PROCEEDINGS

OF THE

41ST ANNUAL

FREQUENCY SYMPOSIUM

1987

Co-Sponsored 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. 87CH2427-3

Library of Congress No. 58-60781

DTIC
ELECTE
JAN 11 1990
S B D

Dunfey City Line Hotel

Philadelphia, Pennsylvania

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, 345 East 47th Street, New York, NY 10017. All rights reserved
Copyright © 1987 by the Institute of Electrical and Electronics Engineers, Inc.

FORTY-FIRST ANNUAL FREQUENCY CONTROL SYMPOSIUM

Co-Sponsored by

U.S. ARMY LABORATORY COMMAND
ELECTRONICS TECHNOLOGY AND DEVICES LABORATORY

and

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.
ULTRASONICS, FERROELECTRICS AND FREQUENCY CONTROL SOCIETY

SYMPOSIUM CHAIRMEN

General Chairman.....	Dr. John R. Vig, U.S. Army LABCOM
Technical Program Chairman.....	Dr. Leonard S. Cutler, Hewlett-Packard Company
Finance Chairman.....	Dr. Thomas E. Parker, Raytheon Research Div.
Publicity Chairman.....	Dr. Raymond L. Filler, U.S. Army LABCOM

TECHNICAL PROGRAM COMMITTEE

David Allan
National Bureau of Standards

Dr Arthur Ballato
U.S Army, LABCOM

Martin Bloch
Frequency Electronics, Inc

Dr Janet Brown
Fisher Controls International, Inc

Dr Leonard Cutler
Hewlett-Packard Company

Michael Driscoll
Westinghouse Systems Center

Dr Lawrence Dworsky
Motorola, Inc

Dr Errol EerNisse
Quartzdynamics, Inc

Dr Terry Flanagan
JAYCOR, Inc

Marvin Frerking
Rockwell International

Dr Michael Garvey
Frequency and Time Systems

Dr Larry Halliburton
Oklahoma State University

Dr Helmut Hellwig
National Bureau of Standards

Charles Jensik
Piezo Crystal Company

Dr Reynold Kagiwada
TRW-ESD

John Kusters
Ball Corporation

Nicholas Lias
Motorola Quartz Corporation

Theodore Lukaszek
U S Army, LABCOM

Dr Donald Malocha
University of Central Florida

Dr Thrygve Meeker
Bell Laboratories (ret)

Dr Thomas Parker
Raytheon Research Division

Vincent Rosati
U.S Army, LABCOM

Lauren Rueger
Applied Physics Laboratory

Stanley Schodowski
U.S Army, LABCOM

Robert Smythe
Piezo Technology, Inc

Dr Samuel Stein
Ball Corporation

Dr Richard Sydnor
Jet Propulsion Laboratory

Dr Jacques Vanier
National Research Council

Dr John Vig
U.S. Army, LABCOM

Rolf Weglein
Hughes Aircraft Company

Dr Joseph White
Naval Research Laboratory

Dr Gernot Winkler
U S Naval Observatory

Dr Nicholas Yannoni
U.S Air Force RADC

TECHNICAL SESSION CHAIRMEN

PLENARY SESSION
Leonard Cutler, Hewlett-Packard Company

RESONATOR THEORY, NONLINEAR
Arthur Ballato, U.S. Army, LABCOM

ELECTRODIFFUSION & POINT DEFECTS
Joseph Balascio, Motorola

ATOMIC STANDARDS I
Jacques Vanier, National Research Council

ATOMIC STANDARDS II
Leonard Cutler, Hewlett-Packard Company

ETCH CHANNELS & DISLOCATIONS
Alfred Kahan, U.S. Air Force, RADC

RESONATOR PROCESSING I
Thrygve Meeker, Bell Laboratories (ret.)

HYDROGEN MASERS & DISTRIBUTION
Lauren Rueger, Johns Hopkins University

DIGITAL TECHNIQUES
Marvin Frerking, Rockwell International, Collins Division

MICROWAVE & MILLIMETER OSCILLATORS
Rolf Waglein, Hughes Aircraft Company

THEORY & SPECIFICATIONS
David Allan, National Bureau of Standards

INSTRUMENTATION & TIME TRANSFER
Richard Sydnor, Jet Propulsion Laboratory

SAW DEVICES
Donald Malocha, University of Central Florida

SENSORS & TRANSDUCERS
Janet Brown, Fisher Controls International, Inc.

FREQUENCY SYNTHESIS
Fred Walls, National Bureau of Standards

RESONATOR PROCESSING II
Robert Smythe, Piezo Technology, Inc.

UHF RESONATORS & OSCILLATORS
Thomas Parker, Raytheon Company

RESONATOR THEORY-LINEAR
Jean-Jacques Gagnepain, Laboratoire de Physique et Metrologie des Oscillateurs du C.N.R.S.

QUARTZ CRYSTAL OSCILLATORS I
Michael Driscoll, Westinghouse Defense & Electronics Center

QUARTZ CRYSTAL OSCILLATORS II
Raymond Filler, U.S. Army LABCOM

TABLE OF CONTENTS

PLENARY SESSION

Award Presentations..... 1

Millisecond Pulsar Rivals Best Atomic Clock Stability..... 2
D W. Allan, National Bureau of Standards

ATOMIC FREQUENCY STANDARDS

Initial Operational Experience with a Mercury Ion Storage Frequency Standard..... 12
L.S. Cutler and R.P. Giffard, Hewlett-Packard Laboratories and P.J. Wheeler and G.M.R. Winkler, U.S. Naval Observatory

JPL Trapped Ion Frequency Standard Development..... 20
J.D. Prestage, G.J. Dick and L. Maleki, California Institute of Technology

Ultra-Sensitive Frequency Discrimination in a Diode Laser Pumped ^{87}Rb Atomic Clock..... 25
M. Hashimoto and H. Furuta, Graduate School at Nagatsuta, Japan and M. Ohtsu, AT&T Bell Laboratories

A Three Dimensional Model of the Gas Cell Atomic Frequency Standard..... 36
J.C. Camparo and R.P. Frueholz, The Aerospace Corporation

Study of Several Error Sources in a Laser Raman Clock..... 42
P.R. Hemmer and B. Bernacki, Rome Air Development Center and V.D. Natoli and M.S. Shahriar, H. Lamela-Rivera, S.P. Smith and S. Ezekiel, Massachusetts Institute of Technology

Cesium Gettering by Graphite..... 47
M.D. Bhaskar, C.M. Kahla, R.P. Frueholz and R.A. Cook, The Aerospace Corporation

New Insights into Causes and Cures of Frequency Instabilities (Drift and Long Term Noise) in Cesium Beam Frequency Standards..... 53
A. DeMarchi, University of Ancona, Italy

An Optically Pumped Cesium Beam Frequency Standard for Military Applications..... 59
T. Mc Clelland, I. Pascaru, J. Zacharski, N.H. Tran and M. Meirs, Frequency Electronics, Inc.

Development of a Rubidium Frequency Standard for the MILSTAR Satellite System..... 66
T. Mc Clelland, I. Pascaru and M. Meirs, Frequency Electronics, Inc.

Hydrogen Masers for Radio Astronomy..... 75
M. Peters, B. Owings and T. Oakley, Sigma Tau Standards Corporation and L. Beno, National Radio Astronomy Observatory

The New Generation of Hydrogen Maser at Shanghai Observatory..... 82
Z.C. Zhai, H.X. Huang, G.X. Jiang, W.H. Luo, J.F. Lu and C.F. Lin, Shanghai Observatory, China

Performance of a Hydrogen Maser with Auto-Tuning Utilizing Cavity Q Modulation..... 87
T.K. Tucker and G.J. Dick, California Institute of Technology

Hyperfine Contribution to Spin-Exchange Frequency Shifts in the Hydrogen Maser..... 91
B.J. Verhaar, J.M.V.A. Koelman, H.T.C. Stoof and O.J. Luiten, Eindhoven Technological University, The Netherlands and S.B. Crampton, Williams College

Surface Interaction of Atomic Hydrogen with Teflon..... 95
E.M. Mattison and R.F.C. Vessot, Smithsonian Astrophysical Observatory and C. Bain, S. Wasserman and G. Whitesides, Harvard University

FREQUENCY STABILITY, DISTRIBUTION AND TIME TRANSFER

Characteristics and Sources of Phase Noise in Stable Oscillators..... 99
T.E. Parker, Raytheon Research Division

Accuracy of Time Transfer in Satellite Systems..... 111
C.M. Will, Washington University

Relating the Allan Variance to the Diffusion Coefficients of a Linear Stochastic Differential Equation Model for Precision Oscillators..... 112
J.W. Chaffee, Systems Control Technology, Inc.

Theoretical Analysis of the Modified Allan Variance..... 116
L.G. Bernier, Ecole Polytechnique Federale de Lausanne, Switzerland

Frequency Stability Characterization of Hopping Sources..... 122
G.A. Kalivas and R.G. Harrison, Carleton University, Canada

A Method for Using a Time Interval Counter to Measure Frequency Stability..... 126
C.A. Greenhall, California Institute of Technology

Positioning and Timing Study of GPS C/A Code Receivers..... 130
Q. Zhuang, W.J. Klepczynski and C.F. Lukac, U.S. Naval Observatory

Estimating Combined Errors Due to Propagation and Ephemeris and their Effect on Time and Frequency Transfer..... 144
D.W. Allan, National Bureau of Standards and L. Ping-ping, Beijing Institute of Radio Measurement, China

Ku-Band Satellite Two-Way Timing Using a Very Small Aperture Terminal (VSAT)..... 149
D.A. Howe, National Bureau of Standards

Reference Frequency Distribution Over Optical Fibers: A Progress Report..... 161
G. Lutes, California Institute of Technology

PIEZOELECTRIC MATERIALS, RESONATORS, SENSORS AND SAW DEVICES

Electrodiffusion or Sweeping of Ions in Quartz..... 167
J.J. Martin, Oklahoma State University

Etch Channels in Single Crystal Cultured Quartz..... 175
G.R. Johnson and R.A. Irvine, Sawyer Research Products, Inc.

Etch Pits and Channels in Swept AT- and SC-Cut Quartz..... 183
J.R. Hunt, Piezo Technology, Inc.

Further Studies on Electrode-Diffusion-Suppressed-Swept Quartz..... 192
J.G. Gualtieri, US Army LABCOM

Effects of Initial Quartz Surface Finish and Etch Removal on Etch Figures and Quartz Crystal Q..... 199
K.H. Jones, Reeves-Hoffman Division of Dynamics Corporation of America

A Study of Dislocations and Inclusions in Alpha Quartz..... 213
A.F. Armington, J.A. Horrigan and M.T. Harris, Rome Air Development Center and J.F. Balascio, Motorola, Inc.

Radiation Effects in Vacuum-Swept Quartz..... 216
A. Kahan, F.K. Euler, and H.G. Lipson, Rome Air Development Center and C.Y Chen and L.E. Halliburton, Oklahoma State University

Dielectric Relaxation and EPR in Quartz Crystals Containing Fe..... 223
S. Keilson, S. Ling and A.S. Nowick, Columbia University and L.E. Halliburton, Oklahoma State University

Transmission X-Ray Topography of Single Crystal Quartz Using White Beam Synchrotron Radiation..... 228
W. Hanson, Piezo Crystal Company and The Johns Hopkins University

Studies of Quartz Resonators by Stroboscopic Topography..... 236
A. Zarka and B. Capelle, Universite P. et M. Curie, France and J. Detaint and J. Schwartzel, Centre National d'Etudes des Telecommunications, France

Cultured Quartz Quality Standards Updated by the EIA..... 241
B. Sawyer, Sawyer Research Products, Inc.

Computing Crystal Orientation from X-Ray Measurements..... 243
J.H. Sherman, Jr., Lynchburg, VA

X-Ray Technology - A Review..... 249
C.A. Adams, Hewlett Packard Company, D.C. Bradley, Colorado Crystal Company and J.A. Kusters, Ball Corporation

An Update of Surface Mount Packages for Quartz Crystal Products..... 258
C. Mercer, Reeves-Hoffman Division

Nonlinear Constants and Their Significance..... 266
J.J. Gagnepain, Universite de Franche-Comte-Besancon, France

Acceleration Effect on the Thickness Vibrations of Doubly Rotated Crystal Resonators..... 277
P.C.Y. Lee and M.S.H. Tang, Princeton University

An Analysis of th Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Rigidly Supported Along the Edges..... 282
H.F. Tiersten and D.V. Shick, Rensselaer Polytechnic Institute

Force Sensitivity of Trapped Energy Vibrations in a Contoured Resonator..... 289
R. Bourquin and B. Dulmet, Ecole Nationale Superieure de Mecanique et des Microtechniques, France

AT Quartz Strip Resonators..... 295
L.N. Dworsky, Motorola, Inc.

A Lagrangean, High Frequency Plate Elemen', for the Static Temperature Behavior of Low Frequency Quartz Resonators..... 303
Y.K. Yong, Rutgers University

An Approximate Expression for the Motional Capacitance of a Lateral Field Resonator..... 311
R.C. Smythe, Piezo Technology, Inc. and H.F. Tiersten, Rensselaer Polytechnic Institute

Energy Trapping in Plane and Corrugated Resonators: Application to Quartz and Berlinite..... 314
J. Detaint, J. Schwartzel and C. Joly, Centre National d'Etudes des Telecommunications, France and E. Philippot, Universite de Montpellier, France

Lateral- and Thickness-Field Coupling in Zincblende Structures..... 325
A. Ballato, T. Lukaszek, M. Mizan and J. Kosinski, US Army LABCOM

New Prospects for Acoustic Sensors: An Overview..... 333
R.M. White, University of California

Theoretical Modeling of Quartz Resonator Pressure Transducers..... 339
E.P. EerNisse, Quartztronics, Inc.

A Reduced Hysteresis, Extended Range Quartz Pressure Transducer..... 344
R.W. Ward and E.P. EerNisse, Quartztronics, Inc.

Study of Liquids in Shear Using a Quartz Resonator..... 350
K.K. Kanazawa and C.E. Reed, IBM Almaden Research Center

A Novel Technique for Trimming the Frequency of a Sealed Surface Acoustic Wave Resonator..... 351
J.A. Greer and T.E. Parker, Raytheon Research Division and M. Rothschild and D.J. Ehrlich, Massachusetts Institute of Technology

Parametric Failure Rate Model for Quartz Crystal Device Aging with Application to Surface Acoustic Wave Filters..... 360
A.A. Feinberg, AT&T Bell Laboratories

Approximation Method for Plate Modes in Surface Acoustic Wave Devices..... 365
J. Hou, Allied-Signal, Inc.

Thin Film Resonator Technology..... 371
K.M. Lakin, G.R. Kline, R.S. Ketcham, A.R. Landin, W.A. Burkland, K.T. McCarron, S.D. Braymen and S.G. Burns, Iowa State University

UHF Oscillator Performance Using Thin Film Resonator Based Topologies..... 382
S.G. Burns, G.R. Kline and K.M. Lakin, Iowa State University

Shear Mode Transducers for High Q Bulk Microwave Resonators..... 388
H.L. Salvo, Jr., Westinghouse Defense and Electronics Center and M. Gottlieb and B.R. McAvoy, Westinghouse R&D Center

Analysis and Design of the Piezoelectric Ceramic Resonator Oscillators..... 391
S. Fujishima, K. Togawa and S. Ohta, Murata Mfg. Co., Ltd., Japan

CRYSTAL OSCILLATORS, MICROWAVE AND MILLIMETER WAVE OSCILLATORS AND FREQUENCY CONTROL CIRCUITRY

The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review..... 398
R.L. Filler, US Army LABCOM

Suppression of Vibration-Induced Phase Noise in Crystal Oscillators: An Update..... 409
V.J. Rosati, US Army LABCOM

Crystal Oscillators for Airborne Applications..... 413
G. Caret, A. Debaisieux E. Gerard and G. Robichon, CEPE, France

The Relationship Between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques..... 420
G.S. Curtis, Hewlett-Packard Company

Fabrication and Methods for Evaluation and Circuit Utilization of Prototype Lateral Field Resonators..... 429
M.M. Driscoll, Westinghouse Defense and Electronics Center and W.P. Hanson, Piezo Crystal Company

A Digitally Compensated TCXO Using a Single Chip LSI..... 435
T. Hara, T. Kudo, S. Uriya, H. Saita, S. Ogou and Y. Katsuta, NEC Corporation, Japan

Analysis of High Performance Compensated Thermal Enclosures..... 439
F.L. Walls, National Bureau of Standards

The Aging of Resonators and Oscillators Under Various Test Conditions..... 444
R. Filler, R. Lindenmuth, J. Messina, V. Rosati and J. Vig, US Army LABCOM

The Design and Analysis of VHF/UHF Crystal Oscillators..... 452
A. Benjaminson, Systematics General Corporation

CMOS Gate Oscillator Design..... 460
T.B. Hills, National Semiconductor Corporation

Review of the Revised Military Specification for Quartz Crystal Oscillators..... 466
S. Schodowski and V. Rosati, US Army LABCOM

Noise in Oscillators Employing Submicron Field-Effect Transistors..... 471
M.S. Gupta, University of Illinois at Chicago

Review of Dielectric Resonator Oscillator Technology..... 478
A.P.S. Khanna, Avantek Inc.

Measurements and Analysis of Cryogenic Sapphire Dielectric Resonators and DRO's..... 487
G.J. Dick and D.M. Strayer, California Institute of Technology

Compact and Simple x 3 (9 to 27 GHz) PLL Frequency Multiplier Using Harmonic Phase Detection..... 492
J. Berenguer, J. Bara and A. Comeron, E.T.S.I. Telecomunicacion, Spain

An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation..... 495
H.T. Nicholas, III and H. Samueli, University of California

The Impact of Digital Signal Processing on Crystal Filter Requirements..... 503
M.E. Frerking, Rockwell International Corporation

Random Noise in Digital Gates and Dividers..... 507
D.E. Phillips, Rockwell International Corporation

Low Noise Frequency Synthesis..... 512
F.L. Walls, National Bureau of Standards and C.M. Felton, Felton Electronics Design

Low Noise Airborne Synthesizer for Frequency Agile Radar..... 519
A. Vulcan and M. Bloch, Frequency Electronics, Inc.

Improvement of the Pull-In Range and Acquisition Time of a Microwave P.L.L. System by Injection Locking the V.C.O..... 524
J. Berenguer, J. Bara, E. Artal, I. Corbella and A. Comeron, E.T.S.I. Telecomunicacion, Spain

LATE SUBMISSIONS

Progress Report on the EIA/P.11 Round Robin Crystal Measurements Experiment..... 527
W.L. Smith, AT&T Bell Laboratories

High Frequency Stable Frequency Sources for Advanced Systems..... 539
A. Harrison, J. Dowsett and D. Sharpe, STC Components, England

A High Linearity SAW Accelerometer..... 544
D. Bower, M. Cracknell and A. Harrison, STC Components, England

Use of Annular SAW for Cutting Quartz Resonator Blanks and Comparison with Other Methods..... 548
J. Dowsett, R.B. Spencer and A.F.B. Wood, STC Components, England

AUTHOR INDEX..... 553

SPECIFICATIONS AND STANDARDS GERMANE TO FREQUENCY CONTROL..... 554

PROCEEDINGS AVAILABILITY INFORMATION..... 555

1987 AWARD WINNERS

The Cady Award

The Cady Award was presented to Virgil E. Bottom "for contributions to fundamental theory and experiments, stimulation of growth of the industry, and education in quartz resonator technology." The award was presented by Charles Adams, Hewlett-Packard Company.

The Rabi Award

The Rabi Award was presented to Louis Essen "for contributions to cesium atomic beam and quartz frequency standards." The award was presented by David Allan, National Bureau of Standards.

The Sawyer Award

The Sawyer Award was presented to John A. Kusters "in recognition of outstanding contributions in engineering, technology development and management relating to quartz crystals and devices." The award was presented by Charles Adams, Hewlett-Packard Company.



John R. Vig, General Chairman; Virgil E. Bottom, Cady Award winner; John A. Kusters, Sawyer Award winner; and Leonard S. Cutler, Technical Program Chairman, after the award presentation. Louis Essen, Rabi Award winner, was unfortunately unable to travel to the Symposium to accept his award.

MILLISECOND PULSAR RIVALS BEST ATOMIC CLOCK STABILITY

by
David W. Allan
Time and Frequency Division
National Bureau of Standards
325 Broadway, Boulder, CO 80303

Abstract

The measurement time residuals between the millisecond pulsar PSR 1937+21 and atomic time have been significantly reduced. Analysis of data for the most recent 865 day period indicates a fractional frequency stability (square root of the modified Allan variance) of less than 2×10^{-14} for integration times of about 1/3 year. The reasons for the improved stability will be discussed; these are a result of the combined efforts of several individuals.

Analysis of the measurements taken in two frequency bands revealed a random walk behavior for dispersion along the 12,000 to 15,000 light year path from the pulsar to the earth. This random walk accumulates to about 1,000 nanoseconds (ns) over 265 days. The final residuals are nominally characterized by a white phase noise at a level of 369 ns.

Following improvement of the signal-to-noise ratio, evidence was found for a residual modulation. Possible explanations for this modulation include: a binary companion (or companions) to the pulsar with approximate period(s) of 120 days and with a mass (or masses) of the order of 1×10^{-9} that of the pulsar; irregular magnetic drag in the pulsar; unaccounted delay variations in the interstellar medium; modeling errors in the earth's ephemeris; reference atomic clock variations in excess of what are estimated; or gravity waves. For gravity waves, the amplitude of the length modulation would be about 5 parts in 10^{19}. Further study is needed to determine which is the most probable explanation.

Introduction

Timekeeping has historically evolved with astrometry; e.g. the rotation of the earth, the orbit of the earth around the sun or the moon around the earth have been fundamental pendula for time keeping. As atomic clocks were shown to be more accurate and stable than those based on astrometry, the second was redefined.[1] It now seems that an astronomical phenomenon [2] may rival the best atomic clocks currently operating. The current best estimate of the period of the millisecond pulsar (PSR 1937+21) is 1.557 806 451 698 38 ms ±0.05 fs as of 6 October 1983 at 2216 UT.[3] This accuracy is such that we could wait over 100 years between measurements of the arrival times of signals from the pulsar before being concerned with which pulse we were counting. The period derivative has been measured as $P = (1.051053 \pm 0.000008) \times 10^{-19}$ seconds per second which is 3.31687 parts in 10^{12} per year. This frequency drift is less than that of a typical rubidium frequency standard and greater than that of a typical cesium frequency standard. However, in the case of the pulsar the drift rate is exceedingly constant; i.e., no second derivative of the period has been observed.[4] The very steady slowing down of the pulsar is believed to be caused by the pulsar radiating electromagnetic and gravitational waves.[5]

Time comparisons on the millisecond pulsar require the best of measurement systems and metrology techniques. This pulsar is estimated to be about 12,000 to 15,000 light years away. The basic elements in the measurement link between the millisecond pulsar and the atomic clock are the dispersion and scintillation due to the interstellar medium, the computation of the ephemeris of the earth in barycentric coordinates, the relativistic transformations because of the dynamics and gravitational potentials of the atomic clocks involved with respect to the reference frame of interstellar space, the sensitivity of the Arecibo Observatory (AO) radio telescope (area of 73,000 m^2 or 18 acres), the accuracy of the Princeton-installed measurement system, the filter bank and data processing techniques for determining the arrival time of the pulses, the transfer of time from the Arecibo Observatory atomic clock to the time from the international timing centers, and, last the algorithms for combining the clocks in the world ensemble to provide the atomic clock reference.

New Measurement Techniques

Since the discovery of the millisecond pulsar by Backer and Kulkarni [2] (14 November 1982), several very significant improvements in the ability to measure the pulsar have occurred. Figure 1 is a stability plot of the residuals over the first two years after all the then-known perturbations were removed. The stability is characterized by a $1/f$ phase modulation (PM) spectral density. The standard deviation of the time residuals over the first two years was 998 ns. In the fall of 1984 the group at Princeton installed a new data acquisition system in conjunction with the filter bank for better determination of the arrival time of the pulses. Nearly simultaneously NBS in cooperation with AO installed a GPS common-view receiver for the link between the Arecibo clock and international timing centers. The white PM noise of the GPS common-view link is less than 10 ns.

The data from the pulsar included measurements made at both 1.4 GHz and 2.38 GHz.[3] The data were unequally spaced with the average sample period varying between about 3 and 20 days depending upon the data segment. Figure 2 and Figure 3 are plots of the raw residuals over the period from the Fall of 1984 to February 1987. Since the data were unequally spaced we analyzed the data in two ways. First, taking the numbers as a simple time series they were analyzed as if they were equally spaced with the assumed spacing τ_0 equal to the average spacing between the data points. Second, we used the actual number of points available, but linearly interpolated

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

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

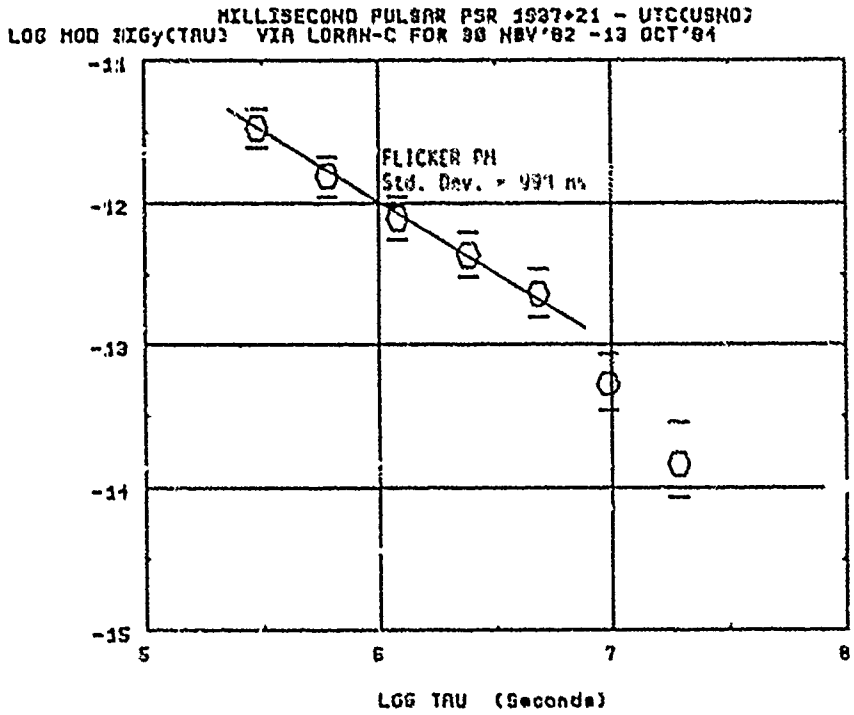


Figure 1. A plot of the square-root of the modified Allan variance, $\sigma_y(\tau)$ as a function of integration time, τ, for the first two years of measurements of the millisecond pulsar timing. Loran C was the time transfer means to relate to UTC(USNO).

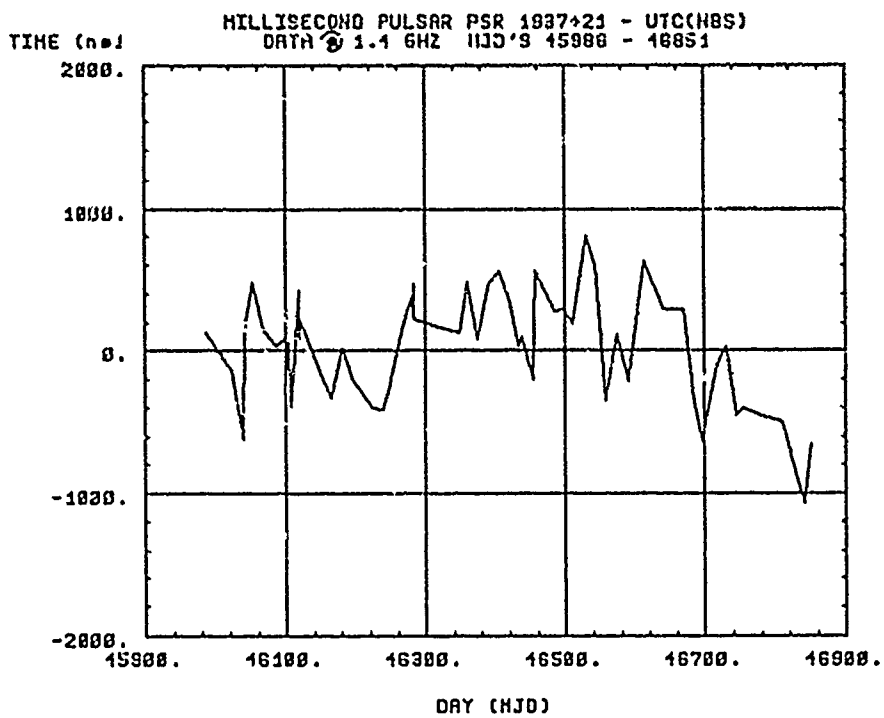


Figure 2. A plot of the residuals of the measurements of the millisecond pulsar time at 1.4 GHz versus UTC(NBS) via the GPS common-view time transfer technique and after installation of the upgraded Princeton measurement system.

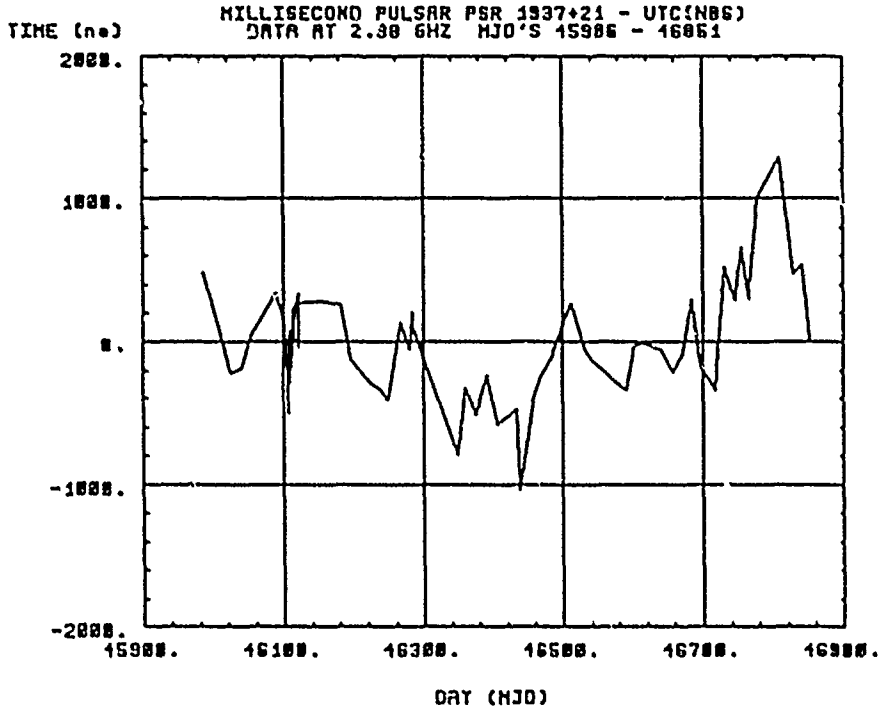


Figure 3. A plot of the residuals of the measurements of the millisecond pulsar time at 2.38 GHz versus UTC(NBS) via the GPS common-view time transfer technique and after installation of the upgraded Princeton measurement system.

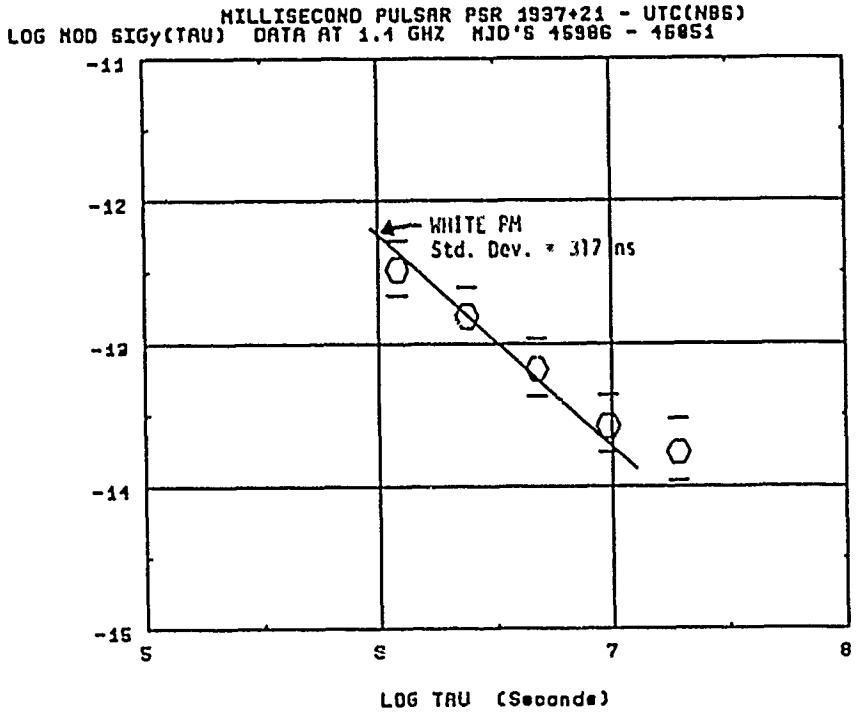


Figure 4. A plot of the square-root of the modified Allan variance, $\sigma_y(\tau)$ as a function of integration time, τ, for the 1.4 GHz data shown in Figure 2.

a data value between adjacent actual values to construct an equally spaced data set. The latter approach had the effect of decreasing the amplitude of the higher frequency Fourier components and increasing the amplitude of the lower frequency Fourier components. The conclusions drawn from the two different methods of analysis were consistent. Since it is the nature of white noise (random uncorrelated deviations) that a measurement is independent of the data spacing, and the modified Allan variance indicated white PM, then that indication is a necessary but not sufficient test to prove that the measurement noise is white PM. On the other hand, if Mod.$\sigma_y(\tau) \propto \bar{\sigma}_y(\tau)$ does not behave as $\tau^{-3/2}$, then this is a necessary and sufficient test that the spectrum is something other than a white noise process. As will be shown later, the latter situation is applicable to our case.

Pulsar Stability Analysis

Figure 4 shows the fractional frequency stability plot $\bar{\sigma}_y(\tau)$, for the 1.4 GHz data against UTC(NBS). There are several significant differences between these data and those taken over the first two years using Loran C. First, the spectral density has changed from a 1/f PM to a more nearly white PM process. Second, the noise level has been reduced significantly. A major part of this reduction is undoubtedly due to the new Princeton measurement system installed in the Fall of 1984. For the shorter integration times the level now is a white PM at 317 ns. A white noise process, if it is normally distributed, can be totally characterized by the mean and the standard deviation. The standard deviation is given by the equation,

$$x_{rms}(\tau_0) = \frac{\tau^{3/2}}{\tau_0^{1/2}\sqrt{3}}\,\bar{\sigma}_y(\tau), \qquad (1)$$

for any τ, for an average data spacing τ_0 and for the white noise PM case. For power law spectra, $S_y(f) \propto f^{\alpha}$, a process with α less than or equal to +1 will have a standard deviation of time residuals which is non-convergent. Hence, the standard deviation is not a good measure of these processes but only of a white noise PM process. If the standard deviation is used in the case $\alpha \leq +1$ then its value is data length dependent. In fact, the ratio of the classical variance to the square of the white PM level (equation 1) is a good measure of the divergence of a process; if the ratio is not 1, the process is not white.[6,7] There is an apparent flattening in the stability plot shown in Figure 4 for the longer integration times. This will be discussed in detail later, but that flattening indicates that the residuals are not random and uncorrelated (white noise). The standard deviation of the residuals around a linear regression for the data in Figure 2 is 386 ns over the 865 days. The ratio of the classical variance to the squared white PM level is 1.44 ± 0.17, which clearly indicates that the residuals are not white noise.

Figure 5 shows the same frequency stability measurement at 2.38 GHz. The noise level is nominally modeled by 325 ns of white PM. Some flattening for the longer integration times can be seen similar to Figure 4. The standard deviation of the residuals around a linear regression on the data in Figure 3 was 378 ns for the 865 days. Again it is apparent that the residuals are not random and uncorrelated as the ratio of the classical variance to the squared white PM level is now 1.42 ± 0.17. The stability plots in Figures 4 and 5 are quite similar.

