, Sept. 18, 1984)

[15] M. K. DaSilva, et al., Patent No. 4,546,331, (Washington, D.C.: U.S. Patent Office, Oct. 8, 1985)

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A NOVEL DESIGN OF A 1.8 GHZ INPUT ODD RATIO FREQUENCY DIVIDER

David Korn, Code 8106, Naval Research Laboratory, 4555 Overlook Ave. S.W., Washington, D.C. 20375

Carl Deierling, Select Electronics, 14730 Beach Blvd., Suite 106, Bldg. F, La Mirada, CA 90638; formerly with GigaBit Logic, Newbury Park, CA at time of authorship

Abstract-Frequency division is an integral part of rf and microwave frequency synthesis. Unfortunately until now, non power-of-2 (e.g. 3,5,6) divider ratios are limited to maximum input frequencies of about 500 MHz ; the few exceptions are at fixed divide ratios. Presented is the analysis and measured results of a new design, which is conceptually and practically simple, that is capable of dividing by any integer with measured results for divide-by-3 at a 1.8 GHz.maximum input. The concept is based on a D flip-flop with a delay line that controls the divide integer in a feedback configuration. The divider is very stable over frequency, power and temperature, is very small, produces extremely low-level spurious frequencies, has moderate residual phase noise and is very manufacturable

INTRODUCTION

Presented is a new technique for building an odd ratio fixed frequency divider at a demonstrated 1.8 GHz maximum input frequency.This technique has been conceived, designed, tested and built by the authors. In this article the word "odd" is used as a mnemonics for a non 2^n integer rather than a non-even number. This divider, which was necessitated by and is intended for rf frequency synthesizer applications, can divide by any integer; however, its novelty is most useful when the integer chosen is a non 2^n number, namely 3,5,6,7 etc, as there are many multi-GHz input 2^n (2,4,8,16 etc) dividers presently available. Both the circuit itself and its intended function, selectable non-2^n ratio and 1 GHz-plus input frequency division, appear to be novel and is most certainly useful. There do not appear to be any dividers available based on this circuit concept and the few dividers which are available with input frequencies above 1 GHz are not capable of dividing by any integer. Consequently it gives the frequency synthesizer designer another design option not currently available. The circuit concept is based on a D flip-flop with a delay Tau fedback from the inverted Q output to the flip-flop clear input with the D input tied high; the divide ratio N is controlled by selecting the delay Tau. The input frequency is applied to the clock input and the output frequency is the Q output. Since this rf circuit is based on a digital ic its features are directly coupled to the anticipated advances in high-speed digital ics, namely increasing maximum input clock frequency and minimizing the residual (additive) phase noise, physical size and dc power consumption. The most important measured inherent characteristics of this divider technique from a frequency synthesis point of view are lack of measurable spurious oscillations as the input rf frequency, rf power, dc voltage and ambient temperature are changed, all dependently or independently of one another, a scarcity of undesired subharmonics, low residual phase noise, very small size, low dc power consumption, low input rf drive and moderate rf power output delivery.

BACKGROUND PERSPECTIVE

Frequency division is an integral part of rf frequency synthesizer techniques. The majority of frequency dividers are realized digitally via edge triggered flip-flops or as analog circuits based on varactor diodes and resonant circuits. The following paragraphs testify to the usefulness of the proposed divider circuit.

Edge triggered digital divider circuits are designed by configuring a flip-flop to change state on a rising or a falling edge but not on both; this in effect is a divide-by-2. A divide-by-2^n (2,4,8 etc.) is made possible by cascading the desired number of flip-flops. If a non 2^n divide ratio is the goal, a feedback circuit between flip-flops can accomplish this. There many ways to do this. Off-the-shelf pre-designed dividers with fixed divide ratios are designed on this technique as are dividers built by the equipment designer using the industry-standard 4-bit counter.

First consider the pre-designed option. Plessey Semiconductor has the most comprehensive selection of pre-designed dividers, all built with silicon, presented are the highest input frequency odd ratio dividers for N < 10. The SP8720 divides by 3 @ Fmax = 300 MHz, the SP8620 divides by 5 @ Fmax = 400 MHz, the SP8740 divides by 6 or 7 @ Fmax = 300 MHz, and the SP8743 divides by 9 @ Fmax = 500 MHz. GigaBit Logic has the highest non 2^n divider, the 10G070, a 2.0 GHz dual-modulus divider; however it divides by 5 or 6 only. If one wants to divide by 3,5,6,7 or 9, these are the options.

US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT

It is worth noting that a number of companies make multi-GHz input 2 dividers. Avantek makes the IFD-01110, a silicon 4.5 GHz divide-by-4, Plessey makes the SP8808A, a 3.3 GHz silicon divide-by-8 and NEC makes the UPG501B, a GaAs 5 GHz divide-by-4 and the UPG506B, a 14 GHz GaAs divide-by-8, just to name a few. These dividers are built with flip-flops as described previously however they do not have asynchronous clear inputs, thereby making them ineligible for use in the proposed divider circuit. As dividers and D flip-flops with asynchronous clear inputs increase in input frequency so will the input frequency of non 2^n dividers built with this technique. If these dividers and flip-flops can be made of silicon instead of GaAs, the improvement of residual phase noise, specifically the approximately 20 dB decrease of the 1/f frequency point, will make this technique better suited for lowest-possible phase noise synthesizers. At this time the chosen D flip-flop, GigaBit Logic's 10G021A-2L, a GaAs device, is the highest input frequency D flip-flop with asynchronous clear that is commonly available.

Next the design-it-yourself option. The numerous 4 bit programmable counters readily lend themselves to 2^n and non 2^n frequency division ratios. These counters are available in cmos, ttl, ecl and GaAs circuit technologies. cmos and ttl counters, manufactured by many companies, have maximum input frequencies at about a 100 MHz. GigaBit Logic makes a GaAs counter, the 10G061, with a 1.3 GHz maximum input frequency, this is the only 1 GHz-plus input frequency 4-bit counter currently available. However due to propagation delays, feedback circuits inherently reduce maximum input frequency, often by a multiplicative factor of 0.60 to 0.80, compared to the same flip-flop configured for straight 2^n divide ratios with no feedback.

In view of the preceding paragraphs one can conclude that the only pre-packaged divider above 500 MHz input frequency is limited to divide ratios of 5 & 6 and the maximum input frequency of a 4-bit counter set for a non 2^n divide ratio is 1 GHz.

CONCEPT AND ANALYSIS

The concept of this novel divider technique is based on a D flip-flop and a delay Tau fedback from the inverted Q output to the asynchronous clear (CLR) input with the D input tied high as shown in the block diagram. The maximum input frequency is limited by the maximum clock (CK) frequency of D flip-flops on the market and the divide ratio is controlled by setting the propagation delay, Tau, of the feedback circuit. The D flip-flop must have an asynchronous CLR input. The circuit analysis, including device-specific propagation delays, is most clearly visualized by referring to the timing diagram of the built divider. The input frequency is 1.4 GHz, 714 picoseconds, the flip-flop is GigaBit Logic's 10G021A, and the desired divide ratio, N, is 3. The cause and consequence of each change of state in either CK, Q, Qbar or CLR with respect to time is marked by a number in the diagram and verbally explained as follows:

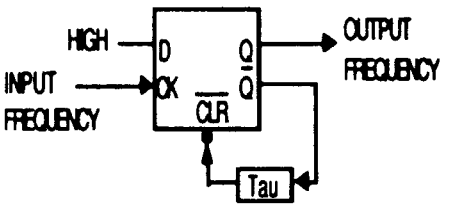


FIGURE 1

-1- On the clock falling edge the D input is strobed into the flip-flop. D is hard-wired logic high and the particular flip-flop is negative edge triggered. Whether the particular flip-flop used is negative or positive edge triggered is inconsequential. The state of Q prior to event -1- is irrelevant.

-2- As a result of event -1- the Q and Qbar outputs go high and low respectively after the CK-to-Q delay, in this case 525 picosecs.

-3- Since the Qbar output is fedback via a delay Tau to the CLR input , a time delay Tau after event - 2- Qbar low appears at the CLR input . There are no changes in Q due to any clock falling edges that might occur between events -2- & -3- since the D input is hard-wired high and Q is already high.

-4- As a result of event -3- the Q and Qbar outputs go low and high respectively after the CLR to Q and Qbar delay, in this case 625 picosecs.

-5- Since the CLR input is merely the delayed Qbar output, a time delay Tau after event -4- the CLR port goes high (deactivated). Remember that between events -3- & -5-, when the CLR input is low, the clock input has no effect on the state of either outputs.

-6- Event -6- is a repeat of event -1-, at that time the D input is strobed into the flip-flop. Q is low at this time.

-7- As a result of event -6-, Q changes state and goes high after the CK-to-Q delay thus completing one frequency division cycle.

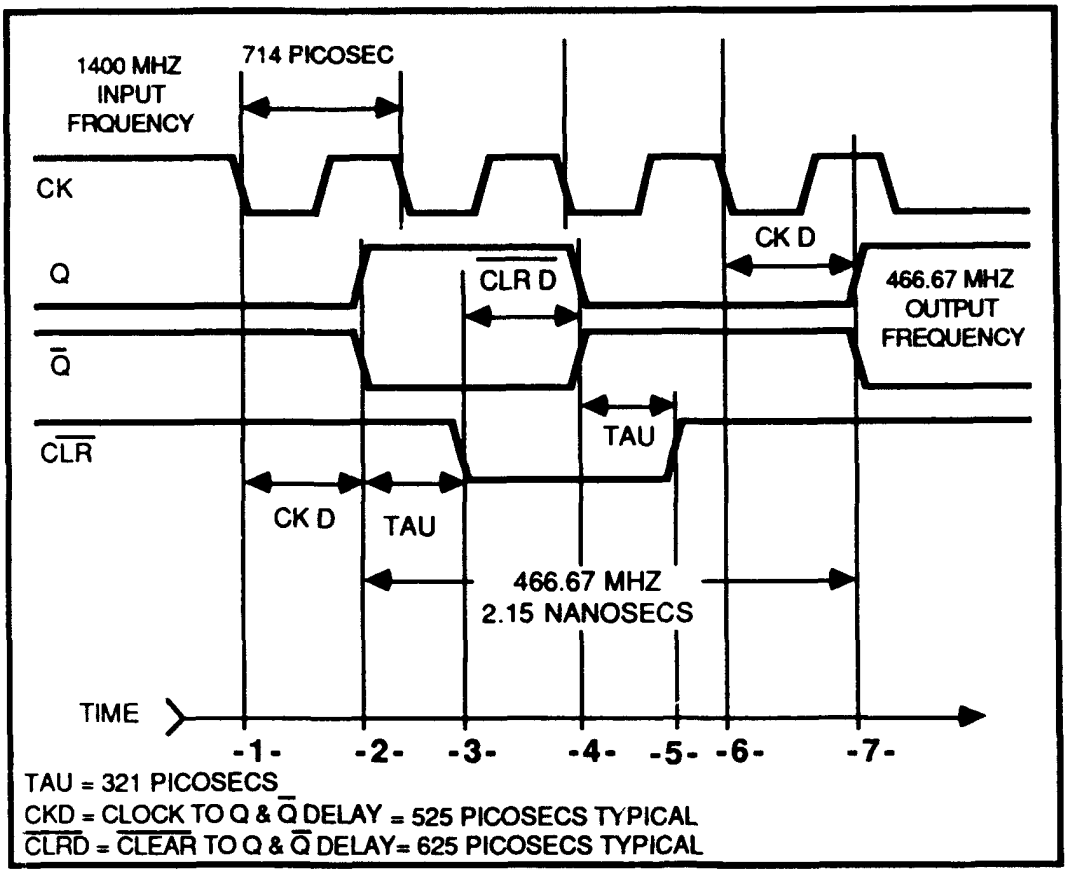


FIGURE 2

Given the preceeding explanation one can realize the following:

The value of Tau is chosen such that 2Tau plus both CK-to-Q and CLR-to-Q delays is greater than N-1 input clock cycles and less than N clock cycles where N is the desired divide ratio; this would be stated in equation form as:

(1CK cycle)(N-1) < 2Tau + CK-to-Q + CLR-to-Q < (1CK cycle)(N) sec

The objective is to solve for Tau given a known (desired) value of N. This is the only pertinent design equation.

Note that the timing diagram reveals that the input to output time delay of this circuit is the D flip-flop internal CK-to-Q delay, in this case 525 picosecs.

This technique can also be used to build a high speed dual modulus prescalar. The only change in the circuit would be to have 2 delay lines, one for and the other for N + 1, connected from Q inverted to a 2:1 mux input, and connect the mux output to the CLR input. The mux control signal would select which delay line is in the circuit thereby selecting N or N + 1.

GENERAL CIRCUIT DESIGN PROCEDURE

The circuit design procedure is fairly straightforward and is summarized as follows:

Select the D flip-flop. The flip-flop must have an asynchronous clear input. Synchronous Q and Qinverted outputs are preferred but not essential. The only relevant ac parameters of the chosen flip-flop are the CK-to-Q and CLR-to-Q delays. The maximum divider (flip-flop plus required feedback) operating frequency is limited by the above mentioned delays, the flip-flop maximum input frequency and the external delay that comprise Tau. The divider maximum frequency has been measured to be about 0.6 to 0.8 times the flip-flop maximum frequency for N = 2 or 3. As expected, the maximum divider input frequency will decrease as the external delay Tau is increased; this must be considered when selecting the flip-flop.

The next step is to determine the value of the external delay, Tau, that must be introduced to acheive the desired divide ratio N. The above stated equation yields the value of Tau. How Tau is implemented depends on what it physically is. For values of N less than 10 Tau can conveniently be realized passively through a coaxial cable, microstrip transmission line or as in this application, a piece of wire. The method the authors used was to consider the passive delay line as a characteristic impedance in series with the shunt CLR input capacitance. These two quantities determine a time constant which in turn determines how quickly the CLR input voltage reaches the threshold voltage. The design example goes through this in detail.

The last step is to determine the circuit design practicalities dictated by the already chosen flip-flop and delay line. No "special" circuits such as tuned or diode clamping circuits

are required. Refer to the detailed circuit schematic and photo of the design example built. The salient features are summarized: Even though this circuit concept is digital at heart (based on and realized with an edge triggered device), consider this an rf circuit and design accordingly. Make sure to use chip capacitors and resistors for rf line terminations, flip-flop output voltage pull downs, ac decoupling and dc blocking. Leaded dc blocking capacitors at the output (Qport of flip-flop) have been determined to cause spurious frequencies near the carrier, the cause of death of many frequency synthesizer circuits.

DESIGN EXAMPLE

The system-imposed specfication that necessitated the invention of this divider dictated that a frequency between 450 MHz and 550 MHz with spurious less than -80 dBc be generated from a 1400 MHz reference. The authors conceived this circuit after confirming that dividing 1.4 GHz by 3 would be the best solution and that there were no pre-designed divide-by-3s available at 1.4 GHz.

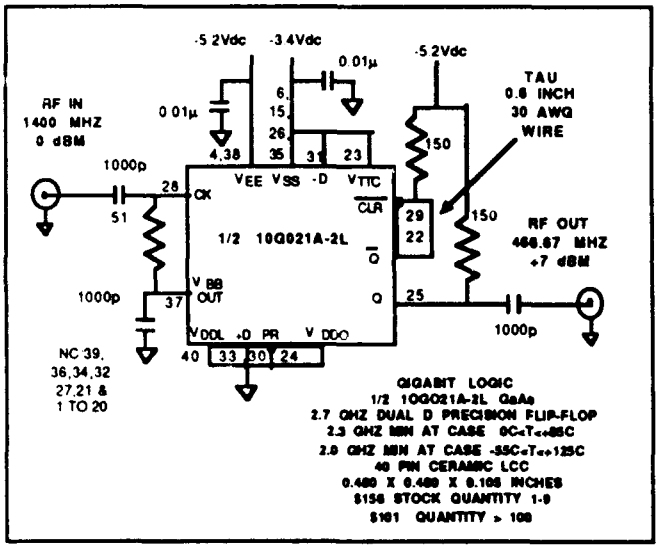


FIGURE 3

The divider was designed with the GigaBit Logic 10G021A-2L in mind, a dual D flip-flop with a 2.7 GHz maximum clock frequency; only one flip-flop of the two in the 40-pin IC was used. The required delay Tau for N = 3 is calculated using the preceding equation:

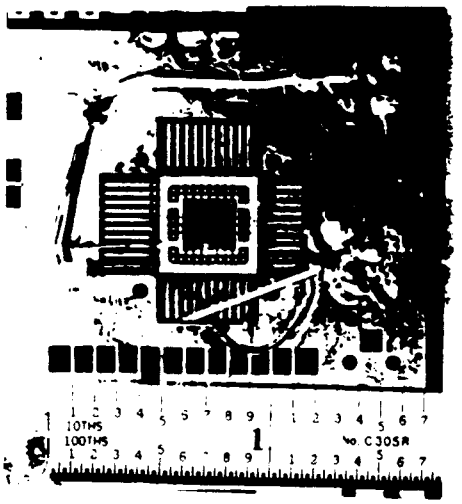
1400 MHz is 714 picosecs, therefore the sum of all 3 delays must be between:

1428 picosecs < 2Tau + CK-to-Q + CLR-to-Q < 2142 picosecs

clock-to-Q delay = 525 picosecs
CLR-to-Q delay = 625 picosecs

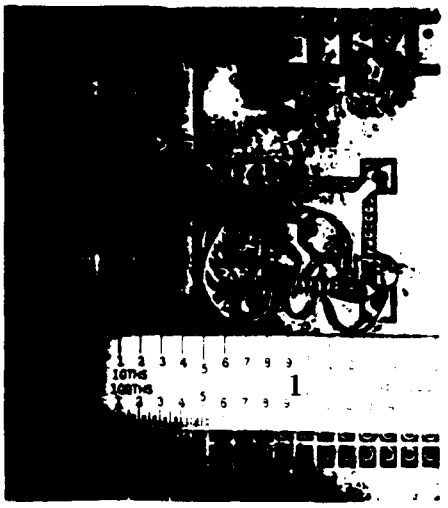
thus: 140 picosecs < Tau < 496 picosecs

The next step was to try to account for all delays that comprise Tau; in order to do this a detailed description of the layout is necessitated. For extra clarity refer to the circuit photo and Tau equivalent circuit. The 10G021A-2L dual D flip-flop is a 40 pin leadless chip carrier (lcc) package that is 0.480 X 0.480 X 0.105 inches. It is directly soldered to the top of a GigaBit Logic prototype board, the 90GUPB, a four layer 70 mil thick board made of G-10 material. The top side of the circuit board has 200 mil long 50Ω microstrip lines and the bottom side has plated through vias, both corresponding to each of the 40 "pinouts". The two middle layers are used for dc voltage distribution only and are not part of the delay Tau. For the sake of soldering convenience Tau was implemented by a piece of wire-wrap wire, 30 awg, and was soldered on the bottom side of the board between the Qinverted output (pin 22) and CLR input (pin 29) plated-through holes. Refer to the circuit photo. The board was subsequently re-built on a 1 oz. copper double sided 63 mil thick epoxy (FR-4) circuit board with no change in performance.



TOP

FIGURE 4



BOTTOM

FIGURE 5

First consider the inherent time delays, the significant ones are summarized as follows:

(1) Assume 10 picosecs delay at each end of the wire due to the path from the LCC through the via to plated-through hole on board bottom. This is 20 picosecs of delay.

(2) The 200 mil long 50Ω line on top of the board is not used or terminated and thus is a capacative open stub driven by a low impedance source. Assuming 2 pf and 8Ω:

One time constant = (8Ω)X(2 E-12pf) = 16 picosecs

(3) The largest inherent delay is due to a combination of the shunt capacitance of the CLR input port and the transmission line that is Tau. A piece of 30 AWG wire laying flush to the board bottom has a characteristic impedance of about 150Ω and the CLR port input capacitance is estimated at 1.5 pf.

One time constant = (150Ω)X(1.5E-12 pf) = 225 picosecs

The following pertains to (2) & (3). One RC time constant is 0.63 of final voltage. Considering that the voltage levels are -0.6V for logic state 1 and -1.7 for logic state 0 :

-0.6V - (0.63)X(1.1V) = -1.29V this is close to the threshold point, -1.3V, of an ECL circuit

Therefore one time constant is equal to the time it takes an ECL level voltage swing to go from a logic 1 state to the threshold voltage. The CLR input is active low, therefore the clear command is initiated at threshold. Summing the time delays due to (1), (2) & (3) yields:

20 + 16 + 225 = 261 picosecs

This is the calculated value of total inherent delay over which the designer has little control. The designer can chose the physical length of line, this is added to the inherent delay. A 0.7 inch long piece of wire-wrap wire was chosen, this causes the following delay:

(0.7in)+(39.3 in/m) = 0.018 m
(0.018 m)+(3 E8 m/s) = 60 picosecs

Therefore the total calculated delay Tau is 261+ 60 = 321 picoseconds. This value is close to the midpoint of the calculated 140 to 496 picosecond range required for N = 3.

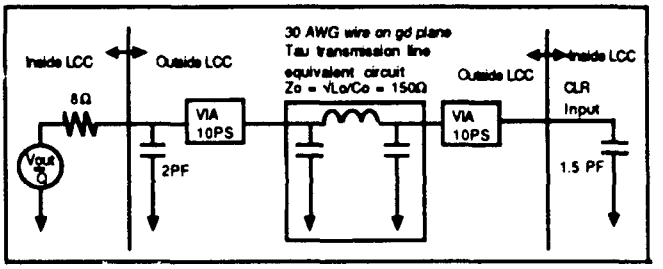


FIGURE 6

Calculating Tau is the only design task other than choosing pull-down resistors and decoupling capacitors, these are standard ecl and GaAs design considerations; thus they merit no further mention here other than to note that the 10G021A-2L accepts and puts out ecl compatible voltage levels.

MEASURED RESULTS

From a circuit design perspective the circuit was an instant success. The delay Tau did not need to be adjusted. The preceding design procedure was followed and the only required re-work was the replacement of a leaded 330 pf chip cap with a non-leaded one functioning as the Q output dc block. The leaded capacitor induced low-level spurious close to the output frequency; a chip capacitor solved that problem. In this article spurious frequencies are defined as frequencies not harmonically or subharmonically related to the input frequency.

The features and measured results are summarized as follows, refer to spectrum analyzer and phase noise plots. The 1.4 GHz source was the HP 8642B, a spectrally pure synthesized signal source, the spectrum analyzer was the HP 8566B and the phase noise was measured on the HP3047 system:

INPUT FREQUENCY (F) OF INTEREST: 1.4 GHz

DESIRED DIVIDE RATIO: N = 3

INPUT FREQUENCY RANGE: (for Tau fixed for N = 3)

1210 MHz < Fin < 1810 MHz	N = 3
Fin > 1810 MHz	No output frequency
1180 MHz < F < 1200 MHz	N = 5 & 7 (unstable)
590 MHz < F < 1170 MHz	N = 2

Note that max operating freq, 1810 MHz, is about 0.67 of specified max clock freq. Also note that as the input frequency is decreased the divide ratio decreases, as predicted by the design equation.

OUTPUT HARMONICS: (c = +7 dBm @ 466.67 MHz)

2d	(933.33 MHz)	-9 dBc
3d	(1400 MHz, input)	-22 dBc
4th	(1866.67 MHz)	-16 dBc
5th	(2333.33 MHz)	-20 dBc
19th	(8.86 GHz)	-70 dBc

Note that the 10G021A-2L puts out an ECL compatible waveform with high odd harmonic content due to the 150 picosecs rise and fall times.

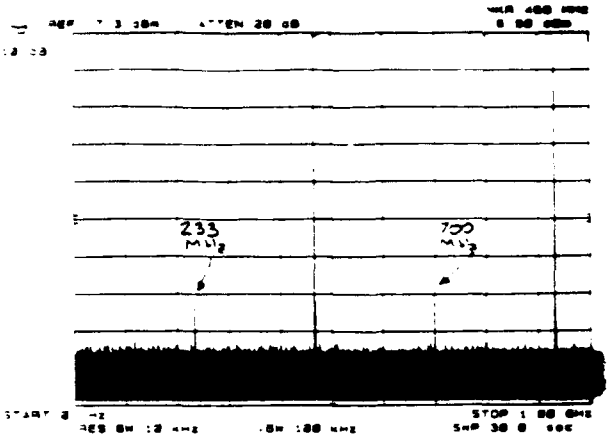
INPUT (F) SUBHARMONICS: (other than integer multiples of F/3)

F/6	(233.33 MHz)	-71 dBc
F/2	(700 MHz)	-69 dBc

No others greater than the spectrum analyzer noise floor, about -85 dBc

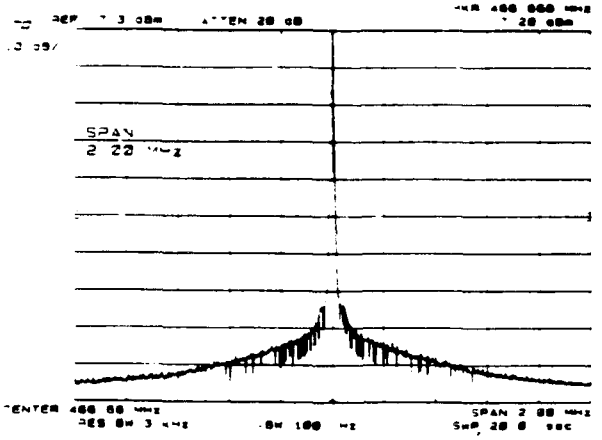
SPURIOUS:

None greater than the spectrum analyzer noise floor, about -85dBc. The phase noise measurement revealed a spur -110 dBc at 600 KHz offset at 1.4 GHz input and spurs -95 TO -100 dBc around 20 MHz offset at 1.7 GHz input.



DC TO 1 GHz

FIGURE 7



466.7 MHz +/- 1 MHz

FIGURE 8

RESIDUAL PHASE NOISE:

In this article residual phase noise is defined as the phase noise generated by any circuit that has a frequency input. These circuits include frequency synthesizers, dividers and mulitpliers, amplifiers, and rf switches. The output phase noise of these devices is the logarithmic sum of the rf source input phase noise and the device generated (residual) phase noise. Since the rf source input phase noise usually predominates over the device generated phase noise, the device generated phase noise is classified as residual. This distinction between phase noise and residual phase noise is not only important for academic clarity, each require a different measurement technique. The residual phase noise measurement is inherently indirect; it involves making two phase noise measurements, one with and one without the circuit-under-test and observing the degradation caused by

the circuit-under-test. Refer to the hp 3047 operating manual and related hp application notes for further details.

The residual phase noise of the divide-by-3s was 10 dB or greater than the measurement system phase noise for most of the 10 Hz to 40 MHz offset span; consequently the measurement system phase noise has negligible effect on the divider residual phase noise. One can conclude that the phase noise plot of the measurements with the dividers is due to the divider residual phase noise only, the measurement system phase noise contribution is negligible. Note that the residual phase noise of each divider is assumed to be 3 dB lower than measured since there are two dividers in the measurement. It is reasonable to assume that both contribute equally to the measured value, hence the 3 dB reduction.

Presented are some observations of the measurent results:

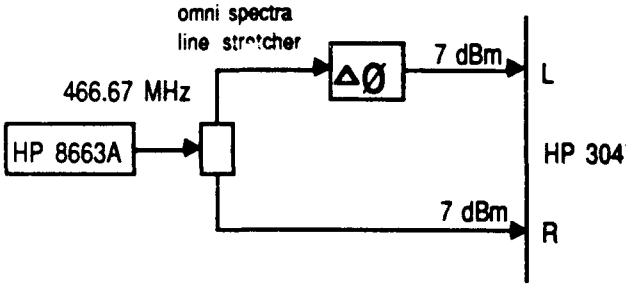
(1) Far-out floor level of about -160 dBc/Hz @ 1.4 GHz input

(2) 1/f (10 dB/decade) slope out to the far-out floor, about 4 MHz offset @ 1.4 GHz input

(3) 10 dB/decade slope increases to about 13 dB/decade below 10 KHz. This increase from 10 dB to 13 dB, theoretically unexplainable, is presumed to be a measurement inaccuracy. It is possible that the resolution bandwidth was wide enough to not detect the ever-present discrete 60 Hz harmonics, consequently the discrete power of these harmonics was "lumped in" with the divider residual phase noise. This statement is an assumption.

(4) The residual phase noise across the entire offset frequency span is about 3 dB lower at 1.4 GHz than at 1.7 GHz

MEASUREMENT SYSTEM PHASE NOISE SET-UP AND RESULTS



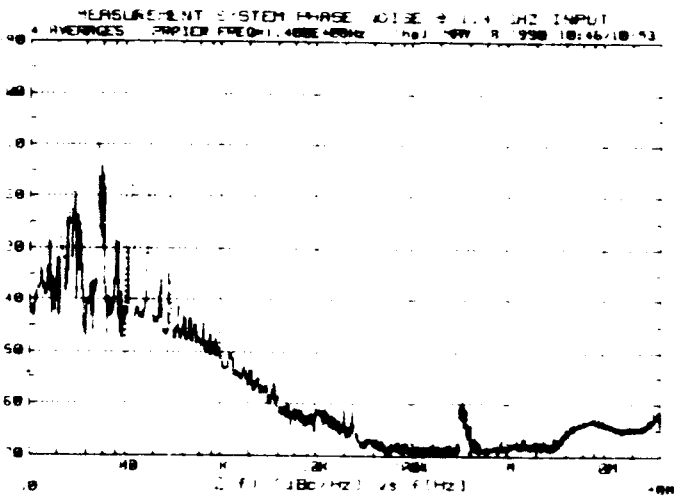
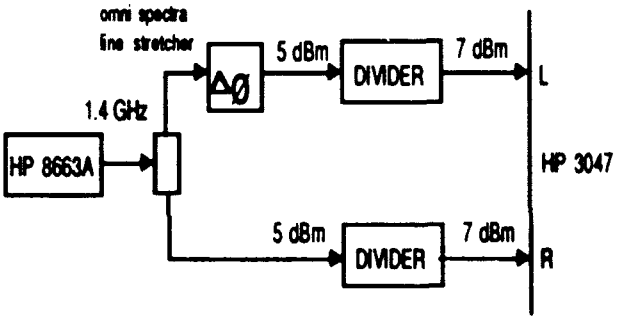
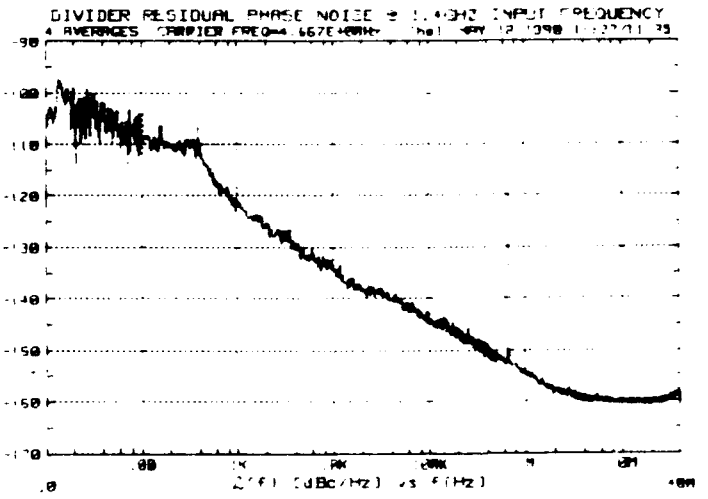


FIGURE 9

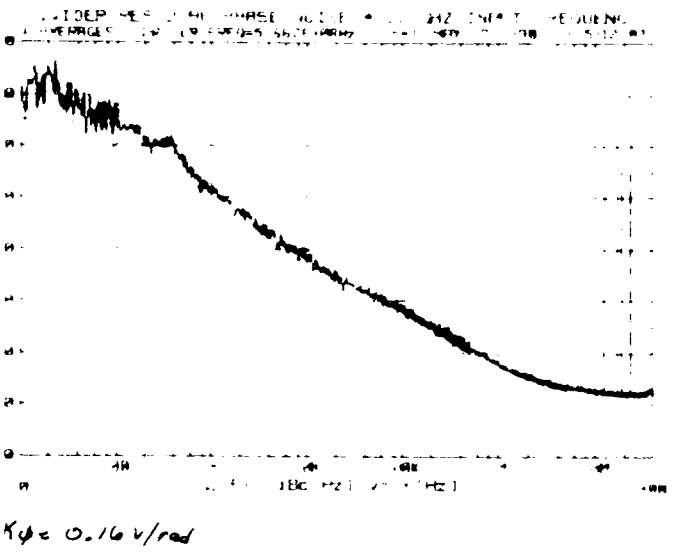
DIVIDER RESIDUAL PHASE NOISE SET-UP AND RESULTS FOR 1.4 GHZ AND 1.7 GHZ INPUTS





$K_\phi = 0.158\ V/rad$

FIGURE 10



$K_\phi \approx 0.16$ v/rad

FIGURE 11

RF INPUT POWER:

+3 dBm minimum
+17 dBm maximum (-4.0 to +0.5V @ 50Ω is 10G021A max)
+5 dBm to +10dBm (about 2.0V pk-to-pk) recommended

OUTPUT POWER INTO 50Ω: +5 to +7 dBm over frequency, temperature and from unit to unit

DC POWER:

-5.2V @ -50mA = 0.26 watts
-3.4V @ -110ma = 0.374 watts
total = 0.634 watts

SIZE: occupies about 0.80 X 0.80 X 0.20 excluding sma connectors

Note that since the 10G021A is dual D flip-flop, a second separate divider can be built with the same lcc resulting in little if any increase in these dimensions and dc power consumption.

OPERATING TEMPERATURE:

All measurements except spectrum analyzer measurements were made at room temperature. The spectrum analyzer measurents consisted of measuring the output power, spurious frequencies and input frequency range. These measurements were made at -55 C, +25C and +85C. There were no spurs and the only degradations at -55 C and +85 C were a slight decrease in output power and input frequency range. The D flip-flop used was specified from 0 C < Tcase < +85 C and is available from -55 C < Tcase < +125 C.

TIME DOMAIN:

A time domain measurement of the output using the HP 54111D digitizing oscilloscope revealed a periodic non-sinusoidal waveform that had the characteristics of a wave rich in harmonics with a 3.4V pk-to-pk output @ 1MΩ impedance and 1.4V pk-to-pk output @ 50Ω impedance with a duty cycle close to 50%. It is omitted from this paper since it reveals very little useful information.

SUMMARY

The most attractive features of this circuit from a frequency synthesis perspective are the extremely low-level spurious frequencies, moderate residual phase noise, close correlation of predicted delay Tau to actual value necessary to realize a desired divide ratio N and circuit simplicity from both theoretical and practical perspectives. The authors would like to thank Sandra Hosley for assembling the circuit, Ike Mathis for generating the drawings and Bill Webster & Jerry Phillips for general support and encouragement of this effort.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

ULTRA-LOW NOISE MICROWAVE PHASE STABILIZER USING SAPPHIRE RING RESONATOR*

G. J. Dick, J. Saunders and T. Tucker

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

Abstract

We present design details for two sapphire "whispering gallery" mode phase stabilizers presently under construction. The sapphire resonators are cooled in individual liquid nitrogen dewars to an operating temperature of approximately 80 K and each operates to stabilize the phase of a 8.1 GHz signal derived from a quartz crystal oscillator operating at 100 MHz. Ultra-low phase noise (-80 dB/Hz at 1 Hz offset) is projected due to the high resonator Q ($3 \cdot 10^7$ at 80 K). The temperature of the thermally uncompensated sapphire resonators is stabilized by RF heating to a value approximately 3 degrees above that of a liquid nitrogen bath. Additionally, we have developed and analyzed novel oscillator configurations which can further reduce oscillator phase noise. These circuit improvements promise crystal oscillator type performance in a room temperature 10 GHz (X-band) oscillator using a whispering gallery mode sapphire resonator with an intrinsic Q of $2 \cdot 10^5$.