Since the stability of UTC(NBS) can be determined independently of this measurement procedure, a very careful analysis of the data taken over 1000 days covering the pulsar analysis period was performed. The stability of UTC(NBS) with coordination entries removed -- denoted AT1 -- was compared in an "N-cornered-hat"[7] procedure against other primary timing centers. This was accomplished using the international NBS/GPS common-view technique, which supplies data to the BIH for the generation of International Atomic Time (TAI).[7] The measurement noise for all of the time comparisons was less than 10 ns for the white noise PM. Figure 6 shows a plot of the estimated frequency stability for the NBS AT1 time scale and indicates that the stability of AT1 is typically less than 10^{-14}.

The following two equations are proposed as a model of the time residuals for the system.

$$X_1 = X_p - X_{D1} - X_A, \qquad (2)$$

$$X_2 = X_p - X_{D2} - X_A, \text{ where} \qquad (3)$$

X_1 is the residual time series at 1.4 GHz,

X_2 is the residual time series at 2.38 GHz,

X_p is the pulsar noise,

X_A is the UTC(NBS) noise,

X_{D1} is the delay variation between X_p and X_A at 1.4 GHz, and

X_{D2} is the delay variation between X_p and X_A at 2.38 GHz.

On a given day X_p and X_A are assumed to be the same in the two equations because of the high Q of the pulsar and the measured dispersion of the atomic clock over the two hour period during which the pulsar is measured at the two frequencies. Taking the variance of the difference between Equations 2 and 3 allows us to study delay variation effects between the two signals:

$$X_2 - X_1 = X_{D1} - X_{D2}. \qquad (4)$$

Figures 7 and 8 are plots of the time residuals and of $\sigma_y(\tau)$ respectively for the difference given by Eq. 4. For Figure 8 the input data were equally spaced to obtain better spectral estimates. The $\tau^{-1/2}$ behavior yields a random walk of the dispersion (white noise frequency modulation, FM), which would accumulate to a level of one microsecond at an integration time of about 2/3 year. Figure 8 demonstrates that the dispersion delay was not constant [3].

A differential delay of 1/2 microsecond over fifteen thousand light years is 1 part in 10^{18},

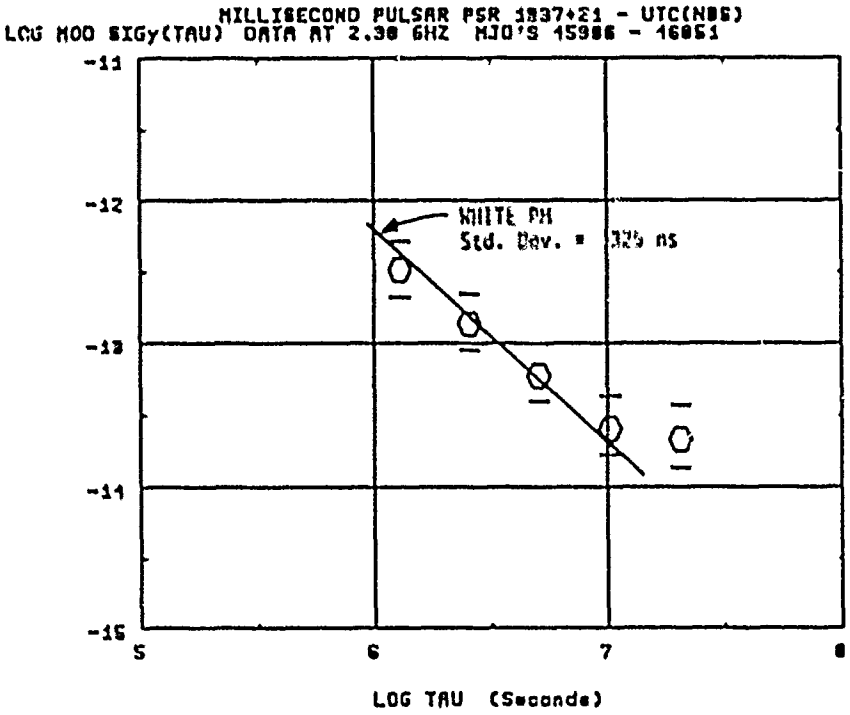


Figure 5. A plot of the square root of the modified Allan variance, $\sigma_y(\tau)$ as a function of integration time, τ, for the 2.38 GHz data shown in Figure 3.

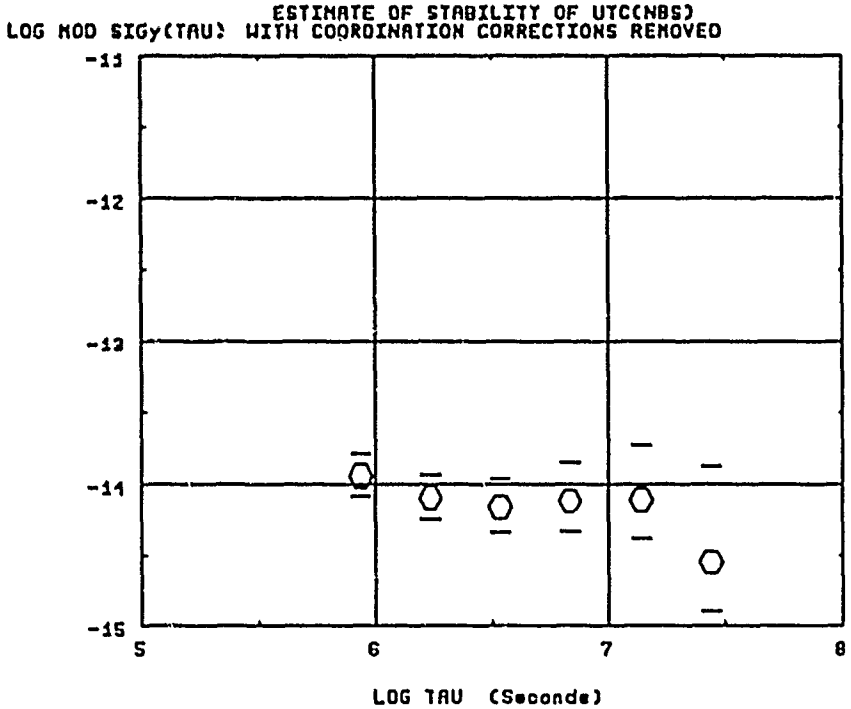


Figure 6. An estimate of the stability of UTC(NBS) with coordination corrections subtracted. The reference used for the estimate was an optimum weighted set of times from all of the international timing centers available via the GPS common-view technique. The NBS algorithm used for this computation is an effort to generate a world's "best clock" as future reference for the millisecond pulsar measurements.

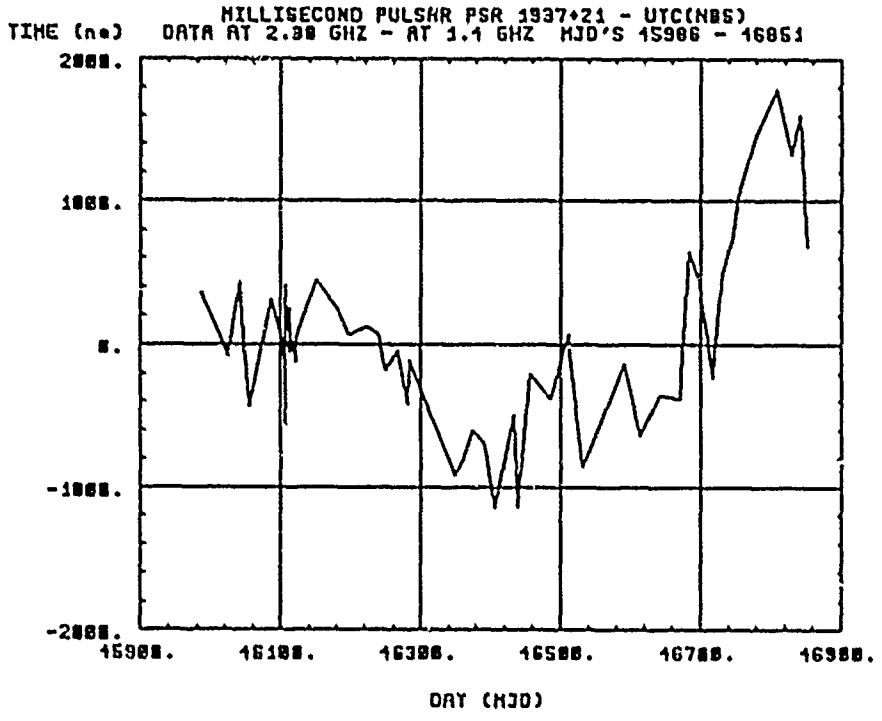


Figure 7. A plot of the difference between the 1.4 and the 2.38 GHz data shown in Figures 5 and 6. This illustrates the apparent random walk of the total electron content in the interstellar medium

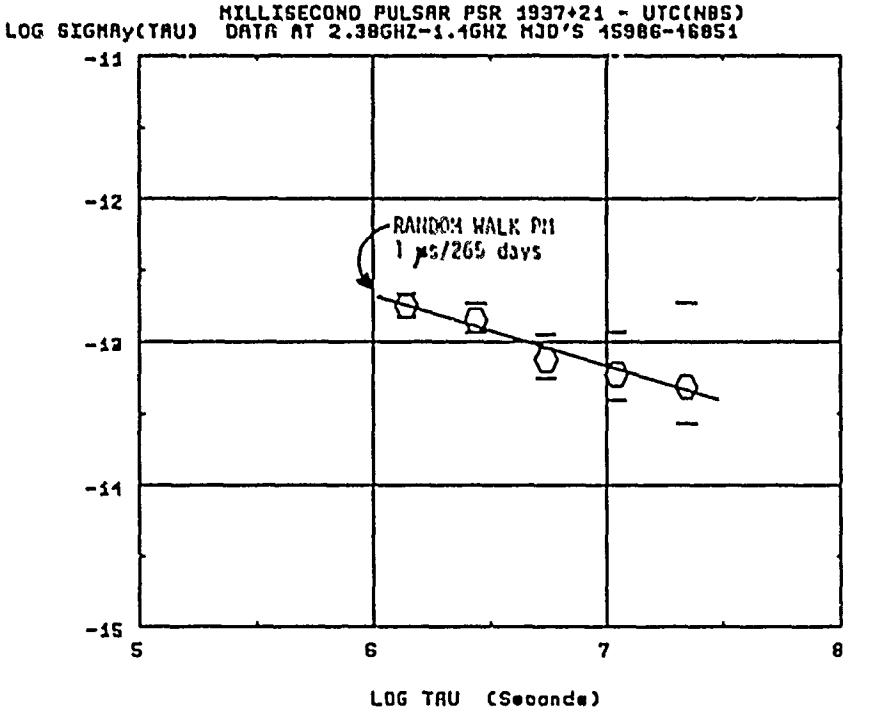


Figure 8. A plot of the square-root of the Allan variance, $\sigma_y(\tau)$ as a function of integration time, τ, in order to estimate the spectral type and level of the differential delay variations between the 1.4 GHz and 2.38 GHz signals received from the pulsar. The $\tau^{-1/2}$ line shown conforms with a random walk, f^{-2}, spectrum of the phase modulation (PM), which is the same as white noise FM (frequency modulation).

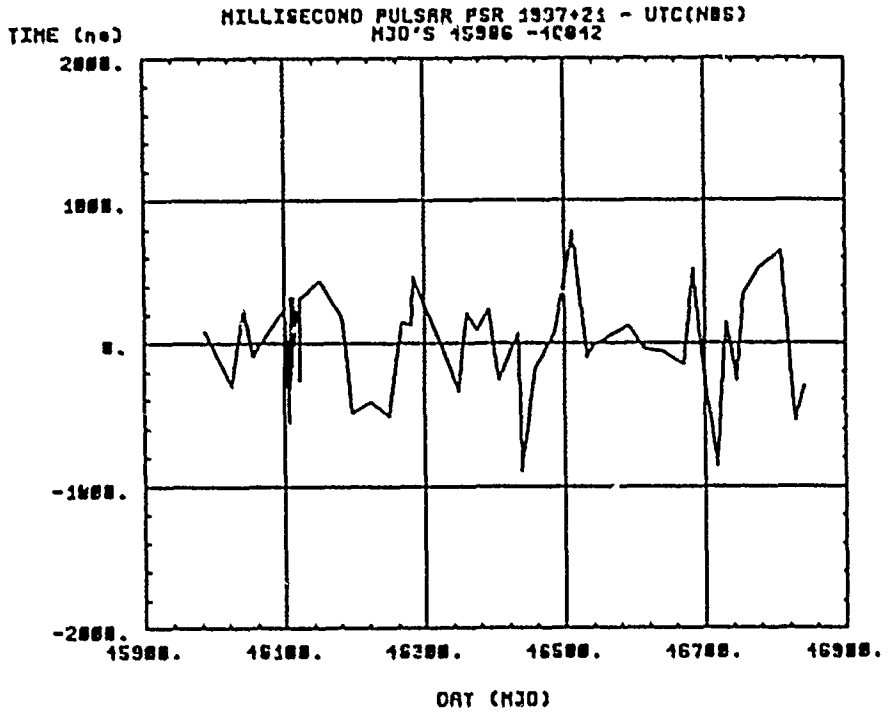


Figure 9. A plot of the residuals after compensation for the variations in the differential delay dispersion -- apparently due to the random walk of the total electron content along the path through the interstellar medium.

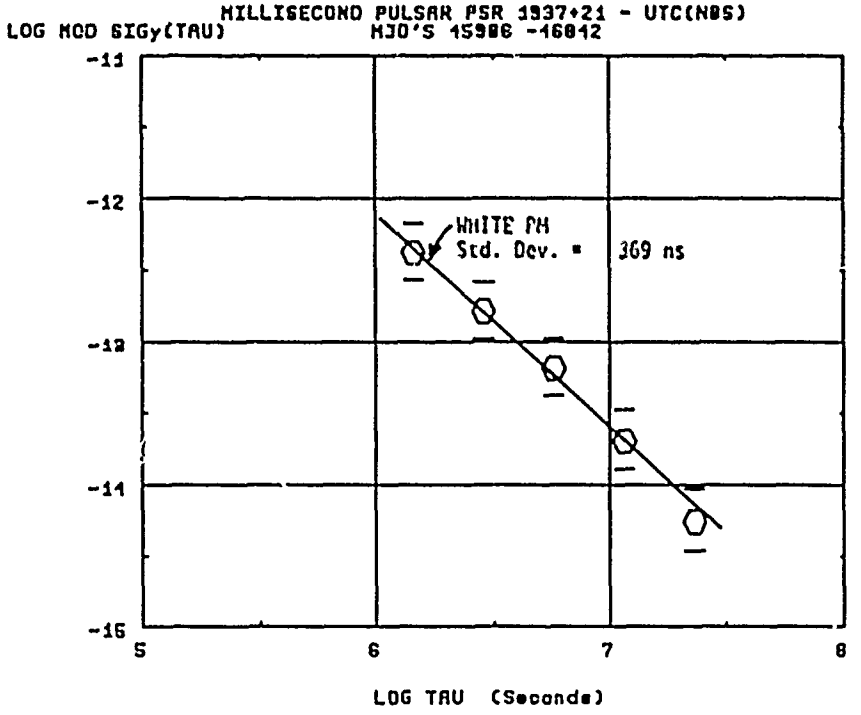


Figure 10. A plot of the square-root of the modified Allan variance, $\sigma_y(\tau)$ as a function of integration time, τ. The residuals analyzed here have been corrected to compensate for the delay due to the variations in total electron content along the path from the pulsar. The residuals appear to be well modeled by white noise PM at a level of about 369 ns as indicated by the line drawn through the frequency stability values.

suggesting that we are dealing with a very good vacuum and that apparently the variations in the total electron content across the interstellar medium can be characterized as a random walk process for this particular path and at integration times of the order of a month and longer. Since this is the first indication of its kind, further study of other paths would be of interest.

The next logical step is to use the differential dispersion delay from (4) as a calibrator. When this is done the combined residuals are plotted in Figure 9 and the frequency stability is plotted in Figure 10. The white PM level is now 369 ns and the standard deviation is 371 ns for the 556 days common to both frequencies. The ratio of the classical variance to the squared white PM level is now 1.01 indicating that the white PM model is a good one. Even though it fits the model well, the lowest frequency stability value (τ=262 days) of 5.5×10^{-15} is probably biased low due to fitting the parameters in determining the ephemeris of the earth. In order to estimate the stability of the pulsar alone, the three-corner-hat technique was employed [7]; the other two corners were TAI and ATI. The best stability estimate for the pulsar was $\bar{\sigma}_y(\tau = 134$ days$) = 1.4 \times 10^{-14}$, which is consistent with a white PM level of 205 ns.

Observation of Correlations

There is evidence of anti-correlation between the variations of X_1 and X_2 as can be seen by visual inspection of Figures 2 and 3. We can study the cross-correlation between the signals by taking a modified Allan variance of Equation 4, solving for the cross term and normalizing it:

$$\bar{\sigma}^2_{(y_2 - y_1)}(\tau) = \bar{\sigma}^2_{y_1}(\tau) - 2\bar{\sigma}_{y_1y_2}(\tau) + \bar{\sigma}^2_{y_2}(\tau), \qquad (5)$$

$$\bar{\rho}(\tau) = \frac{\bar{\sigma}_{y_1y_2}(\tau)}{\bar{\sigma}_{y_1}(\tau) \cdot \bar{\sigma}_{y_2}(\tau)} . \qquad (6)$$

Equations 5 and 6 can, of course, be written in terms of $\sigma_y(\tau)$ as well. The advantage of this cross-correlation analysis approach is that it acts like a high-pass filter with maximum sensitivity at Fourier frequencies centered at $1/2\tau$; i.e. if there are low frequency components or drifts between the signals being cross-correlated these are attenuated. Because of the apparent random walk of the free electrons in the interstellar medium, this approach was useful. In addition the $\bar{\sigma}_y(\tau)$ optimally averages the phase if the measurement noise is white PM, which is also our case.

Plotting $\bar{\rho}(\tau)$ versus τ in Figure 11 shows a very interesting positive cross-correlation coefficient of 0.7 at τ = 60 days. Then the coefficient goes negative for the larger values of τ. The negative coefficient is believed to be processing noise and is due to taking a nominal mean value of the dispersion between the two channels which are random walking with respect to each other as well; hence it appears that one retards as the other advances.

The cause of the positive cross-correlation coefficient is not known other than it is something common to the two channels. This could include heretofore unknown perturbations in the atomic time reference, the ephemeris for the earth, the coordinate or relativistic transformations, delays in the interstellar medium, the pulsar or gravity wave radiation.

In order to better understand the source of these unknown perturbations the whole process was simulated including the assumed white PM measurement noise, the random walk of the free electrons in the interstellar medium and a band of sinewaves of about the right period and amplitude to produce the effect shown in Figure 11. Figure 12 is the result of the simulation.

Analyzing the band of sinewaves with $\bar{\sigma}_y(\tau)$ yields a value of $\bar{\sigma}_y(\tau$ = 60 days) $\cdot 10^{-13}$. Figure 6 suggests that UTC(NBS) is not the cause of the unknown perturbations even though it is about a factor of 2 or 3 less stable than ATI because of the coordination corrections.

Some of the experts [8] in solar system dynamics believe that there are not mismodeling errors at Fourier components of about 3 cycles per year (f = 10^{-7} Hz) that would have an amplitude of the order of 50 meters. The mismodeling errors of the dispersion delay are believed to be below this level, and the variations due to interstellar scintillations are believed to be below the 10^{-14} level, though the dispersion delay and interstellar scintillations need more study.

There are no known transformation errors of the size needed to explain the unknown perturbations [9], so that leaves either the pulsar or gravity wave radiation as the probable cause. Deciding which will be incredibly difficult. If it is the pulsar, possible causes could be: star quakes, as are apparent with other pulsars, irregular magnetic drag, a system of planets or a planet orbiting the pulsar.

If gravity wave radiation is the cause, it can be due to the radiation sweeping over the earth or over the pulsar causing the apparent relative clock rates to fluctuate. In order to distinguish this from other causes one or more pulsars will probably be needed. Fortunately, two more millisecond pulsars are coming up to the horizon, and the stability of PSR 1855+09 is encouraging; it has a period of 5.362 100 452 553 ns ± 69 fs.

If correlated variations now being observed are due to distance modulation between earth atomic clocks and PSR 1937+21, they have an amplitude of about 5 parts in 10^{19}. The calculated levels of the cosmic strings and primordial nucleosynthesis gravitational radiation are in the same vicinity as these unknown perturbations being measured for Fourier frequencies of about 10^{-7} Hz [10]. Lower frequencies, even though theoretically more intense, are more difficult to measure because of the fitting parameters in determining the ephemeris of the earth. e.g. the one cycle per year and two cycle per year terms are not expected in the residuals because of the annual

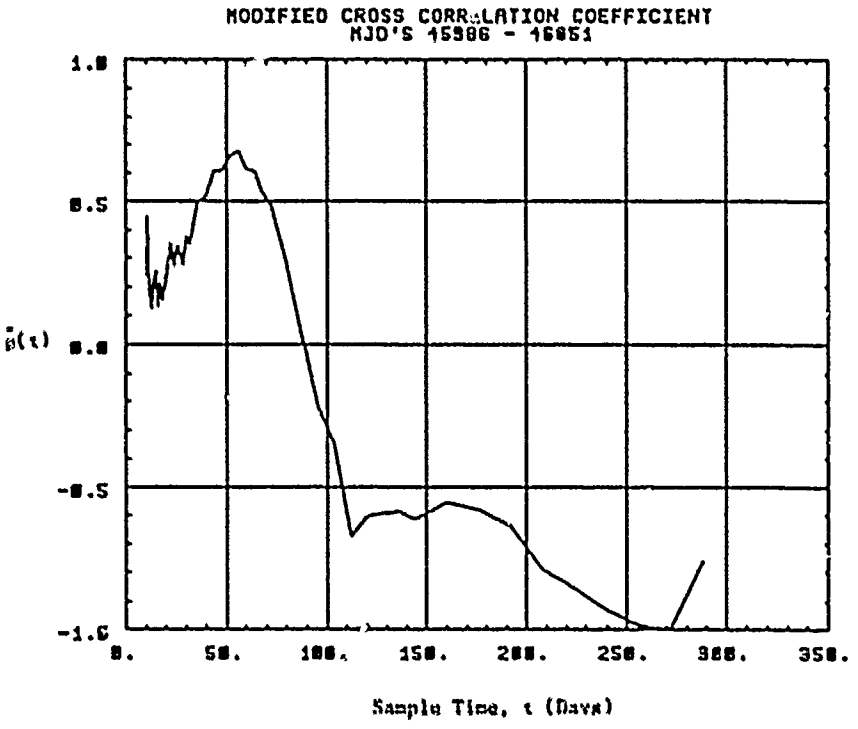


Figure 11. A plot of a cross-correlation coefficient, as defined in equation (6), as a function of integration time, τ. The positive cross-correlation of 0.7 at τ = 60 days corresponds to some unknown instabilities in the over all measurements with Fourier components in the vicinity of three cycles per year (10^{-7} Hz).

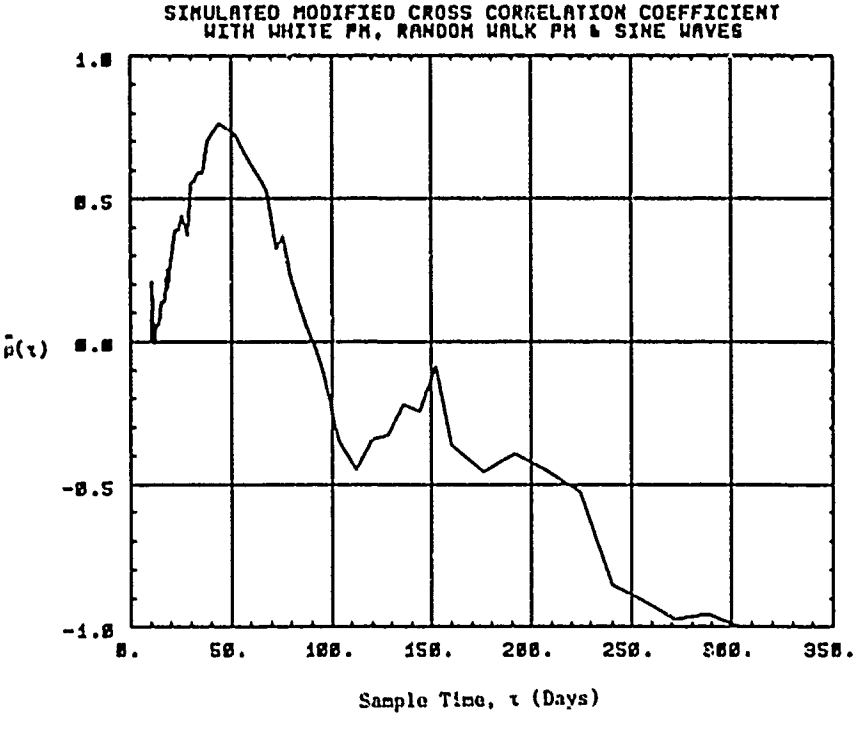


Figure 12. A plot of a cross-correlation coefficient, as defined in equation (6), as a function of integration time, τ for simulated data.

closure and to remove the effects of parallax, respectively.

Conclusion

There are some obvious next steps for improvements in this exciting area of metrology. Work is in progress at Princeton which should decrease the measurement noise and several major timing centers are working to improving the performance of atomic clocks. The BIH and the NAS have made significant progress in combining the best clocks in the world in an optimum weighted algorithm to create the world's "best clock" as a reference. Further studies on models for the interstellar dispersion and its effect on stability of measurements are needed. Another important objective is to find another pulsar in a region of space providing some orthogonality and with adequate stability. This would offer improved opportunity for detection of background gravity wave radiation. PSR 1855+09 holds some promise. Progress on primary reference atomic clocks which might provide a better earth-bound reference is going well. These clocks have improved an order of magnitude every seven years since their introduction in 1949 and we see no reason to believe that this trend will not continue.[11] A mercury ion standard with a transition in the optical region of the spectrum shows theoretical promise for long-term stability of 1 part in 10^{18}, though it will probably be several decades before this potential accuracy is realized.[12] Currently the millisecond pulsar PSR 1937+21 is proving to be a very valuable tool in evaluating long term instabilities in the best time scales in the world.

Acknowledgements

The author is indebted to a large number of people who have made the data available for the herein reported analysis. Because in several aspects this paper is a review the credit list includes several of the authors referenced. In particular the Princeton group (Prof. Joseph Taylor, Prof Daniel Stinebring and Dr. Lloyd Rawley) has been extremely helpful and cooperative in supplying data, suggestions and discussions. In addition, Dr. Michael Davis at Arecibo Observatory, Prof. Neil Ashby of the University of Colorado, Dr. Ronald Hellings of JPL and Dr. Marc Weiss of NBS have provided some very useful comments. Also there are several at NBS to thank who have provided great assistance with data acquisition and analysis -- in particular Mr. Dick Davis, Dr. Judah Levine, Ms. Trudi Peppler and Mrs. Lin Ping-ping (guest scientist).

References

[1] NBS Special Publication 330 The International System of Units (SI), U. S. Department of Commerce, National Bureau of Standards, 1981 Edition.

[2] Backer, D. C., Kulkarni, S. R, Heiles, C., Davis, M. M. & Goss, W. M., Nature 300, pp. 615-618 (1982).

[3] Rawley, L. A., "Timing Millisecond Pulsars," Thesis submitted to the faculty of Princeton University, October 1986.

[4] Blanford, R., Narayan, R., Romani, R. W., "Arrival Time Analysis for a Millisecond Pulsar," Theoretical Astrophysics, California Institute of Technology, Pasadena, CA (1984).

[5] Davis, M. M., Taylor, J. H., Weisberg, J. M., Backer, D. C., "High-precision timing observations of the millisecond pulsar PSR1937+21," Nature 315, pp. 547-550 (1985).

[6] Allan, D.W., "Should the Classical Variance Be Used As a Basic Measure in Standards Metrology?," IEEE Trans. on I & M, Vol. IM-36, No. 2, June 1987.

[7] Weiss, M. A. and Allan, D. W., "An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data," Proceedings of the CPEM Conference, Gaithersburg, MD, June 1986.

[8] Private communication, Dr. Ronald Hellings and Prof. Neil Ashby.

[9] Bertotti, B., Carr, B. J., and Rees, M. J., Mon. Not. R. Astr. Soc. 203, pp. 945-954 (1983).

[10] Thorne, K. S., "Gravitational Radiation," Chapter in 300 Years of Gravitation, edited by S. W. Hawkins and W. Israel, to be published by Cambridge University Press, Cambridge, MA.

[11] Allan, D. W., "National and International Time and Frequency Comparisons," 37th Annual Frequency Control Symposium, pp. 55-60, (1983).

[12] Wineland, D. J., "Trapped Ions, Laser Cooling, and Better Clocks," Science, Vol. 226, pp. 395-400, (26 October 1984).

INITIAL OPERATIONAL EXPERIENCE WITH A MERCURY ION STORAGE FREQUENCY STANDARD

Leonard S. Cutler and Robin P. Giffard
Hewlett-Packard Physics Laboratories
Palo Alto, California 94304
and
Paul J. Wheeler and Gernot M. R. Winkler
U.S. Naval Observatory
Washington, D.C. 20392-5100

Summary

A Mercury-199 trapped ion frequency standard built by Hewlett-Packard Laboratories has been in almost continuous operation at the U.S. Naval Observatory in Washington, D.C. since installation in July, 1986. Brief interruptions in otherwise satisfactory operation resulted from a few electronic malfunctions and from failures of the AC line supply to the instrument. It was necessary to replace the Mercury discharge lamp and the photomultiplier tube after about 6 months of operation. In this paper we describe an extended series of frequency comparisons between the Mercury standard and one of the Hydrogen Maser components of UTC(USNO). The results of the measurements have allowed us to observe the instrument's characteristics, and to assess the potential of such a trapped-ion standard in practical laboratory use.

Introduction

A trapped Mercury 199 Ion Frequency Standard completed at the Hewlett-Packard Physics Laboratories in Palo Alto was installed at the U.S. Naval Observatory (USNO) in Washington, D.C., in July 1986. It has been in near continuous operation at this location since that date. A second, improved unit was delivered to the USNO in May 1987.

These prototype units perform well and a long series of frequency measurements has been made which allows some preliminary conclusions about the performance potential of this type of device in practical laboratory use. This paper describes briefly the details of design and operation and gives an overview of the data collected up to now. The operational experience will be summarized and the potential of the device for timekeeping applications discussed.

Principles of Design and Operational Details

Trap System:

The principles of frequency standards using ion trapping are now well understood (1, 2). We will describe here only the particular features of the prototypes delivered to the USNO and used in the measurements being discussed.

T trap system for our prototypes was developed from the apparatus used for the exploratory experiments described elsewhere (3, 4). The vacuum system has been improved and all potentially ferromagnetic parts removed from the proximity of the trap. The use of a three-layer magnetic shield allows us to work at a field of 100 nT. Isotopic Mercury 199 is provided by continuous thermal dissociation of the isotopically enriched Oxide, and the correct pressure of Helium (1E-5 Torr) background gas is obtained from a pyrex leak. The trap is filled by generating ions with short pulses from an electron-gun. We have found that it is essential to maintain a very clean vacuum system. The temperatures of the trap structure and the innermost shield are regulated to about 0.01° C. The intrinsic hyperfine linewidth has been shown to be much smaller than 0.1 Hz.

Control of the Second Order Doppler Shift

It is well known that in a useful frequency standard the ion macromotion energy must be controlled. As described in a previous publication (5), viscous cooling by a low pressure of Helium gas was used in this work. The effectiveness of this technique was shown by analysis of the thermal Doppler sidebands of the hyperfine line (6).

Once the macromotion temperature has been sufficiently reduced, the ion behavior can be described in terms of a "Cold Ion Cloud" model in which the trapping pseudopotential is strongly modified by the ion space-charge. The characteristic temperature below which this approximation is useful is given for spherical potential wells by

$$T = 0.436\ N^{2/3}$$

where N is the number of ions in the cloud, and the single-ion macromotion frequency is 50kHz. Under the conditions used, N = 2.0x10^6, and T = 6000 K. Since measurements indicate that the macromotion temperature is about 560 K, the cold-cloud approximation is justified.

In the cold-cloud approximation it has been shown elsewhere (4) that the number of ions in the trap, and the single-ion macromotion frequency must be carefully regulated to ensure a given level of frequency stability. In the frequency standard discussed here, both of these parameters are measured on-line, and continuously controlled by feedback loops. For analytical convenience, the shape of the ion cloud is chosen to be as close to spherical as possible.

Principles of Operation

A simplified block diagram of the frequency standard is given in Figure 1. A pulsed mode operation is employed with a cycle time of 2.5s. At the start of the cycle, the ion cloud population is "topped-up" by pulsing the electron-gun, and the initial population distribution is attained by optical pumping. The microwave interrogation signal is then brought to a position on one side of the resonance line, and the pumping lamp is turned off to avoid light-shift. After 1.0s, the microwave frequency is moved off the line, and the pumping lamp is turned on. The photon counter is then activated for 0.5s, and the ion population and frequency measured. This cycle is then repeated on the other side of the resonance line. The frequency error is estimated from successive photon counts on alternating sides of the resonance line by a digital filter algorithm which rejects drift in the background light level.

The mode of operation used for the measurements to be described is known as the "Synthesizer-Null" mode. In this mode, the hyperfine frequency of the Mercury-199 ion resonance is measured in terms of the external reference source by digitally programming the millimeter-wave synthesizer to drive the average output signal from the atomic discriminator to zero. When the external reference is a Hydrogen Maser, the lowest

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

possible level of frequency noise is obtained because noise from the flywheel oscillator is not significant.

The processor handles the primary digital feedback loop, regulation of the trapping parameters, housekeeping tasks, and serial output of averaged data. System parameters may be interrogated or changed as required from the keyboard during normal operation. The output of the standard in Synthesizer Null consists of numerical data corresponding to frequency measurements averaged over contiguous blocks of 1280s.

Rationale for the choice of measurement set-up

The estimation of performance limits for a device which has been designed for superior long-term frequency stability is not a trivial affair. The maintenance of as benign an environment as possible for uninterrupted periods of many months is a well known practical problem. Moreover, it was our intent to use the new standard as much as possible in the present operational environment of the USNO master clock complex to obtain practical experience in long-term laboratory use of the new standard. For these reasons a mode of operation was selected which would allow long-term averaging, as well as the monitoring of critical system parameters (which requires the use of the "monitor mode" in which frequency measurements cannot be made). It was, therefore, decided to operate the unit continuously, but in two alternating modes.

From 1300 UT to 1200 UT on the following day we use the Synthesizer-Null mode described above. For this purpose the 5 MHz output from Hydrogen Maser Nr. 19 is used as phase reference for the quartz oscillator which is internal to the Hg+ device. This phase lock is one of the possible modes of operation. By operating the Hg+ standard in the "Synthesizer Null" mode we effectively measure the external frequency in terms of the Hg+ resonance. Commensurate with the stability of this external signal, averaging over long periods will produce meaningful performance statistics. Of greatest interest, of course, is the level of the "flicker floor" of the device. For the cycle time chosen (2.5s) the hyperfine resonance line has an effective width of 850 mHz. We average 512 of these 2.5s cycles and each of the 1280 second averages is reported to one of the USNO data acquisition systems for permanent recording.

By using the operational Maser as a reference instead of the internal quartz crystal in the conventional way, we gain certain advantages and incur one slight disadvantage; the disadvantage being we have to correct our Hg+ measurements for the daily synthesizer changes in the MASER, adjustments which keep the MASER approximately on extrapolated UTC(USNO). The advantages are that during any one day, an extremely stable source is available for averaging the Hg+ measurements. This allows a good statistical performance evaluation not only in respect to the MASER but also in respect to UTC(USNO). There is no need to take into account any contributions from the internal quartz crystal oscillator. At 1200 UT a MASER synthesizer step may occur, if necessary, and the unit is then operated for one hour in the monitor mode for the measurement of signal and background photon counts and for other parameters. No frequency measurement is, therefore, being made during the time when a synthesizer step may occur (and for a full hour after that as well).

Figure 2 shows the arrangement of the three major components involved in the majority of our measurements. Two VLG-11 Hydrogen MASERS are phaselocked one to the other and are kept to within 10ps rms as measured at 1200 MHz. This is done so that at any sign of trouble, or for adjustments and repairs, we can switch to the standby MASER without noticeable impact on the Master Clock timing. The controller, an HP9915B, reports all synthesizer settings to one of the data acquisition systems for recording. This allows a reduction of all readings to the free running MASERs. The lead MASER, presently H19, receives its synthesizer settings once per day (at 1200 UT) from the data acquisition system which computes the preliminary time scale on the basis of the 24 Cesium clock ensemble.

Measurements

Table 1 is an example for part of a sequence of the 21 minute measurements. The values given are in Hertz and represent the excess of the ion resonance line frequency from the nominal 40.507348GHz of the synthesizer. A good indication of the randomness of the data is the autocorrelation function which is computed for the first 16 delays. As can be seen in this typical example, the successive readings are random within the resolution of the measurements. The last digit of the readings represents units of 100 microHz at 40.5GHz or roughly 2.5 parts in ten to the fifteen (2.5E-15). The standard deviation of the 21 minute averages is within 35 and 70 E-15, strongly correlated with the signal-to-noise ratio. This changes over the period of several months as the mercury lamp ages. At this time the standard is being operated well beyond the normal life of the lamp (6 months) and the variance of the 21 minute measurements is still growing slowly. As long as the data are randomly distributed, averaging over a day should produce corresponding daily frequency measurements with standard deviations of between 4 and 8E-15. This seems indeed to be the case as we can see below.

DATE	HERTZ
.....	.....
.....	.....
6873.298	-3.2834
6873.312	-3.2838
6873.327	-3.2827
6873.342	-3.2852
6873.357	-3.2819
6873.372	-3.2825
6873.386	-3.2836
6873.401	-3.2817
6873.416	-3.2843
6873.431	-3.2877
6873.446	-3.2833
6873.460	-3.2836
6873.475	-3.2849
6873.490	-3.2855

MAX - MIN = 82, -32842 = MEDIAN, SKEWNESS = -.21
SIGMA = 19 -32843 = AVERAGE, 64 RUNS (units in 100 Hz)

AUTOCORRELATION(DEL) =

Delay (units 1280s)	
0	1.000
1	-.017
2	-.011
3	.077
4	-.117
5	-.027
6	-.091
7	-.114
8	-.084
9	-.165
10	.046
11	.022
12	.191
13	-.063
14	-.160

15	.050
16	-.097

Table 1
Part of the Daily Frequency Measurements

For averaging times greater than one day, our measurements have to rely upon the long-term performance of the two MASERs, the USNO preliminary and final time scales based on the 24 Cesium clock ensemble, and the BIH Circular D values for UTC(USNO) which is, on the average, very closely realized by Master Clock #2. An essential question is, therefore, to what degree the MASER drift stays constant. A good performance estimate can be obtained for the two VLG-11's from their difference. This is shown in Figure 3. For the same period, the MASER frequencies are plotted in terms of the Hg+ standard for H18 (Figure 4) and H19 (Figure 5). The assumption of linearity of the drift is justified, albeit only for an environmentally undisturbed period such as the one shown. Over the last 150 days the drift of H19, e.g., has stayed within 2.6 and 3.3E-15 as given in the column 4 fits of Table 3 below.

From the residuals of the linear fits, which remove this systematic drift, a "three-cornered hat" resolution of the individual variances can be obtained. Table 2 summarizes the results which represent estimates (that agree very well with otherwise obtained estimates) for integration times of one day. The systematic "absolute" drift of the device itself must be estimated in reference to the existing laboratory frequency standards. This can be done via the BIH with some caution because of the possibility of a small annual period in that scale. The measurements of the preliminary UTC(USNO) have turned out to be of little value during the period available because of excessive low frequency noise. This is explained below.

Hg+ - H18 Residuals 8.5E-15
H18 - H19 Residuals 4.3E-15
Hg+ - H19 Residuals 7.9E-15

This triad resolved gives for

Hg+ 7.6E-15
H18 ..... 3.8E-15
H19 ..... 2.1E-15

Total number of observations: 115 days
(MJD 46818 - 46933)

Table 2

The better performance of H19 can in part be explained by the fact that its vacuum pump was equipped with new Titanium plates in late 1986, whereas H18 is still operating on its original plates (it was delivered to the USNO in September 1983). The phase locking of H18 adds noise, too. But this contribution is only a few E-16 for a one day integration time. However, both units perform exceedingly well. This has been very fortunate because due to some rearranging of laboratory spaces in preparation for new equipment, most of the Cesiums had to be concentrated in only three clock vaults during the last 9 months. As a result of this rearrangement and various environmental problems the Cesium time-scale has suffered several noticeable rate changes, and has performed unusually poorly during the period under discussion. This is obvious from Table 3 which gives daily measurements of the Mercury standard with respect to the uncorrected frequency of MASER H19 (Column 4), the adjustments made to the frequency of H19 on the basis of the Cesium scale (Column 3), and the actual rate of the Mercury frequency standard with respect to the corrected Master Clock (Column 2).

Hg+ FREQUENCY SUMMARY:

DATE/MJD	Hg+-H19 ADJUSTD	SYNTH/H19	COL2-COL3 Hg+-FREE MASER #19	N	SIG	CHI	DIFF
DAYS	E-15	E-15	E-15		E-15		E-15
6895.000	1190141	1189935	206	63	56	1.0	-10
6896.000	1190125	1189935	190	52	55	1.0	-16
6897.000	1190134	1189935	199	63	48	.9	9
6898.000	1190134	1189935	199	63	48	1.0	-0
6899.000	1190113	1189928	185	63	51	1.1	-13
6900.000	1190097	1189921	176	63	56	.9	-9
6901.000	1190121	1189921	200	63	55	.9	24
6902.000	1190113	1189928	185	63	58	.9	-15
6903.000	1190112	1189935	177	63	44	1.1	-8
6904.000	1190106	1189942	164	63	54	1.1	-12
6905.000	1190113	1189949	164	63	63	.9	-0
6906.000	1190135	1189956	179	63	56	1.0	15
6907.000	1190130	1189963	167	63	58	1.0	-12
6908.000	1190140	1189970	170	63	66	.9	3
6909.000	1190136	1189977	159	63	63	1.0	-11
6910.000	1190154	1189984	170	63	48	1.2	11
6911.000	1190143	1189991	152	63	58	1.0	-18
6912.000	1190159	1189998	161	63	60	.9	10
6913.000	1190152	1189998	154	63	57	1.0	-8
6914.000	1190154	1189998	156	63	67	.8	3
6915.000	1190154	1190005	149	63	64	1.1	-7
6916.000	1190154	1190005	149	63	56	1.0	-0
6917.000	1190152	1189998	154	63	62	1.1	5
6918.000	1190152	1190005	147	63	66	1.2	-8
6919.000	1190144	1190012	132	63	57	1.1	-14
6920.000	1190156	1190019	137	63	69	1.1	5
6921.000	1190152	1190026	126	63	67	.9	-11
6922.000	1190148	1190026	122	63	74	.9	-4
6923.000	1190155	1190033	122	63	70	.8	0
6924.000	1190165	1190040	125	63	65	1.1	3
6925.000	1190153	1190033	120	63	57	1.1	-5
6926.000	1190141	1190026	115	63	71	1.0	-5
6927.000	1190146	1190033	113	63	69	1.0	-2
6928.000	1190117	1190026	91	63	53	1.2	-22
6929.000	1190132	1190033	99	52	66	1.0	7

LINEAR FIT OF COLUMN 4 = F Hg+ - F H19

REF.DATE MJD'	PRED.VALUE E-15	SLOPE E-15/DAY	RMS E-15	N	REJECTED	ALLANDIF E-15
6929.0	106.5	-2.83	7	35	0	7.48

LINEAR FIT OF COLUMN 3 = F H19 - F MC2

REF.DATE MJD'	PRED.VALUE E-15	SLOPE E-15/DAY	RMS E-15	N	REJECTED
6929.0	1190047.6	3.77	10	35	0

LINEAR FIT OF COLUMN 2 = F Hg+ - F MC2

REF.DATE MJD'	PRED.VALUE E-15	SLOPE E-15/DAY	RMS E-15	N	REJECTED
6929.0	1190154.1	.94	14	35	0

Table 3

It is interesting to note from the data in Table 3 that the rms residuals for the fit of the data in Column 4 are equal to the "Allan Difference" from the daily differences in the last column of the table. This suggests that after drift subtraction the data

over a period of 35 days are still randomly distributed, and can thus be usefully averaged over still longer periods. This is a most important result and suggests that Mercury ion standards may be valuable in long-term timekeeping applications. As further support for the absence of flicker noise in the data in Column 4, we note that the residuals for this 35-day period are only slightly smaller than those for the 115-day period given in Table 2.

Discussion

From the above data which represent typical examples, we can see that the measurements can be considered white if the small constant frequency drift of the MASERs is subtracted. In fact, as is to be expected, non-white characteristics occur only as a consequence of severe environmental problems which affect all of the standards used in the measurements. While considerable attention has been paid to protecting the standards from environmental disturbances, several severe failures of environmental control did occur during December 1986, January 1987, and later. The data presented above do not include these periods because they are not typical and were largely a consequence of the installation of new power systems and other equipment. However, one of the more important lessons of the initial operation is indeed the extent of the environmental sensitivities of the Hg^+ device which were initially larger than expected. On the basis of the initial measurements, we changed the shield temperature from 50^0 C to 40^0 C in November 1986. This caused a positive frequency shift but more importantly, it decreased a small frequency drift which had been observed. After about a month, operation became very stable without further interruptions and no further frequency drifts could be detected. If that should continue to be our experience with the Hg^+ frequency standard in its present design, even without further improvements, the impact of the new device on long-term laboratory timekeeping will be considerable. We can obtain estimates of this by integrating our frequency measurements as shown in Figure 6 for the MASER difference and in Figure 7 for MI9 in terms of Hg^+. A random scatter in the daily frequency measurement of 7.6E-15 produces, of course, a random walk of 660ps from day-to-day which is larger than the free-running MASERs (180ps for MI9). However, given the randomness, we can average over several days and get a check on the MASER drift without degrading its day-to-day performance. At present performance levels, runs of 14 days should accomplish this.

A check on the long-term frequency drift via UTC(BIH) as reported in Circular D is shown in Figure 8. This compares the integrated readings of MC#2 in terms of the Hg^+ frequency standard with the BIH values for the same clock. If we divide the period into two parts of 70 days each, then a long-term drift of the Hg^+ device would have to be smaller than about 3E-17 per day to escape detection. However, due to the possibility of a small seasonal term in the BIH scale, we must consider this as very preliminary.

A second unit was recently delivered. No problems were encountered in assembling that unit at the USNO and it, too, is operating normally. Preliminary measurements indicate that the difference in frequency between the two standards is between one and two parts in ten to the thirteen, with comparable noise level. In view of the fact that the construction of the trap in this unit is significantly different, and that neither standard has been calibrated since installation this performance is considered satisfactory.

A few interruptions occurred in our evaluation of the first prototype. These were due to initial problems with the "uninterruptible" power supplies used to drive the instrument. In addition, several electronics problems caused interruptions of a few days. One mercury discharge lamp had to be replaced after about six months of use and the photomultiplier tube was replaced for an intermittent problem. The second lamp has now been operating for 7 months. Figure 9 shows the decay of the photon count over the last 53 days. The last day shown corresponds with day 200 since installation of the new lamp. The decay of the lamp output is gradual and a replacement can conveniently be planned in advance. This decay of photon count can be attributed to three things: (1) a decrease in lamp output, (2) a decrease in PMT sensitivity, and (3) a detuning of the RF. We believe that reason 1 is the predominant one but the exact cause will be determined after the performance of the frequency standard deteriorates to an unacceptable level.

Conclusions

Overall, the experience with the prototypes has been astonishingly good for a device based on entirely new principles. As a result of our operational experience we have discovered the need for some small design changes which should take care of the aforementioned minor problems. In summary, we can state that the Hg^+ frequency standard, even in its prototype configuration, has become a superior instrument for advanced laboratory use. It will establish an entirely new performance standard in long-term timekeeping.

Acknowledgments

It is with great pleasure that we acknowledge the support of Dr. Joseph White, Naval Research Laboratory, and Mr. Jimmie B. Collie, Space and Naval Warfare Systems Command.

References

1. D. J. Wineland, Wayne M. Itano, J. C. Bergquist, J. J. Bollinger and J. D. Prestage in Atomic Physics 9, R. S. Vandyke, Jr., and E. N. Fortson, eds., (World Scientific, Singapore, 1984), p. 3.

2. G. Werth in Atomic Physics 9, R. S. Vandyke, Jr., and E. N. Fortson, eds., (World Scientific, Singapore, 1984), p. 28.

3. L. S. Cutler, R. P. Giffard, and M. D. McGuire (1981) "A Trapped Mercury 199 Ion Frequency Standard" Proc. 13th. PTTI Applications and Planning Meeting, pp. 563-573. (NASA Conference Publication 2220)

4. L. S. Cutler, R. P. Giffard, and M. D. McGuire, FCS 37 1983. (IEEE Cat. No. 83CH1957-0)

5. L. S. Cutler, R. P. Giffard, and M. D. McGuire, Appl. Phys. B 36 (1985)

6. L. S. Cutler, C. A. Flory, R. P. Giffard, and M. D. McGuire, Appl. Phys. B 39 (1985)

$^{199}Hg^{+}$ ATOMIC FREQUENCY STANDARD

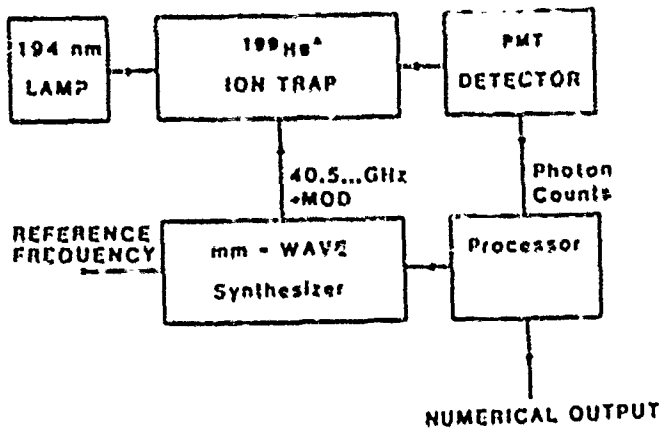
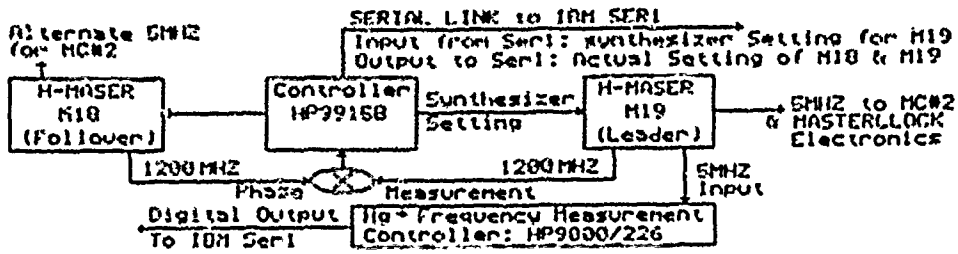


FIGURE 1



MASTERCLOCK FREQUENCY MEASUREMENT SCHEME

Figure 2

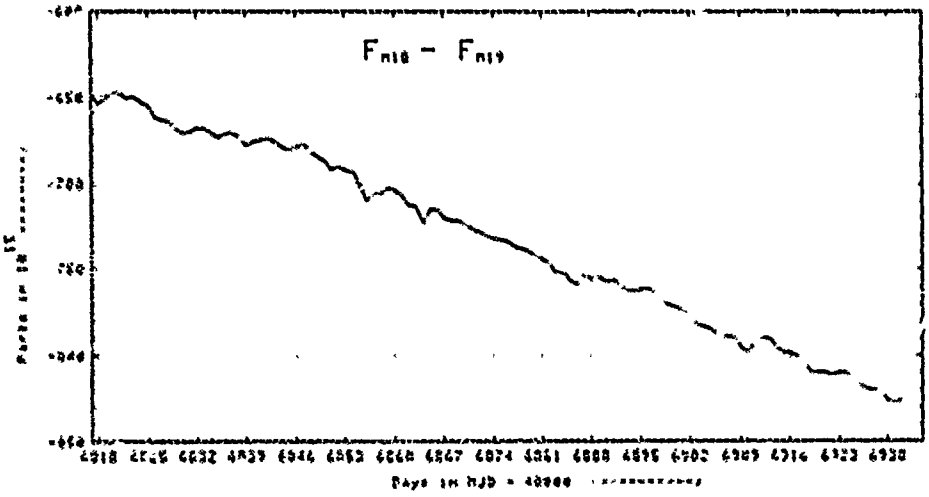


FIGURE 3

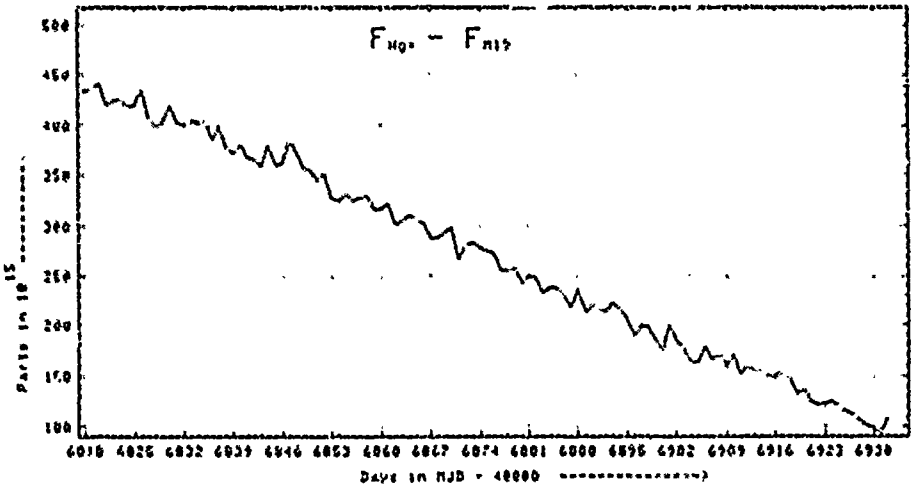


FIGURE 4

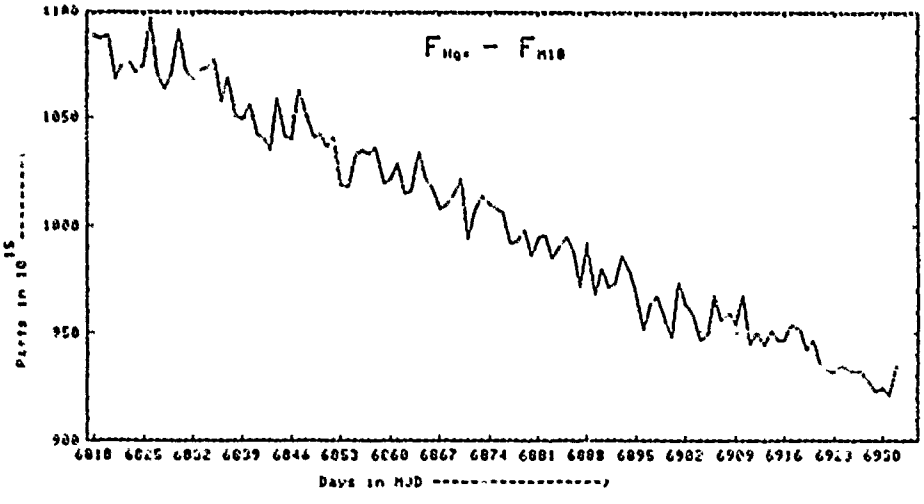


FIGURE 5

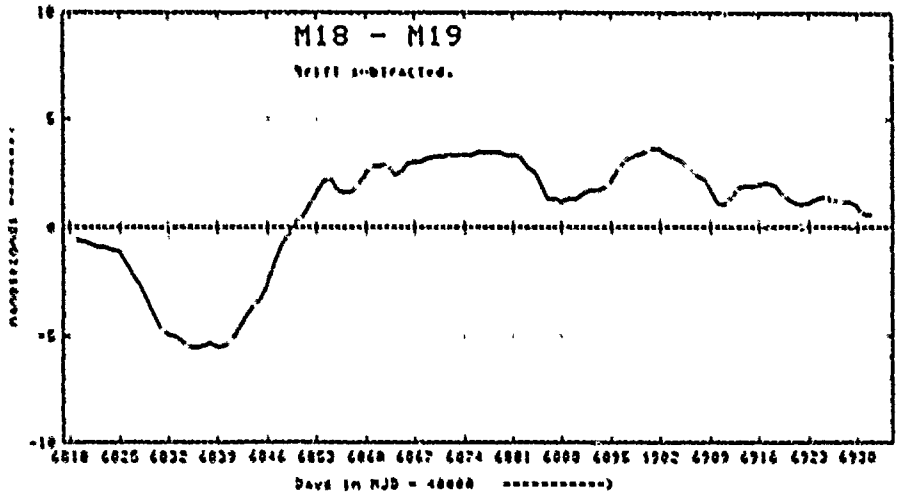


FIGURE 6

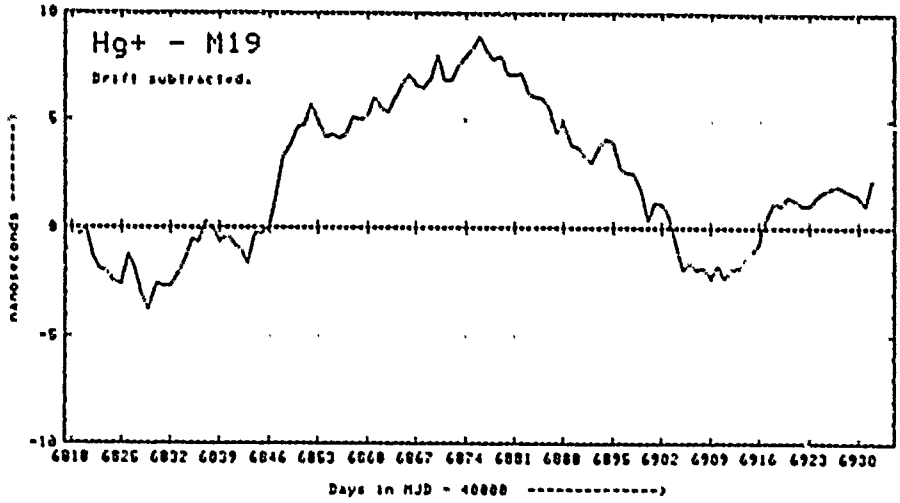


FIGURE 7

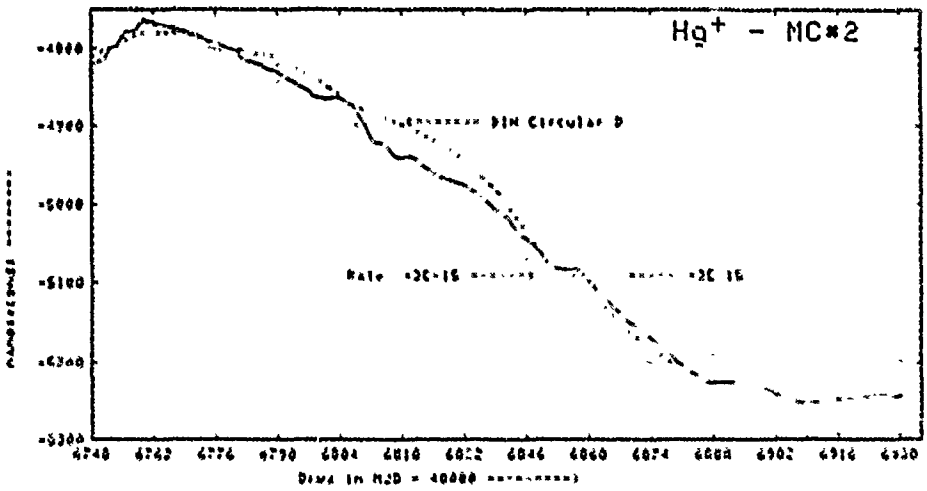


FIGURE 8

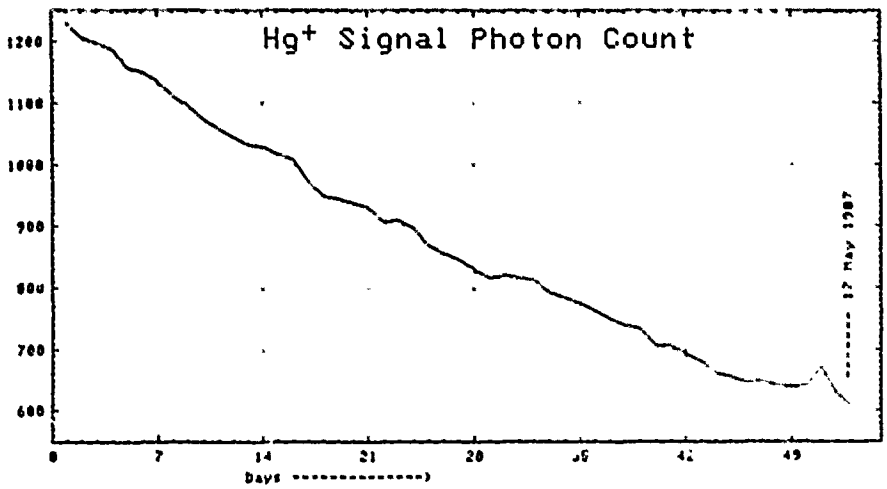


FIGURE 9

JPL Trapped Ion Frequency Standard Development

J. D. Prestage, G. J. Dick, L. Maleki
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Drive
Pasadena, Ca. 91109
(818) 354-4515

SUMMARY

We report the first closed-loop operation of the JPL trapped ion frequency standard. Mercury 199 ions are confined in an RF trap and state-selected by use of optical pumping with 194 nm UV light from a ^{202}Hg discharge lamp. Absorption of microwave radiation at the hyperfine frequency (40.5 GHz) is signaled by atomic fluorescence of the UV light. The frequency of a 40.5 GHz oscillator is locked to a 1.6 Hz wide atomic absorption line of the trapped ions. The measured Allan variance of this locked oscillator is currently $\sigma_y(\tau) = 4.4 \times 10^{-13}/\sqrt{\tau}$ for $20 < \tau < 120$ seconds, which is better stability than the best commercial cesium standards by almost a factor of 2. This first result was achieved without magnetic shielding and without regulation of ion number. This article describes details of this first closed-loop operation.

INTRODUCTION

Recently, there has been much activity toward the development of trapped ion frequency standards because ions confined in an RF quadrupole ion trap are subjected to very small perturbations of their atomic energy levels and weak forces which equalize any population differences among the ground state hyperfine levels. Potentially, the largest source of frequency fluctuation for such a standard stems from the motion of the atoms within the trap via the second order doppler or relativistic time dilation effect. To minimize this perturbation, heavy ions are preferable to light ions since for a given energy a heavy ion will have smaller velocity. For this and other reasons discussed later in this paper, $^{199}Hg^+$ ions have been used for most trapped ion frequency standard work.

There are only two parameters needed to describe the short-term stability of an atomic frequency standard. One is the line Q ($=f/\Delta f$) where f is the resonant frequency of the reference atom and Δf is the width of the atomic resonance. For $^{199}Hg^+$, f is 40.5 GHz and Δf is as small as 0.1Hz, thus trapped ion standards have line Q's which are orders of magnitude higher than other microwave atomic frequency standards. The other parameter that determines stability is the signal-to-noise ratio (SNR) achieved in measuring the atomic resonance. The short-term stability is inversely proportional to the product of Q and SNR.

At present, four groups worldwide are developing trapped Hg^+ ion clocks: Hewlett-Packard, Palo Alto, Ca.;[1] National Bureau of Standards, Boulder, Co.;[2] Universite Paris-Sud, Orsay, France;[3] and JPL/NASA. It should be noted that F.G. Major first proposed the use of trapped $^{199}Hg^+$ ions as a frequency standard while working for NASA at the Goddard Space Flight Center in 1969.[4]

ION TRAPPING

Ions can be confined to a region of space without walls by a combination of electric and magnetic fields appropriate to its mass and charge. The fields may be static electric and magnetic or ac electric. The former is used in dc or penning traps while the latter is used in rf or Paul traps.

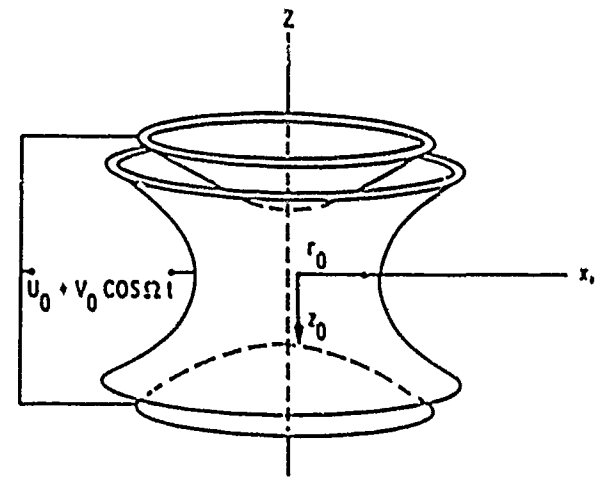


Figure 1

Electrode structure and voltages used in this work

The electrode structure and time-varying voltages for the rf trap used in this work is shown in Figure 1. The electric potential inside the trap when no ions are present is

$$V_T = (U_0 + V_0 \cos \Omega t)(r^2 - 2z^2)/\xi^2$$

where, for the present work, $\Omega = (2\pi)$ 500kHz, $U_0 = 20V$, $V_0 = 660V$, $\xi^2 = (r_0^2 + 2z_0^2)$, and r_0 = inside radius of ring electrode = 1.9 cm = $\sqrt{2}z_0$.

A charged particle moving in the inhomogeneous oscillatory electric field of the ion trap feels a net force (averaged over one cycle of Ω) toward the region of weaker field provided the amplitude of its motion at frequency Ω is small compared with its distance from the center of the trap. The motion under these conditions is a combination of a fast oscillation at frequency Ω (micromotion) and a slower frequency ω, as shown in Figure 2. The action of the RF field in trapping ions is described by the electric pseudopotential energy.[5]

$$\Phi = e^2 V_0^2 (r^2 + 4z^2)/m\Omega^2\xi^4$$

where m and e are the ion's mass and charge, respectively. The DC potential energy,

$$\phi_{dc} = eU_0(r^2 - 2z^2)/\xi^2$$

is added to the pseudopotential giving the total potential energy for an ion in the trap,

$$\phi_T = m\omega_r^2 r^2/2 + m\omega_z^2 z^2/2$$

where, $\omega_r^2 = 2e^2V_0^2/m^2\Omega^2\xi^4 + 2eU_0/m\xi^2$,

and $\omega_z^2 = 8e^2V_0^2/m^2\Omega^2\xi^4 - 4eU_0/m\xi^2$.

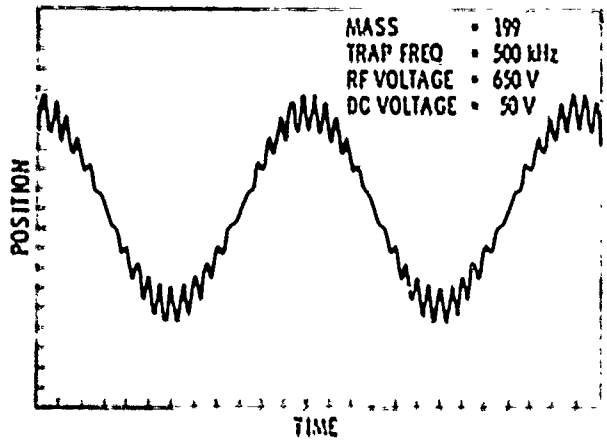


Figure 2

One dimensional motion for ion in combined RF and DC trapping fields

Under the conditions listed earlier about 19 electron volts of kinetic energy is required for an $^{199}Hg^+$ ion at the trap center to reach one of the trapping electrodes.

Our trap is inside a vacuum chamber at a pressure of 1.2×10^{-6} torr. By heating a powder of isotopically enriched mercuric oxide (HgO) to about 100 C, a vapor of neutral ^{199}Hg fills the vacuum chamber to about 10^{-6} torr partial pressure. Electrons from a LaB_6 single crystal filament are injected into the trap (≈ 30μA, 200V) ionizing the neutral ^{199}Hg.

The resulting ion cloud is much hotter than the room temperature neutral vapor in part because ionization of the vapor takes place throughout the trap. It has been found experimentally that the average kinetic energy of the ions is about 10% of the well depth. The resulting 2eV of kinetic energy would produce a fractional second order doppler shift of about 10^{-11}. To reduce this shift the vacuum system is filled to about 10^{-6} torr of ^4He. The Hg^+ ions collide with these room temperature He atoms and are cooled to just above room temperature.

IONIC LEVELS AND QUANTUM ENGINEERING

The magnetic structure of the ground state hyperfine levels of $^{199}Hg^+$ is shown in Fig. 3. The energy difference between the ($F = 0$, $m_F = 0$) and ($F = 1$, $m_F = 0$) levels is used to define the standard frequency, approximately $f_{Hg^+} = 40.507347997$ GHz. The measured frequency, f, depends quadratically on the magnetic field at the position of the ion cloud,

$$f = f_{Hg^+} + 98B^2(Hz/G^2).$$

For comparison, the field dependence for Hydrogen atoms is

$$f = f_H + 2750B^2(Hz/G^2).$$

The ions are state selected by use of optical pumping with light from an ^{202}Hg discharge lamp. The energy levels of $^{202}Hg^+$ and $^{199}Hg^+$ are compared in Fig. 4. Ultraviolet light of wavelength 194.2 nm (≈ 6.4 eV) from the ^{202}Hg lamp, when collected and focused onto the $^{199}Hg^+$ ions will excite the transition $^2S_{1/2}(F = 1, m_F) \rightarrow {}^2P_{1/2}$.

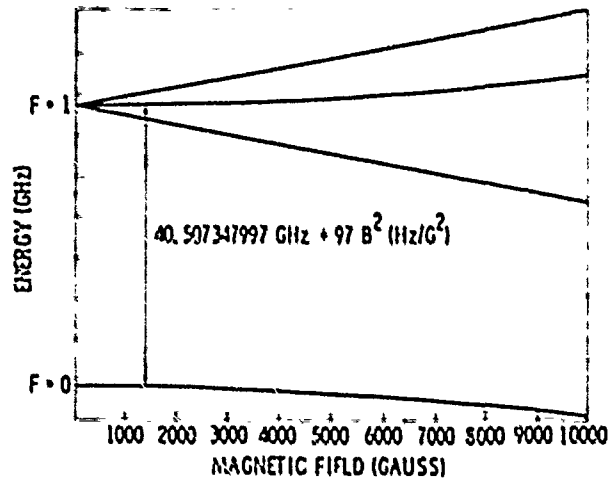


Figure 3

Energy levels of the ground state of $^{199}Hg^+$ versus magnetic field.
The first order field independent transition used in this work is shown.

The $^2P_{1/2}$ state decays after 2 ns lifetime to either $^2S_{1/2}(F = 0, m_F = 0)$, or $^2S_{1/2}(F = 1, m_F)$ thereby scattering a 194 nm photon. Since the transition $^2S_{1/2}(F = 0, m_F = 0) \rightarrow {}^2P_{1/2}$ is not resonant with the light from the ^{202}Hg lamp, the ions are pumped out of the $^2S_{1/2}(F = 1, m_F)$ state into the $^2S_{1/2}(F = 0, m_F = 0)$ state at which time the ions stop scattering UV light. A flux of about 3×10^{12} photons per second per cm^2 passing through the ion cloud will depopulate the $^2S_{1/2}(F = 1, m_F)$ levels in about 1/2 second. In this "prepared" state, an interrogating microwave field (strength 10^{-6} G) at frequency 40.507347997 GHz will transfer the atoms from the ($F = 0$, $m_F = 0$) state to the $^2S_{1/2}(F = 1, m_F = 0)$ state in about 1 second. The ions will then scatter UV light until they are pumped back into the non-fluoresecing $^2S_{1/2}(F = 0, m_F = 0)$ state. The strength of this scattered light is used to determine whether the interrogating microwave signal was correctly tuned to the atomic resonance.