Introduction

Phase noise measurements were previously reported for an 8 GHz (X-band) oscillator stabilized by a room temperature sapphire "whispering gallery" mode resonator[1,2]. Using off-the-shelf components and operating at room temperature, these tests showed noise of -23 dB/Hz at an offset frequency of 1 Hz, a value lower than the best previously reported for a non-cryogenic X-band oscillator[3-7]. On the basis of the results of these measurements and on previous studies of sapphire resonator Q improvement at low temperature, we have designed and are presently constructing two microwave oscillator systems which are designed for ultra-low noise with sapphire resonators cooled by liquid nitrogen. This paper contains a general description of the oscillator systems and the thermal design considerations together with design details and analysis for the microwave aspects. New microwave circuitry is presented which promises further reduction in noise for STAbilized Local Oscillator (STALO) configurations incorporating both room-temperature and cooled "whispering gallery" sapphire resonators.

The dielectric loss in sapphire is rapidly reduced as the temperature is cooled below ambient, showing an approximate T^{-5} dependence on the absolute operating temperature T down to temperatures as low as 60 Kelvin[2,8-10]. While metallic losses limit the achievable Q values in most microwave resonators, recently developed "whispering gallery" mode sapphire resonators isolate the resonant microwave energy to the sapphire itself, and so allow the inherent Q value of the sapphire itself to be realized[2,8-12]. Q values at 8-10 GHz (X-band) of $2 \cdot 10^5$ have been reported at room temperature, increasing to $3 \cdot 10^7$ at 80 K.

Phase noise performance for RF and microwave cavity and DRO oscillators at small offsets ($f < 1000$ Hz) is determined by resonator Q and by the 1/f noise level in the active device. The use of a stabilized local oscillator 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 ($\approx -135dB$

*The work described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

CH2818-3/90/0000-577 $1.00 © 1990 IEEE

Figure 1: Cryogenic components for the two sapphire resonator subsystems.

at 1Hz offset frequency) equals that of active devices at lower frequencies and so essentially eliminates the disadvantage of high frequency operation[3 6]. Together with a projected loaded Q of $Q_l = 1.5 \cdot 10^7$ this mixer performance makes possible oscillator performance of -80 dB/Hz/f^3 (Hz)$^{-3}$ at 8 GHz. This performance improves by 20 30 dB the best that can be attained by the use of frequency multiplied crystal quartz oscillators.

The thermal coefficient of frequency of the sapphire resonators has also been studied, showing a value of $5 \cdot 10^{-5}$ at room temperature reducing to $3 \cdot 10^{-6}$ at 80 K[10]. This very substantial temperature dependence reduces their applicability in this temperature range as frequency stabilizing elements. However, application as a phase stabilizer is not adversely effected for fluctuation frequencies above 1 Hz, where the thermal mass of the sapphire resonator effectively stabilizes its operating frequency.

Further improvement on the phase noise performance of sapphire oscillators, both at room and cryogenic temperatures, appears possible by use of novel circuitry. We have developed and analyzed oscillator configurations which reduce the effect of semiconductor 1/f phase noise sources for stabilized local oscillator circuits. By appropriate use of carrier suppression, a small signal is generated which suffers no loss of loop phase information or signal-to-noise ratio. This small signal can be amplified without degradation by multiplicative amplifier noise, and can be detected without saturation of the detector. These circuit improvements promise crystal oscillator type performance in a room temperature 10 GHz (X band) oscillator using a whispering gallery mode sapphire resonator with an intrinsic Q of $2 \cdot 10^5$, and promise dramatic improvements in the state of the art with cooled resonators and higher Q's.

Figure 2: Sapphire resonator (5 cm diameter) in copper containing can.

Technical

A photograph showing cryogenic components for the two sapphire resonator subsystems is shown in Figure 1. The dewars are commercial units with nominal capacity of 13.6 liters of liquid nitrogen and a hold time of 2 days. The stainless steel vacuum insert and bottom plate are joined by a demountable indium seal. The resonator assembly is cooled by means of a copper heat transfer plate which extends through the bottom plate to make contact directly with the liquid nitrogen.

A sapphire resonator is shown in its copper containing can in Figure 2. Dimensions of the resonator and can are approximately the same as previously reported[2]. We have chosen to use the $WGH_{10,0,0}$ mode of the sapphire wheel resonator which has a resonant frequency of 7.85 GHz at room temperature and 7.95 GHz at 77 K. Tuning to the 8.100 GHz

operating frequency will be accomplished by grinding the flat faces of the resonators to raise the resonant frequency as needed. As previously reported, critical coupling to the $WGH_{10,0,0}$ mode was obtained at room temperature with a WR102 waveguide port directly through the center of the barrel of the containing can. Since such a coupling port would be greatly overcoupled with the higher cryogenic resonator Q, the coupling was weakened by use of WR62 waveguide ports. These ports are operated below their cutoff frequency in order to weaken the coupling appropriate to the 10^7 operating Q. Dielectric inserts in the ports (not shown) are used to fine-tune the coupling.

Thermal Design

Thermal design of a frequency source using uncompensated sapphire resonators is complicated by the large thermal coefficient of frequency and by the very high resonator Q. For example, for an operating Q given by $Q = 10^7$ and a thermal coefficient of fractional frequency variation given by $k_t = 3 \cdot 10^{-6}$/K, a temperature variation of $\Delta T = (Qk_t)^{-1} = .03$ K is sufficient to detune the resonator by one bandwidth. More specifically, the effective thermal conductance due to the change in RF heating with temperature, calculated from the frequency response of the resonator itself, has a maximum value;

$$|K_{rf}| \leq K_{max} = \frac{3\sqrt{3}P_m Q_l k_t}{4} \qquad (1)$$

where P_m is the power absorbed by the resonator at resonance and Q_l is the loaded Q of the resonator. The conductance K_{rf} may be either positive or negative, depending on which side of the resonance the frequency lies. For the case of a single critically coupled port, P_m is equal to the applied power P_o and Q_l is just half the unloaded resonator Q value Q_o. This gives a value for the maximum conductivity of

$$K_{max}|_{cc} = \frac{3\sqrt{3}P_o Q_o k_t}{8} \qquad (2)$$

for critical coupling. For our design values of P_o = 0.010 Watts, $k_t = 3 \cdot 10^{-6}$/Kelvin and $Q_o = 3 \cdot 10^7$, the maximum conductivity given by Eq. 2 is 0.585 Watt/Kelvin. This very large RF-induced conductivity may be either stabilizing or destabilizing, depending on whether the temperature is above or below the nominal value at which the resonant frequency matches that of the local oscillator. A schematic diagram showing thermal aspects of the resonator assembly is shown in Figure 3. Thermal time constants may be calculated for various parts of the resonator assembly from their heat capacities and thermal conductivities. Thermal conductivity of sapphire has been measured to be approximately 8 W/cm/Kelvin at 80 K, and we calculate a heat capacity of 0.0102 J/Kelvin/gram based on the room temperature of 0.174 J/Kelvin/gram and a measured Debye temperature for sapphire of 1047 K[13]. Based on these values, heat capacities of the "ring" and "disk" parts of the sapphire wheel shown in Fig. 3 are 4.2 J/K and 0.1 J/K, respectively, while their characteristic conductivities are approximately given by 70 W/K and 14 W/K. The conductivity of the stainless steel mounting post is only 0.01 W/K, as determined by the design operating temperature of 80 K and power dissipation of 30 mW. From these values thermal time constants can be calculated as 0.06 second for the ring, .3 second for the entire wheel anchored at the center, and 420 seconds for the resonator as cooled by the mounting post, and ≥ 7 seconds as stabilized by the RF induced conductivity discussed above.

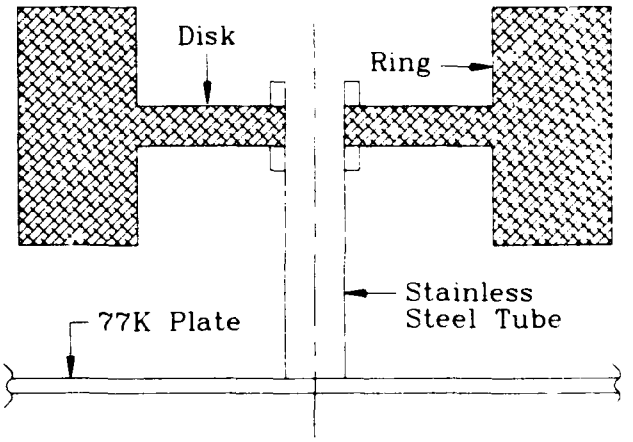


Figure 3: Schematic diagram showing thermal aspects of the sapphire resonator and its mounting post.

Operating procedure for the phase stabilizer will involve heating the sapphire to a temperature above 80 K by means of a heater at its center. Lowering the heater power will then allow the resonator to cool with a characteristic time constant of 420 seconds, and when the resonance condition is reached at 80 K, the resonator will self-stabilize with a time constant of ≥ 7 seconds by RF heating. This relatively long time constant allows undiminished phase stabilizer gain for offset frequencies $f \geq 1$ Hz.

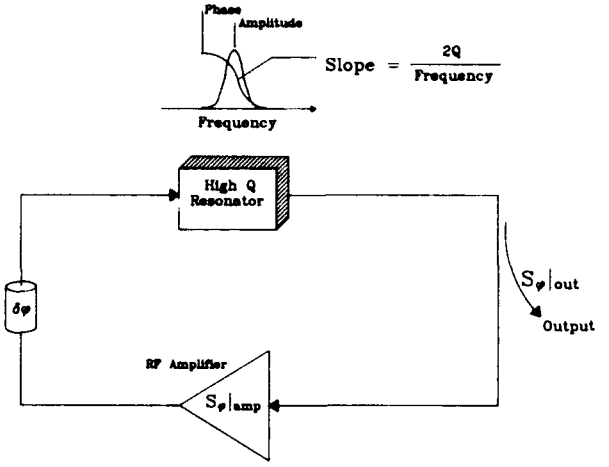


Figure 4: Block diagram of simple oscillator with direct RF feedback. Output phase noise is derived from amplifier noise together with phase slope of resonator. Phase is adjusted to give $2n\pi$ radians around loop at center of sapphire resonator passband.

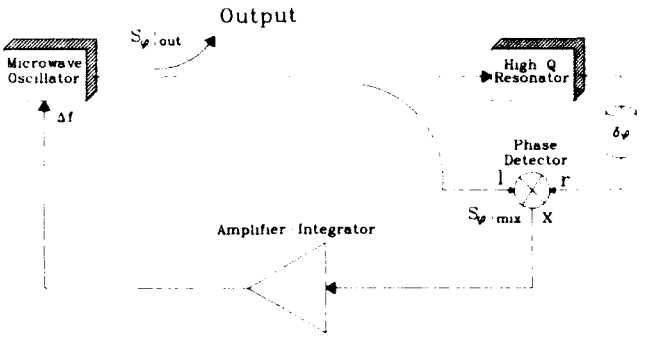


Figure 5: Block diagram of stabilized local oscillator (STALO) with double-balanced mixer type phase detector. Mixer noise plays the same role as amplifier noise in Figure 4. Phase is adjusted to give l and r signals in quadrature.

Oscillator Circuit Design

The units under construction are designed to be used in a frequency feedback configuration, stabilizing the phase of a frequency multiplied 100 MHz crystal quartz VCO with a high-pass feedback filter. The quartz oscillator allows a relatively high (10 kHz) modulation rate which makes possible 50 dB of loop gain in a frequency feedback configuration at an offset frequency greater than 100 Hz. Phase feedback configurations are also being studied. The containing can for the sapphire resonator has been redesigned for reduced RF coupling appropriate for higher (cooled) sapphire Q.

Figures 4 and 5 show conventional microwave self-excited oscillator and STALO configurations together with an identification of the in-oscillator and oscillator output noise spectral densities. In the self-excited oscillator shown in Figure 4, 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

$$S_y(f)|_{out} = \frac{S_\phi(f)|_{amp}}{(2Q)^2} \tag{3}$$

or the mathematically equivalent output phase fluctuations

$$S_\phi(f)|_{out} = \left(\frac{\nu}{f}\right)^2 \frac{S_\phi(f)|_{amp}}{(2Q)^2}. \tag{4}$$

where $y \equiv \delta\nu/\nu$ is the fractional frequency deviation.

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. The phase from the reference loop is adjusted to produce the proper sign of the correction voltage and attain maximum sensitivity by operating in the high slope region of the mixer output vs. reference phase relation. 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). A significant advantage of the STALO is that the properties of the feedback loop are particularly easy to control since the signal is mixed down to baseband (near zero frequency). This allows the use of active filters with narrow bandwidths and sophisticated response shapes which are not possible at microwave frequencies. A second advantage is that the $1/f$ noise for X-band mixers (-135dB/f per Hz

at 10 GHz) [4] is better than that which is available from the best amplifiers (-110 to -120 dB/f) [3-7]. The analysis from the self-excited oscillator can be adapted to the STALO by noting that the only difference is that the phase detection and correction has been moved from the resonator in Figure 4 to the mixer-amplifier-oscillator combination in Figure 5. The phase noise of the RF amplifier of Figure 4 is replaced by that of a mixer. Consequently, the output phase fluctuations are described by

$$S_\phi(f)|_{out} = \left(\frac{\nu}{f}\right)^2 \frac{S_\phi(f)|_{mix}}{(2Q)^2}. \qquad (5)$$

As a consequence, the performance of a direct feedback oscillator with an amplifier having $1/f$ noise of -120 dB/Hz at 1 Hz ($S_\phi(f)|_{amp} = 10^{-12}/f$ radians^2/Hz) is

$$S_\phi|_{out} = \frac{10^{-12}}{(2Q)^2} \cdot \frac{\nu^2}{f^3} \qquad (6)$$

while a STALO using a mixer with $S_\phi(f)|_{mix} = 10^{-13.5}/f$ radians^2/Hz will be 15 dB quieter. These device noise levels represent the quietest components presently available, giving a clear advantage to the STALO configuration.

STALO Configurations for Reduced Noise

Two new STALO configurations are discussed in this section that reduce the effect of mixer noise by increasing its sensitivity by the use of a suppressed-carrier signal. Of these two, the first achieves increased sensitivity without added noise by use of a low level RF amplifier. In the second configuration, increased sensitivity is achieved by operating the high Q resonator at higher power than would otherwise be possible due to mixer power limitations. The second configuration is similar to a method used by Walls[4].

The STALO shown in Figure 6 forms the basis for the new designs. It differs in implementation from Figure 5 in that the signal from the cavity to the mixer is not taken from a second coupling port but is instead taken from the signal reflected from the input port. A circulator separates this signal from the forward driving signal. At critical coupling and on resonance the returned signal is identically zero. However, it is the superposition of two equal signals, one of which emanates from the cavity and a second,

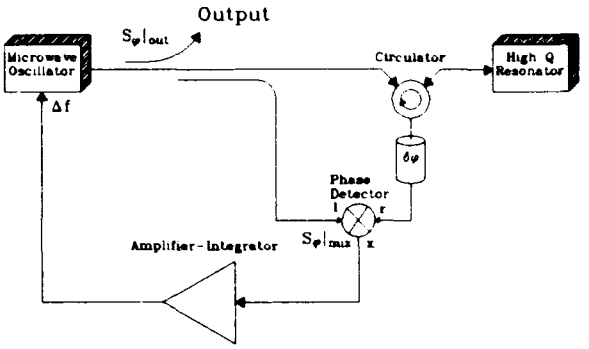


Figure 6: STALO configuration with one-port resonator operation but functionally identical to that shown in Figure 5. Signal returned from resonator is superposition of resonator signal and (constant) reflected signal.

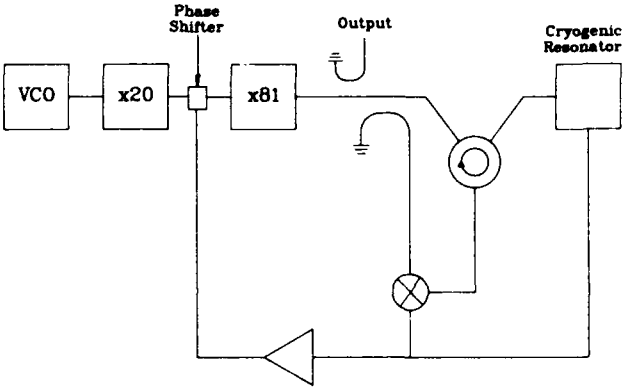


Figure 7: STALO configuration with phase feedback and one-port resonator operation.

reflected, signal which is derived from the driving signal, with a constant phase shift. This reflected signal does not significantly affect the operation of the mixer at resonance since it is in quadrature with the signal at the other mixer port. Thus the two STALOs will have approximately identical performance. While the amplitude goes through zero on resonance, a phase reversal takes place in which the in-phase signal on one side becomes out-of-phase on the other, allowing a linear dependence of mixer output voltage on the frequency error, as required for effective feedback. Instead of viewing the mixer as a phase detector, it is seen as projecting the component of the signal at the "r" input onto the phase of that at the "I" input.

Figure 7 shows a STALO configuration with phase feedback instead of the frequency feedback shown

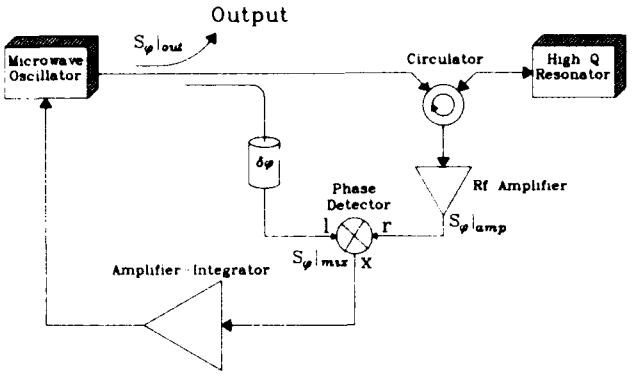


Figure 8: STALO configuration for reduced phase noise. Critical coupling to resonator and operation very near ν_0 allow insensitivity to phase noise of amplifier. Amplifier gain allows reduction of phase detector noise.

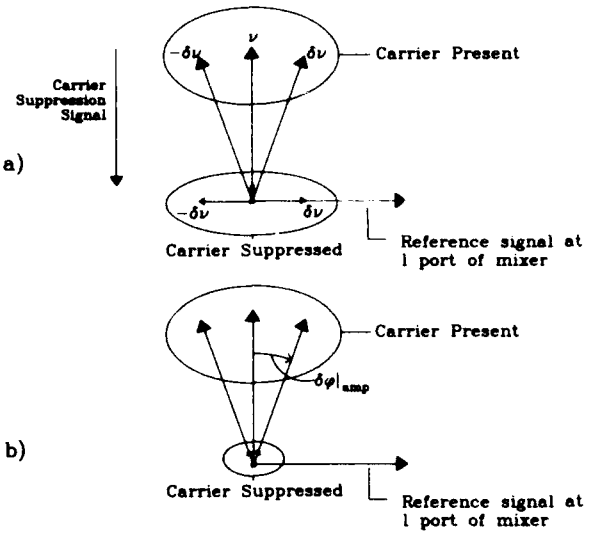


Figure 9: (a) Phasor diagram showing effect of frequency error on RF resonator signal with and without carrier suppression. Also shown are (constant) carrier suppression signal and mixer reference signal. Component of resonator signal in phase with reference is unchanged by 20 dB carrier suppression. (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.)

in Fig. 6. While requiring a somewhat different loop gain character, the two configurations are almost functionally identical. Phase feedback has the advantage of speed over frequency feedback, allowing a higher loop gain bandwidth. Disadvantages include the need for extra components and a maximum phase limit that can be compensated.

STALO Design using RF Amplification

Figure 8 shows further modification of the STALO which can result in improved performance. Here the 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 will be increased by the added gain, an effect which must be compensated for in the design of the base-band amplifier. Secondly, the gain of the amplifier will increase 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.

The third effect of this modification is a little more complicated. Since amplifiers are somewhat more noisy than mixers (-120 dB vs -135 dB at 1 Hz offset as previously discussed) a crucial point is the proper analysis of the contribution of amplifier noise. The kind of noise under discussion is not additive noise, which would be 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 in good amplifiers is insignificant except at offset frequencies $f >\approx$ 10kHz where the $1/f$ multiplicative noise is relatively small.) Figure 9(a) shows in phasor form the cavity signals with and without carrier suppression corresponding to the oscillators shown in Figures 6 and 5, respectively. The added signal due to small frequency variation in the L.O. are also shown. These added signals are detected by the mixer in order to allow feedback circuitry to cancel the frequency variations. The effect of multiplicative phase noise in the amplifier for the two cases is shown in Figure 9(b). It is clear from the figure that this noise source generates signals which are indistinguishable from those caused by actual frequency variations and which are due only to the presence of the coherent carrier. Thus a reduction in amplifier noise is effected which is proportional to the degree of carrier suppression of the microwave signal at its input.

Oscillator phase noise is thus 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 $-135\text{dB}/f/\text{Hz}$ were reduced by 25 dB of amplifier gain to $-160\text{dB}/f/\text{Hz}$, and amplifier noise of $-120\text{dB}/f/\text{Hz}$ were reduced by 40 dB of carrier suppression to the same value, the combined noise of $-157\text{dB}/f/\text{Hz}$ would determine oscillator performance. For a loaded Q of 10,000 (intrinsic $Q = 20{,}000$) this would allow an oscillator phase noise of $S_\phi(f)|_{osc} = -43\text{dB}/f^3/\text{Hz}$, a value superior to any room temperature microwave oscillator to date. For room temperature and thermoelectrically cooled sapphire resonators with Q's of 10^5 and 10^6, performance would be superior to that of any available source at $S_\phi(f)|_{osc} = -63\text{dB}/f^3/\text{Hz}$ and $-83\text{dB}/f^3/\text{Hz}$, respectively.

STALO Design with High Power Resonator Operation

In this configuration, enhanced sensitivity in the phase detector is achieved by means of (relatively) high power in the high Q resonator. It has the advantage of simplicity and absence of any amplifier to introduce added phase noise if carrier suppression is incomplete. A disadvantage is that power limitations in the microwave source or high Q resonator restrict the available improvement factor.

As previously discussed and as shown in Figure 9(a), suppression of the carrier at the r port of the mixer in Figure 6 has only incidental consequence regarding mixer sensitivity, since the suppressed part of the signal is in quadrature with the reference signal at the l port. (The part of the signal due to frequency variations, $\pm\delta$, is in phase with the reference and so is detected in any case.) However, things are not quite identical to the conventional STALO shown in Figure 5. Suppression of the carrier at the r port allows the power to the high Q resonator to be increased without saturating the mixer. This increased power results in an enhanced sensitivity of the mixer output voltage to frequency variations, $\pm\delta$.

Figure 10 describes such a circumstance. Besides the increased power levels in oscillator and resonator, the only difference from Figure 6 is an appropriately weaker coupling to the mixer's l port. Mixers typically saturate at signal levels on the order of 20 milliwatts, while frequency sources and resonators can operate at power levels up to one Watt or even higher. The resultant increase in sensitivity of up to 17dB reduces the consequence of mixer noise by the same factor. For a mixer with flicker noise of

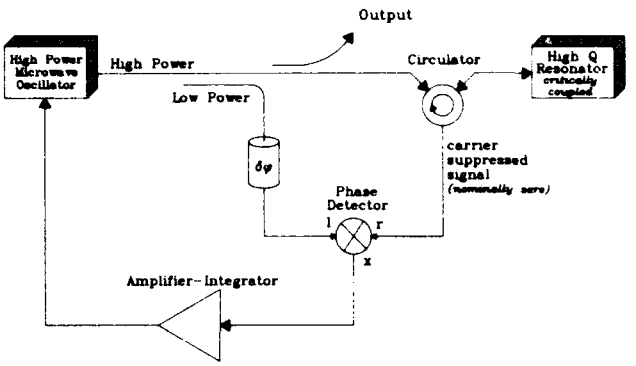


Figure 10: Configuration to allow increased power into the high Q resonator without saturating the mixer. The resultant increase in sensitivity of up to 17 dB reduces the consequence of mixer noise by the same factor.

-135dB/Hz at 1 Hz offset, the effective noise could be reduced to a value of -152dB/Hz.

Performance Aspects

Typical low-noise 10 GHz (X-band) oscillators use a single transistor for excitation, while a more elaborate STALO configuration is used for the lowest possible phase noise. With the development of sapphire "whispering gallery" mode resonators with Q's above 10^5 to 10^7 at 10 GHz (X-band), the possibilities have been raised considerably. Where lower frequency SAW or BAW quartz crystal oscillators previously had far lower noise than their higher frequency counterparts, they are now rivalled by an X-band oscillator using the sapphire resonator. Together with improved oscillator circuits, this resonator may make possible close-in phase noise lower than that of any non-cryogenic frequency source. Furthermore, cooling by means of thermoelectric coolers or liquid nitrogen may make practical frequency sources with greatly reduced phase noise.

New design configurations for stabilized local oscillators presented here allow reduced phase noise in comparison to conventional configurations. By appropriate use of carrier suppression, a small signal is generated which suffers no loss of loop phase information or signal-to-noise ratio. This small signal can be amplified without degradation by multiplicative amplifier noise, and can be detected without saturation of the detector.

Figure 11 shows phase noise calculations for conventional and improved sapphire "whispering

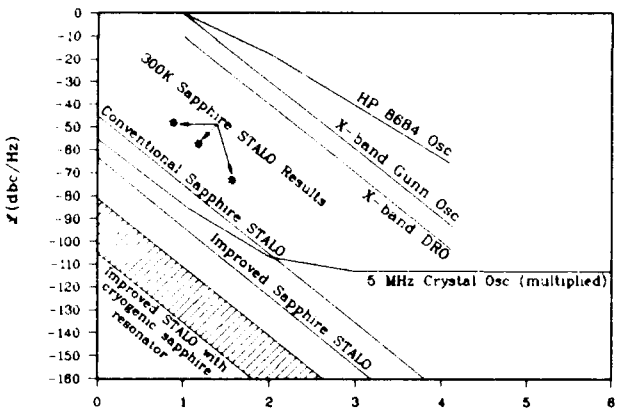


Figure 11: Phase noise calculations for sapphire "whispering gallery" mode STALO sources shown in Figures 8 and 6. Noise plots for various conventional 10 GHz (X-band) frequency sources are also shown. The multiplied 5 MHz crystal oscillator presently represents the best performance available at 10 GHz (X-band).

gallery" mode STALO configurations as shown in Figures 6 and 8. Intrinsic quality factors Q_i are assumed to be $Q_i = 2 \cdot 10^5$ at room temperature, and cryogenic Q's of $Q_i = 2 \cdot 10^6$ at 170 K and $Q_i = 3 \cdot 10^7$ at 77 K. Conventional STALO performance is based on mixer phase noise of -135db/Hz at 1 Hz, and improved values assume 20 dB amplifier gain with 35 dB of carrier suppression. Noise plots for various conventional 10 GHz (X-band) frequency sources are also shown.

Acknowledgements

The authors would like to express their thanks to R. T. Wang for assistance in thermal and cryogenic design, L. E. Primas for feedback loop design, and L. Maleki for many helpful questions.

References

[1] J. Dick and J. Saunders, "Measurement and Analysis of a Microwave Oscillator Stabilized by a Sapphire Dielectric Ring Resonator for Ultra-Low Noise," *Proc. 43rd Ann. Symposium on Frequency Control*, 107-114 (1989).

[2] 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).

[3] 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).

[4] F. L. Walls and C. M. Felton, "High Spectral Purity X-band Source," this conference.

[5] T. E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", *Proc. 41st Ann. Symposium on Frequency Control*, 99-110, (1987).

[6] 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).

[7] R. G. Rogers, "Theory and Design of Low Phase Noise Microwave Oscillators," *Proc. 42nd Ann. Symposium on Frequency Control*, 301-303 (1988).

[8] V. B. Braginsky, V.P. Mitrofanov and V.I. Panov, *Systems with Small Dissipation* (Univ. of Chicago Press, Chicago, 1985), 85-89.

[9] 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).

[10] 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).

[11] 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).

[12] X. H. Jiao, P. Guillon, and L. A. Bermudez, "Resonant frequencies of whispering-gallery dielectric resonator modes", *IEE Proceedings* **134**, Pt. H, 497 (1987).

[13] C. Kittel, *Introduction to Solid State Physics*, (John Wiley & Sons, Inc., New York, 1962) p. 144.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

An Analysis of Oscillation Frequency Characteristics in a CMOS Oscillating Circuit Using a Coupling Quartz Crystal Resonator

Hirofumi KAWASHIMA

Seiko Electronic Components Ltd.
Hiraicho Tochigi-shi 328, Japan

Abstract

This paper describes oscillation frequency characteristics in a CMOS oscillating circuit using a coupling quartz crystal resonator, particularly, a miniaturized GT cut quartz crystal resonator consisting of the vibrational portion and the supporting portions, formed by an etching method. An object of this paper is to clarify a relationship of its resonator's frequency temperature behavior versus load capacitance CL. In analysis procedure, first, since a CMOS inverter which is used in constructing an oscillating circuit can be linearly expressed by utilizing an electric current source and drain conductance in a small signal, the oscillating circuit is shown by an electrical equivalent circuit.

Next, by applying Kirchhoff's law to each loop of the equivalent circuit, an amplitude continuation condition and an oscillation condition are, theoretically, derived, and then, a coupling between both vibrations of a coupling quartz crystal resonator can be regarded as a capacitance coupling, when expressed by an electrical equivalent circuit. From the equivalent circuit of the coupling quartz crystal resonator, an imaginary part is, therefore, easily calculated. As a result, a frequency equation which is given as a function of load capacitance CL, is readily derived from the oscillation condition and the imaginary part of the coupling resonator, so that, by choosing very suitably CL, it is shown that frequency deviation in this CMOS oscillating circuit using the present resonator is less than 1 ppm over a wide temperature range of -30℃ to +70℃ without any temperature compensation.

§1 Introduction

A quartz crystal oscillator which is used in communication equipment and consumer products as a frequency standard, is excellent in frequency temperature characteristics and simultaneously, a further miniaturized oscillator with low consumption power is keenly required. However, because the conventional high precision oscillator needs temperature compensation by thermister and so on, as things are, the electric current consumption is large and its miniaturization is very difficult. A new oscillating circuit has been, therefore, required.

As an oscillating circuit satisfying these requirements, a CMOS oscillating circuit using a GT cut quartz crystal resonator [1]-[3] with the supporting portions at both ends, formed by an etching method is optimum, because its resonator is of comparatively low frequency of 1 to 4 MHz, miniaturized and shows the most excellent frequency temperature characteristics of all quartz crystal resonators, namely, temperature compensation by thermister and so on is unnecessary.

However, amplitude continuation characteristics and oscillation frequency ones of the CMOS oscillating circuit using a coupling quartz crystal resonator are not, theoretically, clarified at all. An object of this paper is, therefore, to clarify these characteristics.

Next, in analysis procedure, first, since a CMOS inverter can be linearly expressed by using a electric current source and drain conductance, an amplitude continuation condition and an oscillation condition of the circuit are, theoretically, derived from this expression and Kirchhoff's low. Moreover, a coupling between both vibrations of a coupling quartz crystal resonator can be regarded as capacitance coupling [5] [6] in an electrical equivalent circuit and the resonator is given as a summation of a real part and an imaginary part. Accordingly, a frequency equation to give resonant frequency is easily calculated from the imaginary part and the oscillation condition. In still more detail, the resonant frequency is given as a function of coupling factor Km and load capacitance CL which is shown by gate capacitance Cg and drain capacitance Cd including stray capacitance.

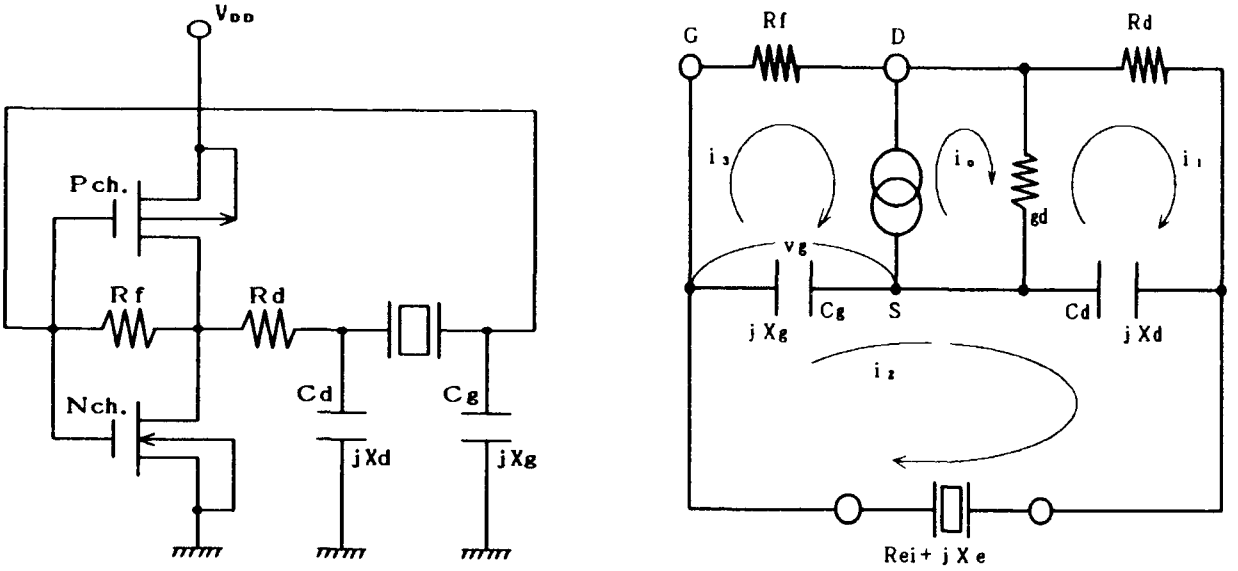
Therefore, a change of the resonant frequency versus CL values and the first, second and third order frequency temperature coefficients by differentiating the resonant frequency with temperature are calculated respectively. These results are then compared with the measured data. As a result, it is shown that, by a selection of the optimum CL value, the CMOS oscillating circuit with frequency deviation less than 1 ppm over a wide temperature range of -30℃ to +70℃, namely extremely stable versus temperature, is substantially obtained. Let us describe concretely below.