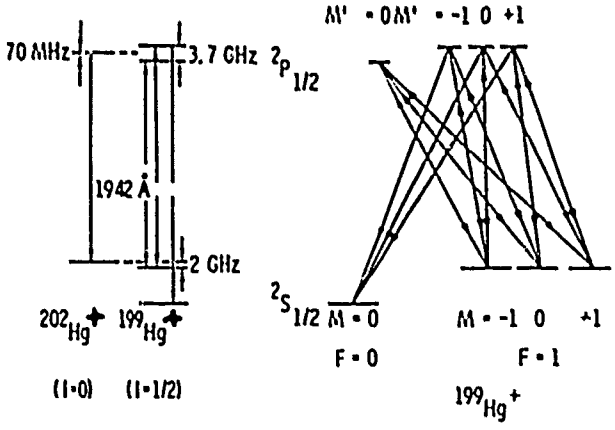


Figure 4

Ground and lowest optically excited states of $^{202}Hg^+$ and $^{199}Hg^+$ are compared showing how light from $^{202}Hg^+$ will pump $^{199}Hg^+$ ions into the $^2S_{1/2}(F = 0, M = 0)$ Level.

UV OPTICAL SYSTEM

The design of the optical system is crucial to the signal-to-noise ratio of the atomic resonance signal in a trapped ion standard. The detected UV fluorescent radiation from the ions, produced after absorbing microwave photons near 40.5 GHz comprises the signal. The efficiency of collection of the light scattered by the ions further determines the signal level. The noise, on the other hand, is produced by a number of sources. These include the optical pumping light reflected from electrodes and windows, radiation scattered by the background neutral mercury atoms, and light due to the hot electron filament.

Figure 5 shows the optical system used to collect and focus UV light from the ^{202}Hg lamp onto the ion cloud. The system consists of an rf excited lamp which produces the UV light, and an ellipsoidal mirror which

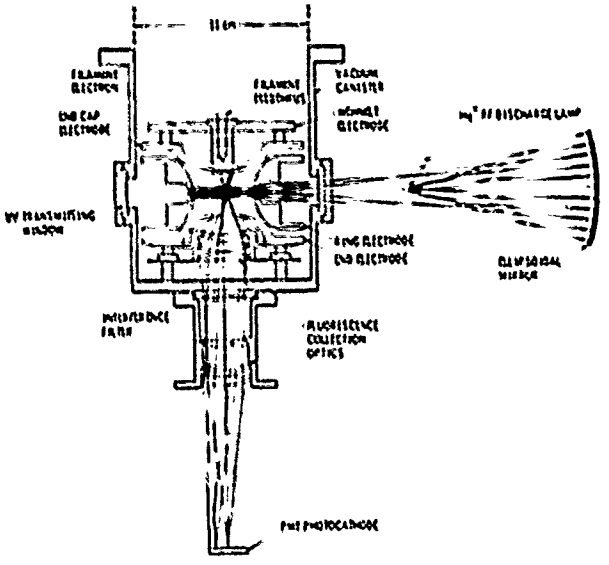


Figure 5

194 nm UV light is scattered by the trapped $^{199}Hg^{+}$ ions after absorption of microwave radiation near 40.5 GHz.

collects and focuses the light into the trap. This single element optical system ensures that the maximum amount of light is collected, while the scattering from the surfaces is nearly eliminated.

The lamp is excited with 15-20 watts of rf power (160 MHz) creating a very bright discharge in the quartz cell containing the ^{202}Hg vapor with about 30 millitorr of argon buffer gas. The useful light from the ^{202}Hg lamp is from the 194.2 nm transition in $^{202}Hg^{+}$ shown in Figure 4.

However, the brightest wavelengths produced in the lamp are from transitions in the neutral Hg atom. Any light detected at wavelengths other than 194.2 nm will degrade the SNR of the measured atomic resonance. There are three ways we suppress the detection of light with wavelength different from 194 nm:

1) The photomultiplier tube (PMT), used to detect UV aromic fluorescence is only sensitive to light that h wavelength between 160 and 320 nm, with peak sensitivity of 12-15% at 210nm. The brightest line coming from the lamp in this bandwidth is the 254 nm line of neutral Hg and is 200 times brighter than the 194 nm line.

2) The ellipsoidal collection mirror has a thin film dielectric coating which maximizes reflection at 194 nm , while keeping the reflectivity at 254 nm at 50% with lower reflectivities for longer wavelengths. The entrance window to the trap region is coated to be 98% reflective at 254 nm while being 90% transmitting at 194 nm.

3) All light collected by the detection optics --stray scattered light plus fluorescent light from the atoms--is filtered with a 194 nm bandpass filter with peak transmission 30-40% and bandwidth 45 nm. The five element lens system used in this first operation is designed to form an image of the ion cloud (assumed to be a sphere of 6 mm diameter) on the PMT. This system optimizes the collection of light emitted by the ions within the constraints of the existing dimensions of the endcap electrode.

The solid angle subtended by the collection optics around the trap center is 5% of the total 4π solid angle. The total efficiency of this detection system is equal to the solid angle (0.05) times the bandpass filter loss (.35) times the PMT sensitivity (0.12) or, 2×10^{-3}.

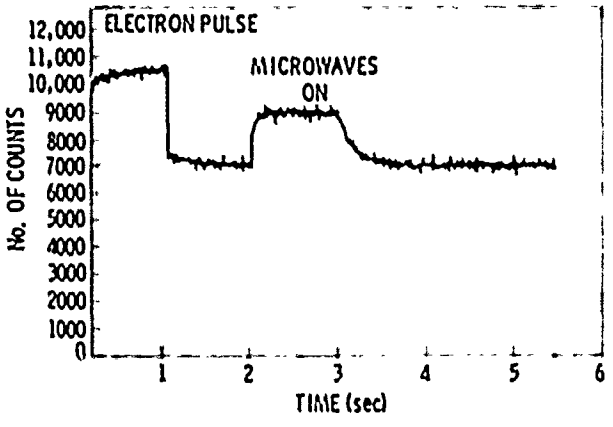


Figure 6

Light scattered from the ions when microwaves (40.5 GHz) are applied is detected together with stray light from electrodes and windows.

The atomic fluorescence and stray scattered light can be seen in Figure 6, showing UV light detected by the PMT versus time as counts in a multichannel scaler. The scan is triggered by the start of the electron pulse, which forms the ions and causes some increase in detected light. After about 1 second the electron pulse is switched off, the ions are pumped into the $^2S_{1/2}(F = 0, m_F = 0)$ level and only stray light is collected. At 2 seconds, the 40.5 GHz radiation is switched on, transferring some of the atoms into the $^2S_{1/2}(F = 1, m_F = 0)$ state where they scatter light as discussed earlier. Finally, at about 3 seconds, the microwaves are switched off and the atomic fluorescence dies away with time constant of about 1/5 second.

CLOSED LOOP OPERATION

The sequence of operations used to carry out a measurement of the $^2S_{1/2}(F = 0, m_F = 0) \rightarrow {}^2S_{1/2}(F = 1, m_F = 0)$ frequency are shown in Figure 7. By repeating this sequence as the frequency of the microwaves is

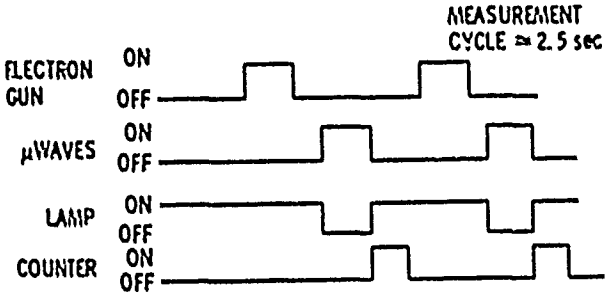


Figure 7

The sequence of operations used to determine whether the 40.5 GHz is correctly tuned to the atomic resonance. Two cycles are shown.

stepped in 0.2 Hz increments, we measure the resonance curve shown in Figure 8. For the 0.5 second square microwave pulse used in this measurement, the smallest linewidth theoretically attainable is 1.6 Hz, which corresponds to $Q \sim 2.5 \times 10^{10}$.

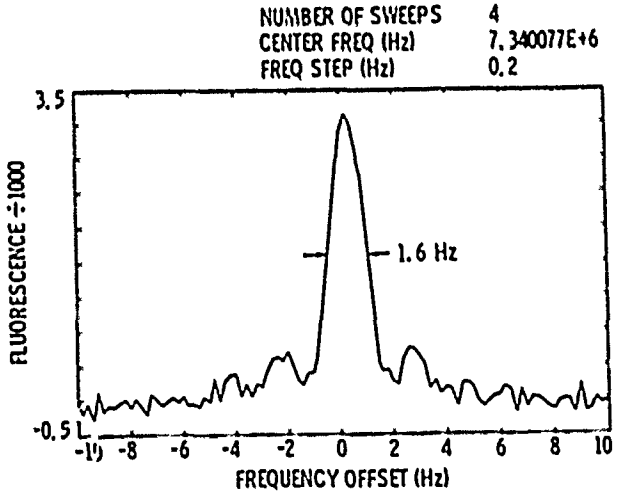


Figure 8

Atomic fluorescence as the frequency of the 40.5 GHz oscillator is swept through the atomic resonance.

The oscillator is locked to the 1.6 Hz wide resonance line by stepping the oscillator 0.8 Hz to either side of the resonance and adjusting the center frequency of this ± 0.8 Hz step to null the difference in fluorescence rates. More precisely, suppose the center frequency of the 40.5 GHz oscillator, F_i, is within one linewidth of the resonance, three measurements of the fluorescence are made on alternate sides of F_i:

C1 at F_i + 0.8 Hz

C2 at F_i - 0.8 Hz

C3 at F_i + 0.8 Hz

The oscillator center frequency is then changed to $F_{i+1} = F_i + (0.8\text{Hz}/T)(C1 + C3 - 2C2)/[2(\text{SIGNAL})]$ where SIGNAL is the height of the fluorescence above background and T is the loop time constant in units of measurement cycle time. The "second difference" [C1 + C3 - 2(C2)] is used because it is insensitive to linear drifts in lamp intensity. If the first difference (C1-C2) had been used to change F_i, a linear drift in lamp intensity would force a linear drift in the 40.5 GHz oscillator's frequency, away from the atomic line center.

The sequence of center frequencies obtained in this closed loop operation is shown in Figure 9. The measurement shown lasted just over 2 hours. A single measurement cycle--electron pulse to load ions in the trap, a wait period while ions are optically pumped, lamp switched off while microwaves drive the transition, and lamp and counter switched on to monitor fluorescence--lasts about 2.5 seconds. The loop time constant T is 5 measurement cycles, that is, about 12.5 seconds.

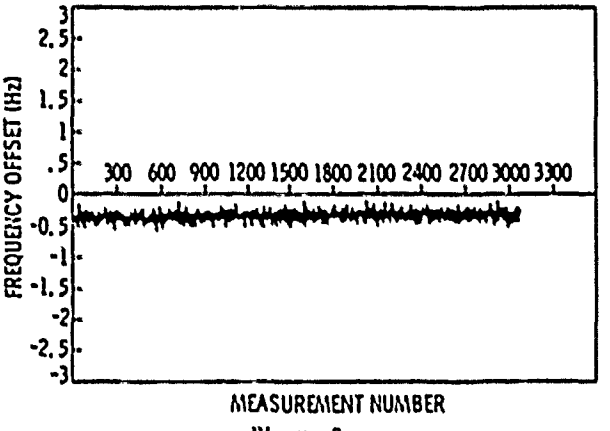


Figure 9

Frequency deviation of the 40.5 GHz oscillator while servoed to the 1.6 Hz wide atomic resonance. Each measurement lasts 2.5 seconds.

The Allan variance derived from the sequence of frequencies (F_i) is shown in Figure 10. The reference frequency for the 40.5 GHz oscillator is provided by a Hydrogen Maser (SAO-21 in the Frequency Standards Laboratory). For times long compared to the loop time constant the Allan variance falls as $4.4 \times 10^{-12}/\sqrt{\tau}$. This short-term stability is nearly a factor of 2 better than that of the best commercial Cesium standards ($8.5 \times 10^{-12}/\sqrt{\tau}$).
These first results were done without shielding the 0.8 G ambient magnetic field in the trap. The residual field dependence at 0.8 G is 157 Hz/G. Frequency stability of 2×10^{-13} at this field sensitivity requires magnetic field fluctuations smaller than 60 μG over the 320 seconds required to reduce statistical error to 2×10^{-13}. For this reason we have not pushed the Allan variance data beyond 320 seconds in this first test.

CONCLUSION

In its first closed loop operation, the frequency stability of the trapped $^{199}Hg^+$ frequency standard has been measured to be $\sigma_y(\tau) = 4.4 \times 10^{-12}/\sqrt{\tau}$ for $20 < \tau < 320$ seconds.

Many improvements are underway to increase the short- and long- term stability of this standard including increased collection efficiency, shielding the ambient magnetic field, and designs for traps which could store up to a hundred times the present ion number.

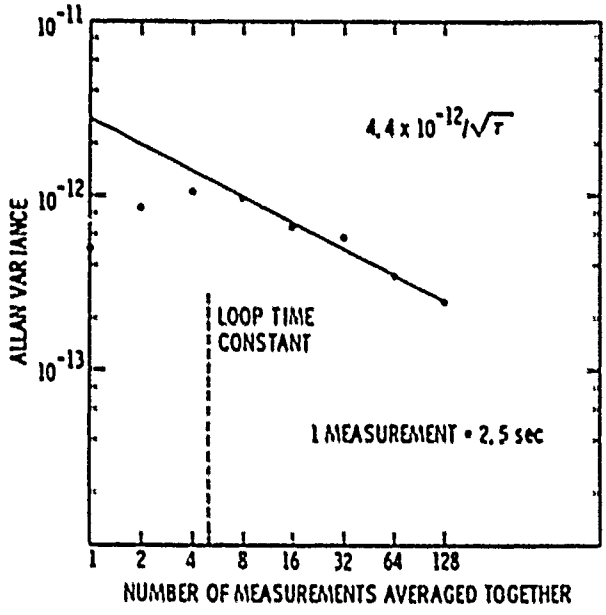


Figure 10

Allan variance of the sequence of center frequencies shown in Figure 9. Multiply by 2.5 seconds to convert the horizontal axis to time.

Acknowledgements

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

We wish to thank T. Tucker for Hg lamp construction and assisting in the development of the 40.5 Ghz oscillator.

REFERENCES

1. Cutler, L. S., Giffard, R. P., Wheeler, P. J., and Winkler, G. M. R., "Initial Operational Experience with a Mercury Ion Storage Frequency Standard", Proc. 41st Ann. Sym. on Freq. Control, Philadelphia, Pa., May, 1987.

2. Wineland, D. J. , Itano, W. M., Bergquist, J. C., and Walls, F. L., "Proposed Stored $^{201}Hg^{+}$Ion Frequency Standards, " Proc. 35th Ann. Sym. on Freq. Control, Philadelphia, Pa., p. 602.

3. Jardino, M., DesaintFuscien, M., and Plumelle, F., "Prospects for a Mercury Ion Frequency Standard," Journal de Physique, C8 (1981), p.327.

4. Major, F. G., "Remarks on the Application of RF Spectroscopy of Stored Ions to Frequency Standards," FAA Util. of Time/Freq. in Collision Avoidance Systems, N69-19856 09-14 (1968), p. 125-126.

5. Dehmelt, H. G., "Radio Frequency Spectroscopy of Stored Ions," Adv. At. Mol. Phys. 3, (1967), p. 53-154.

41st Annual Frequency Control Symposium - 1987

ULTRA-SENSITIVE FREQUENCY DISCRIMINATION In A DIODE LASER PUMPED ^{87}Rb ATOMIC CLOCK

M. Hashimoto, M. Ohtsu*, and H. Furuta

Department of Imformation Processing
Graduate School at Nagatsuta
4259 Nagatsuta, Midori-ku, Yokohama, Kanagawa 227, Japan

*Presently, AT&T Bell Laboratories
Room HOH R237, Holmdel, New Jersey 07733, USA

ABSTRACT

We propose a novel FM sideband technique to reduce drastically the double resonance linewidth in a Rb atomic clock pumped by a stabilized diode laser. This could be achived by the use of the nonlinear susceptibility of three-level Rb atoms when the incident microwave signal is modulated. By using resultant FM sidebands of the laser and synchronous detection, high-order derivative shapes of absorptive and disperaive parts of double-resonance spectral line were measured. The linewidth as narrow as 20 Hz was obtained, which was 1/10 times that of the rf-excited ^{87}Rb atomic clock. Optimum values of modulation parameters were found by the computer simulation. Futhermore, contributions of the laser frequency detuning to the light shifts were also estimated to attain the microwave frequency accuracy as high as 10^{-12}.

I. Introduction

Highly stable microwave frequency oscillators have been required for many applications such as satellite communications , satellite tracking, global positioning system (GPS), and so on. Portable Cesium (^{133}Cs) atomic clocks and Rubidium (^{87}Rb) atomic clocks have been used for these applications, and improvements of their performances are indispensable for the progresses of these applications.

For these improvements , the optical pumping and the optical detection by diode lasers have been proposed for ^{133}Cs atomic clocks[1]. In the case of the optical pumping for ^{87}Rb atomic clocks, the replacement of a rf - excited ^{87}Rb lamp by a diode laser has been proposed , and preliminary experiments on laser pumped ^{87}Rb atomic clocks have been reported[2]. The authors already showed that a quantitative evaluation of light shift and a novel double resonance lineshape with a drastically narrower linewidth can be obtained by utilizing an advantageous usage of a high temporal coherence of the laser[3]. This can be achived by the use of FM laser sidebands induced by the nonlinear susceptibility of three-level ^{87}Rb atoms and the microwave frequency modulation. However, to the author's knowledge, this effect has not been known yet and is employed as an advanced technique for improved ^{87}Rb atomic clocks. Therefore, extremely high stability of microwave frequency can be obtained by using this technique at optimal conditions.

In this paper, a theoretical analysis for this technique and an evaluation of microwave frequency stability for a diode laser pumped ^{87}Rb atomic clock are carried out. In section II, the observation of some novel double-resonance lineshapes with narrower linewidth are demonstrated. In section III, the theoretical analysis for double-resonance by means of the equations of motion of density matrix elements are given. In section IV, optimal conditions for the highest microwave frequency stability are found by the computer simulation. Finally, in section V, both contributions of the laser frequency detuning to the light shifts and effects of laser linewidth on microwave frequency stability are investigated.

II. Observation of Double Resonance Spectral Shapes with Narrower Linewidth

Figure 1 shows an experimental setup for a laser pumped ^{87}Rb atomic clock, in which a commercially available ^{87}Rb atomic clock (Fujitsu, 5407A) was used. The rubidium gas cell with natural rubidium vapor (^{87}Rb/^{85}Rb = 3/7) and buffer gases was installed in the microwave cavity. Figure 2 (a) and (b) show examples of novel optical-microwave double resonance spectrum. The vertical axes of these figures are the output voltage of the phase sensitive detector (P.S.D.) in Fig.1. At first glance, the lineshapes of Fig.2 (a) and (b) seem to be quite different. These

CH2427-3/87/0000-025 $1.00©1987 IEEE

difference, however, only stems from the difference of the modulation index (M), the modulation frequency ($\omega_m/2\pi$), and the phase defference (θ) between the output signal from the post-detector amplifier and the reference signal for P.S.D. Other conditions, for instance, the laser power density and the rubidium gas cell temperature, were equal.

It is worth to mention that the peak-to-peak linewidths at the center of these figures, i.e., the separation between points A and B in Fig.2 (a) and (b), are much narrower than those of rf-excited ^{87}Rb atomic clocks. In the case of Fig.2 (a) and (b), the linewidths were 75 Hz and 110 Hz, respectively. In contrast with them, the linewidth of rf-excited ^{87}Rb atomic clock was 360 Hz as shown in Fig.3. Furthermore, the slope at the center of these lineshapes were too steep, and depended upon the modulation parameters such as M , $\omega_m/2\pi$, and θ . The slope is related to the microwave frequency stability given by[4]

$$\sigma_y(\tau) = \frac{0.2}{Q(S/N)}\,\tau^{-1/2} , \qquad (1)$$

where $\sigma_y(\tau)$ is the square root of the Allan variance of microwave frequency fluctuations[5], τ is the integration time, S/N is the signal-to-noise ratio of the signal from post-detector amplifier, and Q is the quality factor of the frequency discriminator which is given by

$$Q = \nu_{Rb}/\Delta\nu \quad . \qquad (2)$$

In eq.(2), ν_{Rb} is the microwave transition frequency , and $\Delta\nu$ is the linewidth of the derivative shape which is defined as the separation between the points A and B of Fig.2. Eq.(1) also can be expressed by the use of $S_1(=S/\Delta\omega)$, as

$$\sigma_y(\tau) = \frac{0.2\,N}{\nu_{Rb}}\,\frac{1}{S_1}\,\tau^{-1/2} . \qquad (3)$$

Since microwave frequency stability $\sigma_y(\tau)$ is inversely proportional to S_1, large S_1 value improves the microwave stability. Thus, to find optimal conditions for larger slope is useful to improve the microwave frequency stability for a laser pumped ^{87}Rb atomic clock.

III. Analysis of Optical-Microwave Double Resonance

The three - level model of an atom is employed to analyze the optical - microwave double resonance[6]. As showed in Fig.4, two hyperfine levels of the ground state of $(5S_{1/2}, F=1, m_F=0)$ and $(5S_{1/2}, F=2, m_F=0)$, and an excited state $(5P_{3/2})$ correspond to the relevant three levels, which are expressed as $|a\rangle$, $|b\rangle$, and $|c\rangle$, respectively. The magnetic field of microwave and the electric field of a laser are represented as

$$H_M(r,z,t) = (H_{M0}J_0(3.832r/R)\sin(k_M z)/2)\cdot \exp(-i\omega_M t) + C.C. , \qquad (4)$$

$$E_L(z,t) = (E_L/2)\exp(-i(\omega_L t - k_L z)) + C.C. , \qquad (5)$$

where H_M and E_L are the amplitudes of the magnetic and the electric field, ω_M and ω_L are angular frequencies, k_M and k_L are wave vectors, respectively, J_0 is the Bessel function of the 0-th order, R is the cavity diameter, z and r are the positions along the cylindrical microwave cavity axis and the radius, respectively, and C.C. represents a complex conjugate. The transmissivity T of the amplitude of the laser electric field through the gas cell can be derived by solving the equations of motion for the density matrix of the three - level model under the perturbations due to external fields given by eqs.(4) and (5). The solution can be expressed as $T = \exp(-\delta - i\phi)$, where δ and ϕ are the amplitude attenuation and the phase shift, respectively. The δ and ϕ can be described by using nonlinear complex susceptibility $\chi_e^{(3)} (= \chi_e'^{(3)} - i\chi_e''^{(3)})$

$$\delta = \frac{\chi_e''^{(3)}\,\omega L_g}{2c} \qquad (6)$$

and

$$\phi = \left[1 + \frac{1}{2}\chi_e'^{(3)}\right]\frac{\omega L_g}{c} \qquad (7)$$

where, c is the speed of light, L_g is the length of ^{87}Rb gas cell. The macroscopic polarization (P) of ^{87}Rb gas can be described as

$$P = \varepsilon_0 \chi_e^{(3)} E_L = Tr(\mu_e \rho) , \qquad (8)$$

where, ε_0 is dielectric constant of vacuum, μ_e is the electric dipolemoment operator, and ρ is the ensemble-averaged density matrix.

On the other hand, the atomic density matrix obeys the Schrödinger equation

$$\frac{\partial \rho}{\partial t} = \frac{1}{i\hbar}[H, \rho] + \frac{\partial \rho}{\partial t}_{exec} + \frac{\partial \rho}{\partial t}_{relax} \quad , \quad (9)$$

where, t is the time, $\hbar$ is Plank's constant, H is the Hamiltonian including the atomic interactions with the laser light field and the microwave field, ρ_{exec} and ρ_{relax} represent the atomic excitation and relaxation processes in a phenomenological way. The Doppler effect on account of the atomic motion is not taken into account, because the drift velocity (v) of ^{87}Rb atoms in buffer gases is sufficiently low, i.e., $|k_L v|$, $|k_M v| \ll 1$ [7]. The Schrödinger equation (9) of three-level density matrix can be solved by taking account third-order perturbations arisen by external fields. The results, of which details of this derivation will be published[8], from eq.(4)-(9) are

$$\delta = \int_{-\infty}^{+\infty} \frac{\omega_{ca} L_g |\mu_{e(ca)}|^2}{2c\epsilon_0 \hbar} \frac{\xi_{ca}^{(3)} \gamma_{ab}' - \zeta_{ca}^{(3)}(\omega_{ba}' - \omega_M)}{(\omega_{ba}' - \omega_M)^2 + \gamma_{ab}'^2} g(\omega_L) d\omega_L \quad , \quad (10)$$

and

$$\phi = \int_{-\infty}^{+\infty} \frac{\omega_{ca} L_g}{2c} \left[2 + \frac{|\mu_{e(ca)}|^2}{\epsilon_0 \hbar} \frac{\xi_{ca}^{(3)}(\omega_{ba}' - \omega_M) + \zeta_{ca}^{(3)} \gamma_{ab}'}{(\omega_{ba}' - \omega_M)^2 + \gamma_{ab}'^2} \right] g(\omega_L) d\omega_L \quad , \quad (11)$$

where

$$\omega_{ba}' = \omega_{ba} - \frac{\gamma_{ab}(\gamma_{cb} + \gamma_{ca}) + x_L^2}{\gamma_{cb}\gamma_{ca} + x_M^2 - \Delta\omega_L^2} \Delta\omega_L \quad , \quad (12)$$

$$\gamma_{ab}' = \gamma_{ab} + \frac{\gamma_{ca} x_L^2 + (\gamma_{cb} + \gamma_{ca})(\omega_{ba} - \omega_M)\Delta\omega_L}{\gamma_{cb}\gamma_{ca} + x_M^2 - \Delta\omega_L^2} \quad , \quad (13)$$

$$\xi_{ca}^{(3)} = \frac{[(n_a - n_c) - (\gamma_a^{-1} + \gamma_c^{-1})w_L + \gamma_a^{-1} w_M][\gamma_{cb}\gamma_{ab} + x_L^2 + (\omega_{ba} - \omega_M)\Delta\omega_L] - [(n_a - n_b) + (\gamma_a^{-1} + \gamma_b^{-1})w_M - \gamma_a^{-1} w_L] x_M^2}{\gamma_{cb}\gamma_{ca} + x_M^2 - \Delta\omega_L^2} \quad , \quad (14)$$

and

$$\zeta_{ca}^{(3)} = \frac{[(n_a - n_c) - (\gamma_a^{-1} + \gamma_b^{-1})w_L + \gamma_a^{-1} w_M][\gamma_{cb}(\omega_{ba} - \omega_M) - \gamma_{ab}\Delta\omega_L]}{\gamma_{cb}\gamma_{ca} + x_M^2 - \Delta\omega_L^2} \quad , \quad (15)$$

ω_{xy} is the angular frequency of the optical transition between the levels $|x\rangle$ and $|y\rangle$, γ_{xy} (x,y = a,b,c $x \neq y$) is a transverse relaxation constant , γ_x (x = a,b,c) is a longitudinal relaxation constant, x_M and x_L are the Rabi angular frequencies of microwave and laser, respectively, $\mu_{e(ca)}$ is the electric dipole moment between levels $|a\rangle$ and $|c\rangle$, n_x (x = a,b,c) is the population of each level , $\Delta\omega_L (= \omega_{ca} - \omega_L)$ is the laser angular frequency detuning, w_M and w_L are products of the population and the transition probability of microwave and laser, respectively, and $g(\omega_L)$ is an energy density of the laser field which is given by

$$g(\omega_L) = \frac{1}{\pi} \frac{\Delta\omega_1/2}{(\omega_0 - \omega_L)^2 + (\Delta\omega_1/2)^2} \quad . \quad (16)$$

Here, ω_0 is a center angular frequency of the laser, and $\Delta\omega_1$ is a FWHM of the laser.

The second term of the right side of eq.(12) represents light shift, i.e., the microwave frequency shift induced by laser frequency detuning. The second term $\gamma_{ca}x_L^2$ and the third term $(\gamma_{cb}+\gamma_{ca})(\omega_{ba}-\omega_M)\Delta\omega_L$ of right side of eq.(13) represent a power broadening and an inhomogeneous broadening induced by light shift, respectively.

IV. Optimal Conditions for the Highest Microwave Frequency Stability

In order to obtain a frequency discrimination pattern, the microwave signal applied to microwave cavity was modulated as given by

$$\omega_M(t) = \omega_M(1 + M\sin\omega_m t) \quad . \tag{16}$$

Since the phase shift of eq.(11) depends on the microwave angular frequency ω_M, the phase of the transmitted laser light is modulated by means of modulation of microwave signal. As a result of such an induced phase modulation , FM sidebands would appear in the electric field of the transmitted laser light. It can be expressed as

$$E_T = (E_L'(t)/2)\exp[-i(\omega_L t - k_L z)] + \text{C.C.} \quad , \tag{17}$$

where

$$E_L' = E_L \sum_q T_q(\omega_M + q\omega_m)J_q(M)\exp(iq\omega_m t) \quad , \tag{18}$$

T_q is the transmissivity at the laser angular frequency $\omega_L + q\,\omega_m$ ($q = 0, \pm1, \pm2, \pm3$, the q-th order. The light power detected by a photodetector is given by $I_T(t) = c|E_L'|^2/8\pi$. If $|\delta|$, $|\phi| \ll 1$ and the laser FM sidebands up to the third order are retained for calculations. $I_T(t)$ can be written

$$\begin{aligned} I_T(t) = I_0[&A + B\cos\omega_m t + C\sin\omega_m t \\ &+ D\cos2\omega_m t + E\sin2\omega_m t \\ &+ F\cos3\omega_m t + G\sin3\omega_m t] \quad , \end{aligned} \tag{19}$$

where

$$I_0 = \frac{c|E_L|^2}{8\pi} \quad , \tag{20}$$

$$\begin{aligned} A = {} & J_3^2(1-2\delta_{-3}) + J_2^2(1-2\delta_{-2}) \\ & + J_1^2(1-2\delta_{-1}) + J_0^2(1-2\delta_0) + J_1^2(1-2\delta_1) \\ & + J_2^2(1-2\delta_2) + J_3^2(1-2\delta_3) \quad , \end{aligned} \tag{21a}$$

$$\begin{aligned} B = 2[& J_2J_3(\delta_{-3}+\delta_{-2}-\delta_2-\delta_3) \\ & + J_1J_2(\delta_{-2}+\delta_{-1}-\delta_1-\delta_2) \\ & + J_0J_1(\delta_{-1}-\delta_1)] \quad , \end{aligned} \tag{21b}$$

$$\begin{aligned} C = 2[& J_2J_3(\phi_{-3}-\phi_{-2}-\phi_2+\phi_3) \\ & + J_1J_2(\phi_{-2}-\phi_{-1}-\phi_1+\phi_2) \\ & + J_0J_1(\phi_{-1}-2\phi_0+\phi_1)] \quad , \end{aligned} \tag{21c}$$

$$\begin{aligned} D = 2[& J_1J_3(\delta_{-3}+\delta_{-1}-\delta_1-\delta_3) \\ & + J_0J_2(\delta_{-2}-2\delta_0+\delta_2) \\ & + J_1^2(\delta_{-1}-\delta_1)] \quad , \end{aligned} \tag{21d}$$

$$\begin{aligned} E = 2[& -J_1J_3(\phi_{-3}-\phi_{-1}-\phi_1+\phi_3) \\ & + J_0J_2(\phi_{-2}-\phi_2) \\ & + J_1^2(\phi_{-1}-\phi_1)] \quad , \end{aligned} \tag{21e}$$

$$\begin{aligned} F = 2[& -J_1J_2(\delta_{-2}-\delta_{-1}-\delta_1+\delta_2) \\ & + J_0J_3(\delta_{-3}-\delta_3)] \quad , \end{aligned} \tag{21f}$$

and

$$\begin{aligned} G = 2[& -J_1J_2(\phi_{-2}-\phi_{-1}+\phi_1-\phi_2) \\ & + J_0J_3(\phi_{-3}-2\phi_0+\phi_3)] \quad , \end{aligned} \tag{21g}$$

δ_q and ϕ_q are values of the attenuation and the phase shift at the laser angular frequency $\omega_L + q\,\omega_m$, respectively. The principal term of A of (21a) is $J_0^2(1-2\delta_0)$, in which $(1-2\delta_0)$ can be approximated as $\exp(-2\delta_0)$. This term represents the attenuation of the laser power on account of ^{87}Rb vapor. The principal terms of B and C of (21b) and (21c) are $J_0J_1(\delta_{-1}-\delta_1)$ and $J_0J_1(\phi_{-1}-2\phi_0+\phi_1)$, in which $(\delta_{-1}-\delta_1)$ and $(\phi_{-1}-2\phi_0+\phi_1)$ can be approximated as $d\,\delta/d\,\omega_M$ and $d^2\,\phi/d\,\omega_M^2$, i.e., the first derivative of the absorption and the second derivative of the dispersion, respectively. Figure 5 shows the dependences of A, B, C, D, E, F, and G of eq.(21) on the microwave frequency. The measured spectral profile of A, B, and C are shown in Fig.6.

As the profiles of B and C cross the abscissa at the center frequency, they can be used as a frequency discriminator. If a reference signal for the phase sensitive detection of Fig.1 is proportional to $\cos(\omega_m t - \theta)$, the output signal V_{PSD} from the phase sensitive detector is expressed as

$$V_{PSD} = V_0[(B/2)\cos\theta + (C/2)\sin\theta] \quad . \tag{22}$$

Since the ratio of contributions from B and C in eq.(22) is governed by the phase θ , several specific double resonance spectral lineshapes can be obtained by adjusting the value of θ at several values. Figure 7 (a) and (b) show the simulated result obtained by using eq.(22), where the values of γ_{ab}', M, and θ were swept so that the lineshape of Fig.7 fits that of Fig.2. The conditions of the present experimental result of Fig.2 (a) were given below

$$\begin{aligned} & M = 1.8\ \omega_m/\gamma_{ab}' = 2.3 \quad \theta = 234^\circ \\ & S_1 = 4.3 * 10^{-5} \quad (1/\text{Hz}) \quad , \end{aligned} \tag{23}$$

where, S_1 is the slope which is defined as a frequency discrimination sensitivity in eq.(3). It is seen from these figures that the lineshape is composed of several fine structures, and the peak-to-peak linewidth at the center of this shape is much narrower than that of the rf-excited ^{87}Rb atomic clock (,i.e., Fig.3). Such a narrower linewidth was obtained by utilizing the induced FM sidebands of the laser, and from the discussion given above, it can be recognized that a part of this technique is equivalent to that of the FM laser spectroscopy[9]. The minimum linewidth obtained in the present experiment was 20 Hz at the laser power density of 28.6 $\mu W/cm^2$. The linewidth of 20 Hz is 1/18 times that of the rf-excited ^{87}Rb atomic clock of Fig.3.

Values of parameters given in eq.(23) have not been optimized yet to get the steepest slope. Figure 8 shows the results of the dependence of the slope on these parameters calculated to find the optimal value of parameters. Figure 9 shows the dependence of this maximum value on the normalized modulation frequency and the modulation index. From Fig.8 and Fig.9, the optimal conditions for the highest microwave frequency stability are found as follows

$$M = 1.2\ \omega_m/\gamma_{ab}' = 1.0 \quad \theta = 120^\circ$$
$$S_{1(MAX)} = 1.3 * 10^{-4}\ (1/Hz) \quad . \qquad (24)$$

It is seen that the maximum slope is 3 times larger than that of the present experimental results of eq.(23), which means that value of the slope can be expected by optimizing the operating parameters in Fig.2 (a).

It is rather difficult to utilize the second order harmonics D and E in frequency discriminators because they do not cross the absissa at the center frequency. The third order harmonics F and G, however, can be utilized in it. If the reference signal is $\cos(3\ \omega_m t - \upsilon)$, the output signal is

$$V_{PSD} = V_0[(F/2)\cos\upsilon + (G/2)\sin\theta] \quad . \qquad (25)$$

Figure 10 shows the dependence of the slope of F and G of eq.(23) on the modulation and phase sensitive detection parameters. Figure 11 shows the dependence of this maximum value on the normalized modulation frequency and modulation index. From Fig.10 and Fig.11, the optimal conditions for the heighest microwave frequency stability are found by the computer simulation as follows

$$M = 3.6\ \omega_m/\gamma_{ab}' = 0.5 \quad \theta = 270^\circ$$
$$S_{1(MAX)} = 6.1 * 10^{-5} \quad (1/Hz) \quad . \qquad (26)$$

It was confirmed that the maximum slope was only a half that of the fundamental compenents of eq.(24). It is worth to mention, however, that the use of the phase sensitive detection at higher Fourier frequency is more favorable to reduce the effects of noise as long as the system is governed by flicker noise.