§2 Analysis Procedure

2.1 Construction of CMOS quartz crystal oscillating circuit

Figure 1 shows a construction diagram(a) of a CMOS quartz crystal oscillating circuit and its electrical circuit(b). The circuit(a) comprises a CMOS inverter, drain capacitance Cd, gate capacitance Cg, feedback resistance Rf, output resistance Rd, source voltage V_{DD} and a coupling

CH2818-3/90/0000-585 $1.00 © 1990 IEEE



(a) (b)

Fig. 1 Construction of CMOS quartz crystal oscillating circuit (a) and its electrical equivalent circuit (b)

quartz crystal resonator. Since the CMOS inverter can be linearly expressed by employing an electric current source ⓒ and drain conductance gd(=1/rd) in a small signal[4], the circuit(a) can be equivalently expressed by the circuit (b). Where an electric current i_0 of the electric current source is given as $i_0=gmvg$ by utilizing mutual conductance gm and between gate-source voltage vg, when electric currents i_1, i_2, i_3 which flow in each loop, are taken. In addition, a coupling quartz crystal resonator is electrically 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). Accordingly, the oscillation condition and the amplitude continuation condition are calculated from the circuit (b).

2.2 Derivatives of oscillation frequency

In the circuit of Fig. 1(b), from Kirchhoff's second low by taking electric currents i_1, i_2, i_3, i_0 flowing in each loop, the following relations are obtained:

$$\left.\begin{aligned}(rd+Rd+jXd)i_1-jXd\,i_2+\mu vg&=0\\ -jXdi_1+(Rei+jXe+jXg+jXd)i_2-jXgi_3&=0\\ -jXgi_2+(Rf+jXg)i_3&=0\end{aligned}\right\}\qquad(1)$$

Where the sign μ shows an amplification factor, which is given as $\mu=gmrd$. Since the electric currents i_1, i_2, i_3 are not zero, in order to obtain a nontrivial solution in equation (1), the following relation is given as, namely, taking the relation of $vg=jXg(i_2-i_3)$ into account,

$$\begin{vmatrix} rd+Rd+jXd & j(\mu Xg-Xd) & -j\mu Xg\\ -jXd & Rei+j(Xe+Xg+Xd) & -jXg\\ 0 & -jXg & Rf+jXg\end{vmatrix}=0\qquad(2)$$

Therefore, as equation(2) is expressed by a summation of the imaginary part and the real part, calculating Xe from the imaginary part and μ from the real part, the oscillation condition and the amplitude continuation condition are, respectively, given as the following relations:

$$Xe=-\frac{1}{\dfrac{XdXg}{Rf(rd+Rd)}}(Xg+Xd)-\frac{Rei}{rd+Rd}Xd\,\frac{Rei}{Rf}Xg\qquad(3)$$

$$\mu=\frac{Xd}{Xg}+\frac{Rei(rd+Rd)}{XgXd}+\frac{rd+Rd}{Rf}\,\frac{Xg}{Xd}\,\frac{Rei}{Rf}\qquad(4)$$

Furthermore, when angular frequency ω is taken, the signs Xg and Xd are given as $Xg=-1/\omega Cg$ and $Xd=1/\omega Cd$, and substituting these relations into equations(3) and (4), moreover, taking account of Rf=Rei, the oscillation condition and the amplitude continuation condition are simplified as follows:

(Oscillation condition)

$$Xe=-\frac{1}{\omega CL}-\frac{1}{\dfrac{R_L}{Rf(\mu+gmRd)}}\qquad(5)$$

Where $\dfrac{1}{CL}=\dfrac{1}{Cg}+\dfrac{1}{Cd}\left(1+\dfrac{Rei}{rd+Rd}\right)$

$$-\frac{1}{Cg}+\frac{1}{Cd}\left(1+\frac{gmRei}{\mu+gmRd}\right)$$

$$-\frac{1}{\omega^2CgCdrd}-\frac{R_L}{\mu}\qquad(-R_L:\text{Negative resistance})$$

(Amplitude continuation condition)

$$\mu\geq\frac{Cg}{Cd}+\omega^2CgCdRei(rd+Rd)+\frac{rd+Rd}{Rf}\,\frac{Cd}{Cg}\qquad(6)$$

Therefore, by calculating reactance Xe of the quartz crystal resonator, the oscillation frequency can be derived frome quation(5). On the other hand, the amplification factor μ giving the amplitude continuation condition and negative resistance $-R_L$ of CMOS IC must be designed so as

to satisfy the relations of equations(7) and (8).

$$\mu \geq \mu_0 \quad 1/\bar{\mu} \quad (\bar{\mu}: \text{Feedback factor}) \tag{7}$$

and

$$|R_L| > Rei \tag{8}$$

Next, let us calculate the reactance Xe and the resistance Rei of a coupling quartz crystal resonator.

Figure 2 shows the electrical equivalent circuit (a) [5], [6] of a coupling quartz crystal resonator in the case of capacitance coupling and its impedance sign expression(b). Each sign shows principal vibration(L_{11}, C_{11}, R_{11}), subvibration(L_{22}, C_{22}, R_{22}), coupling capacitance Cm and shunt capacitance Co. In addition, providing assumption that electric currents i_4 and i_5 flow in each branch of the principal vibration and the subvibration when voltage v_1 between the terminals A and B is applied. Furthermore, in Fig.2 (b), the signs Z_1, Z_2 represent impedance of mechanical resonance and shunt capacitance C_0. Accordingly, the impedance Z between the terminals A, B is given as

$$Z = \frac{Z_1 Z_2}{Z_1 + Z_2} \tag{9}$$

First of all, let us calculate Z_1. In Fig.2 (a), from the electric currents i_4, i_5 flowing in each branch of the principal vibration and the subvibration, the following relation is obtained:

$$\left.\begin{aligned} R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right) i_4 - j\frac{1}{\omega Cm} i_5 &= v_1 \\ -j\frac{1}{\omega Cm} i_4 + R_2 + j\left(\omega L_2 - \frac{1}{\omega C_2}\right) i_5 &= v_1 \end{aligned}\right\} \tag{10}$$

Where

$$R_{11} = R_1,\ R_{22} = R_2,\ L_{11} = L_1,\ L_{22} = L_2$$

$$\frac{1}{C_1} = \frac{1}{C_{11}} + \frac{1}{Cm},\quad \frac{1}{C_2} = \frac{1}{C_{22}} + \frac{1}{Cm}$$

Accordingly, by calculating i_4 and i_5 from equation(10), and taking account of $Z_1 = v_1/(i_4 + i_5)$ the impedance Z_1 between the terminals A, B is given as

$$Z_1 = \frac{\left[R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right)\right]\left[R_2 + j\left(\omega L_2 - \frac{1}{\omega C_2}\right)\right] + \left(\frac{1}{\omega Cm}\right)^2}{R_1 + j\left(\omega L_1 - \frac{1}{\omega C_1}\right) + R_2 + j\left(\omega L_2 - \frac{1}{\omega C_2}\right) + j\frac{2}{\omega Cm}} \tag{11}$$

Similarly, Z_2 is given as

$$Z_2 = -j\frac{1}{\omega Co} \tag{12}$$

In addition, since vibration losses of a quartz crystal resonator are very small, taking account of $R_1 = R_2 = 0$ approximately, the imaginary port Xe of the impedance Z in equation (9) is given as

$$Xe = -\frac{1}{\omega Co} \cdot \frac{\left(\frac{\omega^2}{\omega_1^2} - 1\right)\left(\frac{\omega^2}{\omega_2^2} - 1\right) - \frac{C_1 C_2}{Cm^2}}{\left(\frac{\omega^2}{\omega_1^2} - 1\right)\left(\frac{\omega^2}{\omega_2^2} - 1\right) - \frac{C_1 C_2}{Cm^2} - \frac{C_2}{Co}\left(\frac{\omega^2}{\omega_1^2} - 1 + \frac{C_1}{Cm}\right) + \frac{C_1}{Co}\left(\frac{\omega^2}{\omega_2^2} - 1 + \frac{C_2}{Cm}\right)}$$

Where

$$\omega_i^2 = 1/L_i C_i \tag{13}$$

$$\omega_i = 2\pi f_i \qquad (i = 1, 2)$$

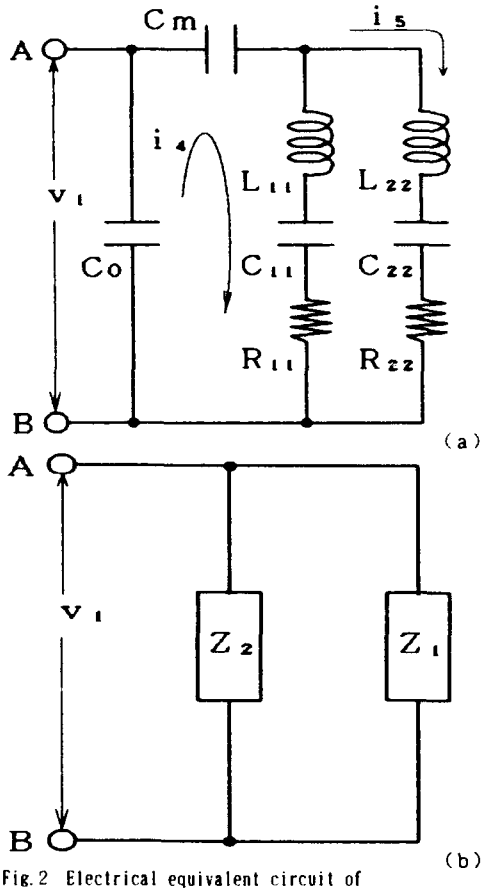


Fig.2 Electrical equivalent circuit of coupling quartz crystal resonator (a) and its impedance expression (b).

Quite similarly, the real part is given as follows:

$$Rei \fallingdotseq Ri\left(1 + \frac{Co}{CL}\right)^2 \qquad (i = 1, 2) \tag{14}$$

Therefore, the oscillation frequency can be easily derived from the obtained equations(5)and(13), so that it is given by the following equation(15).

Equation(15) gives two oscillation frequencies, namely, the oscillation frequency f_+ of the principal vibration for plus sign and the oscillation frequency f_- of the subvibration for minus sign.

$$f_\pm^2 = \frac{1}{2}\left[(f_1'^2 + f_2'^2) \pm \sqrt{(f_1'^2 - f_2'^2)^2 + 4Ks^2 f_1^2 f_2^2}\right]$$

where

$$f_1'^2 = \left(1 + \frac{1}{r_1}\frac{a}{1 + a - h}\right) f_1^2 \tag{15}$$

$$f_2'^2 = \left(1 + \frac{1}{r_2}\frac{a}{1 + a - h}\right) f_2^2$$

$$Ks^2 = Km^2\left(1 + \frac{b}{1 + a - h}\right)^2$$

$$h = \frac{R_L}{Rf\,(\mu + g_m Rd)}$$

$$a = Co/CL \qquad r_1 = Co/C_{11}$$

$$b = Cm/CL \qquad r_2 = Co/C_{22} \qquad Km^2 = \frac{C_1 C_2}{Cm^2}$$

Table 1 Electrical equivalent circuit parameters of the new shape GT cut quartz crystal resonators used in the present experiments.

No.		Frequency (MHz)	$R_{1,2}$ (ohm)	$L_{1,2}$ (mH)	$C_{1,2}$ (fF)	C_0 (pF)	$r_{1,2}$	Q (x10³)
1	Prin.	2.0995	54.3	970	5.93	2.45	431	235
	Sub.	1.8975	630	2680	2.63	2.45	949	50
2	Prin.	2.1000	59.1	1011	5.68	2.43	447	226
	Sub.	1.8937	2620	2910	2.42	2.43	1023	13
3	Prin.	2.0995	61.6	975	5.90	2.46	435	209
	Sub.	1.8958	364	2740	2.57	2.46	975	90

§3 Results and Discussion

First of all, a relationship of variational ratio for coupling factor Km versus load capacitance CL is shown in the present GT cut quartz crystal resonator with coupling capacitance Cm, since the GT cut resonator which is used in the experiments of this paper has the phase difference of 180° between the principal vibration and the subvibration, the coupling factor Km is shown to become large with the increase of CL.

Next, frequency deviation of the principal vibration is shown when the gate capacitance Cg and the drain capacitance Cd vary in the CMOS quartz crystal oscillating circuit, namely, versus load capacitance CL.

In addition, the first, second and third order temperature coefficients α_+, β_+, γ_+ for the principal vibration and simultaneously, the first order temperature coefficient α_- for the subvibration versus the change of load capacitance CL are also shown.

Finally, the frequency temperature characteristics of the principal vibration and the subvibration are shown when CL is taken as a parameter. Let us describe concretely below.

Table 1 shows the resonant frequency and the electrical equivalent circuit parameters of the principal vibration and the subvibration for three new shape GT cut quartz crystal resonators used in the experiments. As is apparent from Table 1, the three resonators show the very satisfactory values as a resonator. Furthermore, capacitance ratio r_i is defined as $r_i = C_0/C_{1,i}$ $(i=1,2)$, here it is, therefore, calculated from $C_{1,i}$ which is obtained from coupling capacitance Cm calculated by employing the coupling factor (See Appendix) derived from the frequency temperature characteristics.

3.1 Coupling factor and resonant frequency

Figure 3 shows the relationship of variational ratio $\triangle Km/Km$ ($\triangle Km$: Ks - Km) of coupling factor Km versus load capacitance CL. Since the present GT cut quartz crystal resonator has a coupling factor of about three percent [5], it has coupling capacitance Cm of about −0.136pF

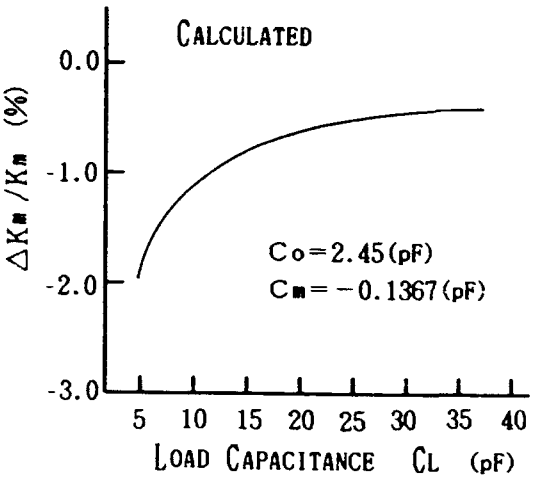


Fig. 3 Relationship between variational ratio $\triangle Km/Km$ of coupling factor Km and load capacitance CL.

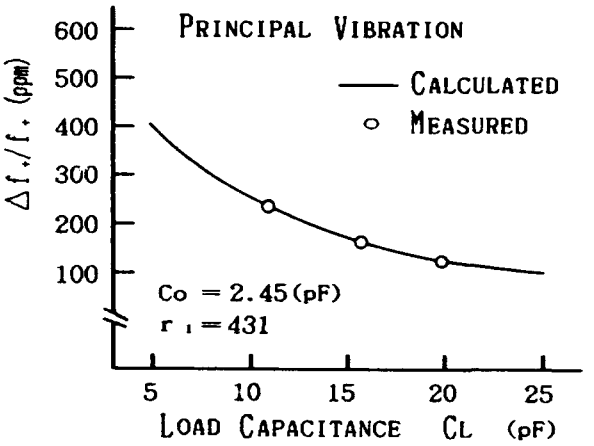


Fig. 4 Relationship of frequency deviation $\triangle f_+/f_+$ of principal vibration to load capacitance CL. (Sample No. 1)

when resonant frequency of the principal vibration is 2.1MHz. In addition, the coupling capacitance Cm has a negative value, because in this GT cut resonator, the electrical coupling by the transformation ratio has a decreasing polarization to the elastic internal coupling, namely, there exists the phase difference of 180° between the principal vibration and the subvibration. Therefore, as $\triangle Km/Km$ versus CL is shown in Fig. 3 when $CL=\infty$ is taken as a standard, $\triangle Km/Km$ becomes small according to decrease of CL, in other words, as the coupling factor Ks (=Km at $CL=\infty$) becomes small, the coupling between the principal vibration and the subvibration is found to get weak.

Figure 4 shows frequency deviation $\triangle f_+/f_+$ of the principal vibration with changeing load capacitance CL, when the GT cut resonator of sample No. 1 is used with shunt capacitance Co=2.45pF, motional capacitance $C_{1,1}$=5.68fF(Capacitance ratio r_1=431), coupling capacitance Cm=−0.1367pF, coupling factor Km=2.89% and resonant frequency f=2.1MHz of the principal vibration, namely, when

the drain capacitance Cd and the gate capacitance Cg is changed simultaneously, and Cg Cd is taken approximately in the CMOS quartz crystal oscillating circuit with h 0 approximately from characteristics and construction of the CMOS inverter. According as CL becomes small, resonant frequency becomes high, the smaller CL becomes, the larger the inclination becomes.

3, 2 Frequency temperature coefficients

Figure 5 shows the first order temperature coefficient α_+ with the change in load capacitance CL when the resonator of sample No.1 is employed with the first, second and third order temperature coefficients α_+ -2.63 $\times 10^{-8}$ /℃, β_+ 1.97 $\times 10^{-10}$ /℃^{2}, γ_+ -7.04 $\times 10^{-14}$ /℃^{3} at no load capacitance. The circles are the measured values and the solid line is the theoretical result calculated by utilizing Km and Cm derived from the experimental values of CL 5.61pF(α_+ = -3.3 $\times 10^{-7}$ /℃) and 15.4pF(α_+ = -3.4 $\times 10^{-9}$ /℃). Where the calculated values of the first order temperature coefficients of Km and uncoupled principal vibration were used in the calculation. Accordingly, it is natural that the calculated values agree very well with the measured ones at CL 5.61pF and 15.4pF. However, a small difference between the calculated and measured results takes place at CL 10.85pF (Measured value: α_+ 1.87 $\times 10^{-8}$ /℃, Calculated value: α_+ 9.34 $\times 10^{-8}$ /℃) and 20.2pF (Measured value: α_+ 7.93 $\times 10^{-9}$ /℃, Calculated value: α_+ 5.19 $\times 10^{-8}$ /℃). This is probably because the theoretical values of the first order temperature coefficients of Km and uncoupled principal vibration used, slightly differ from the measured values. But, the error is extremely small with a change in the order of 10^{-8}. In addition, as CL gets small, α_+ also becomes small, this is because there exists the phase difference of 180° between the principal vibration and the subvibration, so that coupling factor Km becomes small. Simultaneously, as is obvious from Fig.5, the variation quantity of α_+ becomes markedly large as CL gets small, it is, however, understood that its variation quantity becomes extremely small with the change of the order of 10^{-8} at the value of CL larger than 10.85pF. From this result, in order to decrease CL dependence versus α_+, CL value is also found to be large.

Figure 6 shows the first order temperature coefficient α_+ of the principal vibration with the change of load capacitance CL when the resonator of sample No.3 is utilized with the coupling factor Km=2.88%, α_+ = -2.02 $\times 10^{-8}$ /℃ at CL=∞. As well as shown in Fig.5, as CL becomes small, α_+ has a further negative value and its variation quantity gets markedly large as CL value gets small. In case of the present resonator, the first order temperature coefficient α_+ of the principal vibration varies from -7.19 $\times 10^{-8}$ /℃ to -15.6 $\times 10^{-8}$ /℃, namely, at the order of 10^{-8} even if the CL varies from 19.6pF to 9.1pF, while α_+ has a value of -2.89 $\times 10^{-7}$ /℃ and show the marked variation, according as the CL becomes a further small value such as 5.9pF. Thus, it is also understood from this resonator that the smaller the CL value becomes, the larger the change of α_+ becomes.

Figure 7 shows the relationship between the first order temperature coefficient α_- of the subvibration and load capacitance CL when the resonator of sample No.3 is employed. α_- of the subvibration has a value of -34.3 $\times$

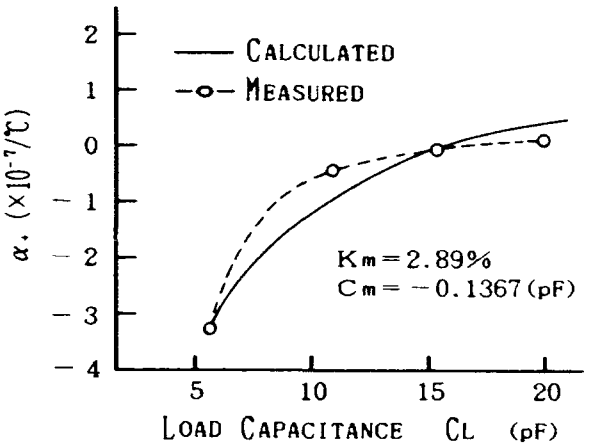


Fig.5 Relationship between the first order temperature coefficient α of principal vibration and load capacitance CL. Sample No.1

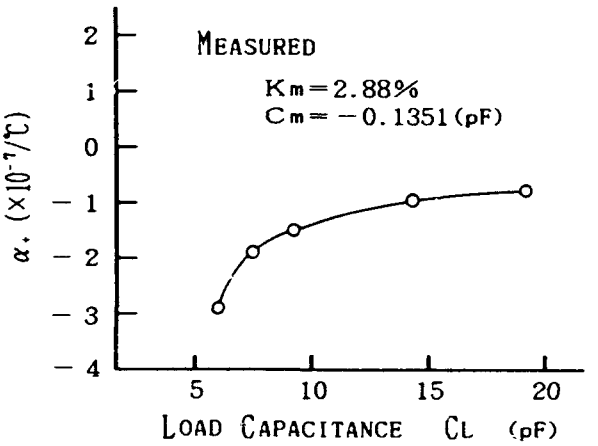


Fig.6 Relationship between the first order temperature coefficient α of principal vibration and load capacitance CL. Sample No.3

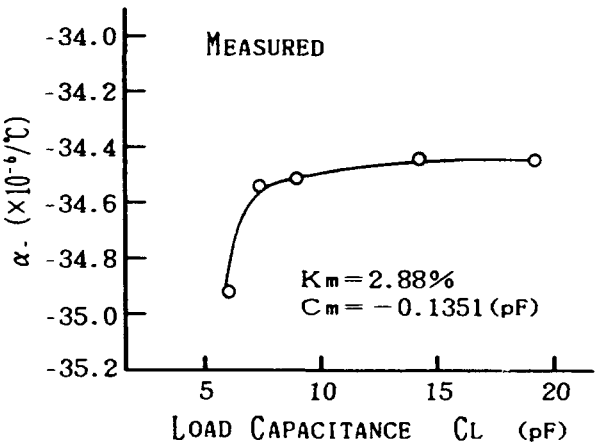


Fig.7 Relationship between the first order temperature coefficient α of subvibration and load capacitance CL. Sample No.3

10^{-5} C and its absolute value shows a very large value, when CL is infinite. Therefore, as shown in Fig. 7, its variation quantity is so small that it can be ignored due to the change of the order of 10^{-7}, even if the CL value varies from 19.6pF to 5.9pF.

Figure 8 shows the second order temperature coefficient β_2 of the principal vibration with the change in load capacitance CL when the resonator of sample No. 1 is employed. β_2 varies at the order of 10^{-10} when the CL value varies from 5.61pF(β_2 1.46×10^{-10} (C^{-2})) to 20.2pF (β_2 2.35×10^{-10} (C^{-2}), while it varies at the order of less than 10^{-11} when the CL changes from 10.85pF(β_2 2.31×10^{-10} (C^{-2})) to 20.2pF. Since its variation quantity is extremely small, β_2 versus the CL value can be regarded as the constant value.

Figure 9 shows the relationship between the third order temperature coefficient γ_2 of the principal vibration and load capacitance CL when the resonator of sample No. 1 is used. Even if the CL value changes from 5.61pF(γ_2 3.11×10^{-12} (C^{-3})) to 20.2pF (γ_2 7.17×10^{-11} (C^{-3})), γ_2 varies at the order of 10^{-12} (the order of less than 10^{-11} to changing from 10.85pF(γ_2 6.99×10^{-11} (C^{-3}) to 20.2pF), its variation quantity is extremely small as well as that of β_2, therefore, γ_2 versus CL can be also regarded as the constant value. From these results, the frequency temperature characteristics versus the CL value have only to consider the first order temperature coefficient α_1.

3.3 Frequency temperature characteristics

Figure 10 shows one example of frequency temperature characteristics of the principal vibration when Cd and Cg are changed in a CMOS oscillating circuit comprising the resonator of sample No. 1, namely, when the load capacitance CL is taken as a parameter. As is apparent from Fig. 10, the frequency temperature characteristics are foumd to vary by the CL value. In particular, as it becomes small, the first order temperature coefficient has a further negative value because the coupling factor remarkably varies and gets smaller. Therefore, in order to lower frequency deviation versus the CL value in frequency temperature characteristics, the oscillating circuit must be so designed that it becomes as large as posible. On the other hand, since the frequency temperature characteristics vary by the CL value, it is substantially possible to take α_1 to reach zero by selection of the optimum CL value. For example, for the present resonator, because α_1 0 approximately at CL 15.4pF, frequency deviation has a value of 0.7ppm over a wide temperature range of 30℃ to +70℃, as shown in Fig. 10, thus, a quartz crystal oscillating circuit with very excellent frequency temperature characteristics has been obtained. From these results, arbitrary frequency temperature characteristics can, therefore, obtained by selection of the CL value, meeting required specifications.

Figure 11 shows another frequency temperature characteristics of the principal vibration in a CMOS oscillating circuit comprising the resonator of sample No. 2 with the coupling factor Km=2.92% and the coupling capacitance Cm= 0.1269pF, when the load capacitance CL is also taken as a parameter. This resonator has a comparatively large series resistance R_2 of the subvibration. As is similar to the example shown in Fig. 10, according as the

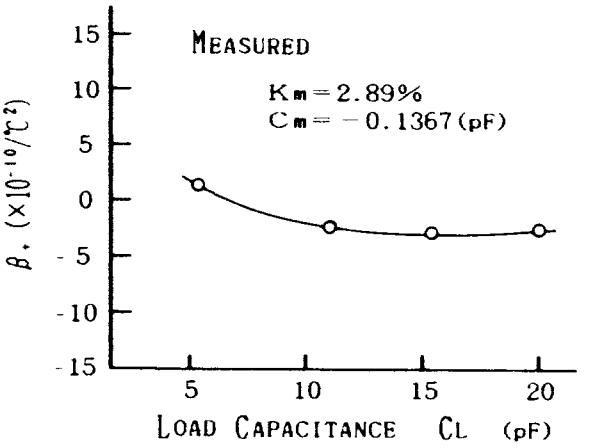


Fig. 8 Relationship of the second order temperature coefficients β_2 of principal vibration to load capacitance CL. Sample No. 1

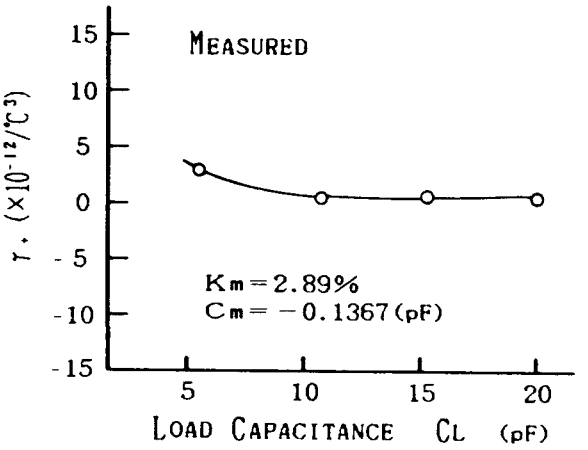


Fig. 9 Relationship of the third order temperature coefficient γ_2 of principal vibration to load capacitance CL. Sample No. 1

load capacitance CL has a smaller value of CL 10.24, 8.87 and 8.12pF, the first order temperature coefficient α_1 has a value of α_1 3.01×10^{-6} (C), β_2 1.15×10^{-10} (C^{-2}), γ_2 3.27×10^{-11} (C^{-3}), 0.645 $\times 10^{-6}$ (C), β_2 1.41×10^{-10} (C^{-2}), γ_2 5.19×10^{-11} (C^{-3}) and 1.59×10^{-6} (C), β_2 1.98×10^{-10} (C^{-2}), γ_2 8.63×10^{-11} (C^{-3}), and it is readily apprehended that α_1 shifts to the further negative value. In addition, in case of employing the present resonator, because α_1 0 approximately at CL 8.87pF, as shown in Fig. 11, it has a value of 0.55ppm in frequency deviation over a wide temperature range of 30℃ to +70℃, thus, the quartz crystal oscillator with very good frequency temperature characteristics has been obtained as well as the result of Fig. 10.

Figure 12 shows another frequency temperature characteristics of the principal vibration when load capacitance CL in a CMOS oscillating circuit comprising the resonator of sample No. 3 is taken as a parameter. As is the same as Fig. 10, the frequency temperature characteristics are found to vary markedly by the CL value. Since

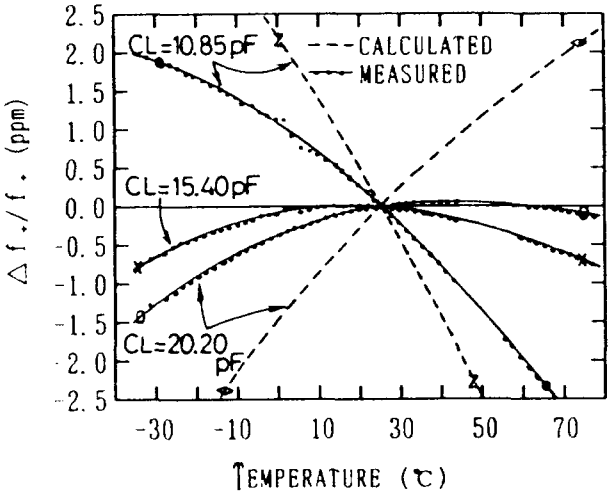


Fig. 10 Frequency temperature characteristics of principal vibration, when load capacitance CL is taken as a parameter. Sample No. 1

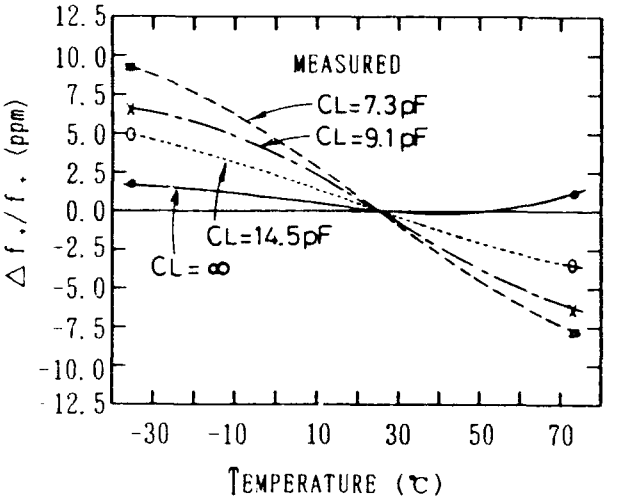


Fig. 12 Frequency temperature characteristics of principal vibration, when load capacitance CL is taken as a parameter. Sample No. 3

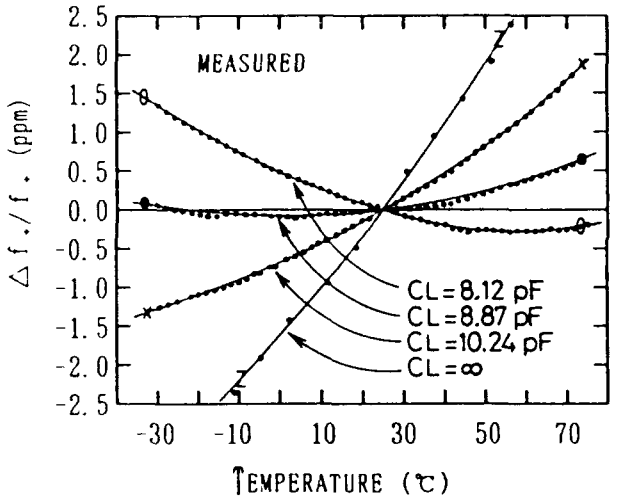


Fig. 11 Frequency temperature characteristics of principal vibration, when load capacitance CL is taken as a parameter. Sample No. 2

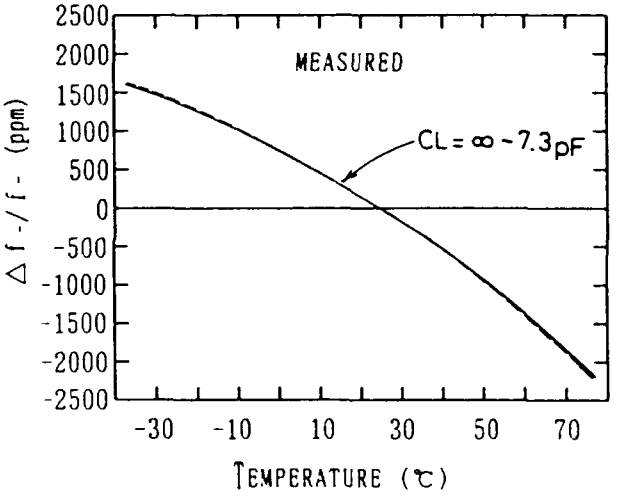


Fig. 13 Frequency temperature characteristics of subvibration, when load capacitance CL is taken as a parameter. Sample No. 3

this resonator from the beginning has a negative value of α_{-}. 2.02 · 10^{-8} /℃ at CL ∞, it is actually impossible to take α_{-} 0 by the CL value. That is, in order to obtain α_{-} 0 by the CL value, it goes without saying that α_{-} must have a positive value at CL ∞. However, the present resonator shows good frequency temperature characteristics because it has a very small value of 1.5ppm in frequency deviation over a wide temperature range of -30℃ to +70℃. Therefore, the resonator shows very good frequency temperature characteristics as CL has a large value.

Figure 13 shows the frequency temperature characteristics of the subvibration when the load capacitance CL in a CMOS oscillating circuit comprising the resonator of sample No. 3 is taken as a parameter. As described in Fig. 7, α of the subvibration has a very large value of 31.3 · 10^{-6} /℃ in absolute value as compared with the value of α_{+-}, in addition, as shown in Fig. 13, α can be regarded as the constant value by the CL value, because the change of α versus the change of the CL value is very small at the order of 10^{-7}.

§4 Conclusions

In this paper, a study has been performed with a view to clarifying the oscillation frequency characteristics in a CMOS quartz crystal oscillating circuit comprising a coupling quartz crystal resonator, especially, a GT cut quartz crystal resonator formed by an etching method.

First, in analysis procedure, since a CMOS inverter can be linearly by an electric current source and drain conductance, an electrical equivalent circuit of the CMOS

quartz crystal oscillating circuit comprising the coupling quartz crystal resonator has been shown. In addition, an oscillation condition and an amplitude continuation condition have been, theoretically, derived from Kirchhoff's second low.

Next, the resonant frequencies have been derived from the oscillation condition obtained previously and the imaginary part of the GT cut quartz crystal resonator which is one of coupling quartz crystal resonators. Besides, since this GT cut quartz crystal resonator can be regarded as a resonator wherein the electrical coupling by the transformation ratio and the elastic internal coupling coexist, and still more, the electrical coupling becomes a decreasing polarization to the elastic internal coupling, the coupling capacitance Cm has a negative value. Therefore, it has been shown that the coupling factor gets small with the decrease of load capacitance CL. Simultaneously, it has been, theoretically and experimentally, clarified that, of frequency temperature coefficients, particularly, the first order temperature coefficient α_+ of the principal vibration remarkably varies by the CL value and α_+ shifts to a further negative value, because the coupling factor gets small according to decrease of the CL value. In other words, the smaller the CL value becomes, the larger the change in the frequency temperature characteristics of the principal vibration to the CL value becomes. From this result, a quartz crystal oscillating circuit with very good frequency stability wherein the frequency deviation is less than 1 ppm over a wide temperature range of −30℃ to +70℃ without any temperature compensation, has been obtained by employing the present resonator. Thus, the good frequency temperature characteristics have been obtained by selection of the optimum CL value.

On the other hand, it has been also clarified that the CL value must be large so as to lower the dependence of frequency temperature characteristics by the CL value. However, in order to adjust frequency of the principal vibration to the nominal frequency by the CL value in the CMOS quartz crystal oscillating circuit and obtain the circuit with very good frequency temperature characteristics, it has been, therefore, understood that it is of great importance to grasp fully the relationship between frequency deviation of the principal vibration of the resonator which is manufactured and its frequency temperature characteristics. For the coming subject, we would like to study a method for adjusting the frequency shifted by CL value to the nominal frequency accurately.

Acknowledgements

I would like to express my thanks to Mr. K. Sunaga for his helping in the experiments and in the calculation, and also Mr. S. Takigawa for his helping in the experiments.

References

[1] H. Kawashima: "A Variational Analysis of a GT Cut Quartz Crystal Resonator with the Supporting Portions at the Ends", Trans. IECE(A), J68 A, 8, pp771-778 (1985-08) (in Japanese)

[2] 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)

[3] 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)

[4] K. Kataoka and M. Ida: "How to employ FET for beginners", Ohmu sha (1984-02) (in Japanese)

[5] H. Kawashima: "Dependence of Load Capacitance for a Coupling Quartz Crystal Resonator", Jpn. J. Appl. Phys. 26, Supplement 26-1, pp. 114-116 (1987)

[6] M. Onoe: "Analysis and Detection of subvibration for quartz crystal resonator", Journal of IECE, vol. 47, 1, pp. 38-45 (1964) (in Japanese)

Appendix

Derivatives of coupling factor Km

The coupling factor Km of a GT cut quartz crystal resonator can be calculated from the experiments of the frequency temperature characteristics. Let us describe its derivatives method concretely below. The first order temperature coefficients α_+ and α_- of the principal vibration and the subvibration from equation(15) are given as

$$\alpha_+ = -\frac{1}{4f_+^2}\left\{2\left(f_1'^2\frac{\dot{f}_1'}{f_1'}+f_2'^2\frac{\dot{f}_2'}{f_2'}\right)+\dot{P}\right\}$$

Where

$$P^2=(f_1'^2-f_2'^2)^2+4Ks^2f_1^2f_2^2 \qquad (I)$$

$$\dot{P}=-2(f_1'^2-f_2'^2)\left(f_1'^2\frac{\dot{f}_1'}{f_1'}+f_2'^2\frac{\dot{f}_2'}{f_2'}\right)+4Ks^2f_1^2f_2^2\left(\frac{\dot{Ks}}{Ks}+\frac{\dot{f}_1}{f_1}+\frac{\dot{f}_2}{f_2}\right)\Big/P$$

$\cdot = \partial/\partial T$, T: Temperature

In addition, the following relation is obtained from equation(15):

$$\begin{aligned} &f_1'^2+f_2'^2=f_+^2+f_-^2 \\ &(1-\varepsilon Ks^2)f_1'^2f_2'^2=f_+^2f_-^2 \end{aligned} \qquad (II)$$

Where

$$\varepsilon=1/\left(1+\frac{1}{r_1}\frac{a}{1+a-h}\right)\left(1+\frac{1}{r_2}\frac{a}{1+a-h}\right)$$

Therefore, from the data of Table 1, it is approximately expressed as $\varepsilon=1$. Furthermore, as the ε hardly influences the frequency temperature characteristics of the principal vibration and the subvibration and as $\dot{f}_1/f_1=\dot{f}_2/f_2$ theoretically from the cut angles of the GT plate, the coupling factor Km from equations (I) and (II) is calculated as follows:

$$Km=\frac{1+a}{1+a+b}\left[\frac{(1-f_{wr}^2)(\alpha_+-\dot{f}_1/f_1)}{(1-f_{wr}^2)(\alpha_+-\dot{f}_1/f_1)+f_{wr}^2(\dot{Km}/Km)}\right]^{1/2} \qquad (III)$$

Where $f_{wr}=f_-/f_+$

Equation(III) is given as unknown factors of Km, b, $\dot{f}_1/f_1$ and $\dot{Km}/Km$, accordingly, Km and b are calculated by employing the CL value, and $\dot{f}_1/f_1$ and $\dot{Km}/Km$ derived theoretically.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

A B.V.A. QUARTZ CRYSTAL OSCILLATOR FOR SEVERE ENVIRONMENTS

Raymond J. BESSON, Marc MOUREY

Ecole Nationale Supérieure de Mécanique et des Microtechniques
La Bouloie - Route de Gray - 25030 BESANCON CEDEX - FRANCE

ABSTRACT

A new B.V.A. quartz resonator (sometimes said "tactical B.V.A." resonator) has been developed for use in severe environments. More than one hundred units have been produced and tested in France and in the U.S. The resonator is a 10 MHz, 3rd overtone, SC cut crystal. The unit fits into an ordinary HC 40 can without pinch-off. Typical numberts are : R_1 = 95 Ω, Q = $1.35\ 10^6$. g maximum sensivity less than 10^{-10}/g, turn over point 85° C . Pressure variation sensitivity : $5\ 10^{-11}$/bar.

The resonator design is rapidly discussed and described. Results of various tests and experimental data on this resonator are given and discussed.

A new type of quartz oscillators that have excellent performances has been studied and developed to take advantage of this resonator for use in severe environments. Oscillators of this design usually demonstrate drift rates of less than 10^{-11}/day. Typical $\sigma_y(\tau)$ at 10 seconds is 5 to 7 10^{-13} and spectral purity usually obtained is 115 dB at 1 Hz and 160 dB at 1000 Hz from the carrier. Power consumption is 10 W during warm up time and 2.5 W in ordinary conditions.

Several results have been obtained indicating that oscillator performance is still dominated by the quartz resonator.

The oscillators are designed to survive shocks and vibrations. They use high reliability components which may be surface mounted components. The oscillating loop has a single transistor and the resonator has a capacitive load. A true A.G.C. is used and the output amplifier has at least a 90 dB isolation. The oven is basically made of a metallic box with the heating transistor onto it. Thermistors are used as sensors and the oven control is proportionnal. Then oven insulation may, in addition, use a titanium dewar. The oscillators are rather compact (less than 250 cm^3) and may use hermetically sealed external enclosures to get rid of hygrometry variations influence.

CH2818-3/90/0000-593 $1.00 © 1990 IEEE

INTRODUCTION

This paper presents results on development of a precision oscillator utilizing the miniature 10 MHz BVA resonator. Roughly 125 of those resonators have been built (85 SC cut 3rd overtone and 40 AT cut 3rd overtone) and tested to date whereas 4 test oscillators have been built and tested.

I - MINIATURE 10 MHz BVA RESONATOR

A new BVA miniature resonator has been developed. The unit fits into the ordinary HC 40 can without pinch-off. A B.V.A.[2] ordinary design **(1)** is used, however the supporting structure makes use of a new system which brings in symmetry, low pressure sensitivity and a good compromise between shock and acceleration sensitivities. In fact, the resonator is symmetrically clamped inside a grooved steel cylinder **(2)(3)**. The resonators are carefully prepared from swept premum Q material (usually S.I.C.N. or SAWYER) and the final operations (last cleaning, bake out and coldweld) are done in the same vacuum. The units are coldwelded under a residual pressure between 3×10^{-8} and 10^{-7} mbar.

I-1/ Resonator technical characteristics :

Nominal frequency : 9.999 983 Hz
Frequency tolerance at T.O.P. : - 10 Hz + 6 HZ
Cut : doubly rotated SC cut : 3rd overtone

Enclosure : ordinary HC 40/U (for instance Houston electronics) without pinch-off.
Serie capacitance : 30 pF (nominal)
T.O.P. : 85° C (- 5° C + 8° C)
Aging (after 5 days) : better than 4×10^{-11}/day
Q factor : $1.35\ 10^{6}$ (typical), R_1 = 95 Ω (typical)
G maximum sentitivity : less than 10^{-10}/g
External pressure sensitivity : 5×10^{-11}/bar (typical).

I-2/ Results of various tests and experimental data on the resonators

85 S.C. cut resonators have been built among which roughly 20 have been tested in ONERA (France) and 15 in the US in various cooperative programs. The other resonators have been tested in our facility at E.N.S.M.M. The tests included all the parameters previously quoted. The g sensitivity has been determined by several method (2 g tip over, slow rotation, centrifuge rotating machine and of course dynamic sideband technique). The pressure variation sensitivity was done from atmospheric pressure to vacuum and the reverse eliminating coexistent temperature changes.

G sensitivity of resonators ranged between 1×10^{-11} /g and 2.7×10^{-10}/g. The g maximum sensitivity $|\Gamma max|$ was 5×10^{-11}/g or better in 20% of cases. The average value of $|\Gamma max|$ was 9×10^{-11}/g and 80% of the production was better than 1.5×10^{-10}/g.

The mean value of pressure variation sensitivity is $5\ 10^{-11}$/bar.

The aging has been tested in 20 oscillators in various programs. In 70 % of cases the aging rate was better than 2×10^{-11}/day, in 40% of cases it was less than 10^{-11}/day.

II - SHORT DESCRIPTION OF THE OSCILLATORS

An important effort has been made to match the resonator's parameters and the oscillator design so as to take advantage of the advances offered by the tactical B.V.A. resonator.

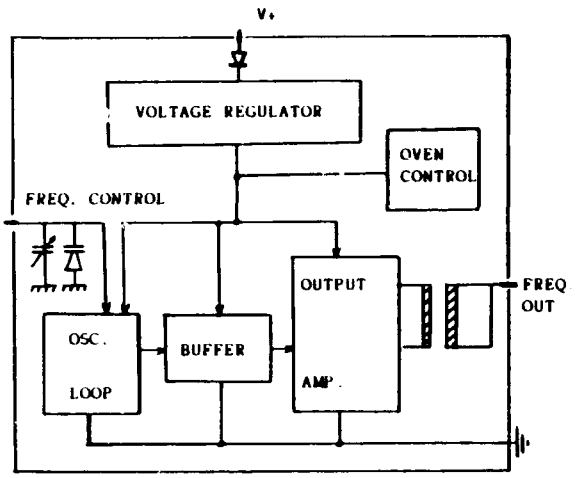


Fig. 1 : Block diagram.

The total oscillator is represented on block diagram of Fig. 1. The oscillating loop uses a single very low noise transistor previously "noise selected". A C.E.configuration is used for the amplifier together with π phase shifting transformer. A selecting resonant circuit with a Q factor approximately of 30 is used to select the C mode frequency when oscillations are started. It may be automatically short circuited by use of A.G.C. signal when oscillations are established so that the amplifier is more aperiodic. The quartz Xtal is connected to ground through a small pure capacitive impedance. As a consequence, the oscillator frequency is not exactly the Xtal resonant frequency. A true A.G.C. is used which involves an F.E.T. and allows oscillation levels on the Xtal between 40 μW and 180 μ W (to be adjusted for minimum phase noise and/or aging rate). A single stage (eventually two stages) buffer followed by a cascode output amplifier yields at least a 90 dB isolation. A frequency control cell with a varicap diode allows frequency adjustements in the 2×10^{-7} range.

The oven is basically made of a copper block in which the Xtal is carefully inbedded. The heating transistor directly sits onto the copper block close to Xtal. The oven control is proportionnal and uses a thermistor (Betacurve 10 K or 100 K type). Oven temperature is adjusted between 75° C and 95° C by means of two resistors. Recently another resistor somewhat sensing external temperature variations has been added (4) but experimental

results with this design have not been obtained yet. Of importance is the fact that all the components (oscillating loop, buffer and output amplifiers, oven control, ...) are temperature controlled. The oven insulation may, in addition, use a titanium dewar and the cylindrical symmetry of the resonator has been kept as far as possible. Special care is devoted to prevention of heat flow through connecting wires. Hermetically sealed external shells are used to get rid of hygromtery variations influence **(5)**.

Of course, the oscillator circuits have first been carefully studied by computer aided design (with p-spice program). Special care has been devoted to temperature influence on oscillating loop and A.G.C. The resonator and the external shell of oscillator have been studied by finite element calculation for heat transfer.

III - PERFORMANCES OF THE OSCILLATORS

Four test oscillators have been built after first debugging of the circuits. The oscillators have been tested for allan variance and phase noise. One oscillator has been tested for temperature variation influence and two oscillators have been checked for aging.

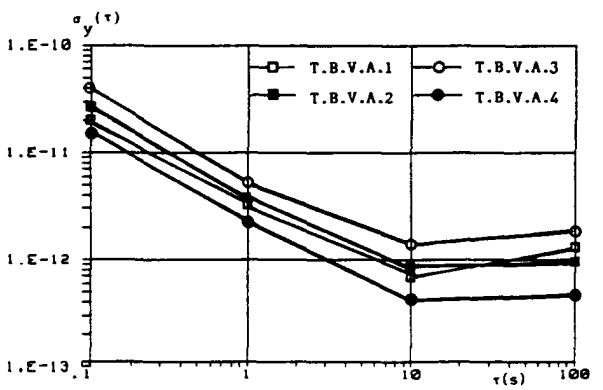


Fig. 2 : ALLAN variance of oscillators.

Fig. 2 shows the Allan variance of oscillators. Best results are obtained at 10 s integration time and the average is 8×10^{-13}.

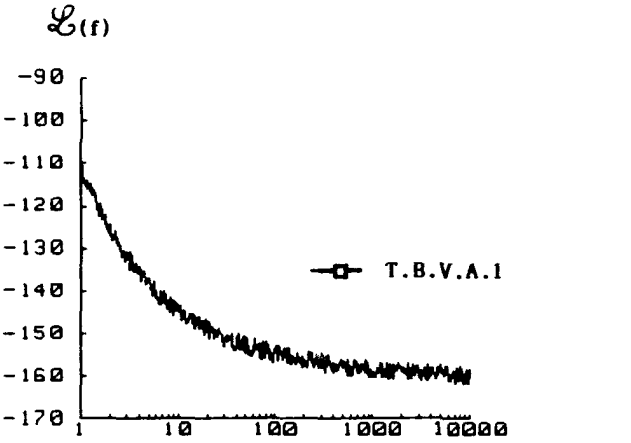


Fig. 3 : Phase noise of 1rst oscillator.

Fig. 3 shows the phase noise result of oscillator $TBVA_1$. This result is consistent with $\sigma_y(\tau)$ observed on Fig. 2. Fig. 4 shows the phase noise observed with the best oscillator $TBVA_4$. In this domain the oscillator performance seems still dominated by the quartz resonator (especially at 1 Hz from the carrier).

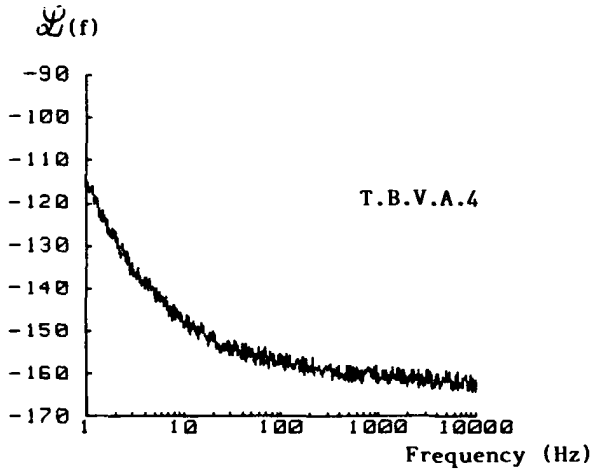


Fig. 4 : Phase noise of 4th oscillator.

Fig. 5 and Fig. 6 show aging result performed during five weeks after starting the units. It can be seen that the aging rate establishes after days and is consistent with aging rates already observed in previous programs. However relative contribution to the aging by the quartz itself and the other components has not been determined yet.

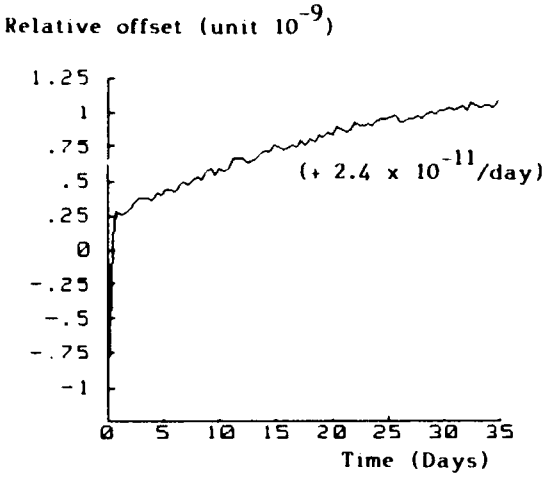


Fig. 5 : Aging data.

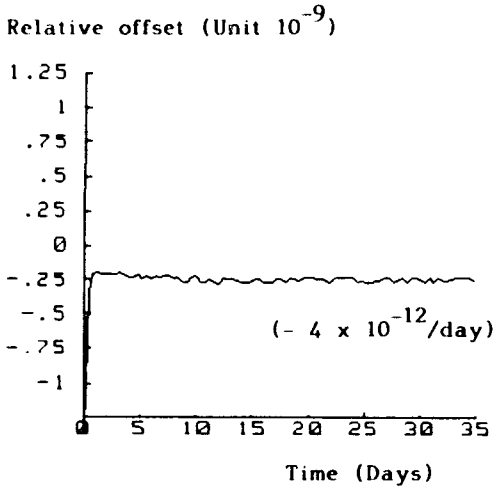


Fig. 6 : Aging data

Temperature influence depends on the type of oven used. It has been found lower than 10^{-10} for an external temperature variation between - 30°C and + 70°C with a unit which used a titanium dewar.

The size of the oscillator is 240 cm^3 i.e., 8.5 cm in height and 6 cm in diamter. The unit consumes 2.3 W at 20°C.

CONCLUSION

A new type of oscillator has been developed for the tactical BVA 10 MHz miniature resonator. First results seem encouraging but many more experiments have to be performed especially dealing with vibration environments and external pressure changes. However, a 10 MHz oscillator with aging in the 10^{-11}/day range, $\sigma_y(\tau)$ at 10 second in the order of 7 10^{-13}, spectral purity of 115 dB at 1 Hz together with a g sensitivity better than 10^{-10}/g worst axis, seems to day feasible on regular scale.

ACKNOWLEDGEMENTS

The authors would like to thank Dr F.L. WALLS for several fruitful discussions.

REFERENCES

(1) R.J. BESSON, "A new electrodeless resonator design", 31st Annual Frequency Control Symposium (A.F.C.S.), proceedings pp 147-152, 1977, and US patent 4.135.108 (1-16-1979).

(2) R.J. BESSON and al, French patent 8509097 and US patent 4.705.982 (11-10-1987).

(3) R.J. BESSON and al, French patent 8814197 (10-28-1988), US Patent and other patents pending.

(4) F.L. WALLS, "Analysis of high performance compensated thermal enclosures", 41st AFCS, proceedings pp 439-443, (1987).

(5) F.L. WALLS, "The influence of pressure and humidity on the medium and long-term frequency stability of quartz oscillators", 42nd AFCS, pp 279-283, (1988).

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

FACTORS INFLUENCING STABILITY IN THE MICROCOMPUTER-COMPENSATED CRYSTAL OSCILLATOR

Albert Benjaminson

General Technical Services, Inc.
105 Clermont Avenue
Alexandria, Virginia 22304

INTRODUCTION

The stability of an MCXO using a Direct Digital Synthesizer combined with phase-locked loop summing is discussed in detail. The characteristics of each subsystem of the MCXO are analyzed and their impact on the stability of the output frequency is discussed. Test results in support of the analyses are displayed. The stability is required to be better than 20 ppb over the range of temperatures from -55°C. to +85°C.

While the MCXO, as developed by General Technical Services, is considerably more accurate than TCXOs and can operate over a wider temperature range than either a TCXO or an OXO, it also exhibits some unique characteristics. These result from the digital techniques used to create this MCXO design.

The accuracy of the MCXO over the military temperature range is also influenced by hysteresis and aging of the two oscillation modes employed in the Dual-mode Crystal Oscillator and by the "goodness" of the fit produced by the polynominal curve-fitting routine.

This report also discusses several interesting anomalies which currently exist in the normal field SC-cut crystals used in the DMXO circuit that drives the MCXO. These anomalies have not been observed prior to the development and subsequent precise temperature cycling required to calibrate and characterize the resonators needed by the MCXO.

A block diagram of the MCXO system is shown in Figure 1. A detailed explanation of this system was presented at the 43rd Annual Frequency Control Symposium[1] and will not be repeated here. Instead several areas of interest have been selected for examination that may help clarify some of the attributes unique to this MCXO.

CH2818-3/90/0000-597 $1.00 © 1990 IEEE

Response to Rapid Temperature Change

The time required by the system's correction cycle limits the MCXO's ability to follow rapid variations in system temperature. A high rate of temperature change may cause noticeable "sawtooth" stepping of the output frequency. While the period counter and the digital mixer circuits used to generate the N1 number (the analog of temperature) can be programmed to cycle as rapidly as ten times per second, the microcomputer is currently limited to about five times per second by the 3.39 MHz clock frequency employed. The clock, derived from the fundamental mode oscillator, was selected to conserve power consumption but can be modified in later models if required, by programming a switch to the higher frequency third overtone oscillator output.

The output frequency of the MCXO is the sum of the frequencies of the overtone oscillator and the DDS-generated correction frequency Df_0. A new value of Df_0, equal to the difference between 10 MHz and the overtone oscillator frequency, is generated at the end of each correction cycle. The new value depends on both the temperature and the size of the temperature change since the frequency-temperature slope varies in a parabolic manner.[1] Little change occurs near the turning point which is normally located at about 20°C. but the greatest changes occur at the temperature extremes, with the maximum slope occuring equally at -55°C and 85°C. The magnitude of these maxima are about 2ppm/°C. In order to maintain an error of less than 20 ppb in a changing temperature environment the maximum temperature ramp must be kept below

$$\frac{\Delta T}{\Delta t} = \frac{20\ \text{ppb} \times 5}{2\ \text{ppm/°C}} = 0.05\text{°C/sec., or } 6\text{°C/min.}$$

at the temperature extremes. At the turning point temperature, of course, the rate of change can be very much larger.

Synthesizer-Induced Limitations and Phase Noise

The Direct Digital Synthesizer (DDS) is driven by the computed value of N2 which has a maximum range of 24 bits. The range is set by the 24 bit adder/register and can be increased further but only at the expense of an increase in power consumption. N2, however is kept well below the maximum value to minimize the jitter inherent to the DDS process, since close-in spurious sidebands are produced as a function of Df_o where:

$$Df_o = N2 \times \frac{f_{in}}{2^{24}}$$

and the maximum sideband levels are:

$$A_s = 20\log \frac{Df_o}{f_{in}} \text{ dBc}$$

The system resolution is set by:

$$\frac{f_{in}}{2^{24}} = 0.0186 \text{ Hz. (1.86ppb)}$$

Since the calculated value of N2 is rounded to an integer, the resolution of Df_o is better than $\pm$1ppb.

The DDS phase jitter produces a small spectrum around the 10 MHz. outut frequency line, which varies with the frequency of Df_o from -70 dBc at 200 HZ to -60 dBc at 700 Hz. Spurs are also produced in the 10 MHz output spectrum offset at the DDS frequency and its harmonics. Both phase jitter and spur lines are greatly reduced by the low-pass action of the PLL and its active lag-lead third-order filter, whose bandwidth is nominally 10 HZ.

Sustaining Amplifier Stability and Hysteresis

The dual-mode oscillator configuration shown in Figure 2 has been discussed previously [1]. This paper constitutes an extension to that work and reviews the progress since last year in achieving an understanding of the factors influencing the performance of resonators and their sustaining amplifiers when subjected to repeated cycling over a wide temperature range.

Inevitably, the question arose as to the influence of circuit components on frequency stability, particularly the reactive components, both inductive and capacitive. In order to resolve this question, the following technique was devised to separate changes in circuit values from the variations in resonator characteristics.

Since the DMXO comprises a pair of series-mode crystal oscillators, they can be made to run as independent L-C oscillators by removing the crystal and inserting a separate stable resistor in its place in each oscillator. By doing this, the stability of the L-C oscillator circuits can be evaluated over the temperature range. Any changes in the L-C oscillators' frequencies, nominally 10 and 3.5 MHz., can be equated through CODA analysis to determine their impact on the crystal-controlled frequencies.

In both the Bridge and Colpitts circuits used in the DMXO, an inductor operates in conjunction with a pair of capacitors. The latter, acting as voltage dividers, provide the positive feedback necessary to sustain oscillations. In each circuit, the inductance resonates with the series capacitance to reduce the net reactance to zero at the series-resonant frequency of the crystal. A change in any reactance will pull the frequency of oscillation. With the crystal present, only a small frequency change occurs before a cancelling reactance is produced by the resonator. Without the crystal, the frequency must change until the net reactance of the tuned L-C circuit is zero.

Through the use of the CODA program, each oscillator was analyzed to calculate the change in crystal-controlled frequency that would result from a 1% change in the critical tuning inductance. (This is not to imply that the inductors are more suspect than the capacitors. The tuning capacitors could have been adjusted instead, but since there are two capacitors in series in each oscillator both would have had to be changed in the correct ratio, a more awkward procedure.) The effect on the L-C oscillator frequencies of a proportional change in inductance (or capacitance) is easily determined since the frequency is inversely proportional to the square root of the inductance or capacitance.

$$\omega^2 LC = 1$$

$$\frac{d\omega}{\omega} = -\frac{1}{2}\frac{dL}{L} = -\frac{1}{2}\frac{dC}{C}$$

By using the CODA program to change the value of the inductance, as indicated above, and as shown in Figures 3 and 4, the value of $d\omega/\omega$ for the 10 MHz. crystal oscillator was found to be $1/10^8/1\%$, yielding a value for $d\omega_{Xtl}/d\omega_{LCO} = 2/10^6$. The corresponding ratio for the 3.5 MHz oscillator is $13/10^6$. The large difference between these values is due to the

differences in the crystal's motional capacitances in the two modes and in the circuit designs.

Figures 5 and 6 illustrate the effects of temperature cycling on the frequencies of a pair of L-C oscillators. The temperature data are from the thermometer in the test chamber, unlike the tests on the DMXOs (discussed later) where temperature information is derived from the beat frequency between the dual-mode oscillators. Figures 5 and 6 show two temperature runs taken almost a week apart on both oscillators. All runs began and ended at 25°C. The dashed curve illustrates the first run where the wait time at each temperature step was only 1 minute. The solid curve illustrates the result of increasing the wait time to 5 minutes. It shows less "apparent hysteresis" than the first run and suggests that the hysteresis is largely due to the thermal inertia of the parts involved.

The left-hand scales show the frequency changes in the L-C connection, while the right hand scales show the calculated impact on the oscillator when under crystal-control (assuming a perfectly stable crystal). The horizontal axis represents the least-squares polynomial fitted to the data, while the curves illustrate the residuals due to drift and hysteresis.

Figures 7 and 8 show similar results from tests run on a slightly different version of the DMXO. The two curves were run on successive days after 4 days of aging cycles. The results indicate that the stability of the sustaining amplifiers is adequate to the task and that the hysteresis and excessive residuals shown by calibration runs on many resonators must be due to the resonators themselves.

Crystal Oscillator Curve Fit, Stability and Hysteresis

Although crystal oscillators have been operated over wide temperature ranges for decades, it has not been possible to examine their performance with any great precision until the advent of the dual-mode oscillator [2, 3, 4]. The current areas of concern to the development of resonators for the MCXO are the evidence of hysteresis and of excessively large residuals remaining after high-order polynomial curve fitting. The relative magnitudes of these two effects vary widely from crystal to crystal and no correlations have yet been made with processing or quality control factors.

Examples of calibration runs showing both low and high levels of residuals and hysteresis can be seen in Figures 9 through 14. The residuals are the differences between the accumulated frequency vs. temperature data and the approximating function, which in this system is a 7th-order polynomial that provides a least-squares approximation to the data. The polynomial which is listed above each plot, relates Df_o to N (where N=N1-NO), and is generated using the 3rd overtone frequency data vs. N. The hysteresis is evidenced by the open loop traced by the data points.

The data has been taken using a computer-controlled test chamber programmed to change the temperature in steps, usually 2 degrees apart, from 25°C to -55°C, then reversing and stepping up to +85°C before returning to 25°C. A complete run takes about 8 hours. The crystal and sustaining amplifiers are both in the chamber. The dual-mode design sustains oscillation on both the fundamental and 3rd overtone C-modes simultaneously. A computer program has been written to calculate the number N1 from the inverse difference between the fundamental and the 3rd overtone divided by 3. This becomes the self-sensing temperature analog used by the MCXO to calculate the required correction frequency Df_o.

The residuals are plotted against N as shown. The temperature scale is nominal, since the N number provides the actual temperature information.

Conclusion

It has been demonstrated that this MCXO system is capable of correcting for predictable frequency-temperature variations smaller than 2 ppb. This potential cannot be realized however until sufficiently stable SC-cut resonators have been developed. More than forty resonators have been examined to date, a typical sampling has been shown by the data in Figures 9 through 14. All of these have been fabricated using the best available processing. The units have been sealed by cold-welding into HC-40 cases. Some of these have been assembled without a final frequency plating to determine if the reduced atmospheric exposure provided any improvement. Others have been fabricated from swept quartz to determine if higher material purity could contribute improved stability. No significant improvements have been seen to date, thus the development of the MCXO presents a challenge to the quartz crystal industry which if met, can offer a serious alternative to the ovenized crystal oscillator.

Acknowledgement

The author wishes to acknowledge the support of the U.S. Army's Electronics Technology & Devices Laboratory under Contract DAAL01-88-C-0804.

References

1. A. Benjaminson, "A Microcomputer-Compensated Crystal Oscillator Using a Dual-Mode Resonator," in Proceedings of the 43rd Annual Frequency Control Symposium.

2. S. Schodowski, "Resonator Self-Temperature Sensing Using a Dual-Harmonic-Mode Crystal Oscillator," in Proceedings of the 43rd Annual Frequency Control Symposium.

3. R. Filler and J. Vig, "Resonators for the Microcomputer-Compensated Crystal Oscillator," in Proceedings of the 43rd Annual Frequency Control Symposium.

4. R. Filler, J. Messina and V. Rosati, "Frequency-Temperature and Aging Performance of Micro-Compensated Crystal Oscillators, in Proceedings of the 43rd Annual Frequency Control Symposium.

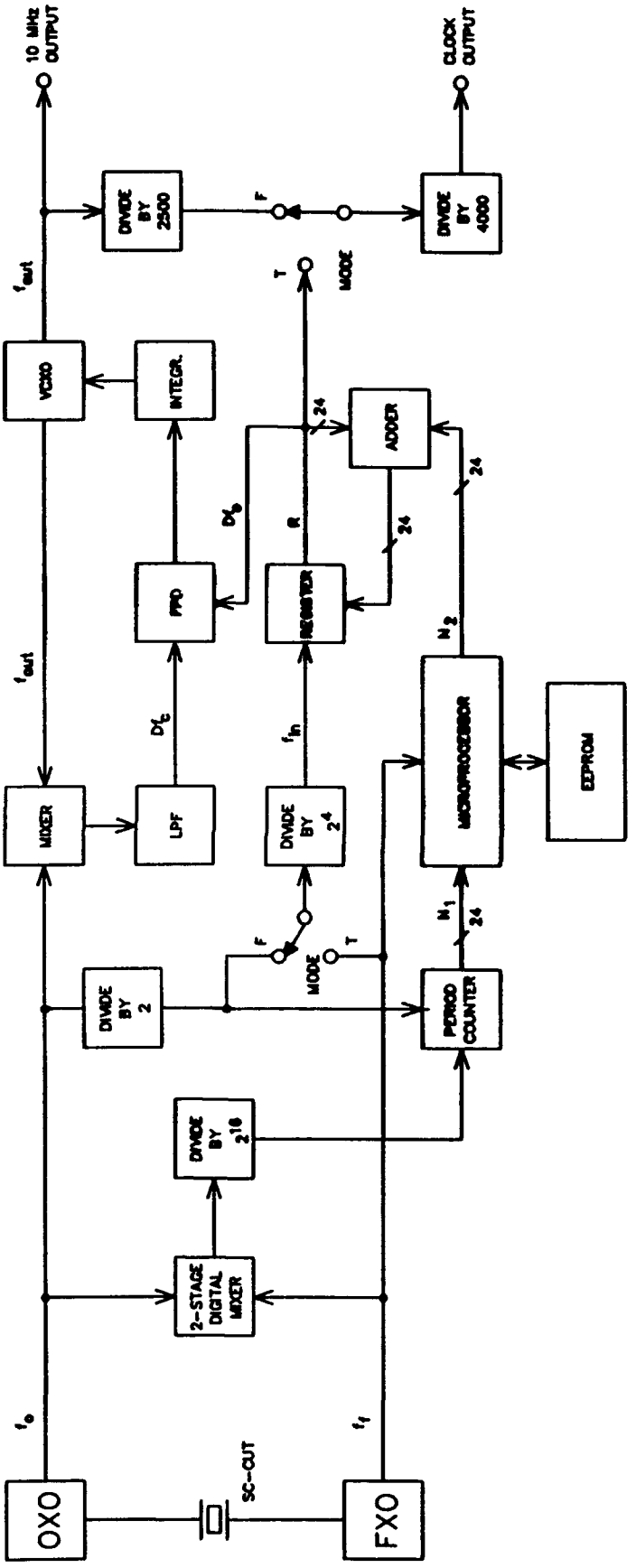

10 MHz OUTPUT
CLOCK OUTPUT
DIVIDE BY 2500
DIVIDE BY 4000
F
T
MODE
VCXO
INTEGR.
ADDER
24
PFD
REGISTER
MIXER
LPF
DIVIDE BY 2^4
MICROPROCESSOR
EEPROM
DIVIDE BY 2
PERIOD COUNTER
DIVIDE BY 2^16
2-STAGE DIGITAL MIXER
OXO
FXO
SC-CUT

Figure 1.