V. Light Shift

The dynamic stark effect by the electric field of pumping light would induce the light shift,i.e., the shift in microwave transition frequency. Since this frequency is used as a frequency reference for ^{87}Rb atomic clocks, the light shift can limit their frequency accuracy as well as frequency stability. Precise evaluations of the light shift have been rather difficult in the conventional ^{87}Rb atomic clocks because of the complicated spectral profile of the ^{87}Rb lamp. However , more precise evaluations can be carried out by using diode lasers with their narrower spectral linewidth, which could be useful to improve the frequency accuracy and the stability of the ^{87}Rb atomic clocks.

Figure 12 shows the relations between the microwave frequency shift ($\Delta\nu_{Rb}$) of the center of the spectrum and the laser frequency detuning ($\Delta\nu_L$) from the center frequency of an optical transition spectral line. It is seen that two curves in this figure exhibit clearly dispersive shapes, which is consistent with theoretical predictions[10]. At the first glance, it can be recongnized that the microwave frequency shift is increased with increasing the laser power density. The origin point of this figure represents the resonance frequency of the microwave transition which is free from the light shift. It means that the laser frequency must be tuned at this point, so that the affection of the light shift can be avoided, for the improvement of the frequency accuracy. For example, in the case of the curve B of Fig.12, laser frequency detuning should be less than 57 kHz to get the microwave frequency accuracy better than 10^{-12}.

Especially, at the laser power density as high as 2880 $\mu W/cm^2$, the discrimination patterns suffered light shift were observed for the study of the spectral lineshape. Figure 13 (a) shows examples of the frequency discrimination pattern correspond to several different detunings of the laser frequency. From this figure, it is apparent that frequency discrimination patterns are extreamly transformed according as laser frequency detunings are far from zero. Good coincidencses of the measured resuts of

Fig.13 (a) and the calculated results . Fig.13.(b) were obtained. The transformations of the amplitude attenuation and the phase shift correspond to the frequency discrimination pattern as a function of laser frequency detuning are also shown in Fig.14. From these figures, it can be concluded that the dynamic stark effect would not cause only energy-level shifts but also transformations of double resonance spectral lineshape. Similar results have been obtained by Camparo et al, but somewhat different model was employed for their analysis[11].

The effects of laser linewidth on microwave frequency stability were theoretically investigated by using eq.(16), by which the amplitude attenuation and the phase shift are calculated by the convolution with the laser oscillation spectral shape. As a result of it, it is seen from Fig.15 that the slope is increased with reducing the linewidth. However, from the fact that the linewidth of the laser used here is as broad as 40 MHz, it can be recognized that any remarkable increase of the slope can not be obtained even if the linewidth is reduced.

VI. Summary

The theoretical analysis for a novel FM sideband technique was carried out to improve the microwave frequency stability of the laser pumped ^{87}Rb atomic clocks. By using this technique, novel double resonance spectral lineshapes with ultra-narrower linewidths were obtained. The minimum linewidth obtained was 20 Hz, which was 1/18 times that of the rf-excited ^{87}Rb atomic clock.

The optimal conditions for the attainment of the best microwave frequency stability are found as follows;

$M = 1.2$ $\omega_m/\gamma_{ab}' = 1.0$ $\theta = 120^{\circ}$.

provided that the frequency discrimination pattern is composed of fundamental components.

The laser frequency detuning should be less than 57 kHz to attain the microwave frequency accuracy better than 10^{-12}.

Acknowledgements

The authors would like to express their gratitudes to Messrs. Kasai, Hashi, and Chiba (Fujitsu Co. Ltd.), Prof. Kano (University of Electro-Communication), Dr. Nakayama (Kyoto Institute of Technology), Prof. Oura, Drs. Kuramochi, and Teramachi (Institute of Vocational Training), for their valuable discussions. This work was partially supported by a Grant-in-Aid for Scientific Reseach from the Ministry of Education, Science and Culture of Japan, and by the Hoso Bunka Foundation.

REFERENCES

[1] J.L.Picque,"Hyperfine Optical Pumping of a Cesium Atomic Beam, and Applications", Metrologia vol.13, pp.115-119, 1977

[2] L.Lewis and M.Feldman, "Optical Pumping by Lasers in Atomic Frequency Standards", Proc. 35th Ann. Freq. Control Symposium, USAERADCOM pp.612-624, May 1981

[3] H.Hashimoto and M. Ohtsu, "Experiments on a Semiconductor Laser Pumped Rubidium Atomic Clock", IEEE J. Quantum Electron., vol. QE-23, pp.446 - 451, 1987

[4] J.Vanier, "On the Signal-to-Noise Ratio and Short-Term Stability of Passive Rubidium Frequency Standards", IEEE Trans. Instrum. and Meas., vol.IM-30, pp.277-282, 1981

[5] D.W.Allan, "Statistics of atomic frequency standards ", Proc. IEEE, vol.54,pp.221-230, 1981

[6] T.W.Hänsch, "Nonlinear Spectroscopy", North-Holland Publishing Company, pp 17-86, 1977

[7] R.H.Dicke, "The Effect of Collisions upon the Doppler Width of Spectral Lines", Phys.Rev., vol.89, pp.472-473, 1953

[8] M.Hashimoto and M.Ohtsu, Submitted to J.Opt.Soc.Am.B.

[9] G.C.Bjorklund, "Frequency-modulation spectroscopy: a new method for measuring weak absorption and dispersions", Opt. Lett., Vol.5, No.1, pp.15-17, January 1980

[10] B.S.Mathur, H.Tang, and W.Happer, "Light Shifts in the Alkali Atoms", Phys.Rev., vol.171, pp.11-19, 1968

[11] J.C.Camparo, R.P.Frueholz, and C.H.Volk, "Inhomogeneous light shift in alkali-metal atoms", Phys.Rev., Vol.27, pp.1914-1925, April 1983

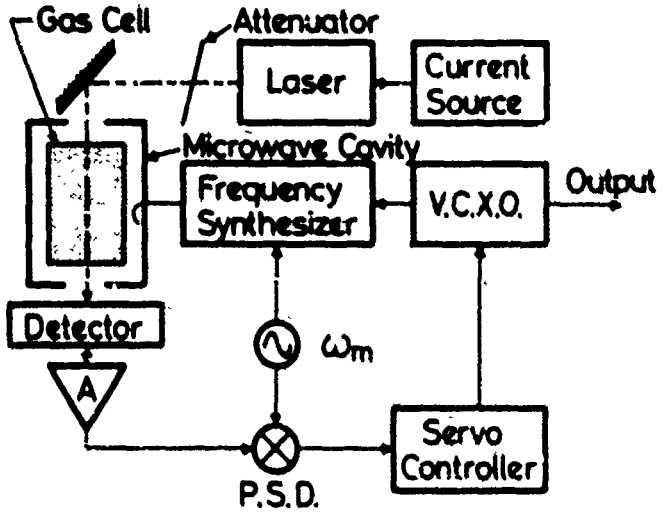


Fig.1 Experimental setup.
P.S.D. : Phase sensitive detector.
V.C.X.O. : Voltage controlled crystal oscillator.
A : Post - detector amplifier.
ω_m : Angular frequency of a low frequency oscillator for microwave frequency modulation.

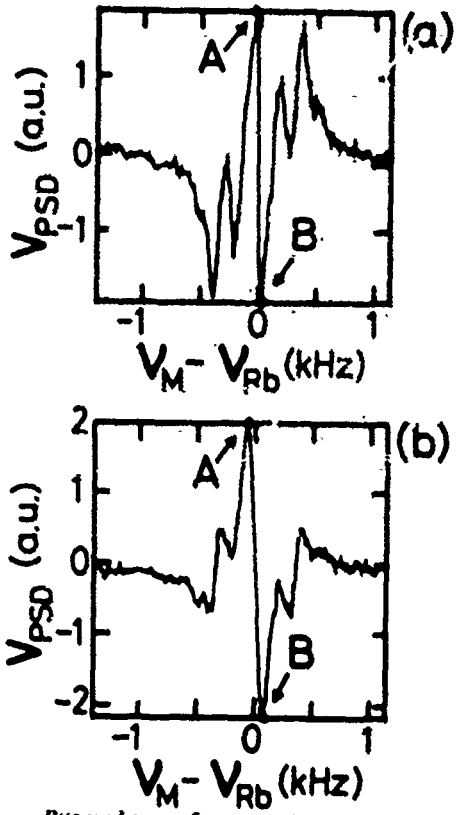


Fig.2 Examples of experimentally obtained lineshape in diode laser pumped ^{87}Rb atomic clock. The modulation frequency ($\omega_m/2\pi$) was fixed at 160 Hz (a) and 200 Hz (b). The laser power density was fixed at 14.4 $\mu W/cm^2$. Other conditions were not explict.
V_{PSD} : Output voltage from the phase sensitive detector.
ν_M : Microwave frequency. ($=\omega_M/2\pi$)
ν_{Rb} : Microwave resonance frequency. ($=\omega_{ba}/2\pi$)

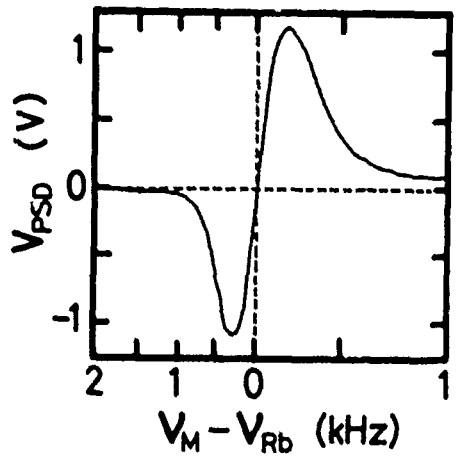


Fig.3 An example of experimentally obtained lineshape in rf-excited ^{87}Rb atomic clock. The modulation frequency was fixed at 160 Hz.

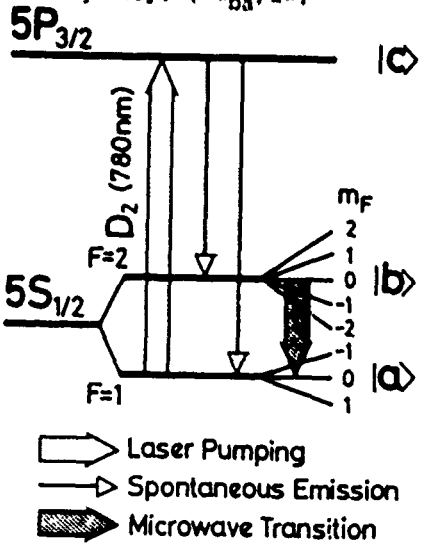


Fig.4 Energy levels of a ^{87}Rb atom relevant to the present study

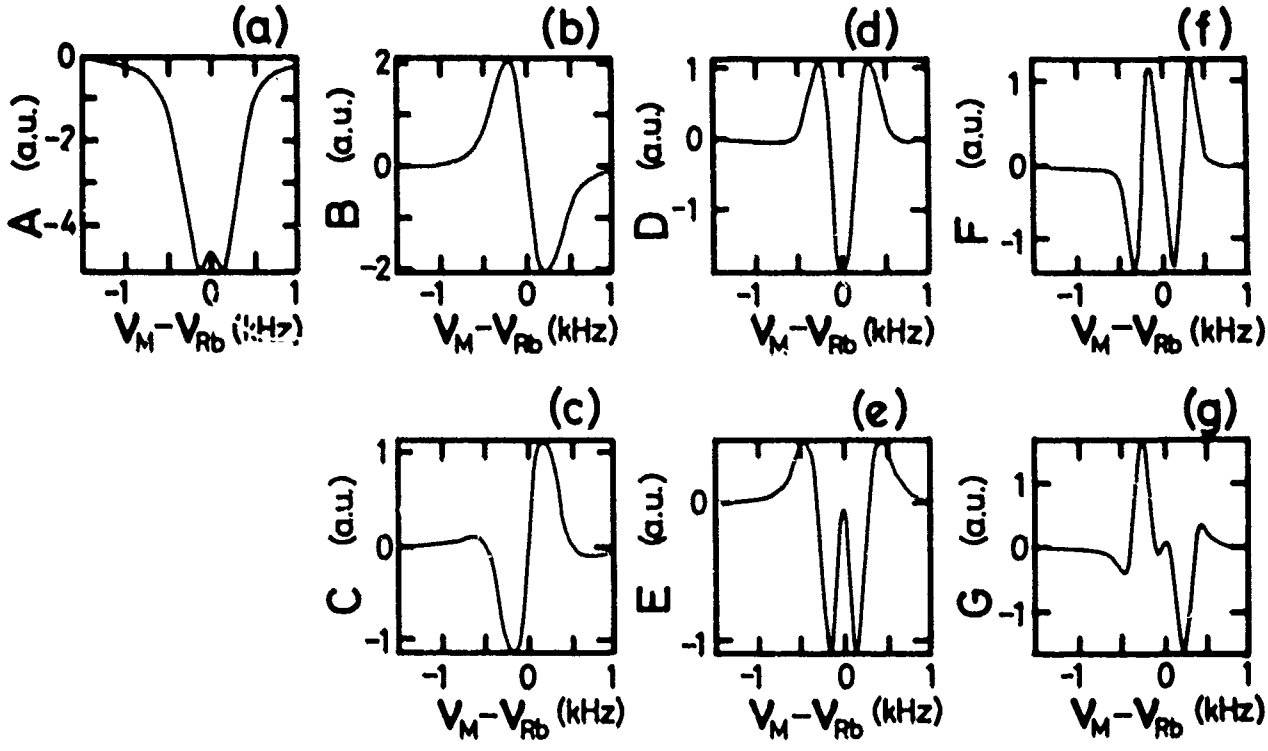


Fig.5 The dependences of the quantities A, B, C, D, E, F, and G of eq.(21) on the microwave frequency $\omega_M/2\pi$. Other quantities were fixed at $\omega_m/2\pi$=160 Hz, $\gamma_{ab}'/2\pi$= 150 Hz, and M=1,8.

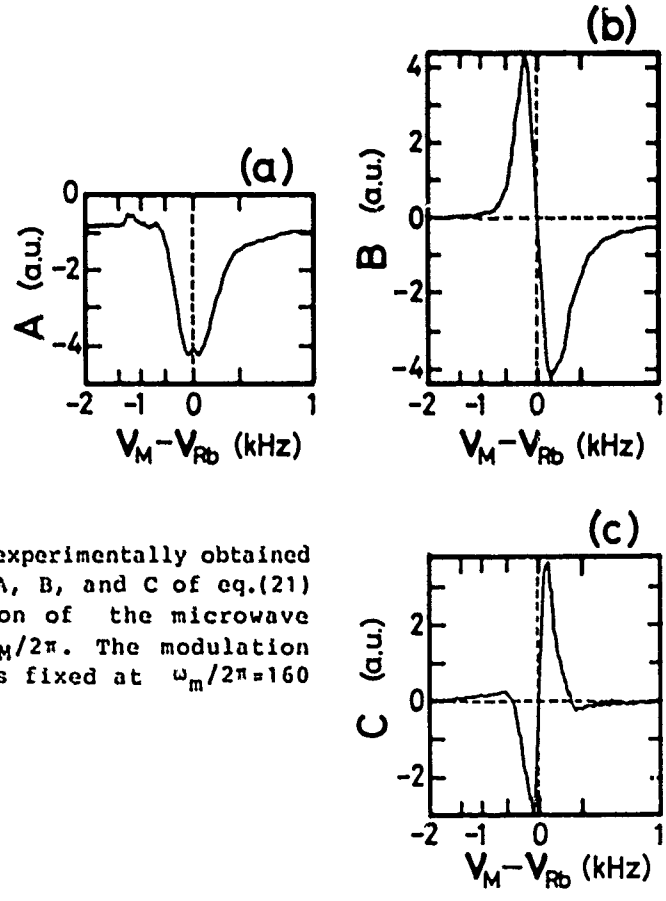


Fig.6 Exmaples of experimentally obtained quantities A, B, and C of eq.(21) as a function of the microwave frequency $\omega_M/2\pi$. The modulation frequncy was fixed at $\omega_m/2\pi$=160 Hz.

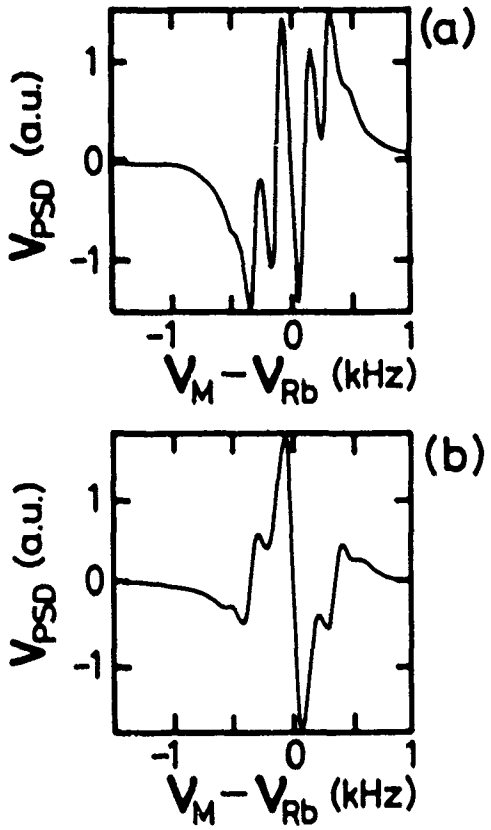


Fig.7 Simulated result obtained by using eq.(19)-(22), where the values of γ_{ab}', M, and θ were swept so that the lineshape of Fig.7(a) and (b) fit that of Fig.2(a) and (b), respectively.
(a)This fitting was obtained when $\omega_m/\gamma_{ab}' = 2.3$, $\theta = 234°$, and $M = 1.8$.
(b)This fitting was obtained when $\omega_m/\gamma_{ab}' = 2.9$, $\theta = 270°$, and $M = 1.6$.

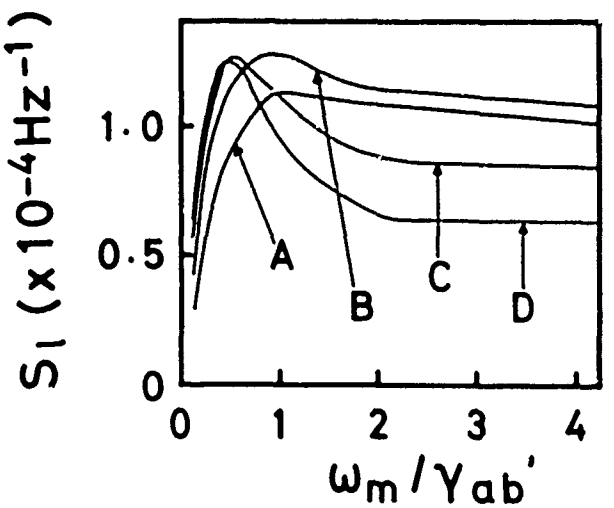


Fig.9 The dependence of the maximum slope of fundamental components on normalized frequency and modulation index. The phase differences were fixed at optimum value.

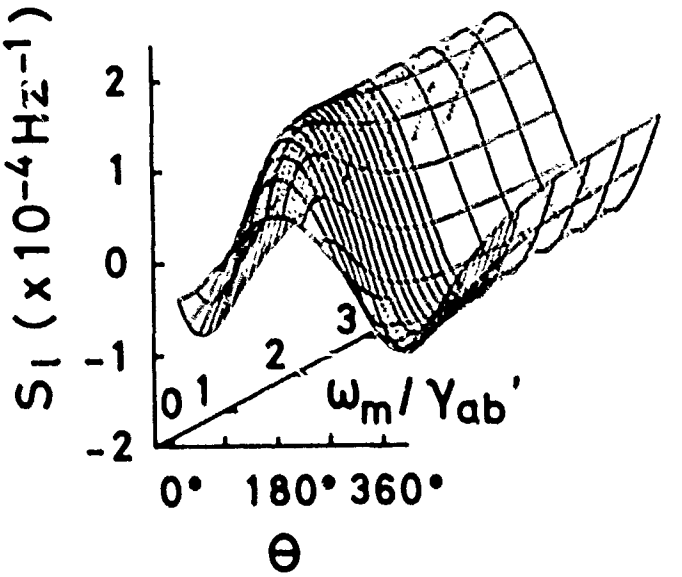


Fig.8 The dependence of the slope of fundamental components on the normalized frequency and phase difference. The modulation index was fixed at M = 1.2.

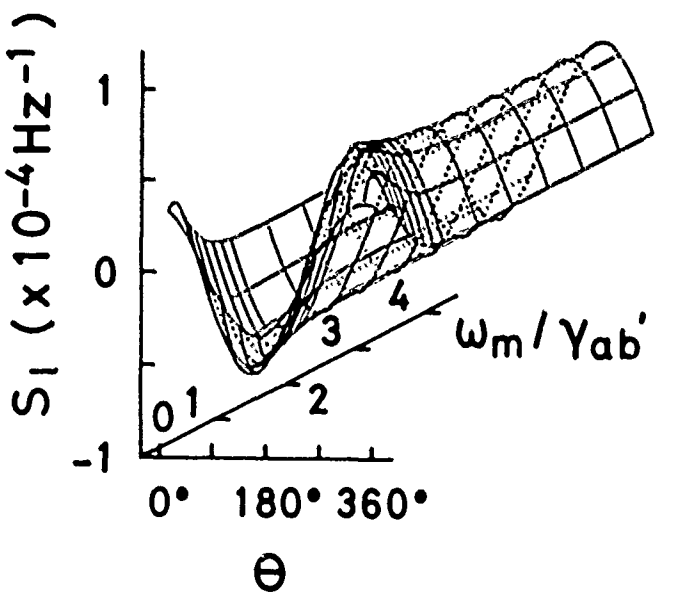


Fig.10 The dependence of the slope of third-harmonics components on the normalized frequency and phase difference. The modulation index was fixed at M = 3.6.

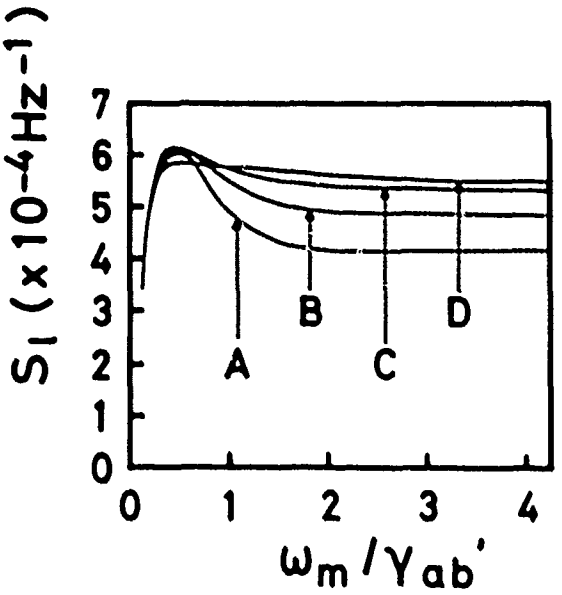


Fig.11 The dependence of the maximum slope of third-harmonics components on normalized frequency and modulation index. The phase differences were fixed at optimum value.

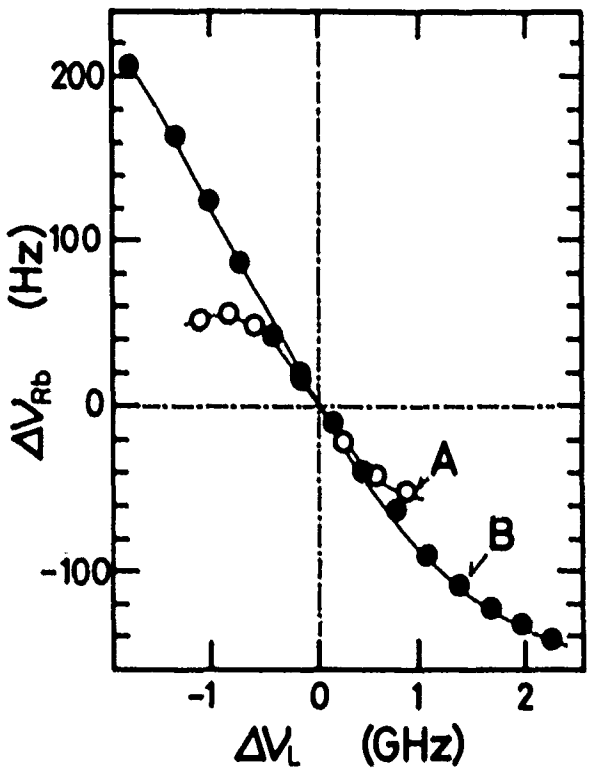


Fig.12 Light shift $\Delta\nu_{Rb}$ induced by laser frequency for the gas cell at 48°C. Since all the curves crossed at a common point, this point was taken as the origin of the axes in this figure. The laser power density was fixed at 57.6 μWcm^2(A,o) and 2880 $\mu W/cm^2$(B,●).
$\Delta\nu_L$: The laser frequency detuning. ($=\Delta\omega_L/2\pi=(\omega_{ca}-\omega_L)/2\pi$)
$\Delta\nu_{Rb}$: The microwave resonance frequency shift. ($=(\omega_{ba}'-\omega_{ba})/2\pi$)

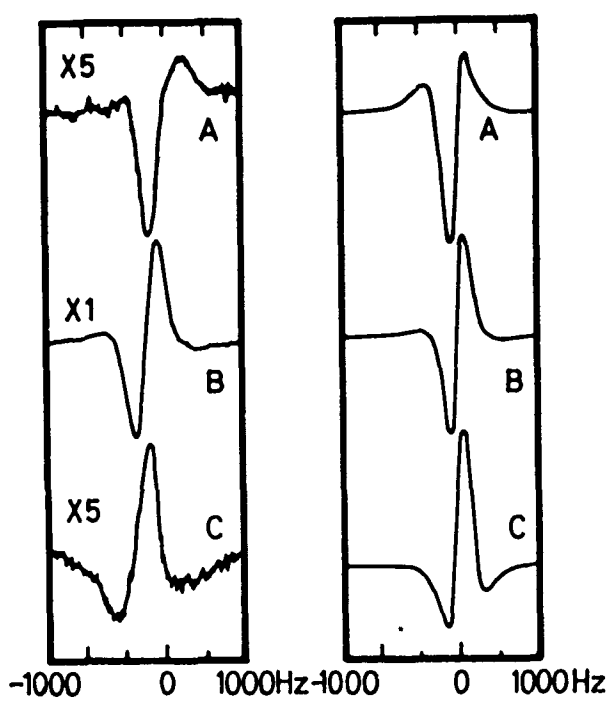


Fig.13 The frequency discrimination pattern suffered light shift on the microwave frequency. (a)The measured results. (b)The calculated results. The laser frequency detuning was fixed at 2070 MHz (A), 0 MHz (B), and -1800 MHz (C). The laser power density was fixed at 2880 $\mu W/cm^2$.

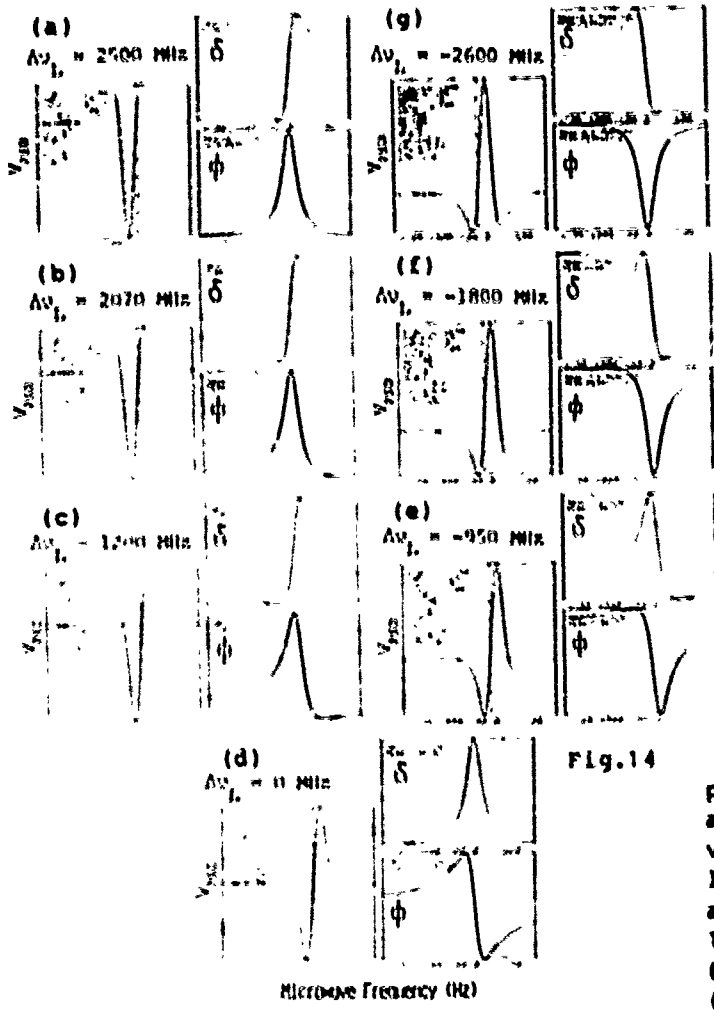


Fig.14 The frequency discrimination patterns, the amplitude attenuation, and the phase shift at various microwave frequencies. The laser frequency detuning was fixed at 2500 MHz (a), 2070 MHz (b), 1200 MHz (c), 0 MHz (d), -950 MHz (e), -1800 MHz (f), and -2600 MHz (g).

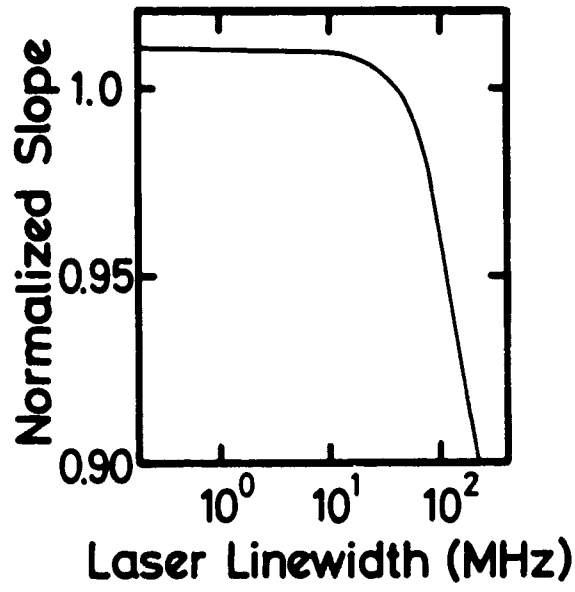


Fig.15 The normalized slope on laser linewidth.

41st Annual Frequency Control Symposium - 1987

A THREE DIMENSIONAL MODEL OF THE GAS CELL ATOMIC FREQUENCY STANDARD

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

Summary

In this paper preliminary calculations from a three dimensional clock model are discussed. In particular we have considered a recent suggestion that short term stability might be improved by varying the microwave power. Though our results support the general conclusion, they show that the degree of sensitivity is less than that predicted by the one dimensional model. The difference in the results of the two models is a manifestation of the more accurate treatment of the Position Shift effect in the three dimensional mode.. This more accurate treatment is highlighted by the three dimensional model's determination of isoefficiency contours (contours showing spatial regions in the clock cavity that have equal efficiency for producing clock signal), and noting their spatial dependence upon microwave power.

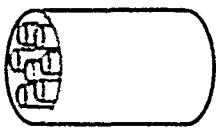
Introduction

Over the past few years a one dimensional, non-empirical model of the gas cell atomic frequency standard has been developed at The Aerospace Corporation.[1] This model analyzes the servo-control feedback circuitry and the atomic physics package's signal lineshape to yield the expected frequency stability of the atomic standard design under study. The model is one dimensional in that it only considers axial variations in: 1) the clock cavity's microwave magnetic field strength and 2) the degree of optical pumping within the clock's rubidium (Rb) resonance cell. However, even with this dimensional limitation, the model has proven itself to be quite useful for analyzing potential frequency stability improvements resulting from the use of a diode laser as the optical pumping light source,[2] and also for comparing the potential frequency stabilities of gas cell standards based on alkalies other than Rb.[3] In essence, the one dimensional clock model has been adequate for addressing these questions, because in these cases it is reasonable to consider the various spatially varying quantities as averages over the clock cavity's and cell's transverse dimensions.

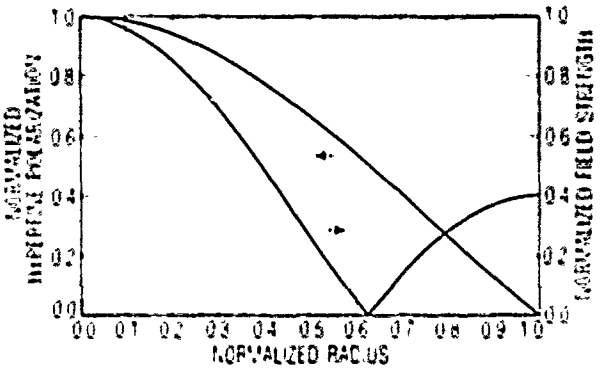
However, in order to analyze a wider range of questions concerning Rb clock performance, in particular in order to analyze potential mechanisms of frequency drift, a one dimensional model is only marginally adequate. Specifically, it has been suggested that frequency drift in the Rb standard may result from a spatial motion of the small region of vapor in the clock cavity that gives rise to the major portion of the clock signal.[4] Consequently, an analysis that averaged va ious quantities over the transverse dimensions of the clock cavity and resonance cell would be unable to investigate this hypothesis. A three dimensional model of the gas cell atomic frequency standard is therefore required in order to properly examine the plausiblity of this "Position Shift" mechanism as a contributor to frequency drift.

The Three Dimensional Model

To create a three dimensional model of the gas cell atomic frequency standard, we consider the clock cavity to be composed of hundreds of tubes (minimally 6800) as shown in Fig. 1a, each of which can be described by our one dimensional model; at present we consider only cylindrical TE_{011} and TE_{111} microwave cavity modes, though other spatial configurations could be modelled without much difficulty. As shown in Fig. 1b the transverse distribution of hyperfine polarization is approximated by considering only the first order radial diffusion mode,[5] and this is included in the model by superimposing this distribution on the microscopic solutions as previously discussed,[1] for simplicity we have assumed in all calculations that the optical pumping light is uniform in the transverse dimensions.



(a)



(b)

Fig. 1. (a) In order to construct a three dimensional model of the gas cell atomic clock we imagine the resonance cell as being composed of hundreds of tubes, each of which can be described by a one dimensional clock model. (b) Here we show the radial distributions of both the hyperfine polarization and the absolute value of the microwave magnetic field for a TE_{011} cavity mode. The radial distribution of hyperfine polarization results from diffusional relaxation, and we consider only the first order radial diffusion mode. The cusp in the microwave magnetic field strength corresponds to a 180 degree phase shift in the field.

Since diffusional relaxation is now included in the calculations by considering the full three dimensional hyperfine polarization distribution that results from this relaxation mechanism, phenomenologi-

cal diffusional relaxation rates are no longer required in the microscopic calculations.[1] For each tube the normalized stimulating microwave magnetic field, which corresponds to a normalized microwave Rabi frequency, is determined by the transverse position of the tube in the cavity and the cavity mode under consideration (see Fig. 1b). The normalizing constant is determined by the cavity Q and the microwave power fed into the cavity mode. Consequently, the dependence of the clock signal on microwave field strength is expressed simply in terms of: 1) the cavity mode, 2) the cavity Q and 3) the input microwave power.