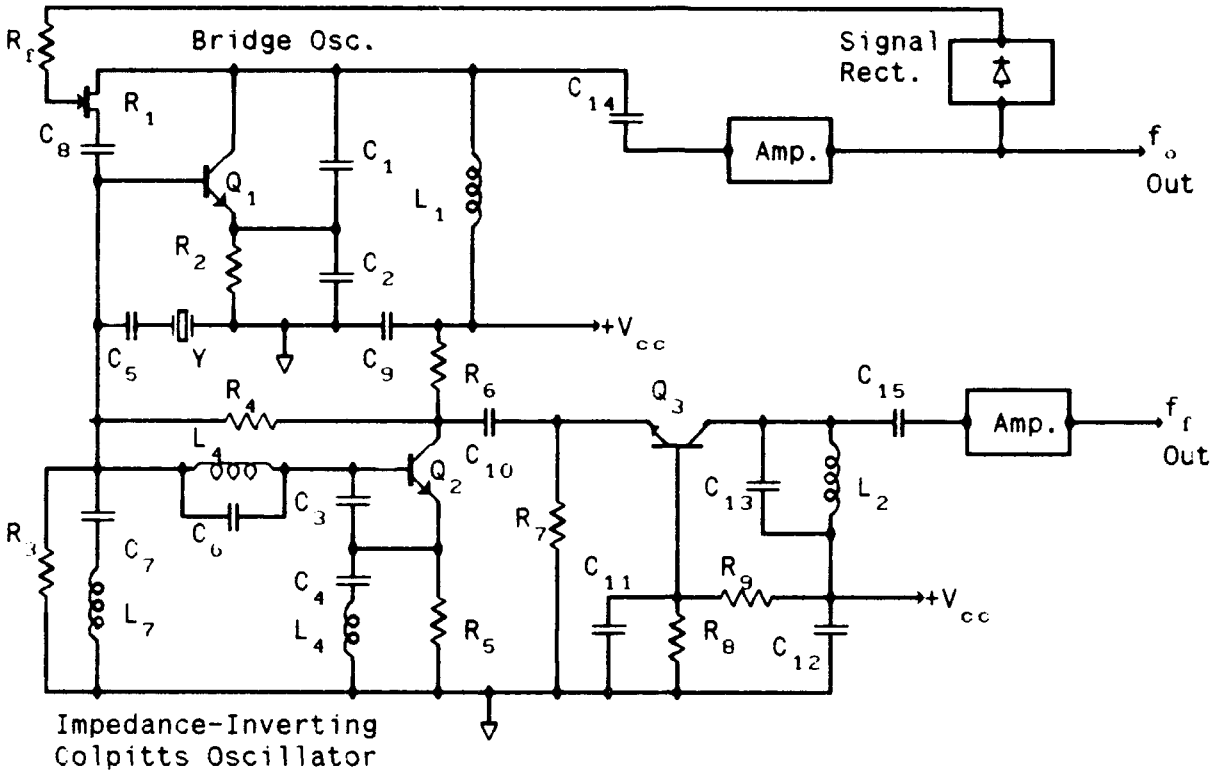
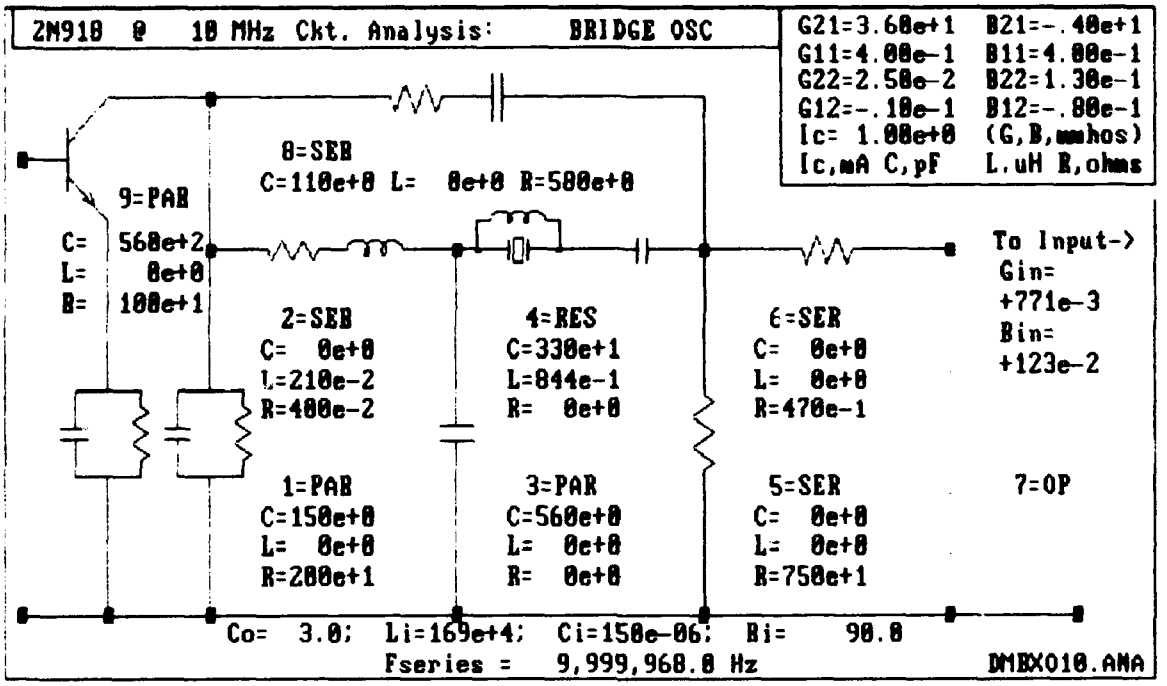
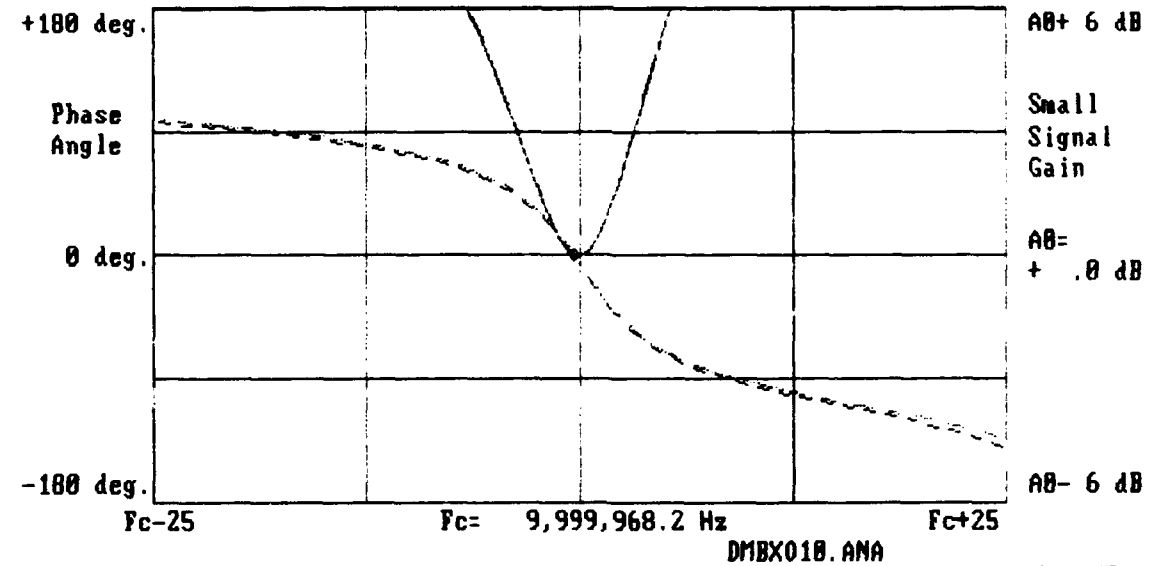


Figure 2. DUAL-MODE OVERTONE CRYSTAL OSCILLATORS



Analyze [A] Change [C] Span [F] Save [S] Design [D] Main [M]



BWu= 8.48 Hz Fosc= 9,999,967.9 Hz Px= 55.94 uW
Qu=1,178,929 Dfo=+ .10 Hz Ic= 1.00 mA DAo=- .1 dB Pi= 1.01 uW
Ql/Qu= 1.34 Ix= 788.4 uA BRIDGE OSC Vo= 415.50 mV Po= 86.32 uW

Reanalyze [R] Span [F] ALC [A] List [L] Noise [N] Main [M]

Figure 3

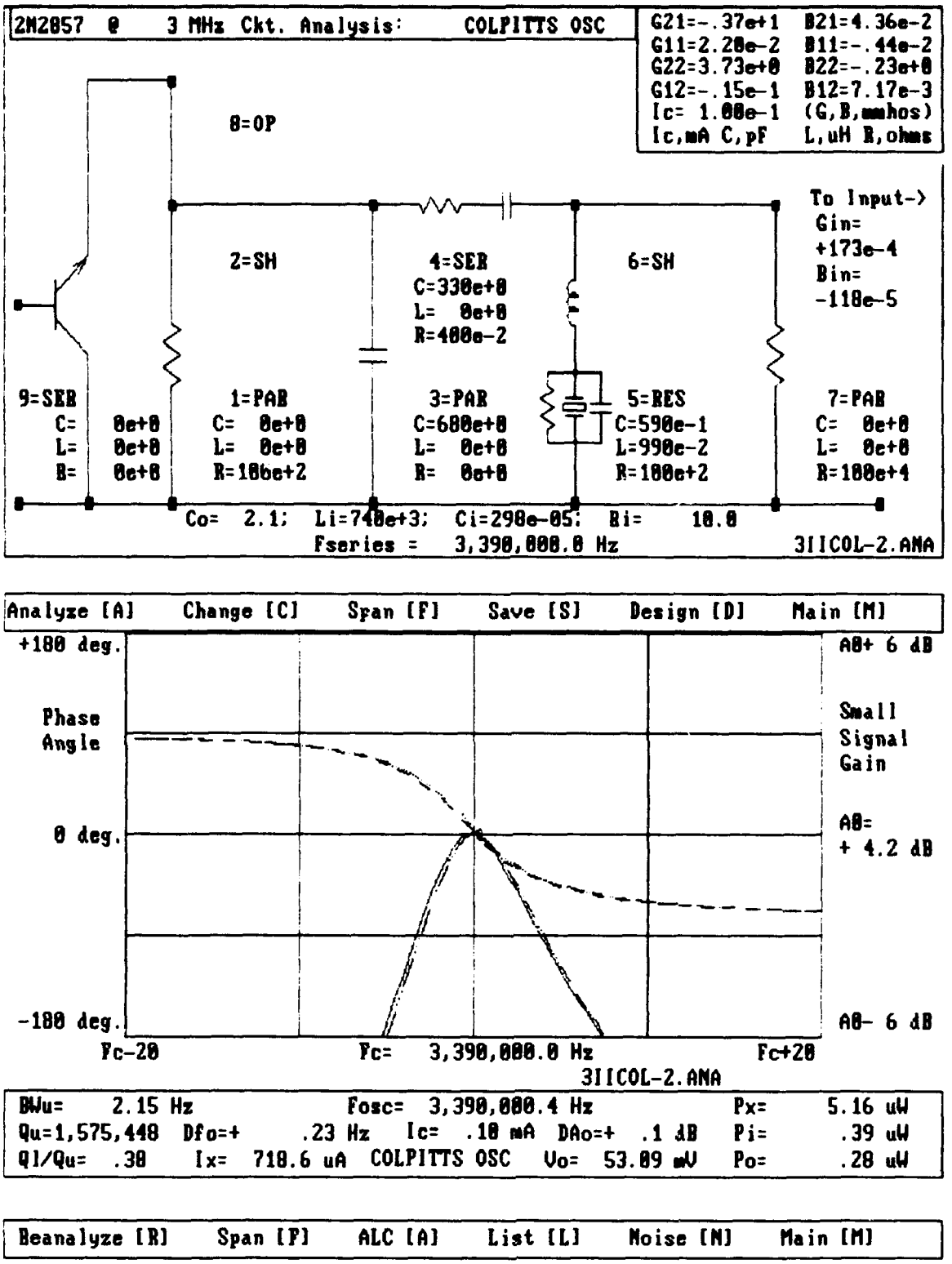

2N2857 @ 3 MHz Ckt. Analysis: COLPITTS OSC
G21=-.37e+1 B21=4.36e-2
G11=2.28e-2 B11=-.44e-2
G22=3.73e+0 B22=-.23e+0
G12=-.15e-1 B12=7.17e-3
Ic= 1.00e-1 (G,B,mmhos)
Ic,mA C,pF L,uH R,ohms
8=OP
To Input->
Gin=
+173e-4
Bin=
-118e-5
2=SH
4=SER
C=330e+0
L= 0e+0
R=400e-2
6=SH
9=SER
C= 0e+0
L= 0e+0
B= 0e+0
1=PAR
C= 0e+0
L= 0e+0
R=186e+2
3=PAR
C=680e+0
L= 0e+0
R= 0e+0
5=RES
C=590e-1
L=990e-2
R=100e+2
7=PAR
C= 0e+0
L= 0e+0
R=100e+4
Co= 2.1; Li=740e+3; Ci=290e-05; Ri= 10.0
Fseries = 3,390,000.0 Hz
3IICOL-2.ANA
Analyze [A] Change [C] Span [F] Save [S] Design [D] Main [M]
+180 deg.
Phase
Angle
0 deg.
-180 deg.
AB+ 6 dB
Small
Signal
Gain
AB=
+ 4.2 dB
AB- 6 dB
Fc-20
Fc= 3,390,000.0 Hz
Fc+20
3IICOL-2.ANA
BWu= 2.15 Hz Fosc= 3,390,000.4 Hz Px= 5.16 uW
Qu=1,575,448 Dfo=+ .23 Hz Ic= .10 mA DAo=+ .1 dB Pi= .39 uW
Ql/Qu= .30 Ix= 718.6 uA COLPITTS OSC Vo= 53.09 mV Po= .28 uW
Reanalyze [R] Span [F] ALC [A] List [L] Noise [N] Main [M]

Figure 4.

*** Polynomial Coefficients, Ff vs Temp. ***

LSA-LCO dmo1m517

Coefficients of Least Squares Approximation
[Where F=A0+A1*T+A2*T^2+A3*T^3+...]

Ao=+3.395243e+06	A1=-1.344766e+02
A2=-5.740010e-01	A3=-5.100818e-03
A4=-4.016435e-05	A5=+3.416258e-06
A6=+1.537282e-08	A7=-3.988309e-10

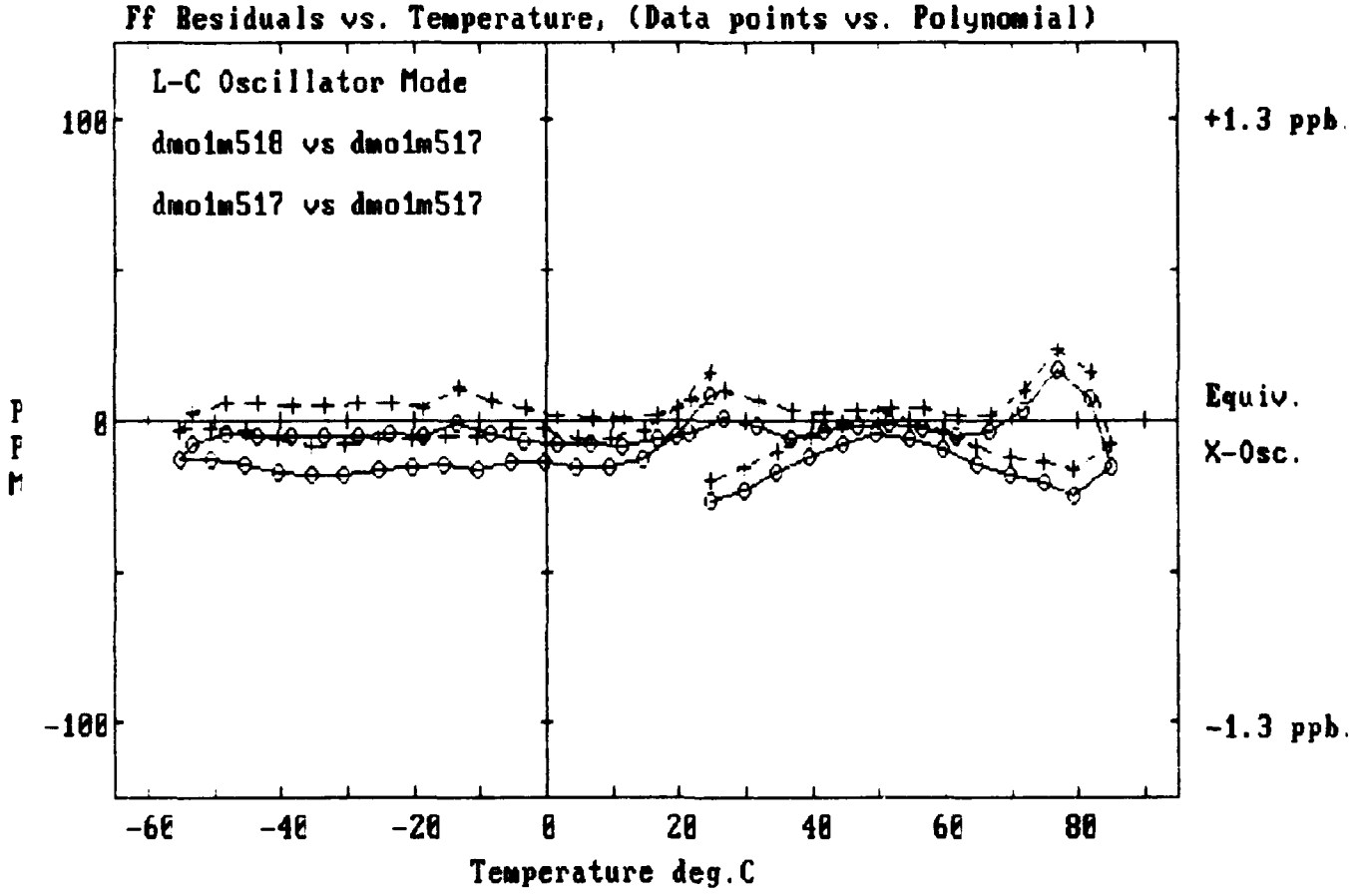


Figure 5.

*** Polynomial Coefficients, Fo vs Temp. ***

LSA-LCO dmo1m517

Coefficients of Least Squares Approximation
[Where F=A0+A1*T+A2*T^2+A3*T^3+...]

Ao=+9.934849e+06	A1=+3.090108e+02
A2=-2.055111e+00	A3=+3.269866e-02
A4=-3.229087e-04	A5=-1.419302e-06
A6=+7.965961e-08	A7=-5.354908e-10

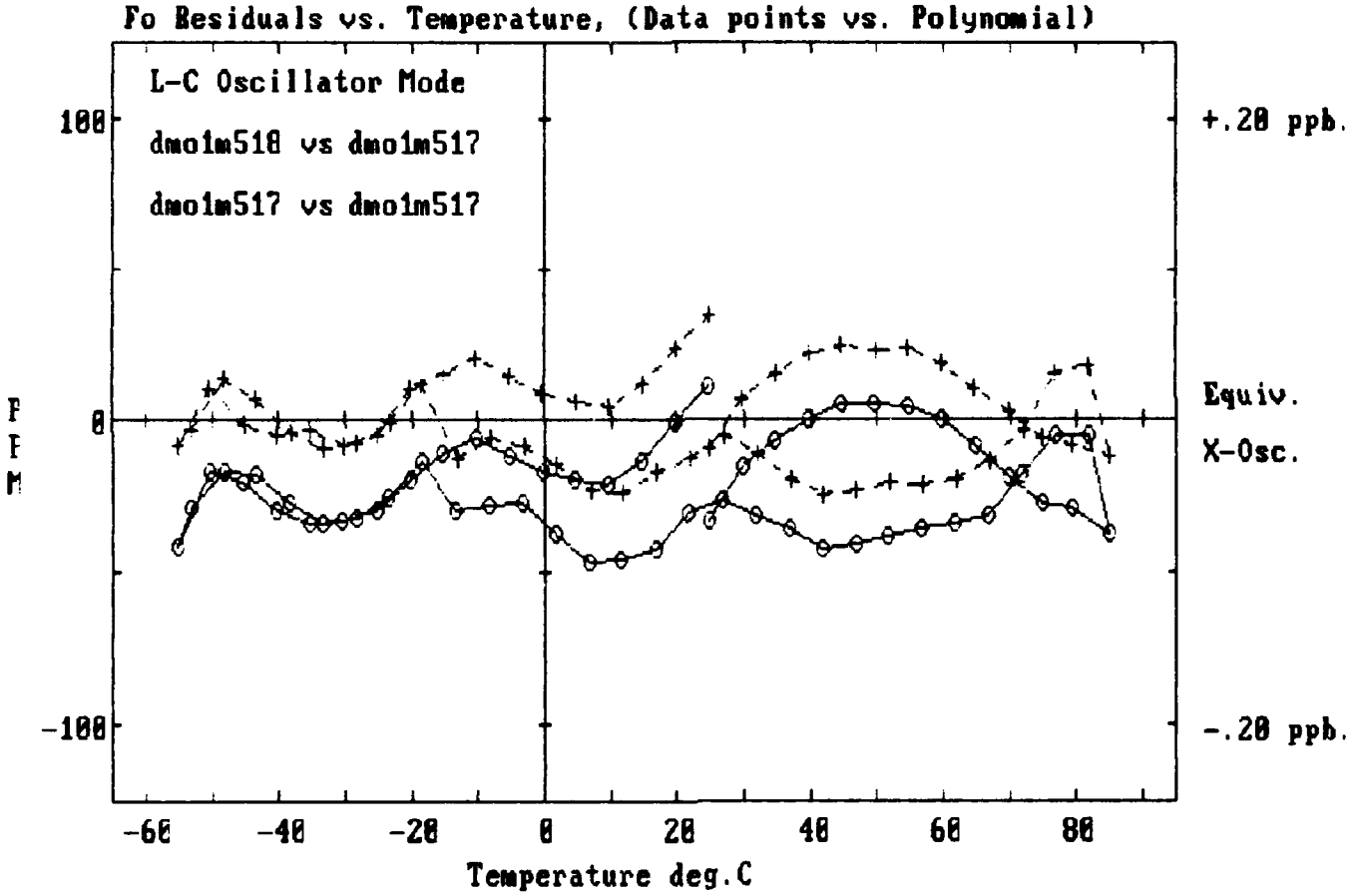


Figure 6.

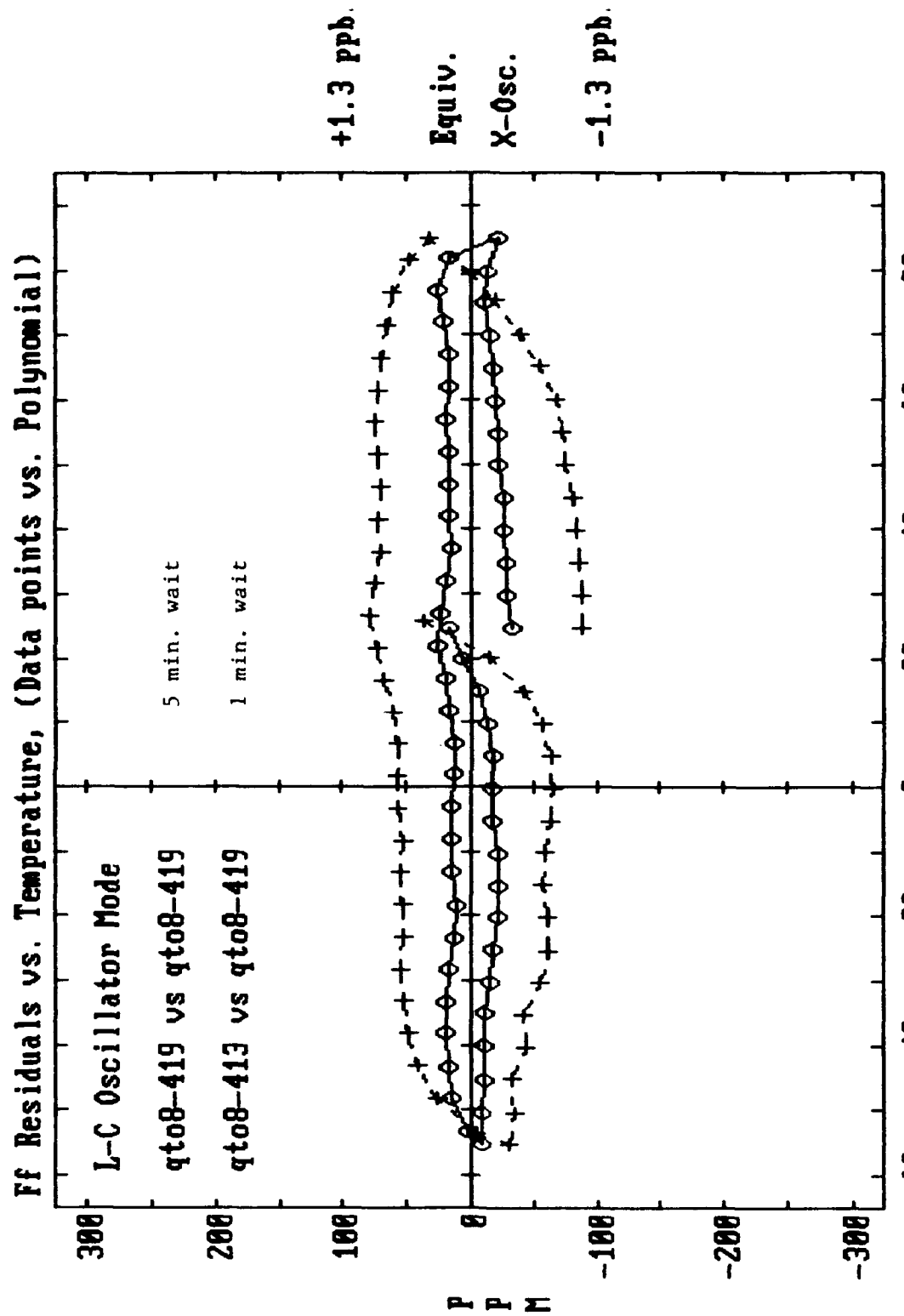

Ff Residuals vs. Temperature, (Data points vs. Polynomial)
L-C Oscillator Mode
qto8-419 vs qto8-419
qto8-413 vs qto8-419
5 min. wait
1 min. wait
+1.3 ppb.
Equiv.
X-Osc.
-1.3 ppb.
P
P
M
300
200
100
0
-100
-200
-300
-60
-40
-20
0
20
40
60
80
Temperature deg.C

Figure 7.

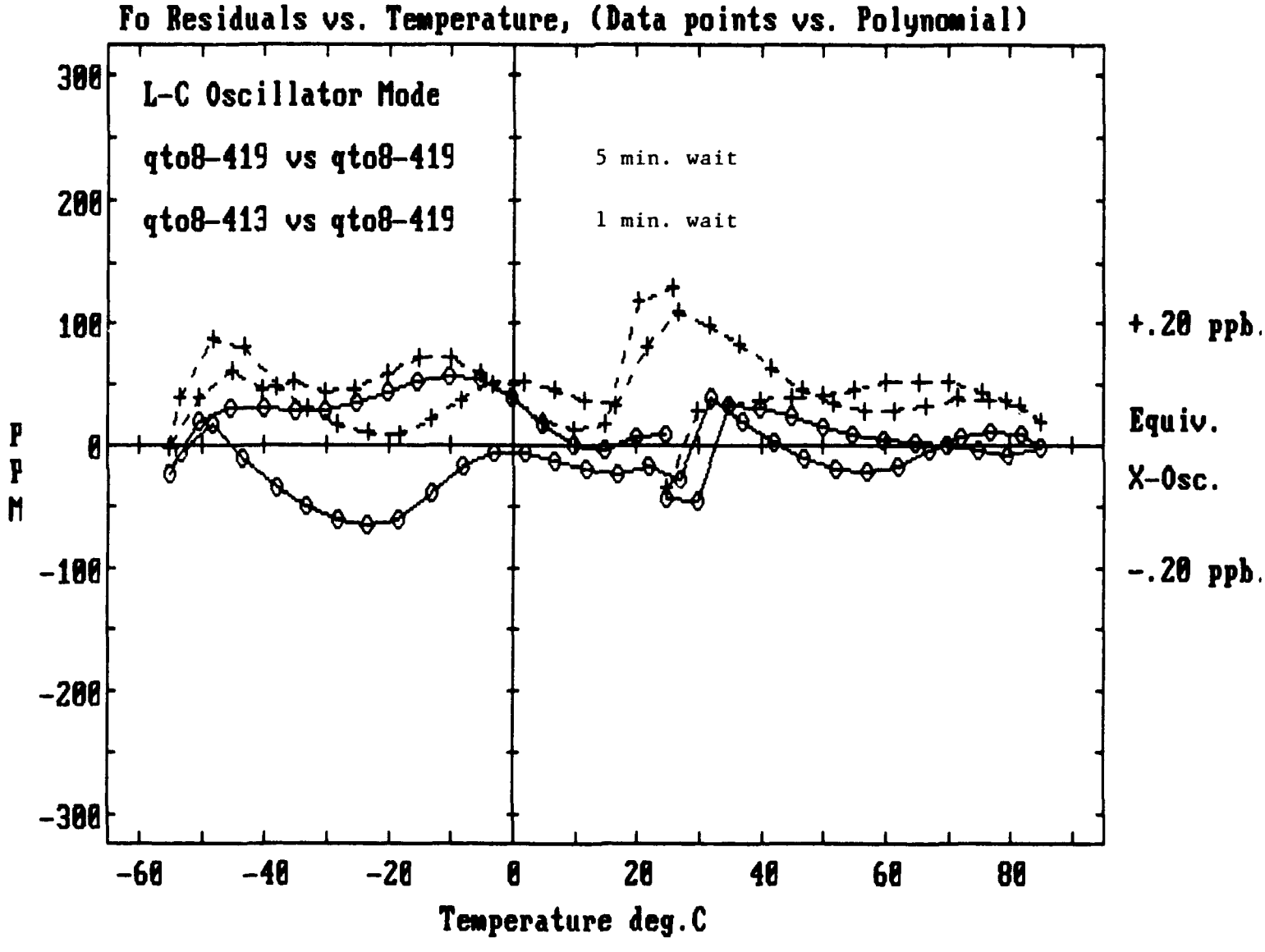

Fo Residuals vs. Temperature, (Data points vs. Polynomial)
L-C Oscillator Mode
qto8-419 vs qto8-419
qto8-413 vs qto8-419
5 min. wait
1 min. wait
300
200
100
0
-100
-200
-300
PPM
+.20 ppb.
Equiv.
X-Osc.
-.20 ppb.
-60
-40
-20
0
20
40
60
80
Temperature deg.C

Figure 8.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S qt1-112

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+2.221207e+02	A1=-8.677494e-04
A2=+1.727012e-06	A3=-4.684418e-11
A4=+3.673765e-16	A5=-3.807997e-21
A6=+1.500552e-25	A7=-6.700403e-31

NO = 2905030

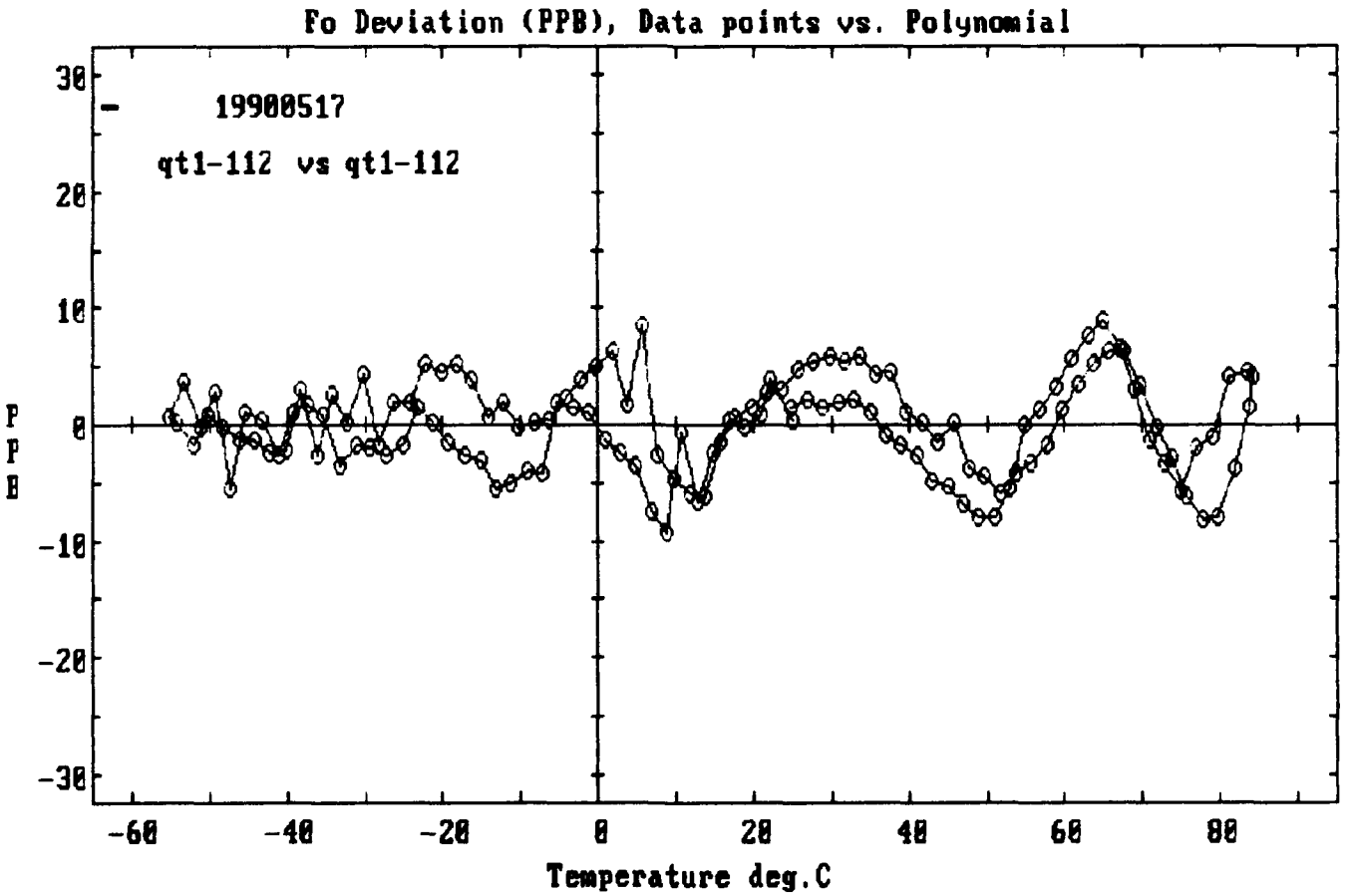


Figure 9.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S sq4-313

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+1.296195e+03	A1=+4.184898e-05
A2=+1.886556e-06	A3=-4.140595e-11
A4=+4.420892e-16	A5=+5.700045e-21
A6=-1.642775e-25	A7=-2.093995e-29

N0 = 3122635

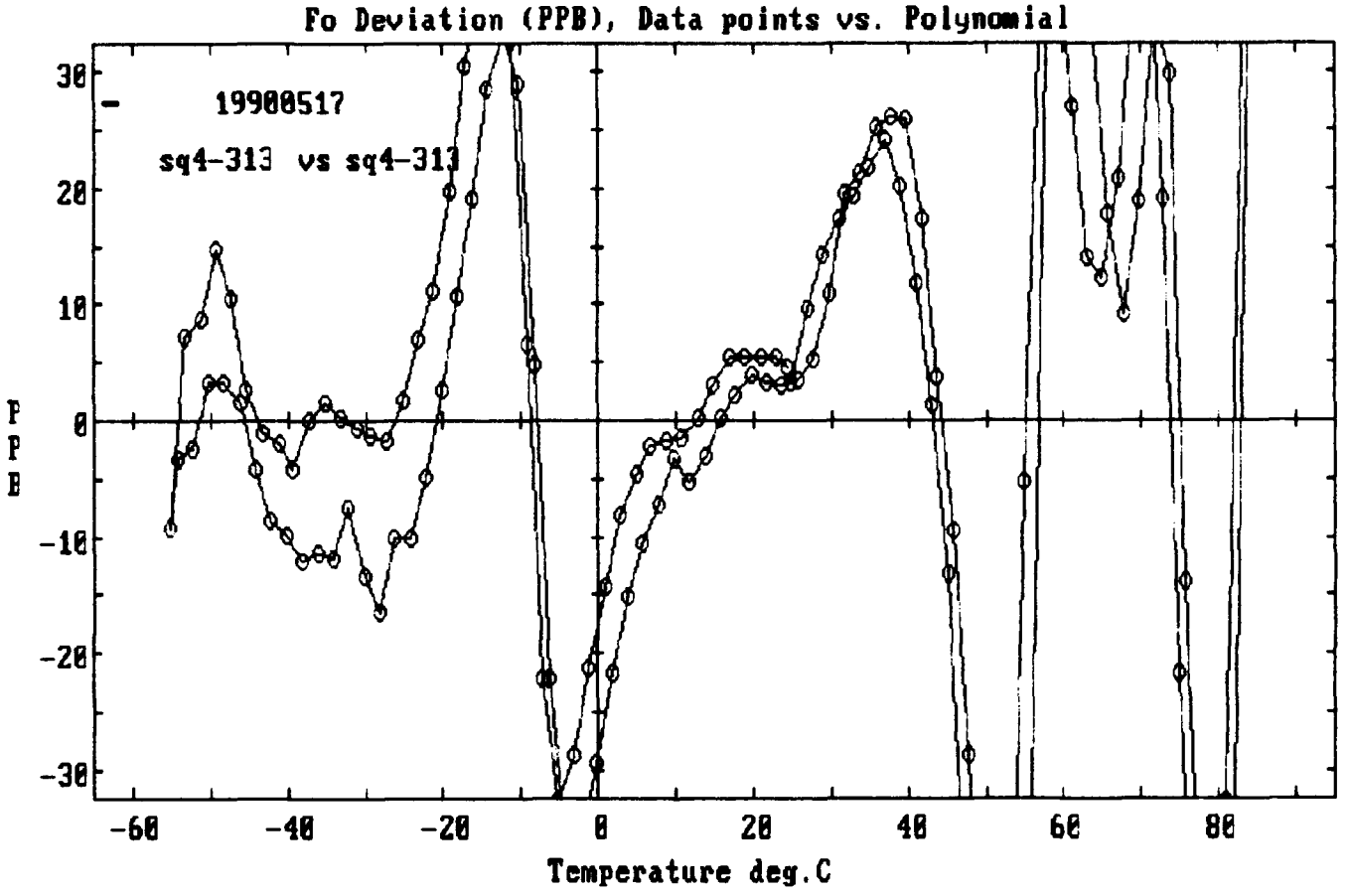


Figure 10.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S sq1-306

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+1.291756e+03	A1=-1.179691e-04
A2=+1.826057e-06	A3=-3.789156e-11
A4=+4.201108e-16	A5=-8.365098e-21
A6=-1.646747e-25	A7=+9.487603e-30

NO = 3183745

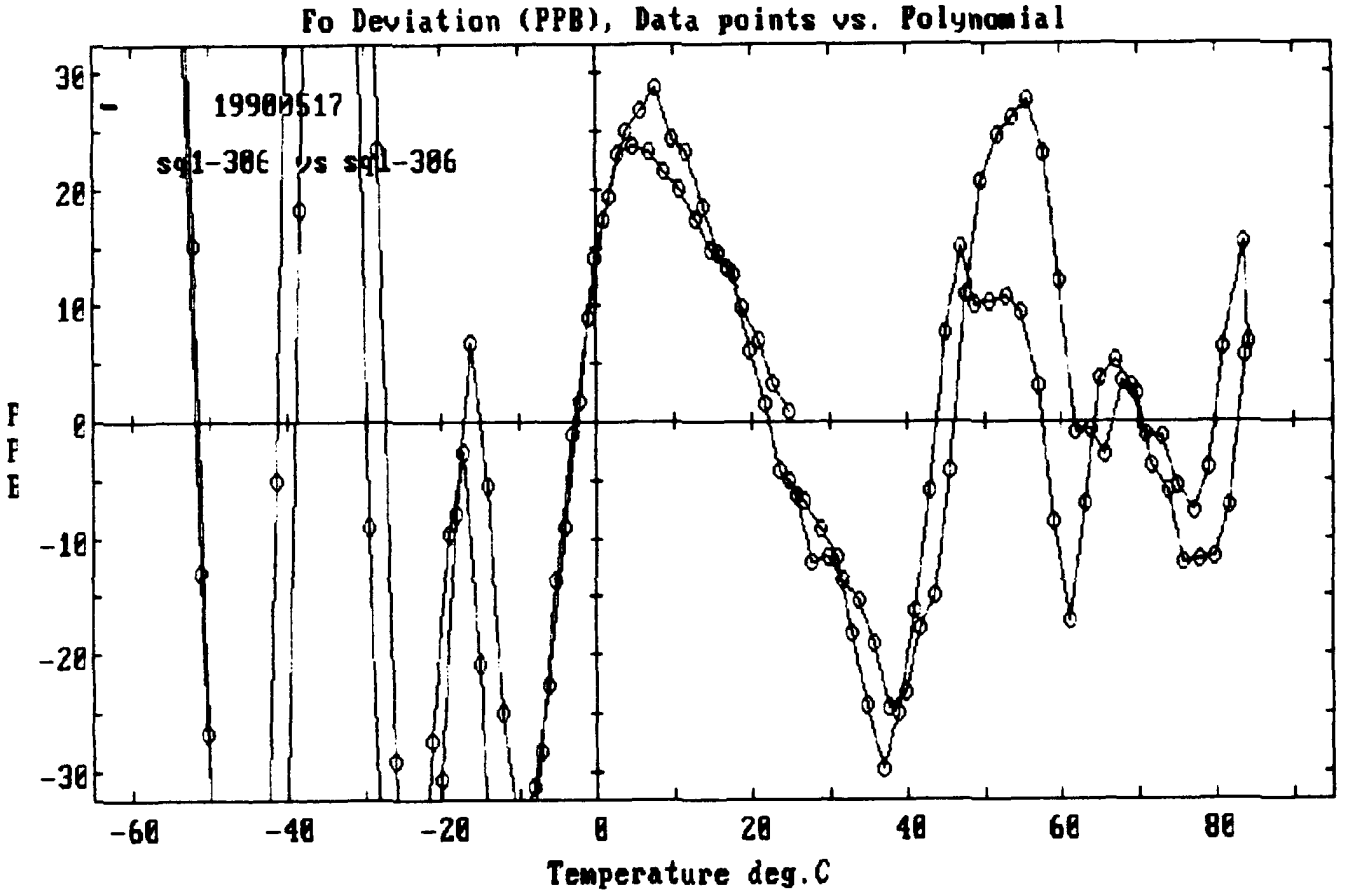


Figure 11.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S pc11-105

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+1.951658e+02	A1=-6.223729e-04
A2=+1.826282e-06	A3=-3.751096e-11
A4=+2.521646e-16	A5=-5.749588e-21
A6=+1.801868e-25	A7=+2.024982e-30

NO = 3150310

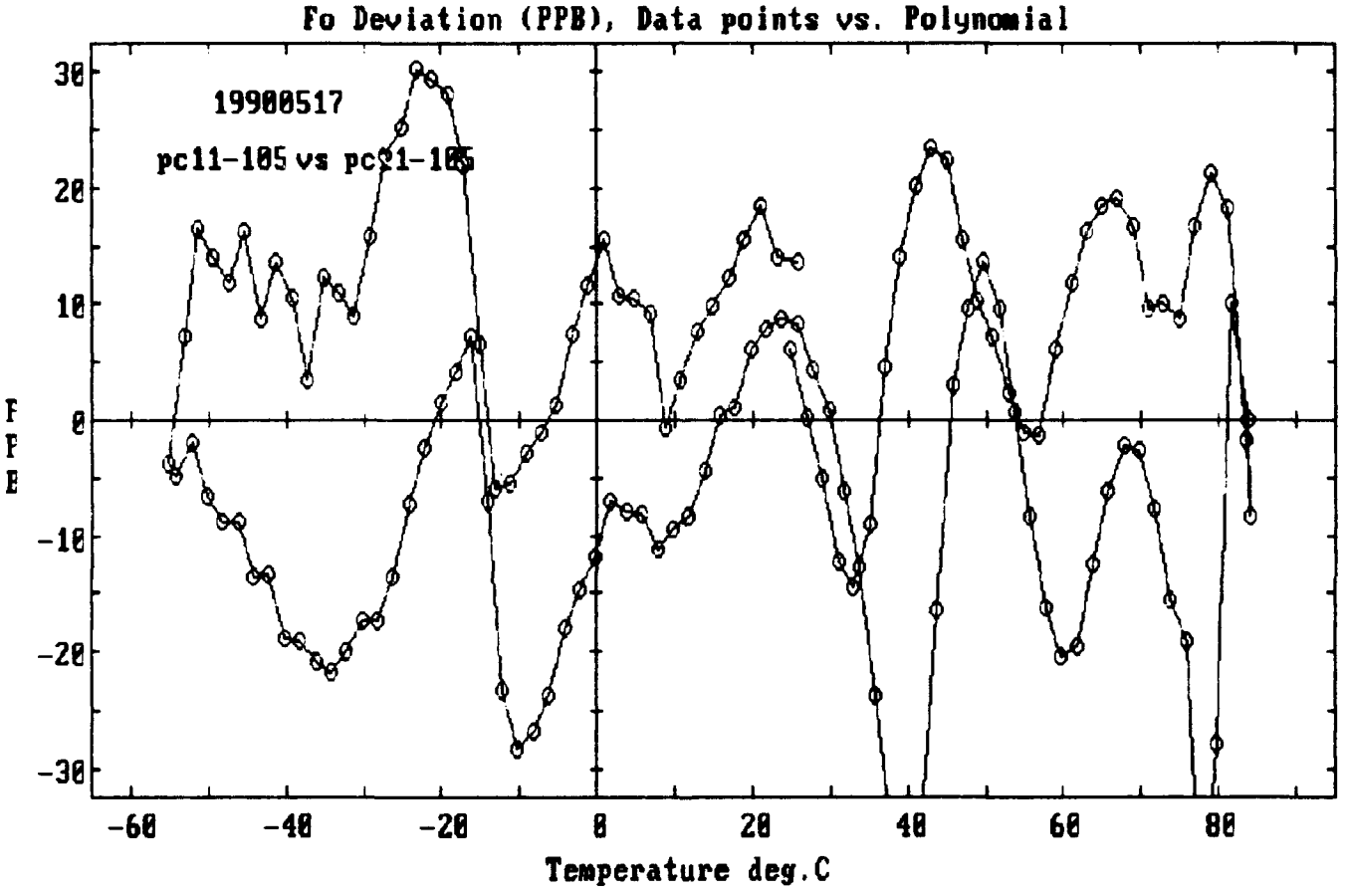


Figure 12.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S pc10-104

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+1.985664e+02	A1=+6.427432e-04
A2=+1.902651e-06	A3=-3.630371e-11
A4=+2.792978e-16	A5=-3.089315e-21
A6=+9.092348e-26	A7=-1.739692e-30

NO = 3240333

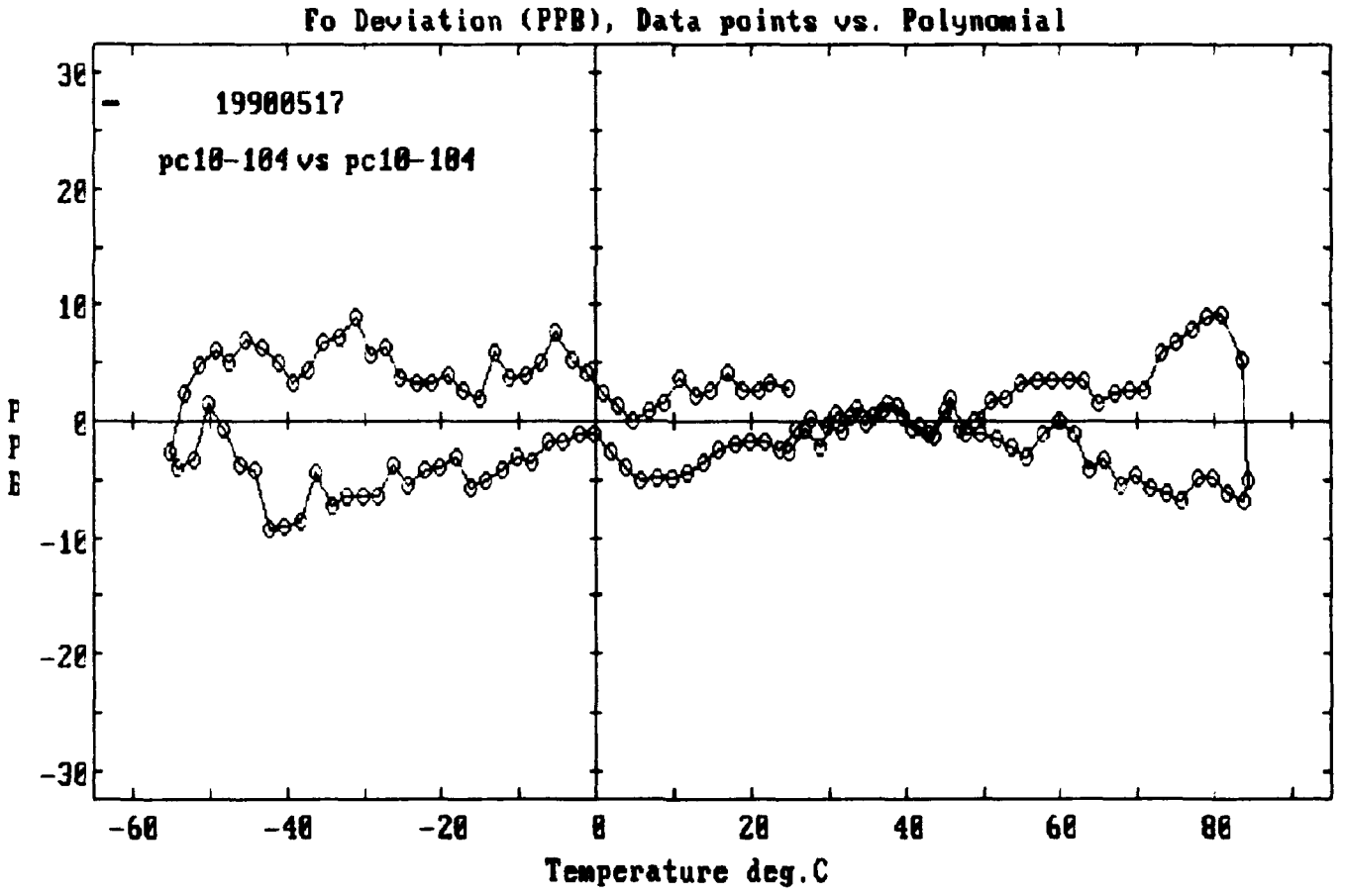


Figure 13.

*** Single-Mixer Based Polynomial, Dfo vs N1 ***

LSA100S cc59-226

Coefficients of Least Squares Approximation
[Where Dfo=fd+A1*N+A2*N^2+A3*N^3+...]

fd=+1.948544e+02	A1=+1.471636e-03
A2=+2.366673e-06	A3=-5.389274e-11
A4=+4.951870e-16	A5=-8.296126e-21
A6=+1.134091e-25	A7=+1.831042e-30

NO = 2876208

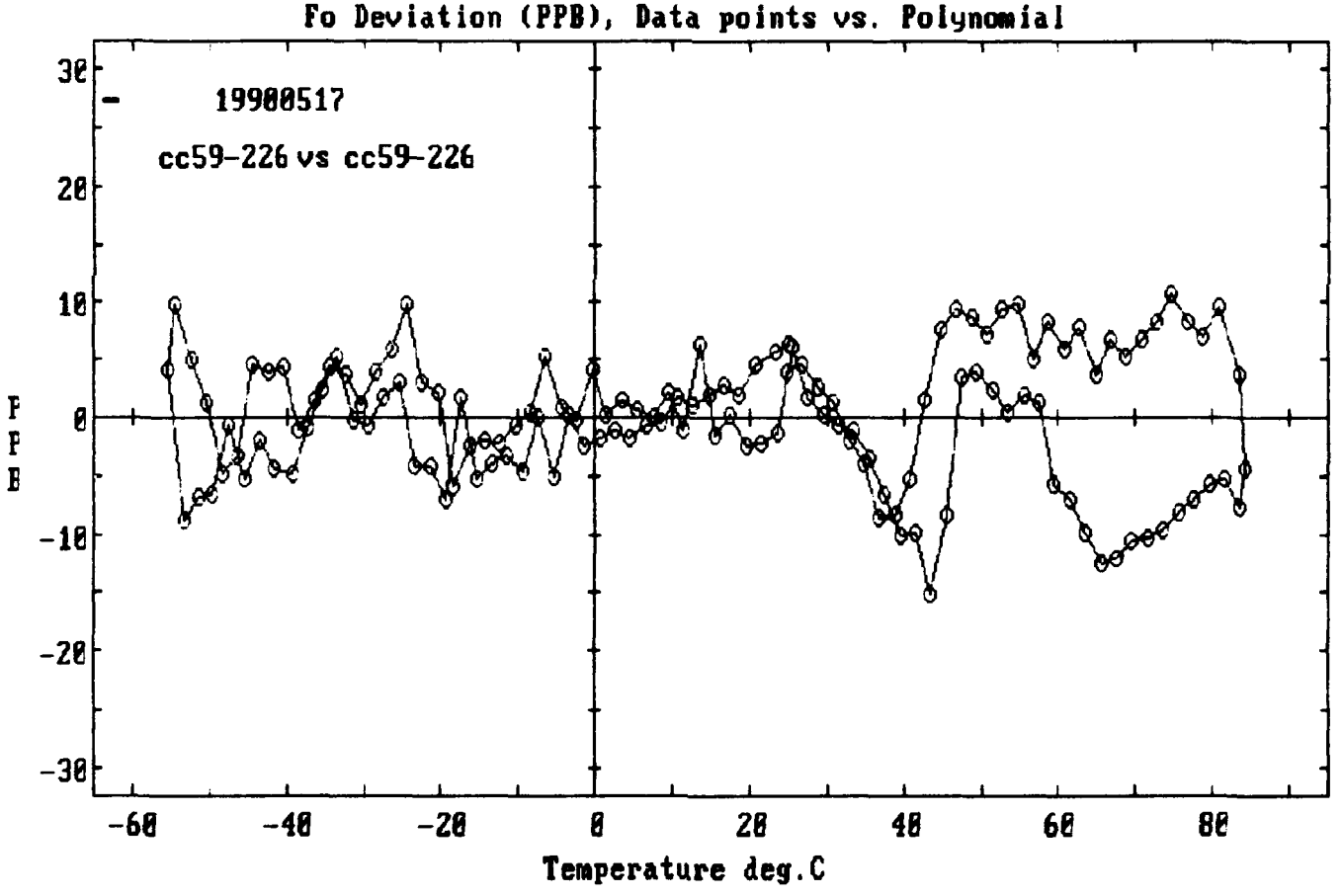


Figure 14.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

AN UPDATE ON THE TMXO

D.A. Symonds & D. Allen

Piezo Technology, Inc.
Orlando Florida 32854-7859

Abstract

The Tactical Miniature Crystal Oscillator (TMXO), is a very small (1.5 cu. in), high stability (< ±5 ppb), fast warm up (as low as 2 minutes), ovenized crystal oscillator using a unique vacuum insulation principle to achieve a very low input power -- 250 mW at -55° C. This paper describes improvements made to the TMXO to enhance performance and improve manufacturability.

Introduction

The Tactical Miniature Crystal Oscillator (TMXO), is a very small, very low power, high stability, fast warm up, ovenized crystal oscillator that is intended for use in a variety of advanced applications. The combination of high stability, small size, and low power consumption afforded by the TMXO is not approached by any other frequency standard.

The TMXO achieves its unique performance as a result of the use of vacuum insulation, hybrid construction, and a special resonator package. Figure 1 shows the mechanical configuration and principal dimensions.

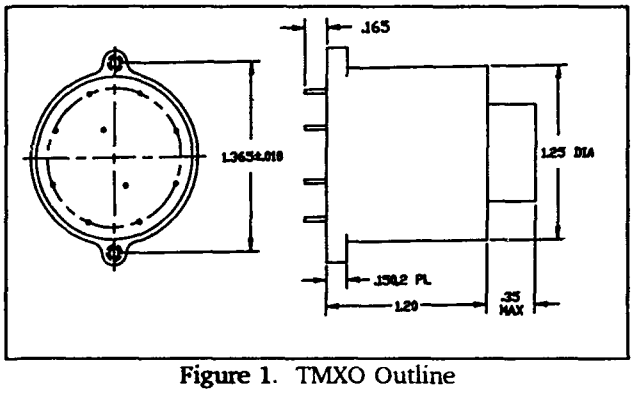


Figure 1. TMXO Outline

The baseline TMXO was developed under LABCOM sponsorship by the Bendix Corporation. It performed well to its original specification, demonstrating the validity of the concept. In the final phase of this effort, Bendix successfully manufactured 100 baseline TMXOs under an MM&T contract.[1,2]

Despite this early success, to make the TMXO commercially available further work was needed. First, certain performance improvements, including the reduction of quiescent phase noise, the elimination of structural resonances, and the addition of radiation hardening were dictated by potential user requirements. Second, manufacturability improvements in order to reduce manufacturing costs were necessary. Third, a manufacturing capability for the TMXO resonator had to be developed. (The crystal unit for the baseline TMXO was a ceramic flatpack SC-cut crystal manufactured at a government-owned facility.[3,4]) In addition, the oscillator was to be re-designed to comply with MIL-O-55310B.

The results to date of these efforts are described in detail in the sections which follow. The TMXO has been radiation-hardened to tactical levels. Phase noise has been reduced to -125 dBc/Hz at 10 Hz offset, with a floor of -160 dBc/Hz. The lowest structural resonance has been increased to 1800 Hz; work on this parameter is continuing. Manufacturability improvements have been numerous and extensive. In addition to a large number of process changes, the thick-film hybrid circuit assembly has been completely re-designed: the package has been improved; the number of thick-film screenings has been reduced from 13 to 2; the circuit itself as well as circuit components have been changed extensively, both to enhance performance and to simplify assembly and alignment. The original header for the TMXO package has been replaced by one which is not only manufacturable but also more reliable and far less prone to failure during oscillator assembly. Finally, the original ceramic flatpack crystal unit has been replaced by a ceramic and metal coldweld surface-mount unit, manufactured by PTI. Principal specifications of the

CH2818-3/90/0000-615 $1.00 © 1990 IEEE

improved TMXO are given in Table 1. Certain of these specifications are in excess of contract requirements to reflect the performance being achieved.

Performance Improvement

Although there is considerable interaction and trade-off between the various tasks they will be described separately.

Circuit design

The TMXO circuit, while adequate for the original specification, was deficient in both phase noise and radiation immunity. The target phase noise specification for the improved design is shown in Table 1 except that the target at 10 Hz was -120 dBc/Hz. The design levels for nuclear exposures were 8 x 10^3 rads (Si) total ionizing dose, 1 x 10^{10} rads (Si)/s peak prompt gamma flux and a neutron fluence of 2 x 10^{12} neutrons/cm^2.

There were several restrictions on the design. Clearly, only semiconductors which are relatively immune to radiation could be used. Transistors were chosen with high values of f_T, to reduce their radiation sensitivity. Because the cost of radiation-immune integrated circuits was excessive, no linear integrated circuits could be incorporated. The voltage regulator therefore had to be made with a zener diode and the temperature controller with discrete transistors. Exposure to tactical nuclear levels irreversibly increases semiconductor leakage currents while reducing gains; therefore, all circuit designs have taken these degraded parameters into account. In addition, to prevent burn-out during prompt gamma radiation, current limiting resistors were placed in all semiconductor power supply paths.

As a further restriction, the hybrid area is very limited, so a large low-pass filter to remove the regulator noise was not possible. Very low power dissipation, less than 30 mW (2.4 mA at 12 V), was required for all the circuits, since the high thermal resistance (approximately 600°C/watt) that results in low oven power also causes the maximum operating temperature to be reduced by 7.2°C per milliampere. Good supply and load immunity were also required.

The oscillator sustaining circuit (figure 2) is a modified Colpitts with base-emitter limiting. To achieve a minimum noise floor, the signal is extracted from the low noise point of circuit[5]. The buffer is a common base stage for good isolation. Critical components have been chosen for low noise and are tested 100%. The total current for the oscillator and buffer is 0.8 mA.

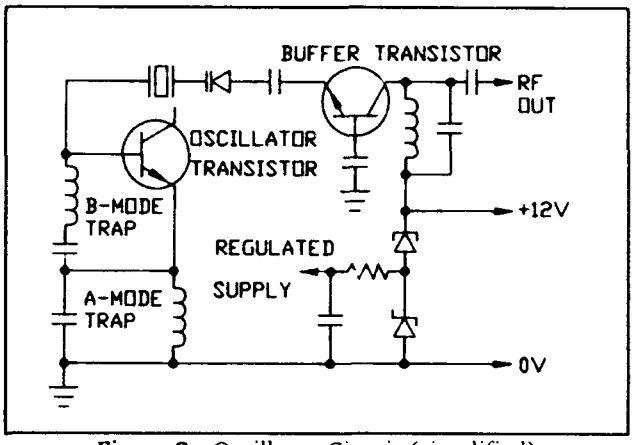


Figure 2. Oscillator Circuit (simplified)

Two output types are available -- sinewave output at 0 dBm and logic output at 5 volt CMOS levels. For the sinewave output version no additional buffer was required and a noise floor of -162dBc/Hz is obtained, figure 3. For the logic output version a discrete transistor squaring circuit is used, which results in a noise floor of about -145 dBc/Hz.

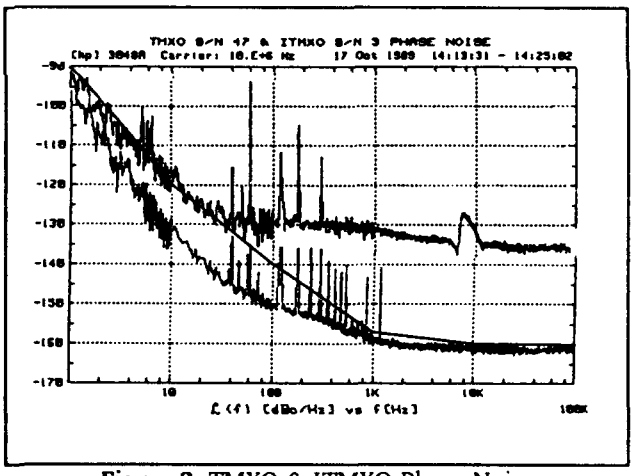


Figure 3. TMXO & ITMXO Phase Noise

The regulator, which supplies the stabilized voltage for the oscillator stage, the temperature controller and the varactor bias, consists of a zener diode with 250 μA of bias current supplied through a current diode. Very low currents are sourced from the regulator to allow the use of an R-C filter to reduce the amount of regulator noise injected to the oscillator/buffer circuit.

Table 1. Principle Specifications

Parameter	Sub-Parameters or Conditions	Specification Value
Frequency Range		4.900 to 12.000MHz Lower with dividers
Input Power	Heater and Oscillator supplies at -55°C Max Warm-up Current Oscillator supply Heater supply Buffer supply (Logic output only)	250 mW max after stabilization 1.0 Amps 12 ± 1.0 volts 12.0 to 18.0 volts 5.0 ± 0.25 volts
Frequency Stability	Temperature:-55°C to +70°C optional to +95°C	$\pm 5 \times 10^{-9}$
	Aging	$\pm 3 \times 10^{-8}$/year, $\pm 2 \times 10^{-10}$/day
	Restabilization at -55°C(warm-up) Retrace	$\pm 1 \times 10^{-8}$ after 3 minutes on time $\pm 2 \times 10^{-9}$ after 4 hours on time, following 24 hours off time
	Supply Voltage ± 5%	$\pm 2 \times 10^{-9}$
	Load Impedance ± 10 pF	$\pm 2 \times 10^{-9}$
	Phase Noise: (Sinewave version)	1 Hz -90 dBc/Hz 10 Hz -125 dBc/Hz 100 Hz -145 dBc/Hz 1000 Hz -157 dBc/Hz 10000 Hz -160 dBc/Hz
	Short Term Stability:	0.1 sec 1×10^{-10} 1.0 sec 2×10^{-11} 1.0 sec 1×10^{-11}
Frequency Adjustment Range	External Resistor or voltage	$\pm 5 \times 10^{-7}$
Output Levels	Sinewave	0 ± 3 dBm in 50 ohms
	Logic	5 volt CMOS
Environmental	Temperature Range, Operating Shock Vibration, Random Acceleration Sensitivity	-55°C to +70°C optional to +95°C MIL-STD-202, Method 213, Cond I, 100g, 6 mS, sawtooth MIL-STD-202, Method 214, Cond 1, Letter C, 9.26g rms $< 1 \times 10^{-9}$/g, total
Mechanical	Size Weight	1.25" diameter, 1.55" height 2 ounces

The temperature controller is fairly conventional except that it is made using discrete transistors and, for radiation hardness, is designed to be able to source sufficient current with degraded gains and has current limiting resistors for gamma pulse protection. The original heater transistor was replaced with a device with higher f_T and good secondary breakdown characteristics. The circuit in the baseline unit was arranged to increase the operating temperature as the set resistor was increased in value. This resulted in burn-up if the externally fitted temperature setting resistor was inadvertently omitted. The circuit is now arranged so that no damage can result from either a short or open. In the baseline design the thermistor had to be selected to suit the operating temperature of a particular crystal. This required a range of tightly sorted thermistors. The circuit is now designed to operate with one thermistor value which simplifies material control and allows for assembly into the hybrid before the crystal is selected. The warm-up time is set by the peak heater current allowed during warm-up. A peak current of 1.0 amps results in a warm-up time of 3 minutes, whereas 1.5 amps gives a 2 minute warm-up.

The resulting frequency-temperature performance is excellent, as shown in figure 4.

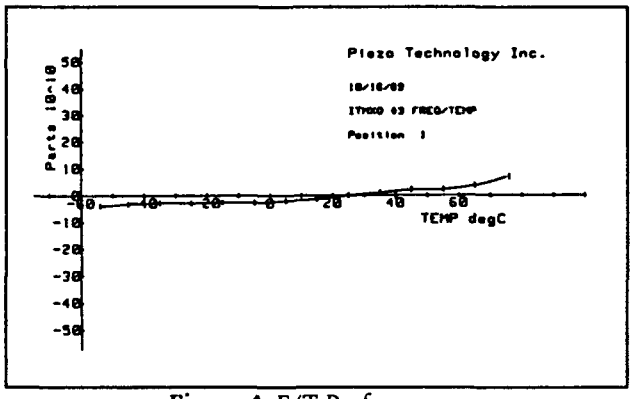


Figure 4. F/T Performance

Nuclear Survivability

In addition to the circuit changes just described, to achieve radiation hardening the TMXO uses a swept, Premium-Q quartz resonator to reduce the effect of ionizing radiation. Sweeping is performed in-house.

Analysis and limited testing of TMXOs of similar design shows that the design will still perform to specification after exposure to the nuclear environment cited above.

Acceleration Sensitivity

Phase noise under vibration is set by the G sensitivity of the resonator[6] and modified by resonances of the structure. A program requirement is that there shall be no structurally-induced degradation below 2000 Hz. The baseline TMXO had resonances at around 700 Hz and 1400 Hz which are believed to be due to the support structure.

To remove these effects structural resonances must be greater than 2 kHz. Therefore the structure must be stiffened, but, to maintain low input power, this must be done without increasing the thermal conductance of the support structure. The method adopted was to replace the solid support wires with thin-walled tubes having the same cross-sectional area, giving a calculated stiffness improvement of 50% (see figure 5).

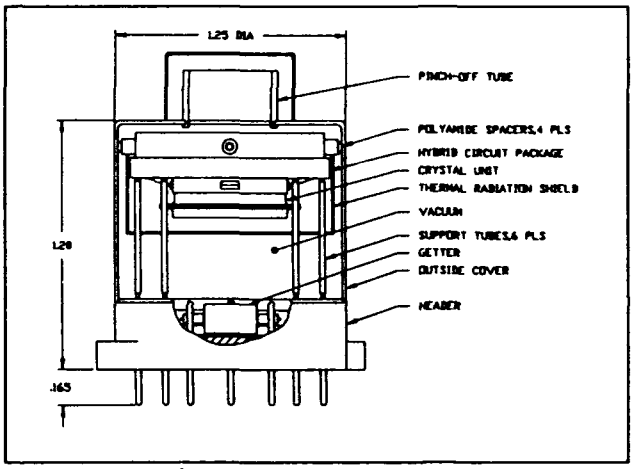


Figure 5. Support Structure

The supported mass was reduced by changing resonator packages, but this reduction is partly offset by the increased mass of the new hybrid package. Vibration measurements show that there is still a resonance at 1800 Hz. Further improvement is still desired. Future plans call for a complete structural re-design.

Producibility

The cost of the TMXO is made up of four main components:

1) Components
2) Assembly labor
3) Yield, especially at certain critical stages
4) Testing of the finished unit.

Of these, the last is determined primarily by requirements of individual customer programs, and will not be considered here. To address the first three, numerous design and process changes have been made.

At the outset of the program, major cost elements had been identified. As manufacturing experience was gained with the baseline design, additional cost elements came to light. Briefly, the price of certain critical components from the suppliers had escalated while the quality had degraded. These components, the crystal package, the hybrid package and the header, were redesigned to reduce costs. A poor yield at the later stages of TMXO assembly was very expensive, since re-work was very difficult and resulted in scrapping of a substantial portion of the assembly. Additional tests were incorporated to ensure that the TMXO works correctly before sealing and the parts were ruggedized to allow for non-destructive rework. The cover attachment process, which had given very poor yields was extensively modified.

Hybrid Package

With the original octagonal package, we experienced continuing problems with pins falling off, often after the hybrid had been completed. Consultation with the vendor led to a re-design to accommodate their standard attachment process using headed pins. In order to keep the pins in the same positions more surface area was required, so the package was changed to a slightly larger circular one. In addition, the surface finish is more closely controlled, and additional metallization has been added to reduce radiant heat losses. The oven power saved is approximately 20 mW at -55°C.

Hybrid Assembly

The number of screenings (13) was extremely high, due mainly to the range of values of screened-on thick film resistors. In the redesign the number of layers is reduced to 2 -- one metallization and one dielectric, reducing the time and cost to print substrates considerably. Thin film chip resistors replace thick film, removing the necessity for trimming resistor values -- a further cost savings -- while reducing resistor-generated noise and allowing resistor values to be changed without making substrate changes. The low-value thick-film resistor which is used to limit the maximum current during warm-up occupied a large area of the hybrid substrate. This resistor was moved to the heat spreader to save space on the substrate and to allow its power loss to be dissipated on the heat spreader.