Dependence of Short Term Stability on Microwave Power

Recently, there has been discussion in the literature regarding the dependence of the Rb clock's short term stability on microwave power.[1,6] In particular, in previous calculations it was shown that for clock designs similar to that of Williams et al.,[7] which are well described by a one dimensional clock model, the short term stability of the standard should be a fairly sensitive function of the microwave power P fed into the clock cavity.[1] Qualitatively, this can be understood from the fact that the short term stability of the standard is shot noise limited,[8] which allows writing:

$$\sigma_y = \frac{\Delta\nu}{S} . \qquad (1)$$

Here, σ_y is the square root of the Allan variance (henceforth we refer to this quantity as the Allan deviation), $\Delta\nu$ is the full width of the atomic hyperfine resonance and S is the atomic resonance signal amplitude. For a single atom, or equivalently an ensemble of atoms all experiencing the same microwave magnetic field strength (in other words for a clock cavity with no spatial field variation), relatively simple expressions relating signal amplitude and width to microwave power may be derived.[9] Considering first the regime of low microwave power (below saturation), S increases linearly with the microwave power, and $\Delta\nu$ is constant. Thus, in this regime the Allan deviation is inversely proportional to P. However, in the regime of high microwave power (saturation regime) S is constant and $\Delta\nu$ increases like the square root of the microwave power. As evidence indicates that gas cell clocks tend to be operated in the saturation regime,[10] one would expect that for clock cavities exhibiting relatively little spatial variation of the microwave field,

$$\sigma_y \sim P^{0.5} . \qquad (2)$$

Consequently, if a clock's normal operating point was at a microwave power level well into the saturation regime, then improvement in the short term stability of the clock could be attained by simply reducing the microwave power to the level corresponding to the onset of saturation.

Typical clock cavities, however, have fields which show considerable variation over the cavity volume. Consequently, since the atoms are effectively frozen in place by the buffer gas during the time intervals over which the clock signal is generated,[11] atoms in different regions of the clock's cavity experience different levels of microwave field strength depending on the clock's cavity mode. This then implies that atoms in different regions of the cavity reach the saturation regime at different levels of microwave cavity input power. Since the clock signal is essentially a sum over all of the individual atomic signals, the clock signal can be expected to have a fairly complicated dependence on the microwave power fed into the cavity.

An example of this dependence as predicted by our three dimensional clock model is presented in Fig. 2, which shows the Allan deviation at one second as a function of microwave cavity input power. In this specific example we have considered a minimum volume, cylindrical TE_{011} cavity with a Q of 100; other parameters used in the calculations of this paper are collected in Table I. Additionally, Figs. 3a and 3b

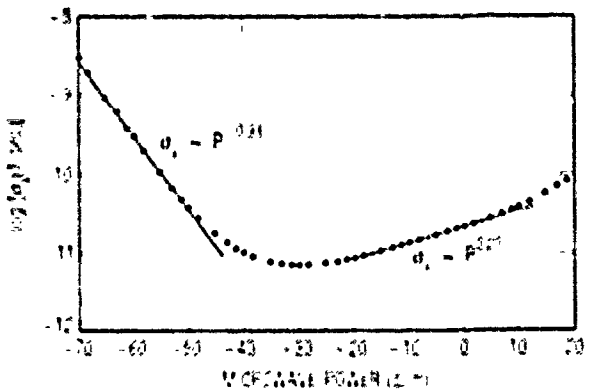


Fig. 2. For a TE_{011} cavity mode the figure shows the log of the Allan deviation as a function of the microwave power fed into a cavity with a Q of 100. For both low and high microwave powers the numerical data is fit to a power law in order to determine the coefficients shown in the figure. Note that at the highest microwave power (roughly 10 dbm) there seems to be a break in the power law scaling, a possible explanation for this effect is suggested in the text.

Table I.
Miscellaneous parameters used in the clock model calculations (see Ref. 1).

Parameter	Value
Optical linewidth	2.0 GHz
Cell temperature	60° C
Photocell response	0.5 A/W
Cavity Q	100
Optical power (D_1)	47 μW
Optical power (D_2)	75 μW

show the theoretical clock signal amplitudes and full widths, respectively, as a function of microwave power for this same case. Curves similar to those shown were obtained for the TE_{011} cavity for various optical pumping light intensities and resonance cell temperatures. Additionally, similar calculations were performed for a minimum volume TE_{111} cavity, and a fictitious, "constant", TE_{011} cavity which was taken

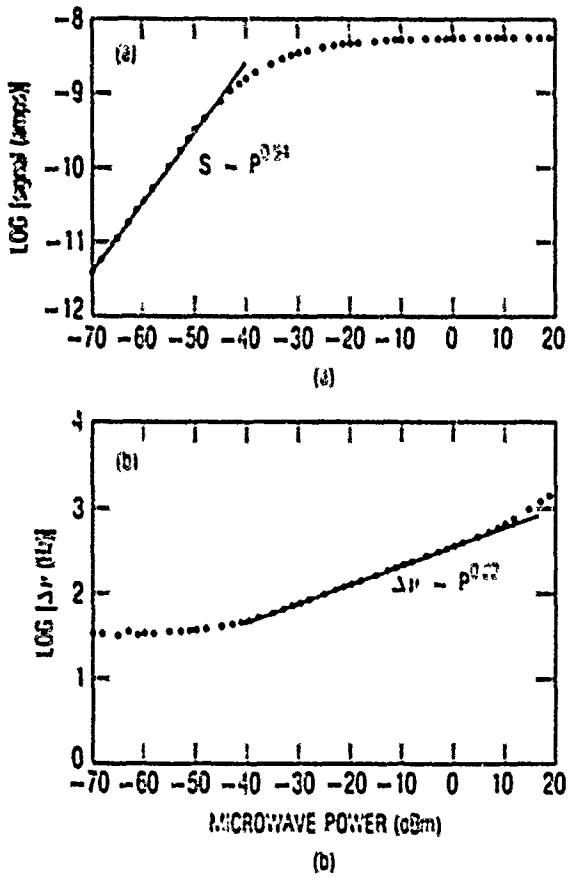


Fig. 3. For the same calculation as Fig. 2, we show the log of both the clock signal and the full width at half-maximum (Δν) as a function of microwave power. Again the numerical data are fit to power laws in order to determine the exponents shown in the figure.

as having no radial variation of microwave field strength (this cavity did, however, have the normal axial field variation).

In order to better understand the Allan deviation's dependence on microwave power, the numerical data for the Allan deviation, the total signal amplitude (i.e., the signal summed over all tubes) and the total signal linewidth are fit to simple power law formulas:

$$\sigma_y \sim P^{\alpha} \tag{3a}$$

$$S \sim P^{\beta} \tag{3b}$$

$$\Delta\nu \sim P^{\gamma} \tag{3c}$$

The goal of this exercise is to determine the power law exponents, and to compare them against the single-atom case. For the range of parameters considered, we find no strong dependence of these exponents on either cavity temperature or optical pumping light intensity. The exponents do appear to be a sensitive function of the cavity mode, and these results are collected in Table II.

It is clear from the table that the Allan deviation's microwave power sensitivity is directly correlated with the microwave power dependence of the clock signal amplitude and linewidth. In the low microwave power regime the linewidth is essentially constant, and variations in the Allan deviation are dominated by changes in the signal amplitude. The signal amplitude, and as a consequence the Allan deviation, displays essentially no difference in the power law scaling for the various field geometries considered. Specifically, the results show nearly the same linear power dependence as the simple single-atom case discussed above. Evidently, the clock signal amplitude is fairly insensitive to the geometry of the exciting microwave field. In the saturation regime, however, where the signal amplitude has essentially attained its maximum value and variations in the Allan deviation result from changes in the signal linewidth, the various cavity modes result in different Allan deviation power law exponents. Furthermore, it appears that as one allows the field to vary over an increasing number of spatial dimensions, the microwave power sensitivity of the signal linewidth decreases. This result can be explained by a correlation between the linewidth power law exponent and the number of degrees of freedom associated with the spatial movement of the signal dominating region. For example, a completely homogeneous field can be said to have no degrees of freedom, whereas the field of the TE_{011} mode can be said to have two degrees of freedom which are associated with movement both axially and radially. As evidenced by the results for the TE_{011} and TE_{111} cavity modes there might be some objection to the general validity of this statement, since the TE_{111} mode has an additional angular field variation, yet shows the same signal linewidth power law scaling as the TE_{011} mode. However, as will be discussed in the fourth section, for the TE_{111} mode the transverse movement of the spatial region that dominates the clock signal is limited to the radial direction. Thus, even though the field of the TE_{111} mode varies angularly, the spatial region which dominates the clock signal does not take advantage of the added degree of freedom.

As a final point we note that Fig. 2 shows a break in the power law scaling of the Allan deviation at approximately 10 dbm, which is reflected in the linewidth data of Fig 3b. It appears that at these high microwave power levels the power law exponent increases to a value near 0.4. A similar break at high microwave power levels was also seen in the TE_{111} mode calculations. Tentatively, we attribute this increase in the power law exponent to a spatial saturation of the clock signal dominating region, that is the region that dominates the clock signal may not move very much for these high microwave powers. Consequently, the power law scaling takes on more of the characteristics of the constant field case. Further calculations, however, need to be performed in order to substantiate this hypothesis.

The results from these calculations thus support the general conclusion that short term stability can be improved by varying the microwave power fed into the cavity. Specifically, since clock signal amplitudes are maximized with operation in the saturation regime, it is likely that for a typical clock one would want to reduce the microwave power to the point where saturation had just set in. The results, however, show that in the saturation regime the short term stability is less sensitive to microwave power than had been previously calculated;[1] this is due to the more accurate treatment of the

Table II.
Exponents for the power law formulas: $\sigma_y \sim P^{\alpha}$, $S \sim P^{\beta}$ and $\Delta\nu \sim P^{\gamma}$ (sat. = saturation regime, bsat. = below saturation regime).

Field Distribution	α(bsat.)	α(sat.)	β(bsat.)	γ(sat.)
Completely homogeneous (single-atom case)	-1.00	0.50	1.00	0.50
TE_{011} "fictitous" (no radial variation)	-0.93	0.45	0.93	0.43
TE_{011} mode	-0.93	0.20	0.94	0.22
TE_{0111} mode	-0.94	0.18	0.96	0.22

microwave field variation in the three dimensional model. Consequently, order of magnitude changes in the microwave power might be required before any appreciable change in the short term stability could be detected.

Inhomogeneity and Position Shift of the Clock Signal Volume

As discussed above, in the typical gas cell standard a buffer gas in the clock's resonance cell effectively freezes the atoms in place. Therefore, individual atoms experience different optical and microwave field strengths, and hence contribute to the total clock signal to varying degrees. Consequently, it is common to imagine the clock signal as being dominated by a small spatial region in the resonance cell (i.e., the resonance cell volume is inhomogeneous with regard to its efficiency in producing clock signal), and for the clock resonance frequency to be dominated by the perturbations experienced by the atoms in this localized region. If clock parameters were to change in such a way as to shift the position of this dominant region, then the possibly different local perturbations of the new region would result in a change in clock frequency. This is referred to as the Position-Shift effect,[12] which is considered to be a likely cause of frequency drift in the Rb standard.[4]

In order to better understand this phenomenon, and also to prepare for a theoretical investigation of the viability of the Position Shift as a mechanism of frequency drift, we have used the three dimensional clock model to map out the regions in the clock cavity which show different degrees of efficiency for producing clock signal. As illustrated in Fig. 1a, the clock cavity is imagined as being composed of hundreds of tubes: each tube transmits some small fraction of the total optical power reaching the clock photocell, and contributes to the total clock signal according to the change in optical power transmitted by the tube as the microwave frequency is varied. Since in our model different tubes can have different cross sectional areas, the efficiency of the various tubes in producing clock signal is compared by examining an individual tube's change in optical intensity (power per area). Tubes that produce the same light intensity change as the microwave frequency is varied are then said to be "isoefficient", and in this way we have been able to establish isoefficiency contours for producing clock signal.

Figure 4 is an example of these isoefficiency contours for a cross sectional slice of a TE_{111} microwave cavity excited by -50 dbm (cavity Q=100).

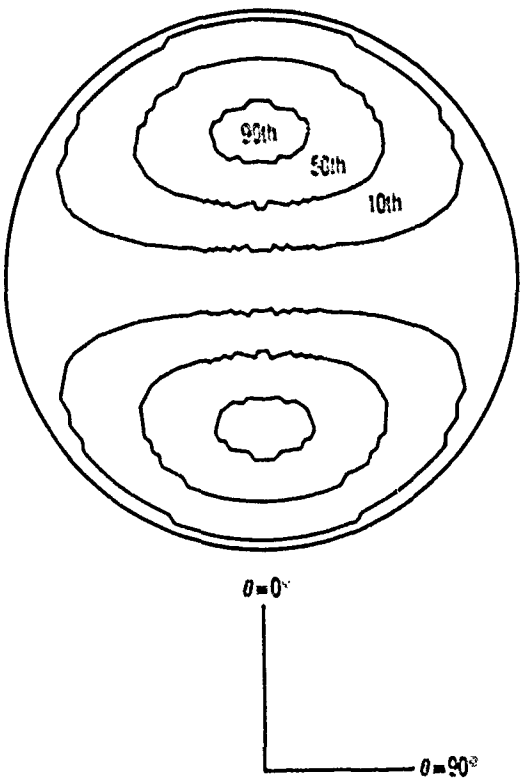


Fig. 4. This figure shows a cross sectional slice of a TE_{111} cavity mode excited by -50 dbm (Q=100). The contours are termed isoefficiency contours, since they correspond to regions in the clock cavity which exhibit the same efficiency for producing clock signal. The region bounded by the innermost contour is the 90th percentile region for efficiently producing clock signal as discussed in the text. Essentially, the 90th percentile region is more efficient than the 50th percentile region in producing clock signal, which in turn is more efficient than the 10th percentile region.

The region bounded by the innermost contour corresponds to the 90th percentile efficiency region (i.e., tubes within this region exhibit a transmitted intensity change that is greater than or equal to 90% of the intensity change exhibited by the most efficient tube), the region bounded the middle contour corresponds to the 50th percentile efficiency region, and the outermost contour corresponds to the 10th

percentile efficiency region. Other quantities associated with these contours are collected in Table III. In particular, the table shows that over 60% of the clock signal comes from roughly only a quarter of the cavity volume.

Table III.
Characteristics of regions bounded by isoefficiency contours.

Region	Fractional contribution to total signal	Fractional cavity volume occupied
90th percentile	14.1 %	4.5 %
50th percentile	62.8 %	26.3 %
10th percentile	96.8 %	66.0 %

In actuality this 60% of total clock signal is coming from less than 25% of the cavity volume, since not all axial regions of the cavity contribute to the clock signal to the same degree. The axial variation of the signal producing efficiency is included in the calculations by modelling the axial distribution of hyperfine polarization and the axial variation of the microwave magnetic field strength, and then numerically integrating the transmitted light intensity over the length of the resonance cell. The result of the numerical integration is a significant improvement in the speed of the calculations. Simultaneously, however, there is a reduction in the facility with which spatial information in the axial dimension may be obtained. Consequently, at the present time we can only say that the 25% is an upper bound to the volume, though we do not expect its true value to differ by more than about a factor of two.

Considering the inhomogeneous nature of the signal volume discussed above, and how one typically imagines the clock signal as being dominated by a small spatial region within the clock cavity, we see that while the three dimensional model provides some justification for this simple description, it does not completely validate it. The model does show that the clock signal derives from a localized spatial region within the clock cavity, but it also shows that this spatial region corresponds to a non-negligable fraction of the cavity volume. With the simple description of the signal volume as being dominated by a small region, there is the implication that the perturbations determining the clock's frequency are fairly well localized; this, however, is not substantiated by the three dimensional model. Considering Fig. 4 and Table III it is more accurate to state that the clock's frequency offset from some nominal frequency corresponds to a weighted average of perturbations over a fraction of the cavity volume.

The isoefficiency contours can also be used to illustrate the fashion in which the Position Shift effect occurs. Fig.5 shows 90% percentile isoefficiency contours for input microwave powers of -50 dbm, -30dbm and -10dbm. Note that as the microwave power is increased the 90% percentile isoefficiency contour shows a macroscopic change in its position in the cavity, moving to the central region of the cavity where the hyperfine polarization is largest. If the atoms in these spatial regions were perturbed to different degrees, then there would be an atomic clock frequency change. Consequently, once the three dimensional model incorporates spatially varying perturbations (e.g., static magnetic fields with gradients), it should be possible to launch a comprehensive theoretical investigation into the Position Shift effect. Such an investigation would serve the purpose of guiding experimental research into drift, thus making the experimental effort more efficient.

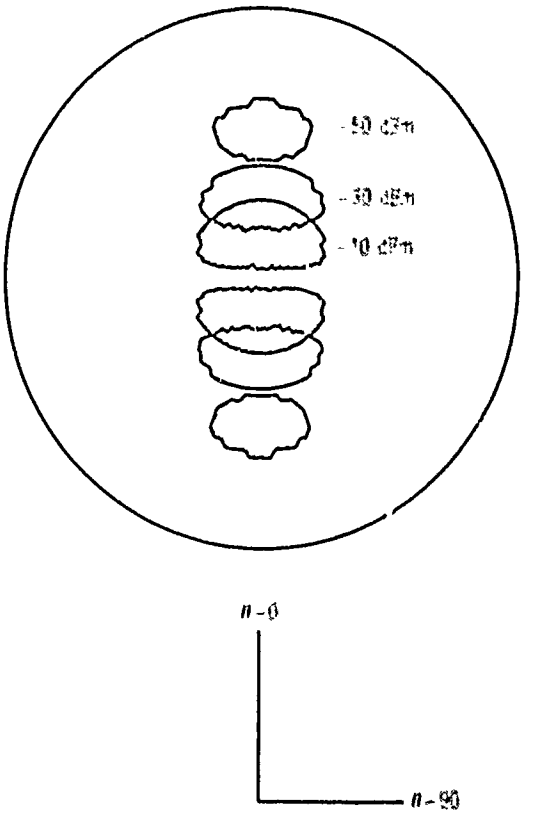


Fig. 5. In order to demonstrate the model's capabilities for analyzing the Position-Shift effect, the figure shows a cross sectional slice of a TE_{111} cavity mode (Q=100), and the 90th percentile efficiency regions as the microwave power exciting the cavity is varied. As the microwave power is increased the 90% percentile region shifts towards the center of the cavity. If the atoms in these different regions experienced different perturbations (i.e., different static magnetic field strengths), then the clock frequency would shift.

Summary

We have developed a three dimensional model of the gas cell atomic frequency standard based on our previous one dimensional clock model, and we are in the process of exploring its capabilities. Results presented here show that the Allan deviation's dependence on microwave power can be reasonably well modelled by power law formulas both below and above the clock signal saturation regime. Additionally, isoefficiency contours can be calculated and used to examine the change in position of the clock signal volume within the microwave cavity. In the near future we plan to incorporate spatially varying perturbations to the microscopic signals, and in this way to calculate atomic clock frequency shifts.

Acknowledgements

This work was supported by the United States Air Force Space Division under contract No. F04701-85-C-0086.

References

1. J. C. Camparo and R. P. Frueholz, "A noncmpirical model of the gas-cell atomic frequency standard," J. Appl. Phys. 59(2), 301 (1986).

2. J. C. Camparo and R. P. Frueholz, "Fundamental stability limits for the diode-laser-pumped rubidium atomic frequency standard," J. Appl. Phys. 59(10), 3313 (1986).

3. J. C. Camparo and R. P. Frueholz, "A comparison of various alkali gas cell atomic frequency standards," to be published IEEE Trans. Ultra. Ferro. and Freq. Control May 1987.

4. J. C. Camparo, "A partial analysis of drift in the rubidium gas cell atomic frequency standard," Proc. 18th Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, 1986.

5. P. Minguzzi, F. Strumia and P. Violino, "Temperature effects in the relaxation of optically oriented alkali vapors," Nuovo Cimento B 46(2), 145 (1966).

6. P. Tremblay, N. Cyr and M. Tetu, "Pumping light intensity transmitted through an inhomogeneously broadened line system: application to passive rubidium frequency standards," Can. J. Phys. 63(12), 1563 (1985).

7. H. E. Williams, T. H. Kwon and T. McClelland, "Compact rectangular cavity for rubidium vapor cell frequency standards," Proc. 37th Freq. Control Symp. (IEEE Press, New York, 1983) pg. 12.

8. J. Vanier and L.-G. Bernier, "On the signal-to-noise ratio and short-term stability of passive rubidium frequency standards," IEEE Trans. Instrum. Meas. IM-30(4), 277 (1981).

9. T. McClelland, L. K. Lam and T. H. Kwon, "Anomalous narrowing of magnetic-resonance linewidths in optically pumped alkali-metal vapors," Phys. Rev. A 33(3), 1697 (1986).

10. See for example the results in: I. Matsuda, N. Kuramochi, N. Shioni and H. Fukuyo, "Signal intensity characteristics of the 87Rb double resonance due to the pumping light," Jap. J. Appl. Phys. 16(3), 391 (1977).

11. R. P. Frueholz and J. C. Camparo, "Microwave field strength measurement in a rubidium clock cavity via adiabatic rapid passage," J. Appl. Phys. 57(3), 704 (1985).

12. A. Risley and G. Busca, "Effect of line inhomogeneity on the frequency of passive Rb87 frequency standards," Proc. 32nd Annual Symposium on Frequency Control (Electronics Industries Assoc., Washington D.C., 1978) pp. 506-513; A. Risley, S. Jarvis and J. Vanier, "The dependence of frequency upon microwave power of wall-coated and buffer-gas-filled gas cell Rb87 frequency standards," J. Appl. Phys. 51(9), 4571 (1980).

STUDY OF SEVERAL ERROR SOURCES IN A LASER RAMAN CLOCK

P.R. Hemmer*, V.D. Natoli, M.S. Shahriar, B. Bernacki*,

H. Lamela-Rivera, S.P. Smith and S. Ezekiel

Research Laboratory of Electronics

Massachusetts Institute of Technology

Cambridge, MA 02139

Abstract

We are investigating the development of a cesium clock using a laser excited resonance Raman interaction in place of direct microwave excitation. Such a scheme, employing only semiconductor laser excitation and exploiting fiber optic and integrated optic technology together with a simple atomic beam design, may lead to the development of smaller, lighter and perhaps cheaper atomic clocks. So far, we have been studying a sodium Raman clock which consists of a sodium atomic beam, a dye laser, and an acousto-optic frequency shifter, used for the generation of the second laser frequency. Recent performance showed a stability of 1×10^{-11} for a 5000 second averaging time. This compares favorably with commercial cesium clocks, when difference in atom transit time and transition frequency are taken into consideration. In this paper, we describe the results of a study of long term error sources unique to the Raman clock, which include: laser beam misalignment effects, errors caused by the laser frequency being slightly off-resonance with the intermediate state, laser intensity effects, and so on.

I. Introduction and Background

We have been investigating the performance of a clock based on laser induced resonance Raman transition in an atomic beam to determine the feasibility of such a scheme and to demonstrate any possible advantages over conventional microwave excited clocks. Although we have been conducting our experiments so far using a sodium atomic beam and dye lasers, this Raman technique is also applicable to a cesium atomic beam employing semiconductor laser excitation and may lead to the development of smaller, lighter and less expensive cesium beam clocks. Among the advantages of such a clock over a conventional cesium clock are the absence of state selection magnets and a microwave cavity, and that all alignments can be made using optics external to the vacuum system.

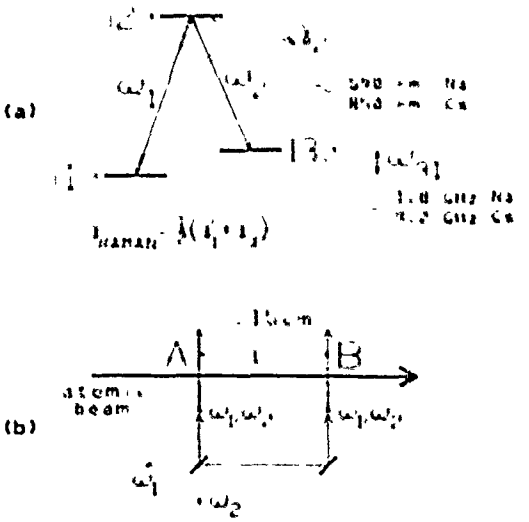


Fig. 1

(a) Schematic of laser induced resonance Raman interaction.
(b) Schematic of separated field excitation.

The stimulated resonance Raman interaction is illustrated using the three level system in Fig. 1(a). Briefly, Raman transitions are induced between states 1 and 3 using two laser fields, at frequencies ω_1 and ω_2, simultaneously resonant with the intermediate state 2. Earlier studies [1,2] show that, for copropagating laser fields interacting with an atomic beam at right angles, the Raman linewidth is determined by the widths of states 1 and 3 only. State 2 greatly enhances the transition probability, but does not contribute to the linewidth. Thus, for long lived states 1 and 3, the Raman linewidth becomes transit time limited, just as for direct microwave excitation.

* Rome Air Development Center,
Hanscom AFB, Bedford, MA 01731

CH2427-3/87/0000-042 $1.00 C 1987IEEE

To obtain a very small transit time linewidth we use Ramsey's method of separated oscillatory fields [3], as illustrated in Fig. 1(b), in analogy with conventional microwave techniques. In separated field excitation, the atom-field superposition states, excited in zones A and B by the two-photon Raman interaction, interfere quantum mechanically. Thus, the resulting interference fringes, as in the microwave case, have frequency spacings which are characteristic of the transit time between the interaction zones.

II. Experimental Setup

The experimental setup [4,5] used to demonstrate Raman clock applications is illustrated in Fig 2 The laser at frequency ω_1 is obtained from a single mode dye laser locked to the sodium D_1 transition, using fluorescence from the atomic beam. The laser field at frequency ω_2 is generated directly from that at ω_1 by an acousto-optic frequency shifter, driven by a quartz stabilized microwave oscillator near the 1772 MHz sodium hyperfine transition frequency. This greatly reduces the effects of laser jitter by correlating the frequency jitters [2] of ω_1 and ω_2 so as to produce a highly stable difference frequency. After leaving the A/O, the laser beams at ω_1 and ω_2 are combined in a single mode fiber, to ensure copropagation, before exciting the atomic beam at the two interaction zones, labeled A and B in Fig. 2.

A magnetic field is applied externally to separate out the different m-levels and right circularly polarized light is used. The m=0 and Δm=0 Raman transition is then selected so that the clock frequency be insensitive to external magnetic field variations to first order.

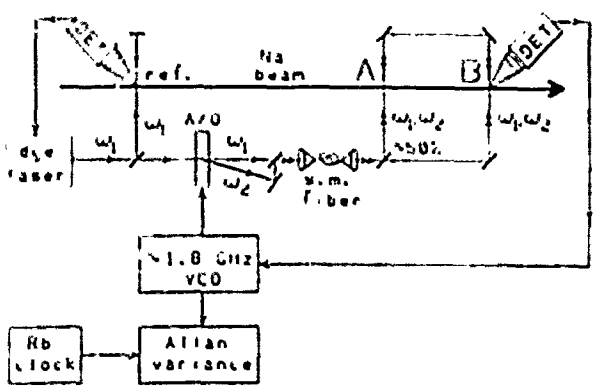


Fig. 2

Schematic of the experimental Raman clock setup.

Typical Raman/Ramsey fringes appear in Fig. 3(a). These fringes are observed by monitoring the fluorescence induced in zone B while scanning the microwave frequency with ω_1 locked to the D_1 transition The central fringe has a width of about 2.6 KHz (FWHM) which is consistent with the transit time for the 15 cm interaction zone separation

To stabilize the frequency of a microwave oscillator to the central fringe in Fig. 3(a), a discriminant is needed This discriminant, shown in Fig. 3(b), is obtained by frequency modulating the microwave source at a rate f_m = 610 Hz and demodulating the zone B fluorescence signal with a lock-in amplifier. The output of the lock-in amplifier is then used in a feedback loop to hold the microwave oscillator frequency at the central zero of the discriminant. The stability of this oscillator is measured by comparing it with a commercial rubidium clock.

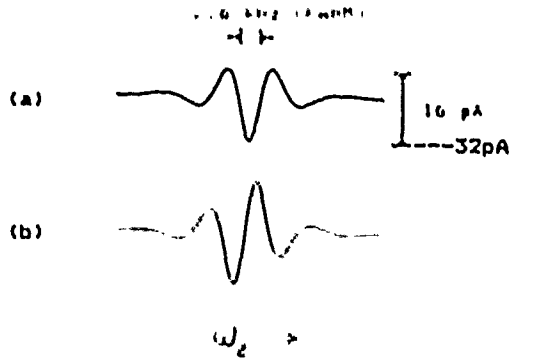


Fig. 3

(a) Typical Raman/Ramsey fringe lineshape for a 15 cm interaction zone separation.
Photomultiplier photocathode current levels as shown.
(b) Discriminant obtained using frequency modulation.

III. Clock Performance

Figure 4 shows a plot of the measured fractional frequency stability of the stabilized microwave oscillator, $\sigma_y(\tau)$, as a function of averaging time, τ. For τ = 5000 secs., the stability is about 1×10^{-11}. The data in this plot is very close to the predicted shot noise limit, shown by the upper dashed line. The lower dashed line in the figure is the projected stability expected if cesium were used in place of sodium, the difference being the result of the larger transition frequency and mass of cesium. As can be seen, the projected cesium results compare favorably with commercial cesium clocks.

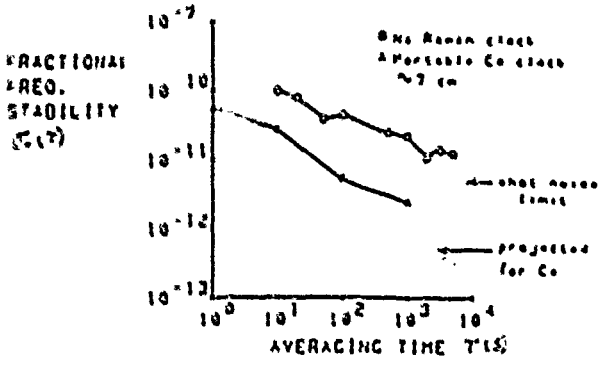


Fig. 4

Fractional frequency stability vs. averaging time for the sodium Raman clock.

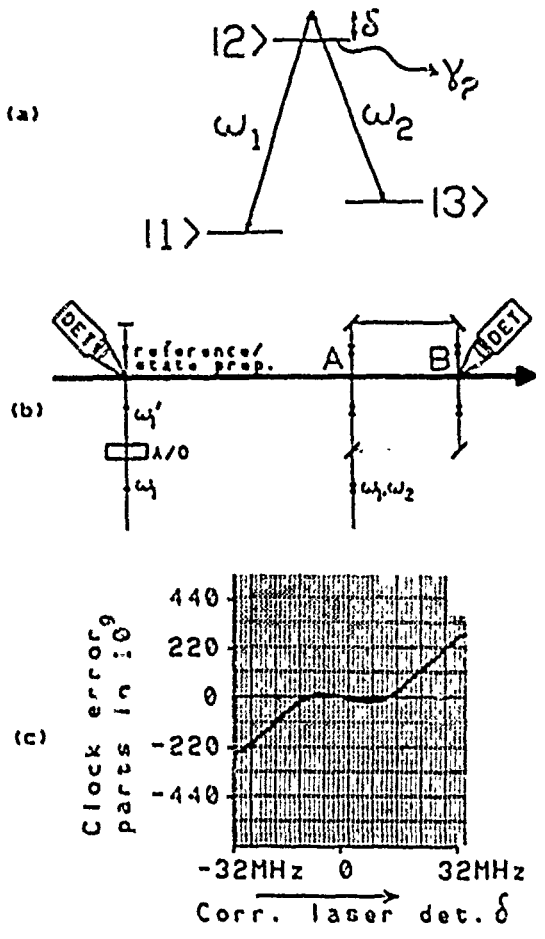


Fig. 5

(a) Schematic illustrating correlated laser detuning δ.
(b) Schematic of experimental technique for generating δ.
(c) Plot of clock error as a function of δ.

IV. Frequency Error Sources

We have been studying potential sources of error that can cause long term frequency drifts in the Raman process. Some error sources are similar to those of microwave clocks. These include the effects of path length phase shift, magnetic field, background slope, atomic beam misalignment, second-order Doppler, as well as errors due to electronics. In addition, there are errors unique to the Raman clock, such as those caused by laser frequency detuning, laser intensity changes, laser beam misalignment, optical atomic recoil, the presence of nearby hyperfine levels, and other smaller effects.

Here we will consider only error sources that are unique to the Raman clock.

(a) Correlated laser frequency detuning effects

Frequency errors can arise when ω_1 and ω_2 are off-resonance (by the same amount) with state 2, as shown in Fig. 5(a), where δ is the correlated detuning. The typical effect of δ on clock frequency error is shown in Fig. 5(c). To scan δ, the reference beam in Fig. 5(b) is frequency shifted with an A/C, so that ω_1 and ω_2 can be tuned with reference to the D_1 resonance. As can be seen, the clock frequency depends strongly on correlated detuning. However, for $\delta \approx 0$ this dependence is small. For the data shown the slope of clock error near δ=0 is 0.43 Hz for a detuning of 1% of the D_1 linewidth (equivalent to a fractional error of 2.4×10^{-10} for the same amount of detuning).

However, this slope depends on the laser intensity in the interaction zones as well as the intensity in the reference beam. For example, Fig. 6 shows the effect of interaction zone intensity on this slope. Here, clock error as a function of δ is plotted for three different interaction zone intensities. As can be seen, the slope is smallest for the interaction zone intensity of 0.8 mW/cm^2. Similarly, we studied the dependence of this slope on the reference beam intensity. The minimum slope achieved so far is, in fractional error, 2.4×10^{-11} for a detuning of 1% of the D_1 linewidth and it should be possible to reduce this further.

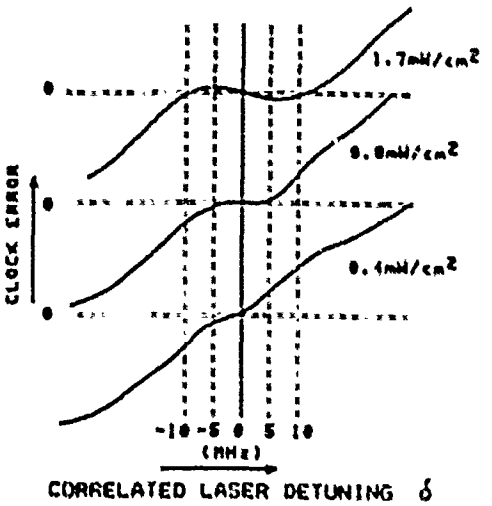


Fig. 6

Plots of clock error as a function of correlated laser detuning δ, for three different interaction zone intensities.

(b) Laser intensity effects

Frequency errors can also arise when the laser intensity changes in ω_1,ω_2 and the reference beam, together or separately. For δ=0, the predicted intensity induced errors are due to any fluorescence background slope, the presence of nearby hyperfine levels, optical atomic recoil, and so on. However, if correlated detuning δ is nonzero, then the amount of intensity induced error is determined primarily by the amount of detuning.

To illustrate this strong dependence of intensity induced error on detuning, we refer to Fig. 6 again. As can be seen, clock error is least sensitive to changes in intensity when detuning is nearly zero. We can find an optimum operating point by adjusting the amount of detuning. Around such a point the minimum slope achieved so far is 2.5×10^{-12} fractional error for 1% change in intensity.

Efforts are underway to study the individual causes of intensity induced errors mentioned above. For example, to determine the effect of nearby hyperfine levels we propose to study other transitions where this effect is expected to be different in magnitude.

(c) Laser misalignment effects

Laser misalignment also causes clock errors indirectly. For example, vertical misalignment causes effective laser intensity changes, as shown in Fig. 7(a), because of the translation of the Gaussian profile of the laser beam with respect to the atomic beam. Horizontal misalignment causes both translational and angular effects, as shown in Fig. 7(b). The angular misalignment creates effective correlated laser detuning due to the Doppler effect. The translational misalignment generates intensity error due to the translation of the laser beam relative to the detector. Misalignment in general can also cause imperfections in the standing waves that are used to reduce the effects of path length phase shifts. Imperfect standing waves cause errors because of changes in effective intensity as well as in path length phase.

To experimentally measure laser misalignment effects, we first find an optimum operating point at which laser detuning, laser intensity changes and path length phase shifts have minimal effects on clock frequency. Path length phase shift errors are minimized by adjusting the path lengths so that the clock frequency is the same for standing and travelling wave excitations. Around this optimum operating point we found that errors due to translations of the beams are much larger than those due to angular misalignments. Typically, the fractional error due to beam translation is about 3×10^{-11} for a 0.1 mm beam displacement.