In the baseline design a tapped coil for mode selection was wound on a plastic form which degraded under continuous heat. In addition, the poor coil design made it difficult to set up the hybrid as the coil inductance changed when the hybrid cover was installed. It was also large and was mounted over some other components, making them inaccessible. The coil was replaced with 2 chip inductors, making the suppression of the unwanted upper and lower resonator modes independent and reducing the sensitivity of the main mode frequency to the mode suppression. As a result of these changes, set-up became very simple.

To set the nominal frequency a multielectrode capacitor was used in the baseline design. It occupied a large surface area and it also required many external connections restricting layout. A capacitor is now selected to set the operating frequency after attaching the resonator but before sealing the hybrid. Fine trimming is obtained with the varactor diode which has increased range; this allows adjustment to final frequency after the TMXO is sealed.

Crystal Unit

The baseline TMXO used a third-overtone, SC-cut resonator housed in a three-piece, all-ceramic package sealed by thermo-compression bonding.[3,4] These crystal units were manufactured in a government-owned facility and were not available commercially. In the initial phases of this effort the technology for this process was transferred to PTI. After evaluation, it was determined that the cost of manufacturing this configuration was excessive, and not warranted by the performance requirements.

Accordingly, a cold-weld surface-mounting ceramic package has been developed, figure 6, enabling the use of a standard sealing process. A disadvantage is

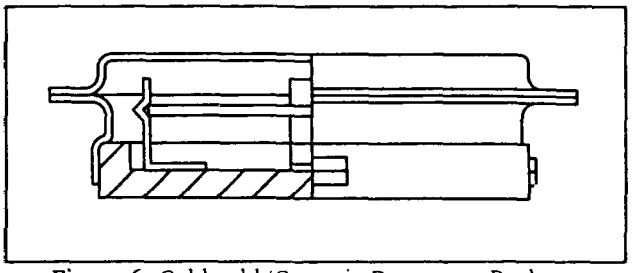


Figure 6. Coldweld/Ceramic Resonator Package

that the new package has only one-sided access to the crystal blank, making cleaning more difficult. Advantages

include low weight and low thermal radiation losses, due to gold plating of the cover. It is capable of high temperature processing. The package appears to work well. For five units, the mean acceleration sensitivity (total vector) was 4 x 10^{-10}/g, with a standard deviation of 0.8 10^{-10}/g. Typical aging (<2 x 10^{-10}/day) and retrace (<1 x 10^{-9}) are shown in figure 7.

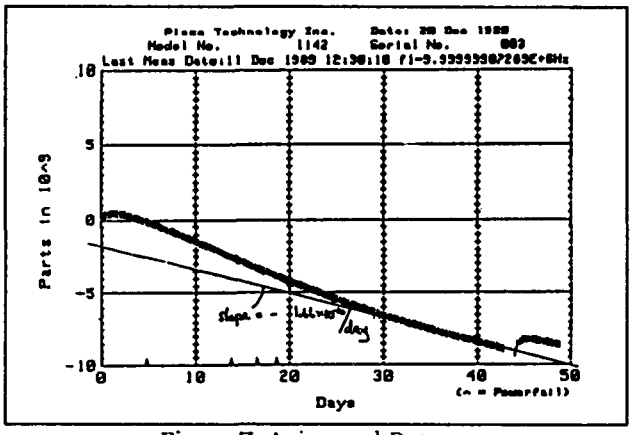


Figure 7. Aging and Retrace

Oscillator Package

Header The TMXO attains its low power mainly as a result of the vacuum insulation. To remain a low power device through its life it must retain the vacuum. A leak rate of around 1 x 10^{-12}/atm cm^3/sec can result in a reduced lifetime for the oscillator so it is imperative that low leak rates be obtained. One of the greatest difficulties in making the TMXO is in obtaining low leakage circuit feedthroughs. In this design the feed-throughs are made at the end of the oscillator with a header which is microplasma welded to the cover. The most expensive part in the TMXO is the vacuum header. At incoming inspection only 50% of the original design of header were surviving the tests; this doubled the effective cost of the header. Moreover, despite pre-acceptance screening the header often failed during and after assembly resulting in a rejected unit. This had, of course, a dramatic effect on the unit cost.

The baseline header consists of a large, circular, ceramic plate fitted in a Kovar hoop. This assembly has high residual stress, making it prone to failure. Further, uneven thermal stresses are produced during sealing with the micro-plasma welder. Still further, completed oscillators tend to be fragile, since any deformation of the base introduces stresses. This was especially significant as the studs were not always quite square and mounting could result in failure. In addition the circuit traces for the buffer were printed on the ceramic plate of the header; the header had to be modified to change the output buffer type.

The header was redesigned to have lower stress and be more rugged. The new header is machined from a single piece of Kovar. Brazed-in individual ceramic feed-throughs are used, as well as a brazed-in ground pin. The studs previously used have been replaced with integral threaded lugs which allow mounting with either screws or studs. A section of the wall has been thinned to reduce thermal shock during plasma welding. The buffer circuit pattern has been replaced with a thin pcb, allowing the buffer to be changed without modification of the header. The header is manufactured by PTI, using a hydrogen/vacuum brazing process to install the ceramic feed-throughs and the ground pin.

Heat sinking during Plasma welding was considerably improved by fixture re-design to reduce thermal shock.

Getter: The getter sometimes became detached during vibration or rough handling. The cause was metal fatigue at the Kovar/tungsten joint. An intermediate metal (inconel) was added as a sleeve around the tungsten wire to prevent the formation of brittle intermetallics. There have been no subsequent getter failures.

Outside Cover: There were two problems caused by the plating of the outside cover: 1.) In the pinch-off region, during pinch-off the plating cracked and produced loose particles, which could cause an electrical short. This was due to the nickel plating under the gold being very brittle. Corrective action was to change from electroless to electrolytic nickel plating which is more malleable. 2.) Pitting occurred in the plasma weld area during the weld, sometimes causing a leak. The cause was traced to phosphorous which was deposited in the electroless nickel plate. Plating was masked from this area and a clean, leak-free weld is now obtained.

Future Plans

The only significant deficiency remaining in the TMXO is its performance under vibration for frequencies greater then 1500 Hz. To solve this problem a structural re-design is being investigated. The current concept is to replace the tubes and polyamide bumpers with a low conductivity cylinder with thick film traces to connect to the header. We presently intend to incorporate active acceleration compensation via external electrical circuitry once the resonances have been removed.

Engineering models of the improved TMXO have been delivered. Twenty Confirmatory Samples containing all the improvements will be built after completing the engineering. Closely similar models are currently being manufactured and delivered to the OEM market.

As a lower-cost alternative, we are developing a larger, foam-insulated version, using the same crystal unit and hybrid circuitry, but a different hybrid package. An early version of this configuration has been made using standard polyurethane foam. We plan to replace the polyurethane with microporous[7] insulation. This alternate configuration is expected to be a substantially lower cost unit, but with somewhat higher power consumption and increased size.

Acknowledgment

The work reported here was supported by the U.S. Army LABCOM under contract DAAL01-86-C-0011. The authors wish to thank Dr. J. R. Vig and Dr. R. L. Filler for their support and encouragement of this effort.

References and Notes

1. D. Brown, E. Laszlo, R. McGill, and P. Stoermer, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillator," Proc. 38th Annual Symposium on Frequency Control, pp. 380-386, 1984.

2. R.L. McGill and F.C. Krochta, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillators: Final Report," Allied Bendix Aerospace, R&D Tech. Rpt. SLCET-TR-82-0277-F, prepared for Electronics Technology and Devices Laboratory, U.S. Army LABCOM, Ft. Monmouth, N.J., March, 1987.

3. P. Wilcox, G. Snow, E. Hafner, & J.R. Vig, "A New Ceramic Flatpack for Quartz Resonators," Proc. 28th Annual Symposium on Frequency Control, pp. 202-210, 1974.

4. J.M. Frank, "Vacuum Processing System for Quartz Crystal Resonators," Proc. 35th Annual Symposium on Frequency Control, pp.40-47, 1981.

5. Robert Burgoon & Robert L. Wilson, "Design Aspects of an Oscillator Using the SC-Cut Crystal," Proc. 33rd Annual Symposium on Frequency Control, pp. 411-416, 1979.

6. R. L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: a Review," IEEE Trans. Ultrasonics,Ferroelectrics, and Frequency Control, v. 35, no. 3, pp. 297-305, May, 1988.

7. Micropore International, Ltd., Hadzor, Droitwich, Worcs, England.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

SUPERIMPOSING LOW-PHASE-NOISE, LOW-DRIFT INSTRUMENTATION TECHNIQUES ON RF DESIGN*

C. M. FELTON
Time and Frequency Division
National Institute of Standards and Technology
325 Broadway, Boulder, CO 80303

Abstract

The information required to do good low-phase-noise design is, for the most part, already in the literature under different titles. Low-noise audio design is concerned with optimizing amplitude signal-to-noise ratio. Instrumentation amplifier design isolates the desired signal, using bridge configurations, (control of common mode rejection ratio (CMRR) and power supply rejection ratio (PSRR)) from interfering noise and line harmonics.[1,2,3] Concerns with operating point stability at the microvolt level lead to awareness of dielectric quality, thermal stability, and bias balance requirements. At a large laboratory there are numerous sources of electromagnetic interference. Switching power supplies, clock and computer radiation, motor start-up surges, and fluorescent light spikes are as hostile an electromagnetic environment as a power utility substation. Critical applications require that such interference be considered in the basic circuit design. A balanced circuit resembles (or can be successfully encouraged to do so by the designer) a parallel, two wire transmission line, providing for rejection of external electromagnetic noise.

The focus of this paper is integrating these considerations into basic designs. A radio frequency (RF) isolation amplifier will be used to illustrate the concepts. I will discuss component selection and circuit details of the completed device and show how these affect both AM and phase noise (PM) performance.

RF design in a phase noise context

It is important to have full-time access to a relatively low-noise source (crystal oscillator), and equipment to measure amplitude and phase noise. Good low-amplitude noise design is not the same as low phase-noise design, although the two are related. To aid my own visualization, I think of phase modulation occurring in series and shunt modes. (Figure 1)

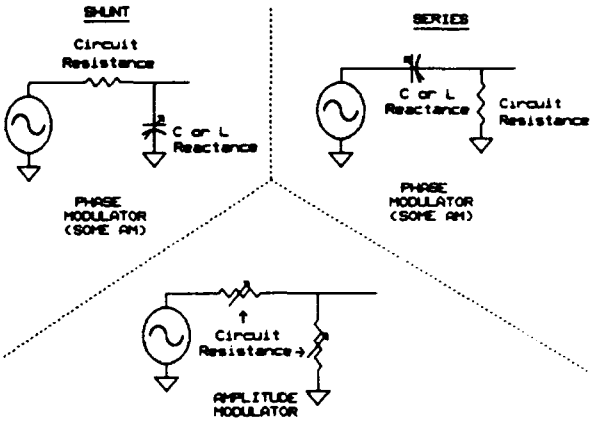


Figure 1. This is an aid to visualization. In a practical circuit, all 3 of these will be occurring simultaneously to one degree or another.

To the extent that a circuit element or stage is nonlinear with frequency, that stage forces a phase shift in whatever signal passes through it. Baseband noise and thermal drift, present with the desired signal, will act as variables in the modulators (Fig. 1) and therefore modulate the desired signal.

Standard microwave radio frequency (RF) amplifier designs, (as presented by computers -- variations on the microstrip matched, grounded emitter, capacitance coupled, single ended (class A or C) configuration), present their basic characteristics as reasons for me to suspect them of being poor phase noise performers.

High-Q ($>=1$) resonant circuitry is undesirable because a major mechanism of phase modulation is the reactive slope at any one point. The steeper the slope is, the higher the modulation efficiency and the higher the temperature coefficient.

Some form of emitter or source degeneration is crucial to obtaining low phase-noise performance. Resistance, emitter to ground, improves operating point or baseband stability and improves signal linearity. Thermal instability is a major contributor to low frequency noise,

*Contribuition of the U.S. Government, not subject to copyright.

below a few hertz, and amplitude nonlinearity reduces power supply rejection as well as introducing an unstable element into the modulators (Fig. 1).

Bipolar transistors have more gain at low frequencies than at radio frequencies. Capacitively coupled bipolar amplifiers can accumulate baseband noise faster than desired signal, with each succeeding stage effecting amplitude modulation (AM) to phase modulation (PM) conversion of the amplified baseband noise.

The single-ended amplifier stage has no low frequency power supply noise rejection, and no common-mode noise rejection (Fig. 5). This is because signal currents share the ground return with operating (B+) currents and input-to-output signal-ground currents (assuming capacitor coupling).

Transmission line transformers are a good solution for interstage coupling at radio frequencies. When impedance matched and compensated, these devices can give a flat reactive slope around design center, while at the same time, effect impedance transformation and phase inversion, necessary for balanced amplification. Because of their transmission line construction, they also cancel external field pickup and radiation.[4,5]

The balanced amplifier configuration allows use of emitter/source-coupled pairs, the basic instrumentation amplifier form which provides isolation from power supply rail noise and common mode ground-plane noise.[2,3] The transformer coupled push-pull pair also allows:

- Direct current paralleling of devices to reduce device generated noise.
- Common mode isolation from input to output.
- Low amplitude-distortion for a given power dissipation.
- Cancellation of even order harmonics.
- Circuit stray capacitances are in series in a push-pull stage and not direct shunts to ground. This allows wider bandwidth for a given impedance level.
- Relatively high reverse isolation from output to input.
- A large resistance or constant current source can be used in the emitter circuit without appearing in the signal path.

Ground path continuity and mass are so important that they almost cannot be used for design trade-off. When we start looking as low as 170 dB below the signal, ground-path resistance and reactance, adequate for lesser dynamic ranges, can allow development of common-mode noise and signals that will compromise the noise-floor. While a balanced RF amplifier can achieve a useful degree of common-mode isolation, this reduces, but does not eliminate, requirements for control of common-mode signals and noise.

Ohm's law (E = IR) describes what happens. I will expand that a little for these purposes:

$$\frac{(CMRR)(E)}{\text{Circuit gain}} = I \times \left(\begin{array}{c} \text{common-mode ground path} \\ \text{resistance + reactance} \end{array} \right)$$

where

E = acceptable limit of interfering signal or desired circuit noise floor

I = common mode noise current that will produce E

CMRR = In this case, this is a complex term including, but not necessarily dominated by, a calculated figure. The amplifier physical ground path, as a percentage of the overall ground path (the path along which common mode signals are conducted), and the angular orientation of that amplifier ground path relative to noise current flow, will determine the percentage of noise signals generated by common mode currents which will be injected into the amplifier. Then, the amplifier's CMRR (if any) will act to isolate the desired signal from this resultant noise. A real number for total CMRR, in this case, is more accurately arrived at experimentally.

Circuit requirements will fix the noise floor and circuit gain. Control of environment will reduce interfering signals to some unavoidable minimum. At radio frequencies, common-mode isolation is limited by relatively unstable stray reactances. After the other terms are optimized, what is left to manipulate is the ground path characteristic.

For example, take three sets of points 5.08 cm (2 inches) apart. One set is connected by a printed circuit trace 1.575 mm (0.062 inches) wide. Another is connected by copper wire of 0.635mm (0.025 inches, #22) diameter. A third set is centrally located on the unetched copper foil of a 7.62cm (3 inch) by 12.7 cm (5 inch) printed circuit card. 100mA of direct current and then 4.5mA RMS at 10MHz was applied between the three sets of printed circuit board pads to obtain approximate values for resistance and reactance.

Conclusive RF measurements across 5.08 cm (2 inches) of a continuous copper ground plane were not

possible in the time available. All indications were that the reactance is very small, approximately 0.1 ohm or less at 10MHz. Experimental estimates of the trace and wire over ground plane were not so difficult because they are much larger – 1.5 ohms and 0.5 ohm at 10MHz. Estimates of the D.C. resistances are 0.0006 ohm for the ground plane, 0.0015 ohm for the wire, and 0.015 ohm for the trace. This is a relatively large trace.

The difficulty of breadboarding these measurements puts these figures in the -50% to +100% range of error. They do, however, give a good indication of the large difference between a single conductor and the continuous ground plane when it comes to controlling common-mode noise.

For example, you've got a single ended amplifier with a gain of X10, and a maximum allowable interfering signal of 1uV. If you have been able to get the current of common mode line harmonics down to 0.1mA, the required total ground plane resistance would be 0.001 ohms.

Active Device Considerations

Available devices require some form of paralleling within the device structure to get a low noise-corner. Large gate area in junction field effect transistors (J-FET) produces the lowest noise but the highest capacitance, limiting frequency response.[3] The U310 FET die is a good compromise between noise and speed. The multiple emitter site, ballasted bipolar transistor configuration, desirable for stability at microwave frequencies, can also be stable and exhibit low noise at baseband.

Since high-Q circuitry will produce a high AM to PM conversion factor and phase/temperature instability, devices with low interelectrode stray capacitances are desirable at RF. Small capacitance values can be absorbed into the compensating capacitance of the transmission line transformers allowing for the flattest possible response.[4] Moderate to low operating impedances are desirable because less inductance is required for a given frequency response in transformers. I have experienced trouble obtaining flat high frequency response at the higher impedances above 400 Ω. The lower the circuit impedance, the less the effect of the unavoidable stray capacitive reactance also. Strays are by their nature unpredictable and unstable.

Junction Field Effect Transistors are limited to the common-gate configuration to keep impedances as low as practical and eliminate adverse effects of reverse transfer capacitance. Microwave bipolar transistors are most useful in the common emitter, emitter-coupled pair configuration now, but better low impedance transformer designs may extend bipolar possibilities by allowing efficient use of common base configurations.

Capacitor Considerations

- Capacitance temperature stability
- Dielectric loss temperature stability,
- Absorption (environmental stability), aging, internal and package mechanical stability (microphonics)
- Capacitor value reactance at frequency of interest, case and lead reactance
- Capacitor reactance at baseband
- For use in signal path, stacked foil Mylar, wound foil Mylar and polypropylene are low inductance and high dielectric quality
- Use solid tantalum only where an electrolytic is required

Inductor Considerations

- Inductance temperature stability
- Control of radiation and rejection of external magnetic fields
- Microphonics
- Resonance effects at signal center frequency and baseband
- The ferrite loaded transmission line transformer optimally applied resolves most inductor considerations and also allows implementation of a balanced, non-resonant RF amplifier.

Resistors

- Because of their superior noise and thermal performance, I use 1% metal film, conservatively rated resistors everywhere.

Power Supply Considerations

Three-terminal regulator broadband-noise output is approximately -135 dBV/Hz as viewed on an audio spectrum analyzer out to 400 kHz. Depending on the power supply rejection ratio and the AM to PM conversion factor of the circuit in question, this could easily set the broadband phase noise floor. Single-ended circuitry almost does not have a PSRR. Class A and C amplifiers and most forms of logic are offenders in this way. (Figure 5) If the power supply situation can accommodate the additional voltage drop and power dissipation, and if the remaining close-in noise is not a problem, a simple resistor/capacitor

(R/C) filter will reduce rail AM noise below where it will be a major contributor to circuit phase noise performance. (Figure 2)

Audio noise at LM317 output (+15V)

1Hz	-112dBV/Hz
10	-118
100	-131
1000	-134
10000	-134
25000	-134

Audio noise after RC filter (+12V)

1Hz	-112dBV/Hz
10	-121
100	-148
1000	-166 M
10000	-170 M
25000	-169 M

M is measurement system limited.

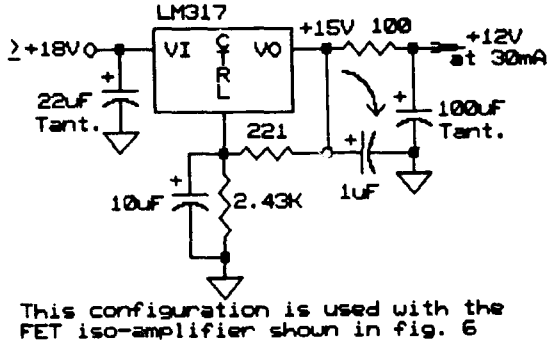


Figure 2.

When output shunt capacitance approaches some fraction of input capacitance or there is reactance on one of the legs, three-terminal regulators can oscillate. They are, after all, gain blocks. This oscillation may not be obvious or even visible with an oscilloscope. But it will get into sensitive circuitry and cause increased noise, instability, and even intermittent thermal shut-down of the regulator. If phase noise is a major consideration, a better solution is to build a discrete regulator using a precision reference, low-noise operational amplifier, and a pass device. This can produce a power supply noise floor of approximately -160 dBV/Hz, but even here, the most critical situations will require some kind of postfiltering. (Figure 3 and Figure 4)

Audio noise at output (+12V at 30 mA.)

1Hz	-134dBV/Hz
10	-150
100	-160
1000	-161
10000	-161
25000	-161

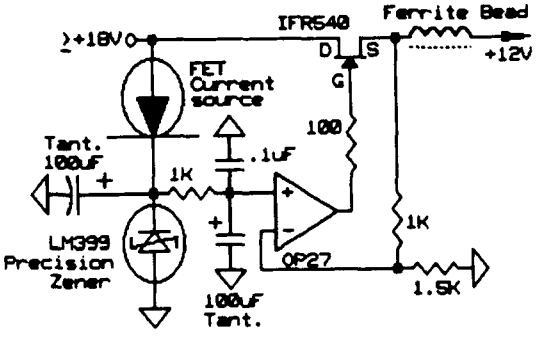


Figure 3.

Simple additional active low-pass filtering can be added after the regulator when current or close-in noise specifications require. This circuit also isolates the regulator from any large shunt capacitor required for a single-ended stage. (Figure 4)

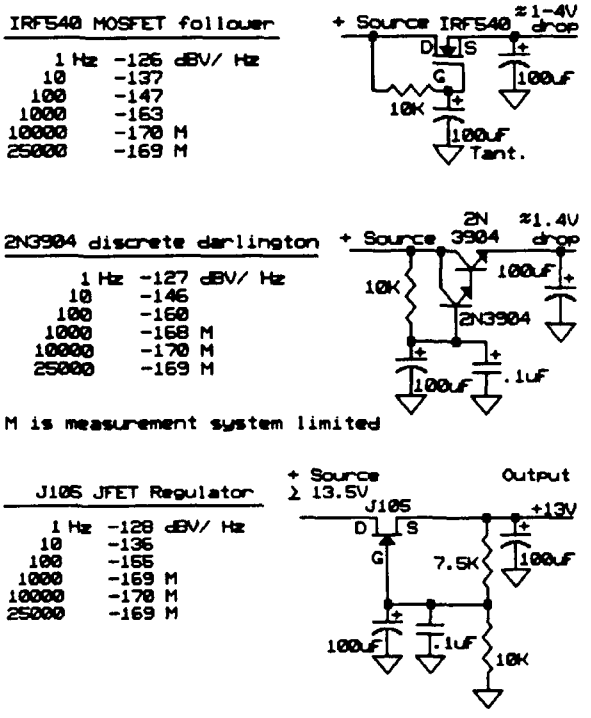


Figure 4. Two low noise clean-up followers and one regulator. Noise measured at 30 mA current.

Baseband resonances in the B+ decoupling networks can generate noise peaks. Common-mode or bias currents circulate equally in the active device, decoupling circuitry, power supply, and ground path. These are all in series for low-frequency noise. Capacitors used as high frequency ground returns and filters will be completely transparent at very low frequencies.
(Figure 5)

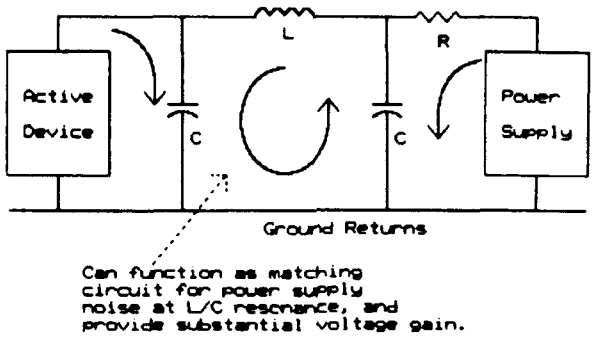


Figure 5.

Figure 6 is a representation of a noise floor peak initially observed during phase noise testing of a circuit. The same peak was observed on the power supply rail. Eliminating an accidental resonance in decoupling circuitry removed both peaks.

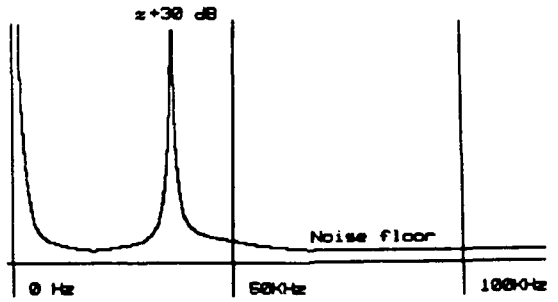


Figure 6.

RF Isolation Amplifier

Here is an isolation amplifier for the lower clock frequencies (1 to 10MHz) using readily available commercial components. (Figure 7)

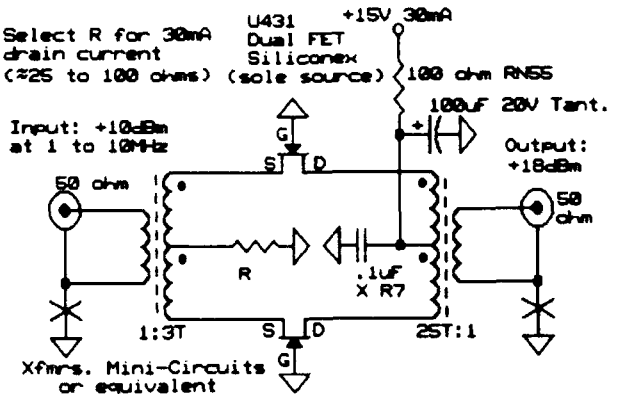


Figure 7

Input and output need not be ground referred; this allows very high input-to-output common-mode isolation.

Substituting optimized custom transformers for the commercial units can substantially expand the bandwidth.

CONSIDERATIONS: (Figure 7)

- Gain: +8dB, 1 to 12MHz (-1dB, 80kHz to 15 MHz; -3dB, 50kHz to 26MHz)
- Reverse isolation: ≥ 34dB, 10kHz to 670MHz
- PSRR: for drain volts, +1V, (at output 18dBM) ≤ +1MV
- Input level for compression at 10MHz: +11dBM/0.1dB, +16dBM/1dB
- Phase change with temperature at 10MHz: ≤ 0.1°/10°C to 60°C
- Phase noise: see Table
- Linearty: see Table

SSB Phase Noise

Input: +10dBm @ 5MHz

Hz from the carrier	dBV/Hz
1	-136
10	-146
100	-156
1000	-166
10000	-174
25000	-174

Harmonic Distortion Data

Frequency	Levels	
Input:		
5MHz	+10dBm	+7dBm
Output:		
5MHz	+18dBm	+15dBm
10MHz	-81dBc	-79dBc
15MHz	-65dBc	-68dBc
20MHz	--Signals below--	
25MHz	measurement floor	

Testing Notes

At the lower signal levels, accidental common-mode ground noise, both coherent line harmonics and broadband noise, will probably be the biggest obstacle to a repeatable test setup. Also, slow line variations will disturb long sweep time measurements. In my own laboratory, I run all equipment involved in a particular test off one line-conditioning ferro-resonant transformer. Then,

in addition to signal grounds, I run chassis-to-chassis ground straps.

After a carrier has been phase-detected, the resultant information is amplitude noise, equivalent to the phase noise. The relationship between the two is the phase-detector sensitivity.[6,7] Almost any indicating device can be used after a low-noise preamplifier. Values indicated by various analyzers or even a single frequency (for example 1 Hz center frequency) operational amplifier filter [8] can be normalized to a 1 Hz bandwidth to provide interchangeability of numbers. There is a limitation in that filter bandwidth must be equal to or less than one tenth of the filter's center frequency.[9] Resultant calibration can be verified using the precision noise source described in NIST paper, "Accuracy Model for Phase Noise Measurements." [10] This and some other papers relevant to phase noise analysis are listed in the bibliography.

To observe the AM-to-PM conversion of an RF amplifier, amplitude modulate the RF signal at some low level (-80 dBc) with a high quality modulator, then measure the resultant phase noise side bands at the amplifier output. This can give real-time feedback on the effects of circuit changes.[6] (Figure 8)

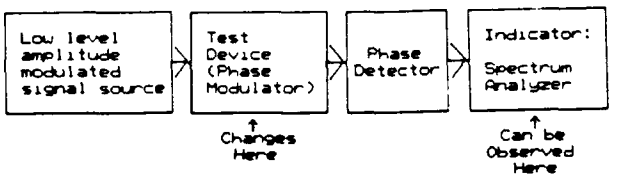


Figure 8

Figure 9 is a block diagram of the test setup used to evaluate the FET iso-amplifier. I used two iso-amplifiers to get the same conditions at both inputs of the phase detector . The resultant figure is the sum of the noise of the two amps. Phase noise coherent to the two amplifiers under test, such as phase noise induced by common power supply noise, will cancel in the phase detector. This has been known to give painfully misleading test results.

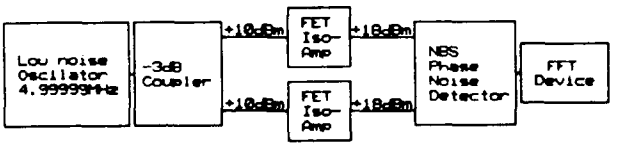


Figure 9

Figure 10 shows a block diagram of a common form of quadrature phase noise measurement system. Measuring original oscillator phase noise is somewhat complicated in that at least two identical oscillators that can be phase locked to each other are required.[6,10] A single oscillator and -3dB coupler will provide the two coherent signals for other tests. However, a second, non-critical oscillator will be required for calibration.[7,10]

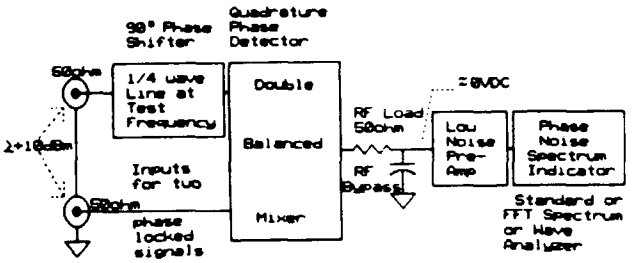


Figure 10

The amplitude noise test results reported for figures 2, 3, and 4, have a repeatability range of ± 1 to 2 dB. The phase noise figures quoted for the amplifier in figure 7 are ± 2 to 3 dB. More time with the same equipment will yield better accuracy.

Low-Amplitude-Noise Preamplifier

Figure 11 is a low-noise video preamplifier using standard parts which can be used in front of a standard or fast Fourier transform (FFT) spectrum analyzer to evaluate amplitude noise. It can also be used after a phase detector for the evaluation of phase noise.

OUTPUT NOISE
(with shorted input)

1Hz	-139dBV/Hz
10	-167
100	-171
1000	-170
10000	-170

All amplitude measurements in this paper used this pre amp. It is the limiting factor on some of them. An "M" by a figure means measurement limited.

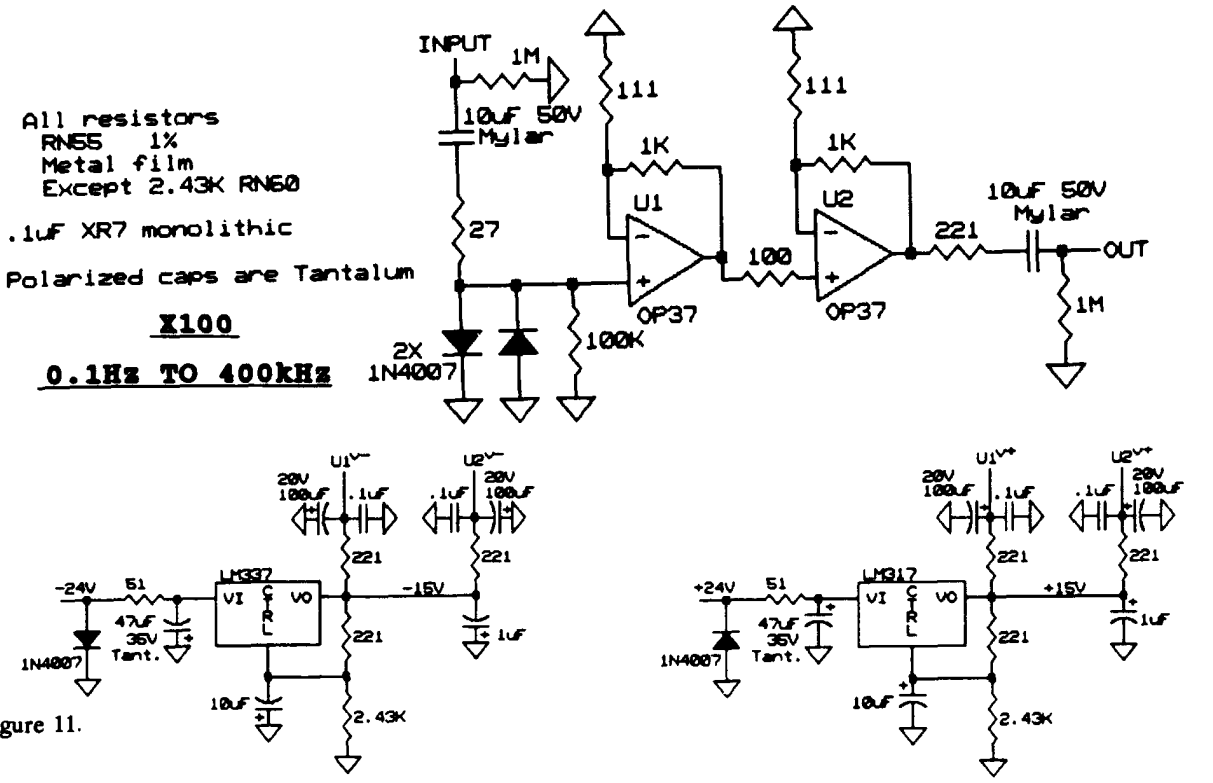


Figure 11.

CONCLUSION

When signal considerations extend 160dB or more below the carrier, calculated incremental design solutions can reach a point of diminishing returns. Attempting to optimize as many design parameters as possible right from the start will save a lot of time. Since amplitude noise is a major source of phase noise, a good beginning is to design the lowest amplitude noise circuit possible. Then make the reactive slopes of that circuit as flat as possible, and allow some form of emitter/source degeneration.

Use of phase-noise measurement equipment to obtain real-time empirical feedback, as regards the effects of design changes, is important in this area of design that is not yet fully understood. Using this paper and the reference publications, fabrication of a basic phase-noise measuring system capable of producing verifiable documentation can be accomplished in the average anaiog laboratory using existing equipment.

Initial design, component selection, layout, power supply requirements, and control of environment are each crucial in a high performance situation. Since phase noise is cumulative (multiplicative) for an entire system, system performance will be somewhat worse than the single worst gain block or device in a system. Achieving low phase-noise system performance requires careful attention to every individual function block in that system.

Acknowledgements

I thank Dr. Fred Walls for a great deal of productive discussion, and the other reviewers who guided this paper to its final form. The regulator of Figure 3 is the result of a conversation with Emmit Kyle, United States Navy.