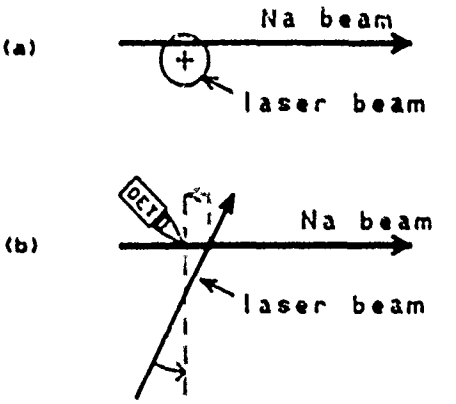


Fig. 7

(a) Schematic illustrating how vertical misalignment causes effective laser intensity changes.
(b) Schematic illustrating how horizontal misalignment changes correlated detuning through angular effects and effective intensity through translational effects.

V. Other Error Sources

Potential frequency-error sources that have not yet been studied include the effects of fluorescence from the interaction zones, the second-order Doppler shifts, limitations of single mode fibers, and so on.

VI. Future Work

As mentioned earlier, one of the more promising applications of the Raman technique is to develop a cesium atomic beam Raman clock, using semiconductor lasers. Such a clock, in which the two frequencies are intensity modulation sidebands, is nearing completion. As projected in Fig. 4, use of cesium is expected to increase the stability by a factor of 16. Also, this setup will have a more compact, recirculating cesium oven, with an atom throughput much larger than that of our present setup. Coupled with planned improvements in fluorescence collection efficiency, this is expected to enhance the signal to noise ratio considerably. With all the beams, atomic and laser, on the same table, misalignment will be greatly reduced. Also, the new beam will allow for much better shielding and control of magnetic fields. All these are expected to enhance our ability to study the remaining error sources with higher resolution.

We are also considering the attractive possibility of extending the resonance Raman technique into the mm-wave region of the spectrum. At these much higher transition frequencies, it may be possible to achieve better clock stabilties and many of the experimental problems associated with exciting mm-wave transitions in an atomic beam could be avoided. Finally, the Raman technique can also be readily applied to slowed or trapped atoms, possibly without greatly increasing the complexity of the experimental setup, since many of these techniques already make extensive use of resonant light.

Acknowledgements

We are grateful to John Kierstead for his able technical assistance. This work was supported by Rome Air Development Center, the Joint Services Electronics Programs, and the National Science Foundation

References

[1] J.E. Thomas, S. Ezekiel, C.C. Leiby, Jr., R.H. Picard and C.R. Willis, "Ultrahigh-resolution spectroscopy and frequency standards in the microwave and far infrared regions using optical lasers," Opt. Lett. 6, 298 (1981) and references therein.

[2] J.E. Thomas, P.R. Hemmer, S. Ezekiel, C.C. Leiby, Jr., R.H. Picard and C.R. Willis, "Observation of Ramsey fringes using a stimulated resonance Raman transition in a sodium atomic beam," Phys. Rev. Lett. 48, 867 (1982).

[3] N.F. Ramsey, Molecular Beams (Oxford U. Press, London, 1963).

[4] P.R. Hemmer, S Ezekiel and C.C. Leiby, Jr., "Stabilization of a microwave oscillator using a resonance Raman transition in a sodium beam," Opt. Lett. 8, 440 (1983).

[5] P.R. Hemmer, G.P. Ontai, and S. Ezekiel, "Precision studies of stimulated-resonance Raman interactions in an atomic beam," Journal of OSA, vol. 3 (1986).

41st Annual Frequency Control Symposium - 1987

CESIUM GETTERING BY GRAPHITE

N. D. Bhaskar, C. M. Kahla, R. P. Frueholz, and R. A. Cook
Chemistry and Physics Laboratory, The Aerospace Corporation
P. O. Box 92957, Los Angeles, California 90009

Summary

In cesium (Cs) beam tube (CBT) clocks only a small fraction of the Cs atoms effusing from the oven travel through the microwave cavity. The remaining fraction is gettered by pieces of graphite and graphite-coated surfaces placed at various locations inside the clock. Should the graphite getters cease absorbing Cs, the increasing Cs background would rapidly obscure the true signal, resulting in rapid reduction in the signal-to-noise ratio. Getter failure is then one of the primary life limiting processes in the Cs beam tubes. We are investigating the gettering properties of graphite in order to be able to characterize the type and the amount of graphite needed to efficiently getter Cs. From a microscopic angle we study the sticking coefficient of Cs on graphite as a function of time, under conditions similar to that encountered in a beam tube. A significant experimental observation is the rapid decrease in the sticking coefficient from an initial value of 1.0 to 0.25-0.30 after several days of exposure to Cs. We are also investigating the bulk gettering properties using various analytical techniques. Of particular interest is the maximum amount of Cs a given graphite can absorb. Preliminary measurements show that graphite can absorb approximately 20% by weight.

Introduction/Background

In the cesium (Cs) beam tube clocks (CBT) only a small fraction of the Cs atoms effusing from the oven travel through the microwave cavity where they undergo the 9192.6 MHz microwave transition. The remaining fraction of the effusing Cs atoms are collimated and gettered by slits and baffles made of graphite placed at various locations inside the clock. This is schematically shown in Fig. 1. The inner surfaces of the beam tube are also coated with graphite (aquadag) to improve the gettering. Should the graphite getters cease absorbing Cs, the increasing background of Cs atoms would rapidly obscure the true signal needed to operate the standard reliably. The clocks will also be rendered unusable if the insulators in the electric feedthroughs get shorted by the conductive coatings of the background Cs atoms. Getter failure is then a life limiting process in the Cs clock. For the clocks to be reliable, the gettering lifetime should be at least as long as the guaranteed lifetimes of the tubes. Standard failure is of particular concern in satellite applications such as GPS where reliable operation over the specified tube lifetime, 7.5 years, is required.

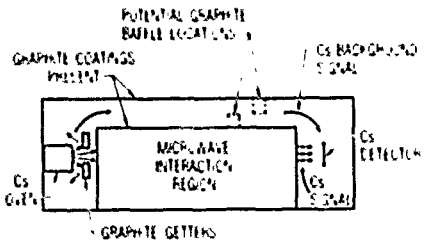


Fig. 1. Schematic representation of a Cs Beam Tube, showing the locations of graphite getters. The exposed metal surfaces inside the tube are also coated with Aquadag to further reduce the migration of Cs atoms to the detector region.

The practice of putting graphite blocks as getters inside the beam tubes dates back to the early days of the development of Cs beam frequency standards. Although the usefulness of graphite as a Cs getter has all along been qualitatively recognized, no systematic study of the Cs gettering properties of graphite has so far been reported. We have been experimentally investigating the gettering properties of graphite in order to be able to reliably characterize the type and amount of graphite to be used in space borne clocks.

Many materials are effective surface and bulk getters[1]. In surface getters the gas molecules collide and react with the surface. There is little diffusion into the material. In bulk getters the efficiency is determined by the diffusion of the impinging molecules into the bulk. In both cases, the characteristics of the surface and the interaction between the bombarding atoms and the surface should be such that the incident atoms stick well to the surface. Stated quantitatively, the prerequisite for efficient gettering is that the sticking coefficient γ be close to 1. Therefore, study of the sticking coefficient enables one to address the problems of gettering from a microscopic view. Another approach is to study the bulk gettering properties of graphite, e.g. how much Cs a particular weight of graphite may absorb. In our laboratory, we are investigating from both the microscopic (sticking coefficient experiments) and the macroscopic (bulk gettering) approaches.

In the second section we present the details of our experimental apparatus for measuring the sticking coefficient of Cs on graphite. The dependence of the sticking coefficient on the sample preparation procedures is also discussed. In the third section the experimental results of our bulk gettering studies are presented. The implications of our experimental findings are discussed in the final section.

Experimental Details and Data

Measurement of Sticking Coefficient (γ)

To measure the sticking coefficient γ of Cs on graphite we use an atomic beam apparatus. The experimental technique is illustrated in Fig. 2. A well collimated beam of Cs atoms effusing from an oven is directed towards a graphite target. The reflected beam at a particular angle θ is measured using a surface ionization detector. We measure the sticking coefficient of Cs on graphite relative to a paraffin coated surface which is known to be perfectly non-sticking. The sticking coefficient γ is defined as

$$\gamma(\theta) = 1 - \frac{\text{reflected signal from graphite at } \theta^{\circ}}{\text{reflected signal from paraffin at } \theta^{\circ}}$$

CH2427-3/87/0000-047 $1.00C1987 IEEE

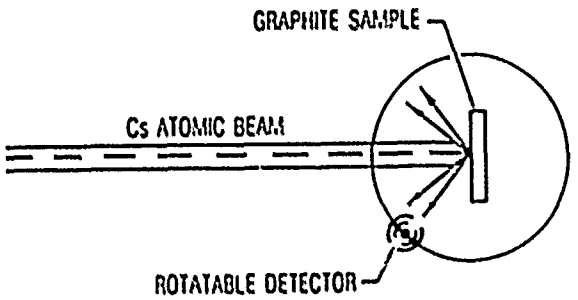


Fig. 2. Experimental arrangement for measuring the sticking coefficient Y of Cs on graphite.

We study the evolution of Y as a function of time for various incident Cs atomic beam intensities. Typical background pressure in the atomic beam apparatus is $< 1 \times 10^{-8}$ torr.

Graphite samples used in the present experiments are first baked externally in a quartz tube furnace at 1000°C under liquid nitrogen trapped rough vacuum for a minimum of two hours by which time the residual gas pressure is found to drop to a few microns. We use glass seals everywhere in order to prevent the contamination of graphite samples. The inner walls of the quartz oven are thoroughly cleaned before the start of baking. The graphite sample is loaded into a copper holder inside the atomic beam apparatus and pumped down to $\sim 1 \times 10^{-7}$ torr. The sample is baked to 350°C in high vacuum for several hours till the background pressure is below 1×10^{-7} torr. The vacuum chamber containing the graphite target is kept isolated from the source chamber containing the Cs oven, except while measuring the reflected beam signal. This is accomplished by placing a gate valve between the two chambers. Physical isolation is necessary because even at room temperature we detect a weak Cs beam effusing from the oven. Inadequate and improper baking of the graphite samples gets reflected in the rapid decrease of the Cs sticking coefficient. When properly degassed, graphite can be made more gas free than most common metals used in high vacuum systems. It is generally believed that graphite degassed at 1000 - 1100°C is "well outgassed."[2]

The residual gas pressure in our apparatus is usually less than 1×10^{-8} torr. This is accomplished primarily by baking the vacuum chambers at about 100°C for about 24 hours. This improvement in the high vacuum resulted in a marked decrease in the background signal recorded by the surface ionization detector. Under best operating conditions, the typical background signal is $< 10^{-15}$ A where as the reflected Cs signal at 70° is $\sim 10^{-11}$-10^{-12}A. The collision rate of the background gas at the residual gas pressure of 1×10^{-8} torr is ~ 27 monolayer/hour which is comparable to the collision rate of Cs on the graphite target. It is conceivable that what we are observing is actually a gas phase reaction catalyzed by the surface-in other words reaction of the impinging Cs atoms with the residual gas atoms on the surface. The evidence counter to this is the fact that our experimental data essentially remains unchanged when the Cs beam flux is increased by a factor of two and residual gas pressure is reduced by about a factor of two.

Short- Cs Exposure Times

Our experimental results are shown in Fig. 3. Cs beam intensities range from 1×10^{12} to 4×10^{12} atoms/cm^2 sec. These intensities are low compared to what would be seen by a getter placed directly in front of the oven in a CBT but comparable to intensities observed by graphite pieces placed at other locations in the beam tube. We have studied different types of graphite--POCO CZR-2, TRA-2 and Ultra Carbon UT6ST--all are known to have high porosity which should lead to superior gettering properties. Our experimental data show excellent reproducibility. For the short duration of six hours, no significant difference in behavior is observed for the different types of graphite studied. It is clear from our data that the rate of decrease of the sticking coefficient with time is nearly the same for the three different intensities of the incident Cs beam. However, care must be exercised in extrapolating our data to long exposure times.

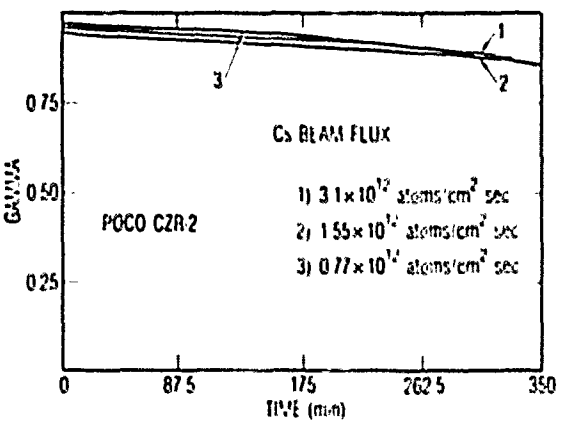


Fig. 3. Experimental data for Y vs t from short exposure times (~ 6 hours). Y is essentially the same for three different Cs beam intensities.

One of the factors that plays a crucial role in ensuring a high gettering efficiency is proper baking of the graphite samples. We have systematically studied the effects of baking on the sticking coefficient and this is shown in Fig. 4. The relevant conditions are shown in Table I.

Long Exposure Times

Our primary focus is to obtain reliable and reproducible data for the gettering efficiency of graphite continuously exposed to a Cs atomic beam for extended periods of time (~ days and months), and under conditions similar to that inside a CBT. This will enable us to develop a good theoretical model for gettering which can subsequently be used to specify and characterize the gettering efficiency of graphite in space borne clocks.

The variation of Y with time for three different types of graphite and an aquadag coated surface is shown in Fig. 5. The percentage porosity of the

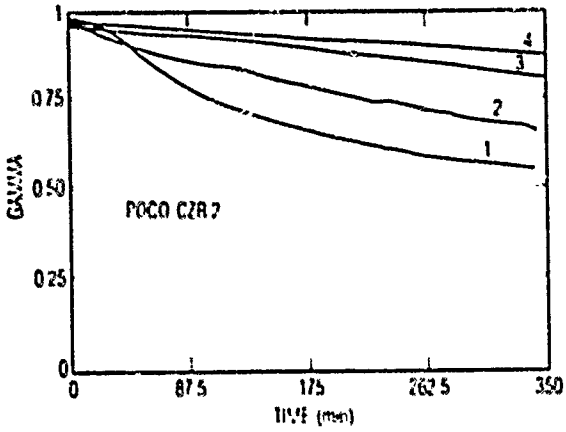


Fig. 4. Variation of Y with time for baked and unbaked graphite targets. Baking is done at 1000°C as explained in the text. See Table I for details.

Table I.
Baking conditions pertaining to the data in Fig. 4.

Curve	1000°C Baking	High Vacuum (Torr)
1	†Contaminated oven	1×10^{-7}
2	unbaked	1×10^{-7}
3	*Clean	1×10^{-7}
4	Clean	$<1 \times 10^{-8}$

†The quartz baking oven had rubber stoppers. During the baking the stoppers released contaminants into the oven thereby poisoning the graphite getters.

*The oven was refurbished with quartz stoppers and tubes to reduce contamination.

respective samples is also indicated in Fig. 5. The sudden steps in the data are due to fluctuations in the incident atomic beam which remains very stable only for about 24 hours. A differential measurement, with continuous monitoring of both the incident and the reflected beams would be more desirable for these types of long term experiments. Suitable corrections for the beam fluctuation were made in computing Y.

Bulk Gettering Studies

Procedure Employed in the Exposure of Graphite to Cesium Vapor

Four types of standard synthetic graphite along with aquadag coated surfaces were exposed to Cs vapor. The graphites selected for study, POCO CZR-2(32%) and TRA-2 (23%) and Ultra Carbon UT-6ST (17%) and UT-8 (17%), have high porosities (the fraction of volume due to voids in graphite, shown in the parenthesis) which we believe should yield superior gettering capabilities. Graphite samples (0.5" ×0.5" ×0.125") were fabricated using greaseless machining tools. These samples weighed approximately 1g. Aquadag was applied to cleaned stainless steel plates (1.0 cm ×1.0 cm ×0.16 cm). Both the graphite and aquadag samples

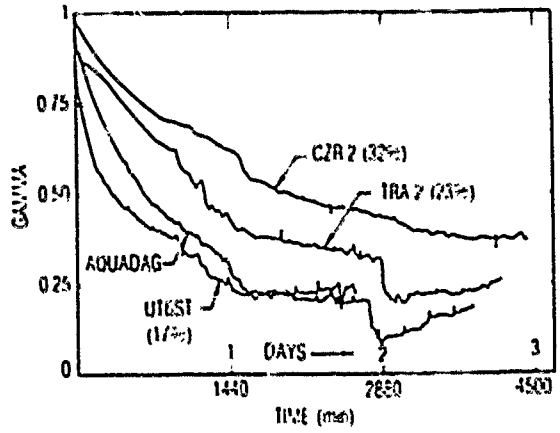


Fig. 5. Long term (60 - 75h) variation of the sticking coefficient as a function of time. At the end of 60-75h of exposure to the Cs beam, the sticking coefficient drops to ~ 0.2-0.4, depending on the type of graphite. The Cs beam intensity used in these measurements is ~ 4 x 10^2 atoms/cm^2 sec. POCO CZR-2 exhibits a distinctly slower rate of decrease of Y than all the other types investigated.

were baked under liquid nitrogen trapped mechanical pump vacuum at 500°C to remove any adsorbed water. The samples were then transferred to glass manifolds of the design shown in Fig. 6. Each graphite-containing manifold holds eight samples, two of each type of graphite being studied. Prior to opening the Cs reservoir the manifold is attached to a glass vacuum system. On the vacuum system the manifolds are baked at 350°C until the system pressure drops below 1×10^{-6} torr, typically 24 hours. After baking, the Cs reservoir is opened and degassed at which point the manifold, still evacuated is removed from the vacuum system. Cs is then transferred from the reservoir to other surfaces as indicated in Fig. 6. This provides a source of Cs vapor with a large surface area. The manifolds are placed in a 50°C oven and the graphite samples are exposed to Cs vapor for the desired periods of time. In these studies exposure times ranged from six to fifteen months, allowing the temporal evolution of Cs absorption to be monitored.

We wish to determine how much Cs the graphite (or aquadag) sample can absorb. Consequently, it is essential to insure that the observed Cs absorption is due to the properties of the graphite and has not been limited by the amount of available Cs vapor. With this in mind, a large surface area of metallic Cs was employed so that copious amounts of Cs vapor would be available. Also the manifolds were fabricated such that there are no serious conductance limitations on the amount of Cs vapor that may pass through the constrictions. If simplistically we assume these atoms impact the graphite surfaces uniformly the incident flux is on the order of 7 $\times 10^{13}$ atoms/cm^2 sec. This is close to the intensities seen by graphite collimators placed near the oven in CPS Cs beam tubes and larger than the intensities graphite surfaces would be exposed to in other locations.

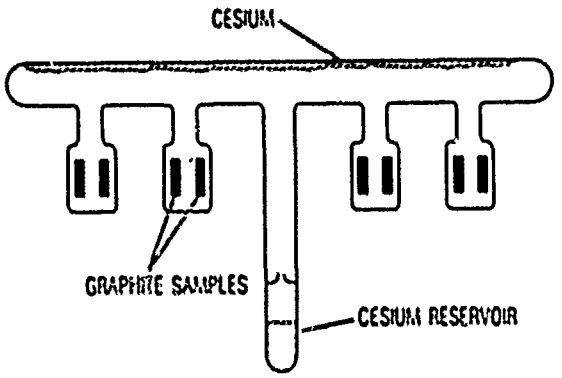


Fig. 6. Glass manifold used to expose graphite samples to cesium vapor. After evacuation the manifold is placed in a 50°C oven.

Determination of the Amount of Absorbed Cesium

1. Gravimetric Procedures

After investigating several techniques we have found that the most reliable method to determine the amount of Cs absorbed by a graphite sample is a simple gravimetric procedure. Prior to placement in the glass manifold graphite samples are accurately weighed. After Cs exposure the glass manifold is opened and the samples reweighed. The increase in sample weight is attributed to Cs absorption. Cs reacts with water and the reaction product, CsOH, is hygroscopic. As a result care is taken to minimize exposure of the sample to water vapor during manifold opening and sample weighing. Additionally, experiments were performed to determine how rapidly Cs impregnated graphite absorbs water. Graphite samples were placed on scales and the weight increases measured as a function of the exposure time to laboratory air. One example of these studies was a sample of POCO CZR-2 that showed a weight increase due to Cs exposure of 21%. After exposure to laboratory air for 30 minutes the weight increase had risen to 23%. After two hours of laboratory air exposure the weight increase had further risen to 25.5%. These results actually have an uncertainty as they depend on the humidity of the laboratory air. Three manifolds were opened and the samples transfered to a desiccator within a glove bag containing dry nitrogen. Their only exposure to the laboratory atmosphere was during the weighing process. Our studies indicated that an insignificant amount of water is absorbed during the three to four minutes needed for weighing. Two additional manifolds containing graphite samples and a single manifold holding the aquadagged samples were opened in air and the samples transferred to the desiccator. The water vapor absorption experiments indicate that a graphite sample's weight should increase by approximately three to five percent due to the water vapor-Cs reaction which might have taken place during the manifold opening. Consequently we consider the weight increases from the last three manifolds to be less reliable than those of the other three manifolds. The percentage weight increases of the graphite samples after Cs exposure are summarized in Table II.

The sample-to-sample variations in weight are greater than the uncertainties associated with the weighing process. At this point it is not clear whether the variations are indicative differences in the gettering abilities of various samples or result from a subtle systematic limitation of our exposure procedure. In spite of the variations, several observations may be made. It appears that the absorptive capability of graphite increases with increasing porosity. Based on the results of the manifolds opened under dry nitrogen, fractional increases in graphite weight resulting from Cs saturation may range from 4 to 20% depending on the graphite type. POCO CZR-2 graphite displayed the highest absorption capability (~ 20% weight increase). Aquadag covered surface also displayed high absorption capabilities, approximately 30% by weight. This result may be artifically high as the aquadag manifold was opened in laboratory air. It must also be noted that these coatings have only limited mass per surface area and hence limited absolute absorptive ability. In our studies the coatings, resulting from a single application of the aquadag solution, displayed surface densities of approximately 10 mg/cm^2. This would indicate each square centimeter of painted surface could at most absorb about 3 mg of Cs. Using the information supplied in Table II, manufacturers will be able to reasonably estimate the minimum graphite needed to absorb the initial Cs charge in the beam tube.

2. Other Measurement Techniques

Several other techniques, scanning Auger, atomic absorption spectroscopy, and acid-base titrations, were applied to the measurement of Cs absorbed by graphite. Of these, the acid-base titation was simpliest to perform and had the highest intrinsic accuracy. This procedure is based on the reaction between Cs and water,

$$Cs + H_2O \rightarrow CsOH + 1/2\ H_2. \qquad (1)$$

In an excess of water the CsOH dissociates producing OH^- ions and as a result a normally neutral water solution becomes basic. For each dissolved Cs atom an extra OH^- ion is produced. In the acid-base titration an amount of acidic solution, which supplies H^+ ions sufficient to neutralize the excess OH^-, is added to the basic solution. When neutrality is regained the titration end point has been reached and the number of H^+ ions added, will equal the initial number of dissolved Cs atoms. In performing this titration one hopes to obtain a completely independent measure of the amount of Cs absorbed by graphite. This measurement will also be insensitive to any water vapor absorbed by the graphite sample.

After weighing, several graphite samples taken from the fifteen month manifold were crushed and soaked in distilled water. Using hydrochloric acid solutions, with precise hydrogen ion concentrations, the titrations were performed. To identify the titration end point a phenolphthalein indicator was added to the solution. When neutrality is approached this indicator undergoes a dramatic color change. Comparison of the amounts of absorbed Cs indicated by the titration with that of the gravimetric analysis is also given in Table II. The titrations consistently show lower amounts of absorbed Cs, approximately 10 to 25% of that found by the gravimetric technique. A requirement for the acid-base titration to be accurate is that all of the absorbed Cs atoms undergo reaction (1). Apparently, a major portion of the absorbed Cs is not readily dissolved in water. This finding is also consistent with our atomic absorption measurements that detect atomic Cs. Prior to the absorption measurement the absorbed Cs must be similiarly dissolved in a water solution. All atomic absorption measurements indicated Cs absorption was less then that found by the gravimetric analysis, typically only

TABLE II. Summary of Gravimetric Measurements

		Manifolds Opened Under Dry Nitrogen					Manifolds Opened Under Laboratory Air		
EXPOSURE TIME (MONTHS)		6(1)	9	15	SATURATION LEVEL (2)	POROSITY (%)	6	10	9
GRAPHITE SAMPLE		Fractional Weight Gain (Acid-Base Titration Result) (%)					Fractional Weight Gain (Acid-Base Titration Result) (%)		
Ultra Carbon UT-65T(3)	#1	3.1	4.2	2.6 (1.1)	4 ± 2	17	16.9	25.0	
	#2	3.7	5.4	4.7			11.4	17.7	
Ultra Carbon UT-8(4)	#1	4.7	9.5	6.7 (1.7)	8 ± 2.5	17	7.9	16.1	
	#2	3.8	11.0	5.7			10.4	15.0	
POCO TRA-2	#1	4.2	12.2	12.1 (1.2)	11 ± 2	23	6.9	8.1	
	#2	2.6	8.1	10.5			7.9	6.6	
POCO CZR-2	#1	5.2	23.5	17.9 (4.7)	20 ± 3	32	17.9	14.2	
	#2	3.5	8.0	19.9			4.8	12.2	
Aqua-Dag Coatings	#1								32
	#2								33
	#3								40
	#4								26 (6.5)

(1) Graphite samples baked at 1100°C, all other samples baked at 500°C
(2) Average of all 9 and 15 month data except for CZR-2 in which 8% value was rejected on statistical grounds
(3) Grain size 0.04 mm
(4) Grain size 0.02 mm

approximately 10% of the gravimetric result. These findings give insights into the Cs absorption process and will be discussed subsequently.

A final comment concerns scanning Auger measurements made upon Cs exposed, graphite samples. Auger measurements produce signals proportional to the Cs atom concentration in the graphite. Unfortunately, it is very difficult to obtain reliable absolute concentrations. The Auger technique was applied to cross sections of graphite samples giving valuable penetration profile information but not absolute concentrations.

Discussion and Conclusions

Our experimental results represent the first quantitative investigation of the Cs gettering capabilities of graphite used in beam tube standards. We have demonstrated the feasibility of making long term studies of the gettering properties of graphite under conditions similar to that encountered in Cs beam tubes. As expected Y decreases with time. The most surprising aspect of our data is the rapid decrease in Y from the initial value of 1 to about 0.2-0.4 after just a few days of exposure to the Cs atomic beam. Undoubtedly Y will continue to decline and eventually Y will be zero. The fact that Y, after just a few days of exposure is merely 0.2-0.4 has implications on the long term tube reliability and provides insights into how gettering occurs within the CBT. To put things in perspective, we compare the actual Cs beam flux that the graphite block placed in the vicinity of the oven is exposed to with the beam flux that we have used in our studies. For graphite blocks placed at a distance of 2 to 3 mm from the oven, the effective Cs beam flux can be much more than an order of magnitude larger than the flux used in our experiments. Our studies indicate that more than 60% of the Cs atoms incident on this graphite block will not stick. This would imply that, at the very early life of the beam tube this block of graphite would loose most of its gettering capabilities. This would also mean that a large part of the gettering would have to be efficiently done by all the graphite coated surfaces and graphite blocks placed at other locations.

Our bulk gettering data provide for the first time a quantitative way of estimating how much graphite should be placed in a beam tube to getter the initial Cs charge of the oven. This is particularly important for Cs beam tubes, such as those used on GPS satellites, upon which stringent reliability requirements are placed. Estimates of the amount of graphite needed in a beam tube based on our results should be considered absolute minimum amounts. The potential sample-to-sample variation in absorptive capability of a given graphite indicates that not all pieces may be able to absorb the same amount of Cs. Also as a getter piece begins to approach Cs saturation the behavior of the sticking coefficient is not known. If the sticking coefficient is extremely low the Cs background pressure could reach deleterious levels

prior to complete saturation. Both of these concerns require further experiments and analysis to determine their importances.

A difficult problem to address experimentally is getter performance after a number of years of Cs exposure. Theoretical approaches may be the most productive when analyzing long term gettering, and in particular sticking coefficient evolution. While a rigorous model of Cs gettering by graphite does not yet exist, the present experiments are of value in arriving at simple models. Graphite is composed of coke grains which are held together with pitch residue. Cs atoms that stick to the graphite surface must penetrate into the graphite or the sticking coefficient would rapidly go to zero. The penetration of Cs into graphite is typically treated as a "grain boundary" process.[3] Processes of this type are characterized by two diffusion mechanisms proceeding at dramatically different rates. Between the grains diffusion proceeds rapidly. Penetration into the grains proceeds at a much slower rate. However, due to the large grain surface area a major fraction of Cs should be able to penetrate into the grains in reasonable periods of time. This is consistent with results of the acid-base titration that indicated 70-90% of the absorbed Cs could not be dissolved in water. Apparently this Cs has penetrated into the grains. At present we are using bulk exposure and sticking coefficient data to develope a theoretical model of the gettering process. This should allow accurate prediction of sticking coefficient behavior even after many years of Cs exposure.

Acknowledgement

This work is supported by the United States Air Force Space Division under contract No. F04701-85-C-0086.

References

1. J. F. O'Hanlon, A User's Guide to Vacuum Technology, pp. 206-214, J. Wiley and Sons, Inc. New York (1980).
2. G. A. Beitel, The Use of Graphite in High and Ultrahigh Vacuum: A Revieco, J. Vac. Sci. Technol. 8, 647 (1971).
3. D. Chandra and J. H. Norman, "Diffusion of Cesium through Graphite", J. Nucl. Mat. 62, 293 (1976).

41st Annual Frequency Control Symposium - 1987

NEW INSIGHTS INTO CAUSES AND CURES OF FREQUENCY INSTABILITIES (DRIFT AND LONG TERM NOISE) IN CESIUM BEAM FREQUENCY STANDARDS *

Andrea De Marchi
University of Ancona, Ancona, Italy.

* Work performed at the National Bureau of Standards, Boulder, Colorado, with the support of the US Naval Observatory.

Abstract

Extensive observations and valuable measurements have been performed over the years at the National Bureau of Standards on commercial Cesium Beam frequency standards with long term stability problems [1]. A strong correlation of frequency instability and drift with instability and drift of the microwave power was found, and a connection with variation in time of the velocity distribution of the atoms in the beam was recognized. these phenomena were attributed to transducing capabilities of distributed phase shift, which would map microwave power into end-to-end average cavity phase shift. No conclusive solution was found at the time to these problems, short of active servo stabilization of the microwave power.

In this paper experimental evidence is reported, which confirms previous findings, but unambiguously identifies Rabi pulling and cavity pulling as the major transducing effects that turn power variations into frequency variations. It is shown that C-field values exist, for which the standard's output frequency is unaffected by microwave power variations, and that operation of the standard at one of these zero-crossings represents a possible cure to all long term instability problems which derive from instability of the microwave power.

Extensive observations and valuable measurements have been performed over the years at the National Bureau of Standards on commercial Cesium Beam frequency standards with long term stability problems [1]. A strong correlation of frequency instability and drift with instability and drift of the microwave power was found, and a connection with variation in time of the velocity distribution of the atoms in the beam was recognized. these phenomena were attributed to transducing capabilities of distributed phase shift, which would map microwave power into end-to-end average cavity phase shift. No conclusive solution was found at the time to these problems, short of active servo stabilization of the microwave power.

In this paper experimental evidence is reported, which confirms previous findings, but unambiguously identifies Rabi pulling and cavity pulling as the major transducing effects that turn power variations into frequency variations. It is shown that C-field values exist, for which the standard's output frequency is unaffected by microwave power variations, and that operation of the standard at one of these zero-crossings represents a possible cure to all long term instability problems which derive from instability of the microwave power.

1 Introduction

It is well known that the stability of Cesium beam frequency standards available on the market often deviates from the $\tau^{-\frac{1}{2}}$ slope beam-shot-noise limited behaviour as early as at a few days of averaging time and at a level not much better than 10^{-13}. When this happens it is considered a problem because for many of today's applications it would be desirable to have a stability in the low 10^{-14} range.

This problem had been addressed in the early seventies at the National Bureau of Standards from an experimental viewpoint, and valuable information was obtained as to what are the parameters which affect its insurgence. The results obtained at that time pointed at variations of microwave power and velocity distribution in the beam [1]. Plenty of evidence was found of the coupling between these quantities and long term frequency instability. Understanding what bias effects take part in transducing their variations into frequency variations is a question that calls for an analysis of the accuracy budget of commercial standards and of the physical parameters that may possibly affect each bias. In the above mentioned work end-to-end and distributed cavity phase shift were indicated as the main bias effects responsible for this correlation. In this paper we report data which support the theory attributing to Rabi pulling and cavity pulling the major role in producing long term frequency instabilities related to variations of power and velocity distribution.

A quick review of the main bias effects is first given, with attention at showing why phase shifts cannot explain the observed instabilities. Then experimental results are reported which quantify Rabi pulling and cavity pulling, show that no other major biases exist in the observed standards, and that microwave power independent alignments can be obtained. Operation at these points improves the long term stability and the environmental sensitivity of the standards.

The evidence supporting this conclusion is based on data obtained with three high-performance dual beam tube standards. Although the data is limited and relative to devices of only one make, it is the author's feeling that the obtained results should be considered at least a typical scenario for all commercial standards, possibly with some caution when the servo loop scheme used is based on other than sinewave or slow square wave frequency modulation.

2 A short review of frequency biases

A list of the major biases affecting the frequency of a commercial Cesium beam frequency standard is reported in Table 1. All effects known to be liable to cause shifts greater than few parts in 10^{-14} in a state of the art device are listed together with the typical order of magnitude of the associated shifts.

Some of the biases are intrinsecally stable at the 10^{-15} level. These are reported only for completeness. They are marked with the comment "stable" and will not be discussed further.

Other biases are usually smaller than 10^{-13} and stable unless the electronics are misaligned, which happens. These will be discussed here shortly, if only to dismiss them as a normal cause of unstable behavior.

The offset due to spectral impurities has been analyzed in a number of studies. The most complete of them [2] shows that even with 1% unbalance between symmetric sidebands at 60 Hz 30 dB below the 9.2 GHz

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

Accuracy Budget:

Perturbations due to physics	Quantization field (C-field)	$\sim 10^{-10}$	
	Black Body radiation	2×10^{-14}	stable
	Gravitational red shift	$\sim 10^{-14}$	stable
	Doppler 2nd order	$\sim 10^{-14}$	stable
Effects of interrogation method	Doppler 1st order	$< 10^{-14}$	stable
	Phase shift (end to end/ distributed)	$< 10^{-13}$	
	Cavity pulling	$< 10^{-12}$	
	RF spectrum	$< 10^{-13}$	
Effects of servo system	Neighbouring transitions (Rabi pulling)	$< 10^{-11}$	
	Integrator offset	$< 10^{-13}$	
	2nd harmonic distortion	$< 10^{-13}$	

carrier, the offset should be expected to be well below 10^{-13}. Its stability in the 10^{-14} range can be postulated for devices with good electronics.

The integrator offset causes a frequency bias which can be serious. However, in modern chopper-stabilized instrumentation operational amplifiers the input bias can be very stable. With a typical input signal to the integrator of 1 Volt per 10^{-9} frequency offset, the commercially available specifications of $<1\mu V/K$ and $<1\mu V/month$ for the temperature coefficient and the ageing of the input bias are adequate for keeping frequency variations below 10^{-14}. A bigger problem is the fact that the beam signal decreases with Cesium consumption as the tube ages. Avoiding frequency drifts from this effect calls for very careful initial compensation of the offset and/or automatic loop gain control. The latter is usually not included in analogic servo loops.

Other frequency biases due to the electronics, like distortions and offsets in the modulator/ demodulator system, have been analyzed in depth in [3]. Specifications on the electronics are quite tight for a guarantee that these effects be smaller than 10^{-14}. (The burden on the electronics is great if one wants to accurately split the line to 3×10^{-7}!!) Analysis of the electronics in the standards under test suggests that the long term instabilities of these effects should not cause problems at the level of this investigation in well aligned instruments.