References

[1] "Amplifier Techniques," R331-351 and "Electrical Characteristics," 61-66, C. OP-AMP Cookbook 3rd ed., Walter G. Jung, (Howard W. Sams and Co. 1986).

[2] "Emitter Coupled Differential Amplifier," pp. 111-120 and *"Common Mode Rejection Defined,"* 115, GE Transister Manual, 7th ed. (General Electric Co. 1964).

[3] "An Introduction to FETs," 7-11 and "FETs in Balanced Mixers (U431)," 55-64, Siliconix FET Data Book, (Siliconix Inc. 1986).

[4] Jerry Sevick, "Transformer Analyses," and all of Transmission Line Transformers, (American Radio Relay League 1987).

[5] Broadband Transformer Design Information, Iron-Powder and Ferrite Coil Forms, (Amidon Associates Inc. 1989).

[6] F. L. Walls, C.M. Felton, A. J. D. Clements, "Accuracy Model for Phase Noise Measurements," Proc. of the 21st Annual Precision Time and Time Interval (PTTI) Meeting, Redondo Beach, CA, Nov. 28-30, 1989, pp. 295-310.

[7] D. A. Howe, *"Frequency Domain Stability* Measurements: A Tutorial Introduction," NIST Tech Note 679, 1976.

[8] Don Lancaster, "Multiple-Feedback Bandpass Circuit," Active Filter Cookbook, (Howard W. Sams and Co. 1988).

[9] F. L. Walls and S. R. Stein, "Accurate Measurements of Spectral Density of Phase Noise in Devices," Proc. of the 31st Annual Frequency Control Symposium, 1977, pp. 335-343.

[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. of the 42nd Annual Frequency Control Symposium, Baltimore, MD, June 1-4, 1988, pp. 432-441.

FORTY-FOURTH ANNUAL SYMPOSIUM ON FREQUENCY CONTROL

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)

S. R. Stein*
Ball Communication Systems Division
P.O. Box 1235
Broomfield, CO 80020

John Evans
Defense Communications Engineering Center
1860 Wiehle Ave
Reston, VA 22090

Introduction

The goals of this study were to develop and demonstrate techniques for obtaining optimum timekeeping performance from precision station clocks. Kalman filters and ARIMA models were used to study the predictability of precision clocks based on quartz, rubidium and cesium technology. The software performing the time prediction error analysis utilized either Kalman filters or ARIMA models to estimate the time, frequency, and frequency aging of a clock over an initial calibration interval and then forecasted into the 'future' for intervals from one minute up to one year. The forecasts were compared to the actual time of the clock to determine the forecast error. After the rms forecast error was updated for each interval, the starting point was advanced and the process was repeated. This technique was used to study the performance of thirteen clocks over a two year time period. Adaptive versions of both the ARIMA and Kalman approaches were developed using real-time estimators of the filter/model parameters. Although the Kalman and ARIMA approaches performed forecasting equally well, the Kalman filter technique seems most advantageous for operational use in the DCS.

Kalman Filter Approach

The Kalman filter used for this study was based on the work of Jones and Tryon [1]. The clocks are assumed to be perturbed by continuous noise described by the traditional power law spectral density model. In addition to the usual white PM, white FM, and random walk FM noises, random walk frequency aging is included for some clocks in order to describe the time-varying frequency aging often observed in quartz oscillators and rubidium frequency standards. The equations of motion of the clock are written in matrix form

$$\vec{x}(t+\delta) = \Phi(\delta)\vec{x}(t) + \Gamma\vec{s}(t+\delta|t) + \Phi(\delta)\vec{p}(t)$$

to describe the evolution of the time, frequency and frequency aging from time t to time $t+\delta$. The clock's state vector, $\vec{x}$, is

$$\vec{x}(t) \quad \begin{bmatrix} u(t) \\ x(t) \\ y(t) \\ w(t) \end{bmatrix}$$

where $u(t)$ is the time, $x(t)$ is a dummy variable, $y(t)$ is the frequency and $w(t)$ is the frequency aging. The noise vector contains the noise perturbing each clock state during the interval δ.

$$\vec{s}(t+\delta|t) \quad \begin{bmatrix} s'(t+\delta|t) \\ \epsilon'(t+\delta|t) \\ \eta'(t+\delta|t) \\ \alpha'(t+\delta|t) \end{bmatrix}$$

The state transition matrix, Φ, describes the dynamics of the states. Each higher state is the integral of the lower state plus an additive noise term which causes a random walk. The final time has an unintegrated white noise to account for additive white noise of the clock.

$$\Phi(\delta) \quad \begin{bmatrix} 0 & 1 & \delta & \delta^2/2 \\ 0 & 1 & \delta & \delta^2/2 \\ 0 & 0 & 1 & \delta \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The Γ matrix describes the coupling of the noise from the internal phase state, $\vec{x}$, to the observable state, $\vec{u}$.

$$\Gamma(\delta) \quad \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Since the station clocks may be steered to a reference for purposes of network synchronization, the vector $\vec{p}$ is included to describe the control inputs to the clock. Finally, the observations are related to the states of the clock by the measurement equation

*Work supported by Defense Communications Agency, Contract No. DCA100-88-C-0072

US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT

$$\vec{z}(t) = \mathbf{H}(t)\vec{x}(t) + \vec{v}(t),$$

where $\vec{z}$ is the vector of observations and $\vec{v}$ is the white noise perturbing the measurements. In this study, the observations were individual clock time differences with respect to the USNO master clock and the measurement equation reduces to a scalar relationship.

From this point on, the clock state estimation problem may be solved using Kalman's prescription. The plant covariance matrix and recursion for the Kalman gain are not repeated here since they have already been published [2]. An excellent general description of Kalman filters appears in Gelb [3].

The Kalman filter has a degree of robustness due to the use of a physical model to describe the time evolution of the states. The robustness was enhanced by deweighting suspicious observations, *i.e.*, outliers. A judgment is made whether each observation is consistent with the noise processes included in the model. The squared difference between the measurement and its forecast, the innovation, is compared to the covariance of the innovations. When the ratio is too high the Kalman gain is reduced smoothly to zero. Thus, the information content of suspicious observations is gradually rejected. This avoids any discontinuity in the state estimates that would be introduced by simply rejecting the suspected outlier. Such an approach is beneficial since the discontinuities that arise from abrupt data rejection schemes can introduce instabilities in complex systems.

ARIMA Modelling Approach

The ARIMA model is characterized by the three parameters (p, d, q). It is written

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d z(k) = (1 - \theta_1 B - \cdots - \theta_q B^q)a(k),$$

where the backward shift operator B is defined by

$$Bz(k) = z(k-1).$$

The backward difference operator, ∇, defined by

$$\nabla z(k) = z(k) - z(k-1) \ = (1 - B)z(k),$$

is also used.

The coefficient p is the order of the autoregressive operator, d is the number of integrations, and q is the order of the moving average operator. The elements of the observed time series are labeled $z(k)$, and the $a(k)$ are random shocks drawn from an ensemble having a normal distribution. It is assumed that the observations are spaced equally in time. The parameter values are determined from a study of the autocorrelation function of the observed data. The autocorrelation analysis also produces initial estimates for the coefficients of the model. Once the proposed model is obtained, the best values for the coefficients may be calculated using the maximum likelihood technique. It can be shown that the clock model used in this study for Kalman filter analysis has an equivalent ARIMA model. The derivation starts with the measurement model.

$$z(t) = u(t) + v(t)$$

Next, the state equations are used to eliminate the time of the clock, $u(t)$ and then each remaining state of the clock in turn until an equation is obtained for the observations in terms of the noise perturbing the clock. For the most general clock model used here (a clock with white phase noise, white frequency noise, random walk frequency noise, and random walk frequency aging noise) the equation for the time observations is

$$\nabla^3 z(k) = \nabla^3[\beta'(k-1) + v(k)] + \nabla^3\epsilon'(k-1) + \delta\nabla\eta'(k-2) + \delta^2[\alpha'(k-3) + \frac{1}{2}\alpha'(k-2)].$$

Examination of this equation leads to the conclusion that there are no autoregressions, there are three integrations and the noise inputs involve zero through three lags. The equivalent ARIMA model is (0,3,3).

$$\nabla^3 z(k) = a(k) - \theta_1 a(k-1) - \theta_2 a(k-2) - \theta_3 a(k-3)$$

The relationship between the coefficients $\theta_1, \theta_2, \theta_3$, and σ_a^2 on the one hand and $\sigma_v^2 + \sigma_{\beta'}^2$, $\sigma_{\epsilon'}^2$, $\sigma_{\eta'}^2$, and $\sigma_{\alpha'}^2$ on the other is derived by equating the autocorrelation functions of $\nabla^3 z$ for zero through three lags for the two representations.

Adaptive Approaches

It is possible to estimate the noise spectral densities required by the Kalman filter by supplementing the usual recursion. This was accomplished by analyzing the variance of the Kalman filter innovations. The innovation is the difference between what was measured and the predicted measurement – it is simply the one step ahead time prediction error. Like the Allan variance analysis, which is performed on unprocessed, uniformly sampled measurements, the innovation analysis requires only a limited memory of past data. However, the forecasts provided by the Kalman filter allow the computation to be performed at arbitrary intervals once the algebraic form of the innovation variance has been calculated. Also, the Allan variance is not defined for noise processes more divergent than random walk FM, but the variance of the innovations is well behaved because the state equations account for the appropriate degree of nonstationarity.

Details of the method have been published previously [2]. Simulation testing was performed to demonstrate the accuracy of the method and it was used in all the Kalman forecast analysis. Figure 1 shows the real-time estimate of

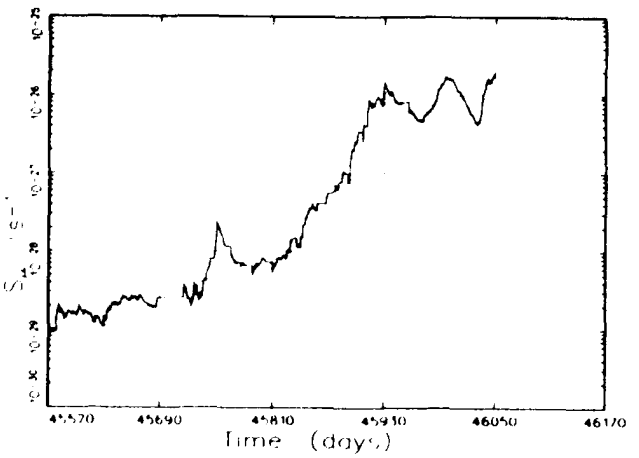


Figure 1: The random walk FM of clock 9 increases 3 orders of magnitude in 2 years.

the random walk FM of clock 9. The result is startling because of the three order of magnitude increase in noise level over the two years of observation. The increase was confirmed by Allan variance analysis of the time measurements. In an operational situation such a dramatic change in the noise of a clock could be used to trigger appropriate maintenance.

A real-time technique was also developed to estimate the parameters for ARIMA modelling [4]. Random shocks are calculated from the inverse model using the current estimates of the parameters. Negative feedback is applied to update the estimate of the n^{th} moving-average parameter by an amount proportional to the product of the current random shock and the one n samples earlier. Since the true model reduces the random shocks to a white sequence, the autocorrelation function is zero for all lags when the parameter values are correct and the negative feedback drives the parameters to these values. The technique works well when the parameters are not too close to the boundary of the region of invertibility. Unfortunately, this was not the case for the clocks and sample rates chosen in this study and the previously described Kalman filter technique was used to overcome this difficulty.

Model Identification

Quartz and rubidium oscillators are perturbed by various levels of white phase noise, white frequency noise, random walk frequency noise, and random walk frequency aging noise. The random walk of the frequency aging is often ignored, but it has been included in this study to take into account the changes in frequency aging which are always seen in these types of oscillators. In general, the more parameters and states that must be estimated in the modelling or filtering process, the worse the parameter estimates and the time forecasts will be. The biggest challenge in model identification is to eliminate unnecessary states and parameters. Thus, it was found that white frequency noise was masked by the high level of white phase noise for all of the quartz and rubidium clocks. In addition, the random walk frequency noise dominated the white phase noise in two of the quartz oscillators. The most general model for a cesium frequency standard includes white phase noise, white frequency noise, and random walk frequency noise. There is no frequency aging in healthy cesium clocks and no random walk frequency aging noise is needed.

Model identification was performed using a combination of Box-Jenkins [5] ARIMA analysis techniques, Kalman filter analysis, and model independent analysis techniques such as the Allan variance. The ARIMA models were inverted to calculate the random shocks which would have produced the observations if the model were valid. These shocks were tested for whiteness using both the cumulative autoperiodogram and spectrum analysis. Kalman filter real-time parameter estimation was used to test for the existence of candidate noise types and to determine nominal values. In the end, the forecast analysis was the final determinant – the best model was defined to be the model which yielded the smallest time prediction errors.

Table 1 lists the nominal physical models for all the clocks. The table lists value of the spectral density or variance of each noise type. S_ξ is the spectral density of the white FM; S_μ is the spectral density of the random walk FM; S_ζ is the spectral density of the random walk of frequency aging; and σ_τ^2 is the variance of the white PM. For simplicity, all the white PM is attributed to the measurement process and β', the clock white PM, is assumed to be zero. The relationship of the spectral density parameters in Table 1 to the traditional h coefficients and the Allan variance is given in Table 2. The physical models for all the quartz and rubidium clocks were characterized by nonzero random walk FM and random walk frequency aging. The white FM which one would normally expect to see in the rubidium frequency standards was below the level of the other noises and was omitted from all the rubidium and quartz models. A white PM like noise should be expected as a result of the one nanosecond resolution of the measurement system used to collect the data. It was measured for all the clocks except the two quartz oscillators which were sampled at one hour intervals. The random walk FM of the quartz oscillators was sufficiently large to dominate the white PM at this long sample time. The physical models for the cesium clocks included white PM, white FM and random walk FM. Forecast analysis was degraded when random walk frequency aging was included in the cesium clock model. The best results were obtained with the frequency aging of the cesium clocks permanently constrained to be zero. Table 3 lists the parameters of the preferred ARIMA models for each clock. They were calculated from the physical models using the previously described technique.

Table 1: Nominal Kalman filter parameters

Clock No.	Type	Interval (s)	Aging	σ_ϵ^2	S_ξ	S_μ	S_ζ
1	Rb	60	yes	5.9×10^{-19}	0	5.0×10^{-27}	4.0×10^{-44}
2	Qz	60	yes	4.0×10^{-18}	0	1.0×10^{-27}	1.0×10^{-40}
4	Rb	900	yes	6.0×10^{-19}	0	1.0×10^{-29}	4.0×10^{-44}
5	Qz	900	yes	1.0×10^{-17}	0	2.5×10^{-27}	6.0×10^{-40}
7	Rb	900	yes	5.0×10^{-19}	0	2.5×10^{-30}	4.0×10^{-44}
8	Rb	3600	yes	7.5×10^{-19}	0	2.3×10^{-29}	1.0×10^{-42}
9	Rb	3600	yes	3.0×10^{-19}	0	1.0×10^{-29}	5.0×10^{-42}
10	Qz	3600	yes	0	0	1.5×10^{-27}	2.0×10^{-40}
11	Qz	3600	yes	0	0	1.0×10^{-27}	1.0×10^{-40}
12	Rb	3600	yes	6.8×10^{-19}	0	3.7×10^{-30}	1.0×10^{-43}
15	Cs	3600	no	2.0×10^{-18}	1.0×10^{-21}	2.0×10^{-30}	0
18	Cs	3600	no	9.0×10^{-19}	2.2×10^{-22}	4.0×10^{-34}	0
19	Cs	3600	no	8.7×10^{-19}	2.0×10^{-22}	2.1×10^{-33}	0

RMS Forecast Error Results

Forecast errors for the clocks were analyzed using both Kalman filters and ARIMA models. All the Kalman filter analyses used adaptive models. A minimum calibration period was used to determine starting estimates for the time, frequency, frequency aging, and noise spectral densities. Typical calibration periods were 10 days for quartz oscillators and between 10 and 30 days for rubidium and cesium clocks. Because the Kalman analysis used real-time parameter estimation, the calibration periods for each state and each parameter adjusted themselves to optimum values determined by the estimated noise levels. Following the initial calibration, forecasts were made for periods of one minute to one year in the future. Then the filter was updated to the next available observation and new forecasts were made. All the forecast errors corresponding to a prediction interval were used to obtain the rms forecast error, or time prediction error, for that interval. Although the individual forecasts are not expected to be independent of one another, especially for long forecasts, this method produces unbiased results and improved confidence compared to the method of nonoverlapping estimates.

The ARIMA analysis of the rms forecast errors was performed in a similar manner. The parameters were estimated using both real time parameter estimation and the maximum likelihood method. No degradation was detected when real-time estimates were used. Two approaches were evaluated for the clocks which had random walk frequency aging noise - the use of the ARIMA (0,3,3) model or the the use of a prefilter to estimate aging followed by the ARIMA (0,2,2) model to account all other noise. The use of the ARIMA (0,3,3) model was always problematic for these clocks. In most cases the model parameters were extremely close to the singularity in the likelihood function which occurs at the boundary between the invertible and noninvertible models. Parameter estimation was not feasible for these clocks. In those cases where parameters could be estimated, the forecast errors which resulted from the use of the (0,3,3) model were always significantly worse than those obtained by removing the effects of aging with the prefilter. Thus the smallest forecast errors that were obtained using ARIMA modelling corresponded to the ARIMA models listed in Table 3.

Table 2: Relationship between spectral densities and h coefficients

Spectral Density	h Coefficient	Allan Variance	Units
$S_\beta(t)$	$(2\pi)^{-2}h_2(t)/2$	$\tau^2\sigma_y^2(\tau$; white PM$)/6f_h$	seconds^3
$S_\xi(t)$	$h_0(t)/2$	$\tau\sigma_y^2(\tau$; white FM$)$	seconds
$S_\mu(t)$	$(2\pi)^2h_{-2}(t)/2$	$3\sigma_y^2(\tau$; RW FM$)/\tau$	seconds^{-1}
$S_\zeta(t)$	$(2\pi)^4h_{-4}(t)/2$	not defined	seconds^{-3}

Table 3: ARIMA (0,2,q) models calculated from the nominal Kalman filter parameters.

Clock	Interval	ARIMA	θ_1	θ_2	Aging Prefilter	σ_a
1	1 m	(0,2,2)	1.707	−.744	yes	8.9×10^{-10}
2	1 m	(0,2,2)	1.879	−.886	yes	2.1×10^{-9}
4	15 m	(0,2,2)	1.53	−.62	yes	9.9×10^{-10}
5	15 m	(0,2,2)	1.11	−.36	yes	5.2×10^{-9}
7	15 m	(0,2,2)	1.65	−.70	yes	8.4×10^{-10}
8	1 h	(0,2,2)	.69	−.15	yes	2.2×10^{-9}
9	1 h	(0,2,2)	.67	−.14	yes	1.4×10^{-9}
10	1 h	(0,2,0)			yes	1.4×10^{-8}
11	1 h	(0,2,0)			yes	1.1×10^{-8}
12	1 h	(0,2,2)	1.04	−.33	yes	1.4×10^{-9}
15	1 h	(0,2,2)	1.14	−.24	no	3.7×10^{-9}
18	1 h	(0,2,2)	1.3990	−0.4019	no	1.5×10^{-9}
19	1 h	(0,2,2)	1.4028	−0.4096	no	1.5×10^{-9}

The rms forecast errors obtained using the Kalman and ARIMA methods were in excellent agreement as illustrated in Fig. 2. This result is expected theoretically since both methods are minimum squared-error techniques. The influence of calibration interval on the forecast errors was studied using the ARIMA approach. There was negligible affect, as shown in Fig. 3, for forecast intervals from 15 minutes to 43 days. One expects this result theoretically as long as the calibration interval exceeds the time at which the clock has its best stability. This condition was true for clock 7, a rubidium frequency standard. Finally, Fig. 4 summarizes the estimated rms forecast errors all thirteen clocks. It can be seen that the variation of forecast errors for the clocks of one model was almost always less than the difference between models. For times longer than one day, the cesium standards had the smallest forecast errors. They were followed in order by one model rubidium standard, the other model rubidium standard, and finally the quartz oscillators. One of the cesium clocks performed in a similar fashion to the best rubidium clock.

Conclusions

Kalman filters and ARIMA models were used to study the predictability of precision clocks based on quartz, rubidium and cesium technology. Some of the results of the study are: improved theoretical models for clock predictability; empirical evaluation based on the performance of thirteen clocks over a two year time period; comparison of the advantages and disadvantages of Kalman filters and ARIMA models for performing clock state estimation and forecasting; and the development of adaptive clock models.

Kalman filters and ARIMA models were expected to yield the same forecast errors since they are both minimum squared-error estimators and this conclusion was confirmed empirically. The actual forecast errors also agreed well with the theoretical predictions. Adaptive versions of both the ARIMA and Kalman approaches were developed using real-time estimators of the filter/model parameters. These adaptive versions were found to exhibit time prediction errors comparable to the use of maximum-likelihood parameters. To make it possible to use these techniques in practical operational situations, robust methods of outlier detection and time and frequency step detection were implemented for both the Kalman and ARIMA approaches.

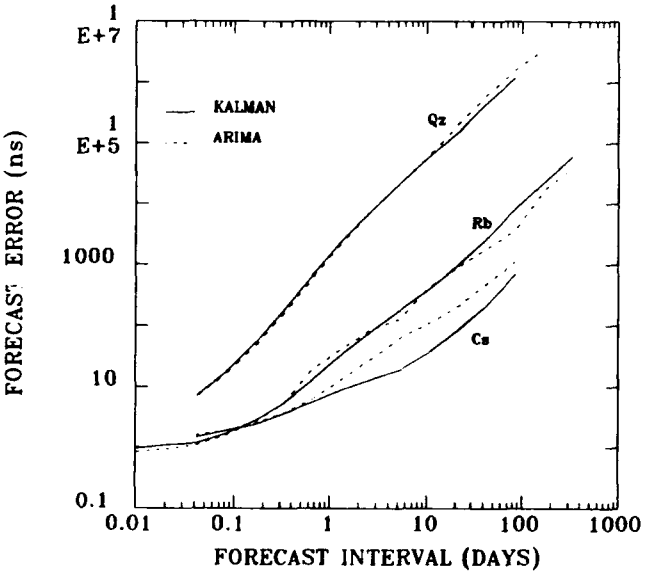


Figure 2: Kalman and ARIMA techniques produce equivalent rms forecast errors.

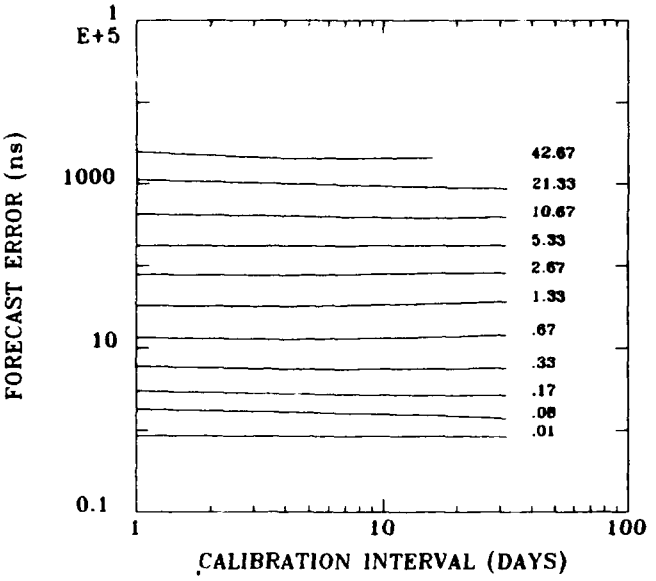


Figure 3: Forecast errors are nearly independent of the calibration interval.

Although the Kalman filters and ARIMA models performed forecasting equally well, the Kalman filter technique seems advantageous for operational use. The Kalman filter used in this study does not require equally spaced data, nor does the associated real-time parameter estimation technique. On the other hand, the Box-Jenkins ARIMA modelling and its real-time parameter estimator assume the data are uniformly spaced and must be patched to deal with typical operational scenarios, especially under adverse conditions. In addition, the Kalman filter uses the known clock dynamics. It therefore produces optimum results during startup and in response to transients, when employed with time and frequency step detectors. Finally, the Kalman real-time parameter estimator automatically adjusts to maintain maximum sensitivity to each parameter estimated. The standard ARIMA modelling methodology would require significant extension to have all these capabilities: the Box-Jenkins approach assumes steady state operation; it is easy to end up with non-physical models; batch parameter estimation is complicated by singularities in the maximum-likelihood function; and the ARIMA method of real-time parameter estimation shows measurable biases.

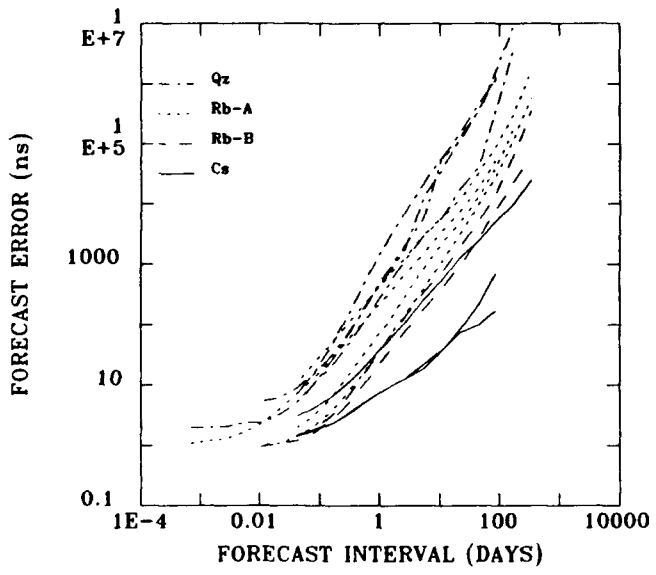


Figure 4: Minimum forecast errors for the 13 clocks included in the study.

References

[1] Continuous Time Series Models for Unequally Spaced Data Applied to Modeling Atomic Clocks, R.H. Jones and P.V. Tryon, SIAM J. Sci. Stat. Comput., vol. 8, 1987.

[2] Kalman Filter Analysis of Precision Clocks with Real-Time Parameter Estimation, S. R. Stein, Proc. 43rd Annual Symposium on Frequency Control, 1989, pp. 232-236.

[3] Gelb, A. ed. "Applied Optimal Estimation," M.I.T. Press, Cambridge, 1974.

[4] J. A. Barnes, An Adaptive Algorithm to Evaluate Clock Performance in Real Time, Proc. of the 20th Annual Precise Time and Time Interval Applications and Planning Meeting, pp. 205-218.

[5] Box, G.E.P. and Jenkins, G.M. "Time Series Analysis, Forecasting and Control." Holden Day, San Francisco, 1970.

AUTHOR INDEX

AUTHOR	PAGE
Allan, D.W.	107, 127, 151
Allen D.	615
Austin, J.	140
Babitch, D.	511
Ballato, A.	309, 444, 488
Barnes, C.E.	39
Bass, S.M.	117
Beaton, S.P.	555
Benjaminson, A.	597
Besson, R.J.	259, 593
Bidart, L.	349
Blisnuk, K.L.	251
Bourgeois, C.	367
Calhoun, M.	474
Camparo, J.C.	39
Capelle, B.	337, 416
Carru, H.	337
Chan, Y.C.	71
Clark, R.L.	493
Conley, B.	559
Conley, W.R.	207
Croitoru, N.	323
D'Albaret, B.	349
Daly, P.	127
Davis, D.D.	107
Dehmelt, H.	12
Deierling, C.	569
Delcamp, S.B.	39
DeMarchi, A.	34
Detaint, J.	337, 416
Dick, G.J.	82, 89, 577
Dokmeci, M.C.	394
Drullinger, R.E.	34, 76
EerNisse, E.P.	185
Estrick, V.	549
Evans, J.	630
Felton, C.M.	542, 622
Filler, R.L.	176
Foise, J.W.	216
Frank, J.	222
Frueholz, R.P.	39
Fukugawa, S.	59
Galliou, S.	259
Garvey, I.F.	511
Garvey, R.M.	117
Golan, G.	323
Golosovsky, M.	53
Griffel, G.	323
Gualtieri, J.G.	222, 228, 251
Guo, X.	468
Hansen, L.	98
Hanson, W.P.	478
Hart, D.W.	222
Heishman, L.C.	478
Hirama, K.	372
Hironaka, N.	107
Hisadome, K.	44
Horine, B.	316
Horton, W.H.	437
Horwitz, S.	332
Irvine, R.A.	216
Ishihara, N.	59
Jaduszliwer, B.	71
Jensik, C.	294
Johnson, G.R.	216
Johnson, W.A.	94
Joly, C.	337
Jones, R.	549
Karuza, S.K.	94
Kawashima, H.	378, 585
Kihara, M.	44
Kitching, I.D.	127
Kodama, S.	59
Korn, D.	569
Kosinski, J.	228, 444, 488
Krishnaswamy, S.	332
Kroupa, V.F.	498
Kuhnle, P.	474
Kusters, J.A.	165, 185
Lackey, R.	294
Lee, P.C.Y.	468
Lefevre, R.	337
Levine, J.	107, 151
Lukaszek, T.	444, 488
Makabe, T.	59
Maleki, L.	82
Martin, J.J.	222
Martin, T.D.	542
Mattison, E.M.	66
Mayorov, A.D.	277
McGowan, R.	444
Meeker, T.R.	478
Mihailov, D.	277
Miller, B.	559
Mizan, M.	444
Mogilevski, A.N.	277
Montress, G.K.	522
Moore, R.	332
Morley, P.E.	193
Mourey, M.	259, 593
Murray, R.A.	228, 251
Nakamura, K.	372
Nakazato, M.	378
Nakazawa, M.	358
Nemesh, A.	145
Okayama, D.	107
Oomura, Y.	363
Panitz, J.K.G.	207
Parker, T.E.	522
Parzen, B.	536
Peled, E.	145
Peppler, T.K.	127
Poll, H.-U.	238
Prestage, J.D.	82
Prevedelli, M.	555
Ramsey, N.	3
Roberts, G.E.	267
Rosenbaum, J.	332
Sarrica, R.W.	98
Sauerland, F.	246
Saunders, J.	577
Schreiter, S.	238
Schwartzel, J.	337, 416
Seidman, A.	323
Sekimoto, H.	358
Shimizu, H.	372
Shirley, J.H.	34
Sillioc, G.	349
Smith, D.	222
Smits, J.G.	309
Smythe, R.C.	437
Soderkvist, J.	406
Sorrell, P.A.	161
Spassov, L.	238, 277
Staples, A.G.	267
Stein, S.R.	630
Stern, A.	53, 145
Straka, E.R.	98
Sydnor, R.	474
Symonds, D.A.	615
Thomas, B.	285
Tiersten, H.F.	452, 461
Toliver, S.	267
Tucker, T.	577
Vale, C.	332
Valentin, J.P.	259
Vessot, R.F.C.	66
Vianco, P.T.	207
Vig, J.R.	165, 201
Vilar, F.	140
Vlitas, P.	117
Voit, F.J.	94
Walls, F.L.	542, 555
Wan, K.W.	140
Wang, R.T.	89
Watanabe, Y.	358, 363
Weiss, M.A.	107
Williamson, R.J.	193, 424
Yamagata, S.	298
Yankov, D.	238, 277
Yasuike, R.	372
Yong, Y.K.	387
Yossifov, E.	238
Zaharieva, R.	277
Zarka, A.	337, 416
Zelitzki, M.	145
Zellers, R.	294
Zheng, Y.	337, 416
Zhou, Y.S.	452, 461

SPECIFICATIONS AND STANDARDS RELATING TO FREQUENCY CONTROL

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS (IEEE)

Order from: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1987 (ANSI/IEEE) Standard on Piezoelectricity (SH11270)

177-1966 Standard Definitions & Methods of Measurements for Piezoelectric Vibrators

180-1986 (ANSI/IEEE) Definitions of Primary Ferroelectric Crystal Terms (SH10553)

319-1971 (Reaff 1978) Piezomagnetic Nomenclature (SH02360)

1139-1988 Standard Definitions of Physical Quantities for Fundamental Frequency & Time Metrology (SH12526)

DEPARTMENT OF DEFENSE

Order from: Military Specifications and Standards
700 Robbins Ave., Bldg. 4D
Philadelphia, PA 19111-5094
(215) 697-2667/2179 - Customer Service
(215) 697-1187 to -1195 Telephone Order Entry System

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for

MIL-C-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

PROCEEDINGS ORDERING INFORMATION

NO.	YEAR	DOCUMENT NO.	OBTAIN FROM*	COST MICROFICHE / HARDCOPY
10	1956	AD-298322	NTIS	$8.00 / $53.00
11	1957	AD-298323	NTIS	8.00 / 60.00
12	1958	AD-298324	NTIS	8.00 / 60.00
13	1959	AD-298325	NTIS	8.00 / 67.00
14	1960	AD-246500	NTIS	8.00 / 45.00
15	1961	AD-265455	NTIS	8.00 / 39.00
16	1962	PB-162343	NTIS	8.00 / 45.00
17	1963	AD-423381	NTIS	8.00 / 60.00
18	1964	AD-450341	NTIS	8.00 / 60.00
19	1965	AD-471229	NTIS	8.00 / 60.00
20	1966	AD-800523	NTIS	8.00 / 60.00
21	1967	AD-659792	NTIS	8.00 / 53.00
22	1968	AD-844911	NTIS	8.00 / 60.00
23	1969	AD-746209	NTIS	8.00 / 39.00
24	1970	AD-746210	NTIS	8.00 / 39.00
25	1971	AD-746211	NTIS	8.00 / 39.00
26	1972	AD-771043	NTIS	8.00 / 39.00
27	1973	AD-771042	NTIS	8.00 / 45.00
28	1974	AD-A011113	NTIS	8.00 / 45.00
29	1975	AD-A017466	NTIS	8.00 / 45.00
30	1976	AD-A046089	NTIS	8.00 / 53.00
31	1977	AD-A088221	NTIS	8.00 / 60.00
32	1978	AD-A955718	NTIS	8.00 / 53.00
33	1979	AD-A213544	NTIS	8.00 / 67.00
34	1980	AD-A213670	NTIS	8.00 / 53.00
35	1981	AD-A110870	NTIS	8.00 / 67.00
36	1982	AD-A130811	NTIS	8.00 / 53.00
37	1983	AD-A136673	NTIS	8.00 / 53.00
38	1984	AD-A217381	NTIS	8.00 / 39.00
39	1985	AD-A217404	NTIS	8.00 / 53.00
40	1986	86CH2330-9	IEEE	70.00 / 70.00
41	1987	AD-A216858	NTIS	8.00 / 53.00
42	1988	AD-A217275	NTIS	8.00 / 53.00
43	1989	89CH2690-6	IEEE	68.00 / 68.00
44	1990	90CH2818-3	IEEE	TBD

*NTIS - National Technical Information Service
5285 Port Royal Road, Sills Building
Springfield, VA 22161
Tel: 703-487-4650

*IEEE - Inst. of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854
Tel: 800-678-4333 or 201-981-0060

An index to all the papers in the Proceedings from 1956 to 1989, an author index, 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.