The remaining bias effects from Table 1 will now be analyzed singularly in somewhat greater depth.

Phase shift

This effect has been indicated in the NBS study mentioned above [1] to be possibly a major responsible for frequency instability caused by microwave power sensitivity. As already said the importance of the stability of power and velocity distribution in the beam was well documented. However the mechanism which connected these quantities to frequency was still misterious. It was proposed that end-to-end and distributed phase shift would play a major role in this process. This must be understood as an hypothesis whose validity was very difficult to assess. In fact it is not possible to reverse the beam to measure phase shift in a sealed tube. It was later contended that in well manifactured cavities phase shifts should be smaller than 10^{-13}, and that diatributed phase shift in particular can hardly be expected to cause power induced frequency variations greater than 10^{-13} [4]. In fact the phase variation across the beam is very small in such devices, with the beam grazing the end shorts of the cavity.

Furthermore, even in the unlikely event of tubes with an end-to-end phase shift of several 10^{-13}, it has been shown [5] that its variation with microwave power cannot be expected to be relevant. In fact the formula traditionally used to evaluate this shift ($\delta\nu_{\phi}/\nu = \delta\phi/2Q$) does not take into account the operation of the servo loop. It would seem from it that as the line Q changes with microwave power via velocity selection, the associated shift should change with inverse proportionality. This is the argument which was put forth in [1]. However it is shown in [5] that a factor A_{ϕ} must be introduced in the formula to summarize the sensitivity of the servo to the end-to-end phase difference. The complete formula is:

$$\left.\frac{\delta\nu}{\nu}\right|_{\delta\phi} = A_{\phi}\,\frac{\delta\phi}{2Q} \qquad (1)$$

It is shown in [5] that the effects of microwave power variations on Q and A_{ϕ} tend to compensate, so that the power dependence of end-to-end phase shift turns out to be very small.

C-field

This is by far the greatest bias in Cesium beam standards. Its magnitude is

$$\delta\nu_c = 8.70262 \times 10^{-4}\, f_z^{\,2} \qquad (2)$$

where f_z is the Zeeman frequency in kHz, defined as the distance between neighbouring transitions. Variations of the C-field produce frequency variations which depend on f_z and are given by:

$$\delta(\delta\nu_c)/\nu = 2 \times 10^{-13}\, f_z \delta f_z \qquad (3)$$

Although the focus of this work is on sensitivity to microwave power, it must be remembered that C-field variations may occur, due to changes in the external magnetic field, in the shields, or in the exciting current. It appears from (3) that, at least for the latter, it is an advantage to operate the standard at

as low as possible a C-field value, in order to decrease the sensitivity to its variations.

Cavity pulling

This effect deserves special attention because it was very early indicated as a major responsible for temperature sensitivity in commercial standards [6].

The complete expression for cavity pulling is [5]:

$$\left.\frac{\delta\nu}{\nu}\right|_c = \Lambda_c \left(\frac{Q_c}{Q}\right)^2 \frac{\delta\nu_c}{\nu} \qquad (4)$$

where Q_c is the cavity's and Q is the Ramsey line's Q. $\delta\nu_c = \nu_c - \nu$ is the cavity mistuning from the Cesium resonance frequency, and Λ_c is a proportionality factor which depends on microwave power and modulation parameters. In fig.1 a typical behaviour of Λ_c with microwave power is shown, as calculated for an existing tube from its velocity distribution and actual modulation parameters. This behaviour can be quite different for different modulation parameters or for a different tube. Sinewave modulation was assumed for this calculation. In fig.1 Λ_H is also shown, which is the limit value of Λ_c in the static approximation, as calculated in [7]. It should be underlined here that while Λ_H vanishes at the optimum power P_{opt}, Λ_c usually crosses zero at some different power level. As a consequence it is clear from (4) that cavity pulling in reality does not necessarily vanish at P_{opt}, but rather at a power level which depends on velocity distribution and modulation parameters. This causes a frequency variation as the cavity resonance moves due to ageing or to its own temperature coefficient. It is shown in [8] that the latter can be of the order of 10^{-14}/K.

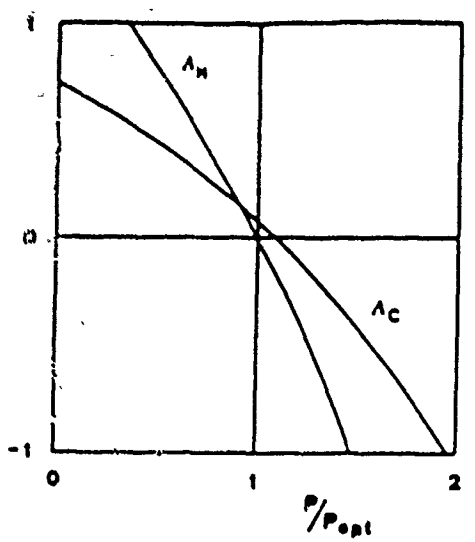


Fig.1 Calculated values of Λ_H and Λ_c as afunction of microwave power. The velocity distribution measured in a real tube was used for these calculations. The values of Λ_c are based on modulation parameters similar to those actually used in commercial standards.

It is also clear from (4) and fig.1 that cavity pulling causes microwave power sensitivity, unless the cavity is very well tuned to Cesium frequency ($\delta\nu_c$=0). Since it is very unlikely that the cavity mistuning should be greater than the half cavity width, the order of magnitude of cavity pulling should be expected to be

$$\left.\frac{\delta\nu}{\nu}\right|_c < \Lambda_c \frac{Q_c}{Q^2} \qquad (5)$$

In short tubes it is usually $Q \leq 3\times10^7$. With $Q_c < 10^9$ and $\Lambda_c(P_{opt}) \leq 0.2$, the upper limits obtained from (5) for cavity pulling and its variation with power are $\delta\nu/\nu|_c < 10^{-13}$ and $\delta(\delta\nu/\nu|_c)/\delta P < 2\times10^{-13}$/dB. For a carefully tuned cavity these figures can be a factor of 20 smaller. However it must be kept in mind that temperature and ageing induced variations of the cavity tuning may reintroduce pulling and power sensitivity at some level even after a careful tune-up.

Rabi pulling

This effect has been studied in [9] and a useful expression for the bias is

$$\left.\frac{\delta\nu}{\nu}\right|_R = \Lambda_R \frac{P}{P_{opt}} R(f_z) \qquad (6)$$

where $R(f_z)$ is a damped oscillating function of the Zeeman frequency with several zero crossing points, and Λ_R is a proportionality factor which depends on power, modulation parameters and velocity distribution. For a typical tube it is found in [5] that Λ_R is roughly proportional to the square root of P/P_{opt} around P_{opt}, as shown in fig.2. This makes Rabi pulling $\propto (P/P_{opt})^{3/2}$ around optimum power.

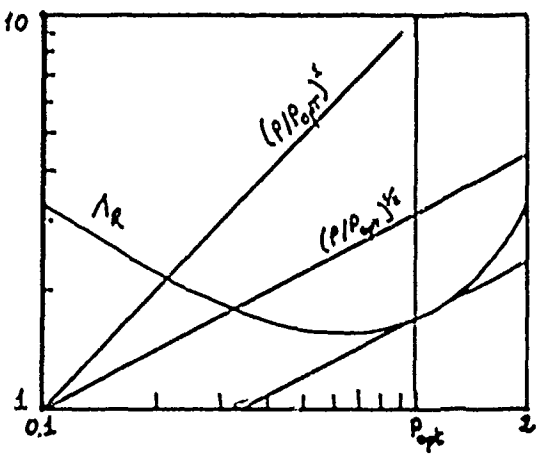


Fig.2 Behaviour of Λ_R as a function of microwave power as calculated for typical velocity distribution and modulation parameters.

3 Power sensitivity

Because phase shift and second order Doppler effect are so small in commercial standards, and in any case exhibit such a small dependence on microwave power [5], the only bias effects which can couple microwave power instability to frequency variations are cavity pulling and Rabi pulling. In this work experimental measurements are reported, which support this conclusion by the analysis of power sensitivity as a function of C-field.

In fig.3 typical experimental results are shown of frequency measurements for two different power levels as a function of the Zeeman frequency. These results were obtained prior to cavity tuning. After the cavity was carefully tuned, by maximizing the beam signal below P_{opt}, the same measurements were taken again on the same standard. The results obtained in this case are given in fig.4.

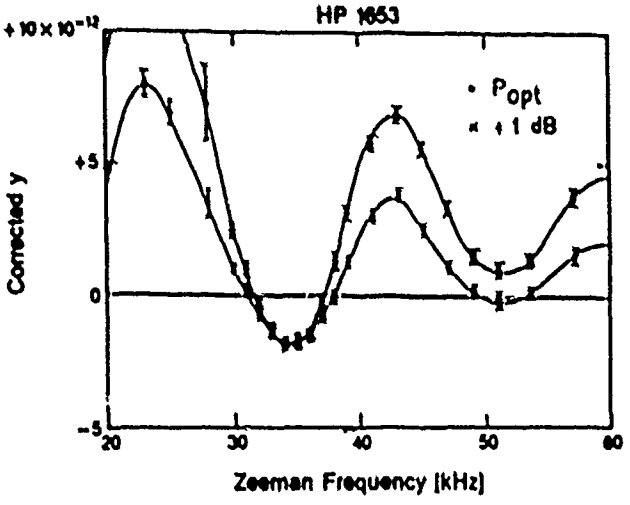


Fig.3 Experimental results for the Rabi pulling of one of the standards analyzed, before cavity tuning. Reported relative frequency data y are measured residual differences from AT1(NBS) once C-field and synthesizer offsets are removed. Dots are points taken at optimum microwave power, and crosses are points taken at about 1dB above optimum power.

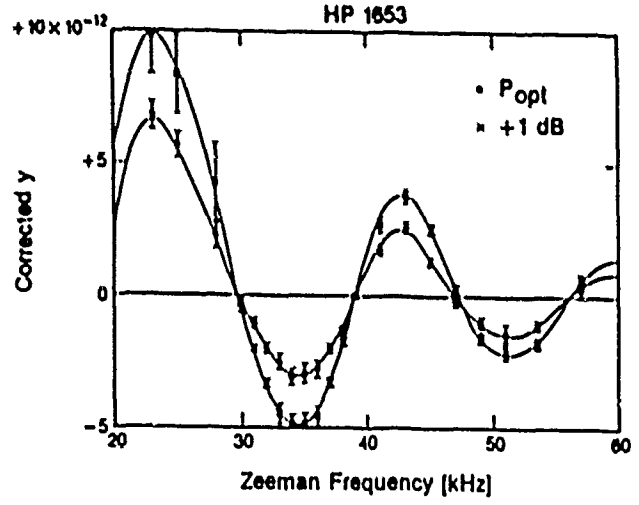


Fig.4 Same measurements as in fig.3, taken after careful tuning of the cavity.

As shown in [8], C-field values for which the power shifts due to Rabi pulling and cavity pulling compensate exist if there are solutions to the equation

$$R(f_z) = -\frac{\partial\Lambda_c/\partial p}{p\,\partial\Lambda_R/\partial p + \Lambda_R}\left(\frac{Q_c}{Q}\right)^2\frac{\delta\nu_c}{\nu} \qquad (7)$$

where $p=P/P_{opt}$. Assuming that these two effects are the only causes of power sensitivity, the latter vanishes at the C-fields identified by (7). In fig.3 $\delta\nu_c$ is such that the two curves at P_{opt} and +1dB barely touch at an extremum around 35 kHz. In fig.4 $\delta\nu_c$=0 and the two curves cross where $R(f_z)$=0.

It is interesting to point out that in fig.4 the frequency at these zero crossing points is within few parts in 10^{13} of AT1(NBS). Boy, would I like to go skiing this weekend! This confirms the Accuracy budget given in Table 1, showing that the algebraic sum of all biases other than C-field offset, Rabi pulling and cavity pulling is not substantially greater than 10^{-13}.

The results shown in fig.3 and 4 are relative to a particular standard which has been studied in depth, but results relative to all the standards which have been tested and realigned are quite similar.

4 Long term stability measurements

Long term stability measurements were taken at different settings (power sensitivity levels) with the standard in a normal laboratory environement (no active temperature stabilization). Temperature variations in the room were estimated to be about a Kelvin rms for the dayly excursions, but maybe half that much in the month range, without change of season, during the measurements.

In fig.5 the flicker floor results obtained for the standard of fig.3 and 4 are reported as a function of the measured power sensitivity. The C-field corresponding to each point is indicated by the Zeeman frequency in kHz which labels it. The point at 53 kHz is marked by an asterisk to remind that it was obtained

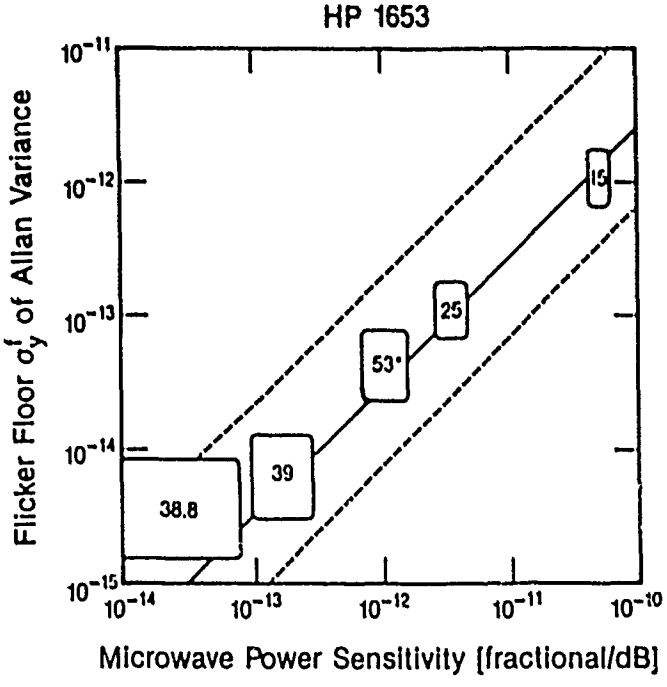


Fig.5 Measured flicker floor levels of the Allan Variance of one of the standards analyzed, as a function of microwave power sensitivity. The corresponding Zeeman frequency is indicated in kHz on each point. The dotted lines above and below the solid line are examples of where flicker floor values may be for microwave power generators more and less stable than the one of the standard under test.

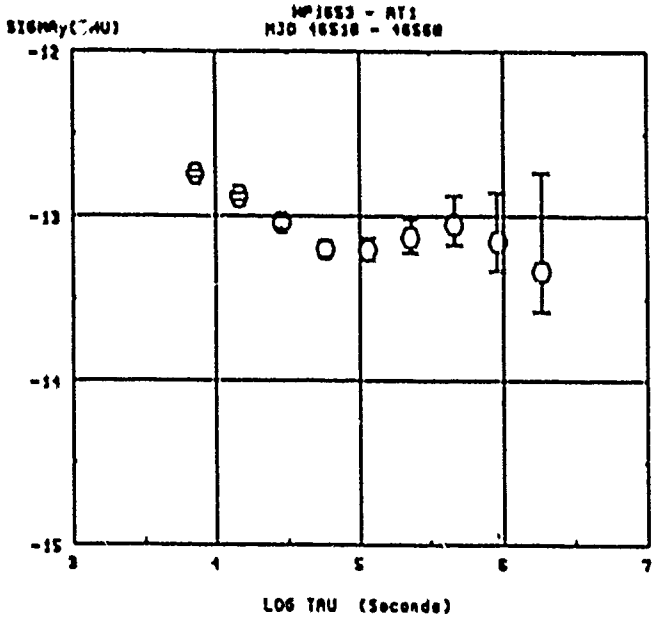
before careful tuning of the cavity, and is plotted against the corresponding power sensitivity measured prior to cavity tune-up. Within the experimental errors the stability points appear to fit nicely on a straight line, suggesting proportionality to microwave power sensitivity into the 10^{-15} range, and indicating therefore long term power variations as the main cause for long term frequency instabilities. No other processes appear to introduce excess long term instability, at least for the units under discussion, with the resolution afforded by the measurement time.

The dotted lines drawn in fig.5 above and below the experimental points are to remind that different harmonic generators, with a different flicker level and/or temperature dependence of the microwave power, may produce better or worse results for the long term frequency stability. In fact, improving the power stability by active stabilization has been proposed [1] as a mean to obtain better frequency stability, before the causes of power sensitivity were understood. Power stabilization is still a viable solution, however the present work shows that power sensitivity can be reduced by proper alignment to a point where existing harmonic generators in the free running mode don't introduce instabilities much greater than 10^{-15}.

The improvement in long term stability, with respect to its initial conditions, obtained by operating it near a null power sensitivity point is dramatically shown in fig.6a,b. In fig.6a is the stability of the standard before tune-up, at the field corresponding to f_c-53 kHz, and in fig.6b is the stability plot after tune-up, at 39 kHz, where a zero crossing of Rabi pulling happens to be (see fig.4).

The best stability obtained from any of the realigned standards is shown in fig.7. The reference for this measurement was AT1(NBS), to which the standard under test was not contributing at the time.

It may be worth pointing out that the frequency of this standard at this C-field with this cavity alignment was very close to absolute Cesium frequency.



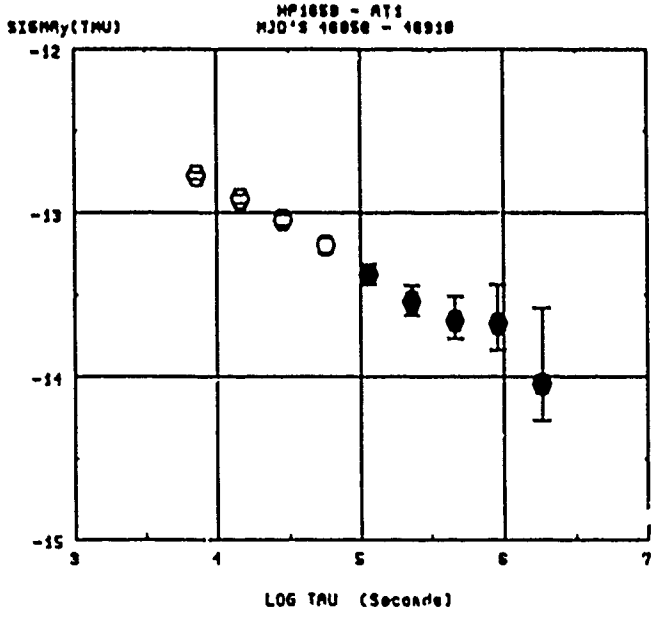


Fig.6 Measured Allan Variance at two different C-field settings for the standard of Fig.3 and 4. The curve of a) corresponds to 53kHz, before tuning the cavity, and the curve of b) to 39kHz, after cavity tuning.

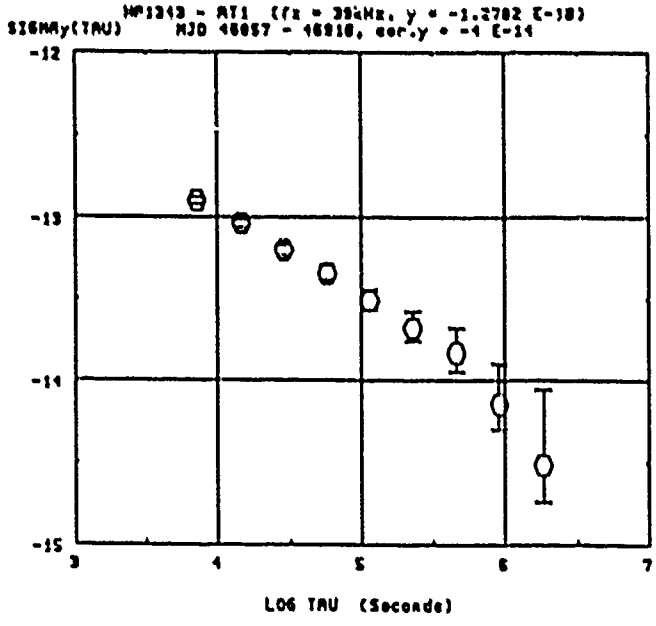


Fig.7 Measured Allan Variance for a different standard from that of Fig. 3 and 4, carefully aligned at an optimum operating point. The temperature was stabilized to better than 1 K during this measurement. The reference was the ensemble of NBS clocks, to which the standard under test was not contributing at the time.

5 Conclusions

In this paper a theoretical and experimental analysis of state of the art commercial Cesium beam frequency standards is reported, which unambiguously identifies Rabi pulling and cavity pulling as the main effects which can cause long term frequency instability.

It has been shown that a careful alignment of the C-field and of the cavity tuning can reduce to a very small level the power sensitivity of the standard and improve its long term stability into the 10^{-15} range. This shows that no other parameters contribute greater instabilities.

Previous studies indicating the stability of the microwave power as the critical parameter are confirmed by this work, although the role played by phase shifts in producing power sensitivity is shown here to be minor, at least in short tubes, contrary to the tentative conclusions of [1].

That Rabi and cavity pulling be the important effects is fortunate, because it means that considerable improvements in long term stability can be achieved by reducing power sensitivity. This would not have been possible if phase shift had been the effect responsible for it. In fact the phase shift cannot be changed once the tube is sealed.

Careful alignment of Cesium beam standards as done in this work was performed in order to demonstrate that today's theoretical understanding of such devices is adequate, and that a major improvement in long term stability can be obtained. The adoption of this kind of alignment as a routine procedure to improve standards in the field can nevertheless be advised, although the actual implementation may be somewhat cumbersome.

The drawbacks of such a solution to the long term stability problem must also be kept in mind. Here are the most relevant:

- Operation at a particular Zeeman frequency inhibits the freedom of tuning the output frequency of the standard by adjusting the C-field.
- Long term instabilities of the C-field not only change the corresponding bias, but also reintroduce power sensitivity.
- Variations with temperature and time of the cavity tuning produce a cavity pulling change, and reintroduce power sensitivity.
- Long term variations of the velocity distribution in the beam change the Rabi pulling curve and reintroduce power sensitivity.
- Whenever the tube is changed, the standard must be realigned for best performance.

As a result of the first four points it is to be expected that a long term stability in the 10^{-15} range obtained in the first year after careful alignment may subsequently slowly deteriorate back into the 10^{-14} range. However it is the author's feeling that it is always well worth applying the alignment procedure to standards which perform in the high 10^{-14} range.

Acknowledgements

The author would like to acknowledge all the scientists who have been supportive and helpful throughout this work. In particular he would like to thank D.Allan, R.Drullinger, J.Gray, L.Hollberg and F.Walls for making available their time, expertise and equipment.

A special thank goes also to the US Naval Observatory for the continuing support.

References

[1] D.W.Allan, H.Hellwig, S.Jarvis Jr., D.A.Howe, R.M.Garvey: "Some causes and cures of frequency instabilities (drift & noise) in Cesium beam frequency standards". Proc. 31st Freq.Contr.Symp., p.555 (1977).

[2] C.Audoin, M.Jardino, L.S.Cutler, R.F.Lacey: "Frequency offset due to spectral impurities in Cesium beam Frequency standards". IEEE Trans. on Instr. and Meas., IM 27, p.325 (1978).

[3] F.L.Walls: "Errors in servo systems using sinusoidal frequency (phase) modulation". IEEE Trans. UFFC, (Nov. 1987).

[4] A. DeMarchi, G.P.Bava: "On cavity phase shift in commercial Cesium beam frequency standards". Metrologia 20, p.33 (1984).

[5] A.DeMarchi, G.D.Rovera, A.Premoli: "Effects of servo loop modulation in atomic beam frequency standards employing a Ramsey cavity". IEEE Trans. UFFC (Nov. 1987).

[6] J.H.Holloway, R.J.Rorden: US Pat.# 3354307 (1967).

[7] J.H.Holloway, R.F.Lacey: "Factors which limit the accuracy of Cesium atomic beam frequency standards". Proc. 7th Int. Conf. Chronom., Lausanne, Vol.1, p.317 (1964).

[8] A.DeMarchi: "Understanding environmental sensitivity and ageing of Cesium beam frequency standards". Proc. 1st Europ.Freq.and Time Forum (1987).

[9] A.DeMarchi, G.D.Rovera, A.Premoli: "Pulling by neighbouring transitions and its effects on the performance of Cesium beam frequency standards". Metrologia 20, p.37 (1984).

AN OPTICALLY PUMPED CESIUM BEAM FREQUENCY STANDARD FOR MILITARY APPLICATIONS

T. McClelland, I. Pascaru, J. Zacharski, N.H. Tran and M. Weirs,
Frequency Electronics, Inc., Mitchel Field, NY 11553

Abstract

An optically pumped Cesium beam frequency standard is being developed, based on the Frequency Electronics 7101 Cesium Tube. The "A" and "B" magnets of the 7101 tube have each been replaced by semiconductor laser diodes. The microwave resonance is detected by monitoring the fluorescence from the "B" laser, which is servo-locked to a hyperfine line of the D2 transition of Cs.

This prototype device represents the first commercial optically pumped atomic beam frequency standard. Short and long term stability performance of this device is discussed, as well as predicted performance for the next generation device. Laser requirements are also discussed with emphasis on techniques used to utilize commercially available semiconductor lasers.

Introduction

Optically Pumped Cesium Beam Frequency Standard

In this paper we report on the development of an optically pumped cesium beam frequency standard (OPCS), in which the conventional state selection ("A"), and state detection ("B") magnets are replaced by optical state selectors and state detectors, respectively. (See Figure 1.) This concept has been investigated by a number of workers, for several years.[1-3] Theoretical predictions, as well as initial experiments, indicate that improvement of several orders of magnitude in the performance of Cs beam frequency standards is possible with such a technique.

It is our goal to develop a commercially viable frequency standard based on this optical pumping concept, which can be used in military applications. In order to meet this goal, we must go further than a laboratory demonstration. In addition, we must develop a self contained, rugged instrument, capable of operating in severe environments. Minimization of size, weight and power consumption, while maximizing reliability, are important design goals. Perhaps most importantly, we must be careful to utilize technologies which are readily adaptable to a manufacturing environment.

The performance goals for the optically pumped cesium beam frequency standard being developed are summarized in Table 1.

Table 1

Accuracy:	+/- 3 x 10^{-12} from -28°C to +65°C
Short term stability:	$\sigma_y(\tau) = 3 \times 10^{-12}/\sqrt{\tau}$
Vibration:	MIL-E-16400 MIL-STD-167-1
Shock:	MIL-S-901C MIL-E-5400 Class 1 (30g)

Program Background

The development program which we report on here had its inception in 1983 with a joint effort between Frequency Electronics, Inc., and the National Bureau of Standards, Time and Frequency Division. This joint effort started with experiments at the NBS facility in Boulder, CO, in which commercial Cs beam tubes manufactured by Frequency Electronics were optically excited. Two Cs beam tubes, designated OPCS-1, and OPCS-2 were tested at NBS.

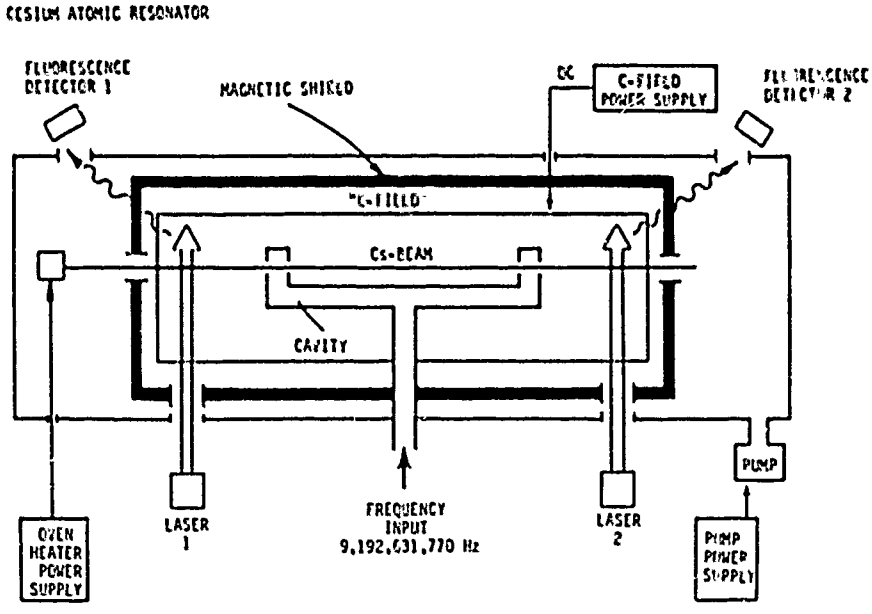


FIGURE 1

OPTICALLY PUMPED Cs BEAM RESONATOR

ATOMS ARE CONVERTED FROM ONE GROUND STATE LEVEL TO THE OTHER, SO NO ATOMS ARE LOST FROM THE BEAM.

CH2427-3/87/0000/059 $1.00 C1987 IEEE

The OPCS-1 beam tube was a commercial tube, modified by the removal of the A and B magnets, and the addition of copper gasket vacuum flanges. The flanges allowed for the attachment of windows so that optical pumping radiation could be externally provided and detected.

The OPCS-2 beam tube differed from OPCS-1, in that sapphire windows were hard sealed to the Cs tube. The hard-seal technology was considered necessary for commercial devices in which good vacuum must be provided for a number of years.

Experiments on both OPCS-1 and OPCS-2 were performed at NBS, in which the Cs beam was excited from various laser sources.[4] In particular, laser diode excitation, in which the laser linewidth was narrowed by means of external optical elements, was performed.

With such a system operating as a laboratory frequency standard, short term frequency stability (Allan variance) of $< 1 \times 10^{-11}/\sqrt{\tau}$ was achieved for $3 \leq \tau \leq 10^4$ seconds.

Design Approach

In order to satisfy the requirements discussed above, we have established several design guidelines which have been adhered to in the present development effort. These guidelines have a substantial impact on the results, but were deemed necessary in order to meet the military design objectives.

It was decided that only modified, Frequency Electronics, Inc. Cs beam tubes, Model No. 7101, would be used. These tubes are compatible in size with a 19-inch rack mount system. Furthermore, the 7101 tube has undergone extensive military environmental testing, and therefore represents a proven design.

As a second guideline, it was decided to use only commercially available semiconductor laser diodes as optical pumping and detection light sources. Other laser sources, such as dye lasers were not considered compatible with a compact, transportable system. Similarly, developmental semiconductor lasers, without any proven production history, were considered to be impractical for a manufacturable OPCS device.

Finally, it was decided to try to minimize the optical components between the laser light sources and the Cs atomic beam. In particular, any active optics which by design affect the lasing dynamics of the laser diodes, were considered undesirable. For example, external mirrors or gratings, which effectively define a laser cavity external to the laser diode chip, can result in beneficial changes in the laser linewidth, noise, and tuning range; however, this is at the expense of mechanical stability, since the position of the elements relative to the chip must be maintained to an extremely high degree of accuracy, in order to produce the beneficial results. It was decided that design performance based on such mechanisms, although relatively easy to demonstrate in the lab, would be difficult to duplicate under the required environmental stresses.

Theoretical Considerations

The theoretical operation of OPCS systems has been described in detail by other authors.[2,3,6] In this section we review only those aspects relevant to the present development effort.

Optical Pumping

We consider excitation of Cs atoms in an atomic beam using D_2 radiation to optically pump the ground state hyperfine levels. An energy level diagram showing the relevant transitions is presented in Figure 2. Although other excitations (such as D_1 radiation) are possible, it has been shown that D_2 excitation is the most promising.[5] It is assumed that whatever light source is used, it is able to resolve each of the allowed transitions between hyperfine levels.

A number of optical pumping schemes have been suggested, for both pumping and detection in OPCS systems.[2,3,6] The schemes are summarized in Table 2.

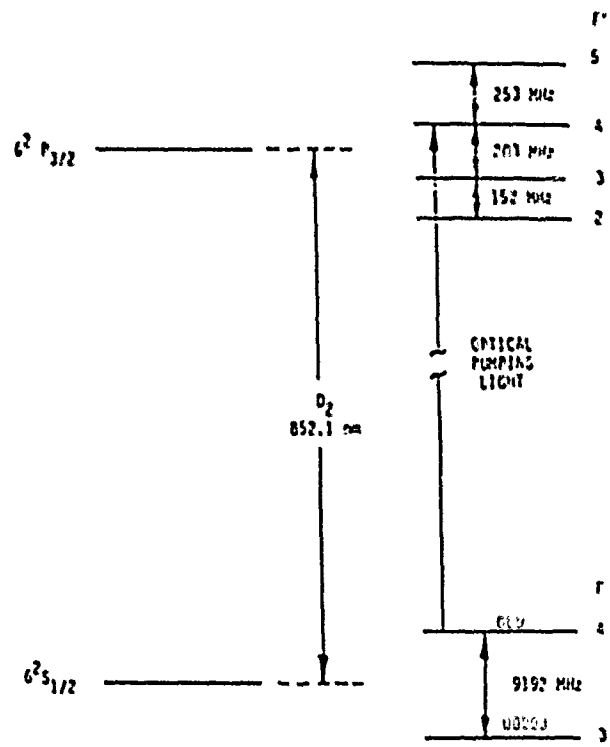


FIGURE 2

ENERGY LEVELS OF Cs^{133} ASSOCIATED WITH D_2 RESONANCE LINE

In single laser schemes, the same laser is used to both pump and probe the cesium beam. It is important to note that a single laser scheme using a laser tuned to the 4 -> 5 (cycling) transition is quite possible, even though no hyperfine optical pumping can occur for such a transition.[5] Such an arrangement works because of the strong Zeeman alignment produced by linearly polarized radiation tuned to this transition. Figure 3 shows typical Zeeman sublevel distributions for atoms excited by 4 -> 5 radiation. Substantial population differences are created between the m = 0 sublevels of the F = 3 and F = 4 hyperfine multiplets, and hence a clock transition signal is possible.

Table 2

No. of Lasers	Pump Laser Transition (F->F')	Probe Laser Transition (F->F')
1	4 -> 5*	4 -> 5
	4 -> 4	4 -> 4
	4 -> 3	4 -> 3
	3 -> 4	3 -> 4
	3 -> 3	3 -> 3
	3 -> 2	3 -> 2*
2	4 -> 4	4 -> 5*
	4 -> 3	4 -> 5*
	3 -> 3	3 -> 2*
	3 -> 4	3 -> 2*
3	3 -> 3 and 4 -> 4x	3 -> 2*
	3 -> 4 and 4 -> 4x	3 -> 2*
	4 -> 3 and 3 -> 3x	4 -> 5*
	4 -> 4 and 3 -> 3x	4 -> 5*

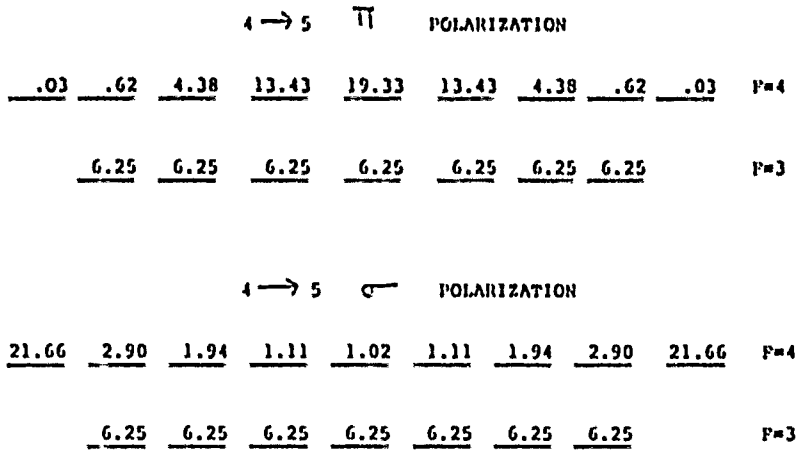
*Cycling transition

Figure 3 also suggests a potential problem, since the two orthogonal linear polarizations of exciting radiation produce population differences between the m = 0 sublevels of opposite signs. Hence a light source containing an admixture of polarizations could result in a cancellation of pumping effects, and accordingly a much degraded clock signal.

Other single laser schemes, as well as the 2 and 3 laser schemes, depend on hyperfine optical pumping to produce a population imbalance between the m = 0 sublevels of the F = 3 and F = 4 multiplets of the ground state. However, in these cases it is also necessary to consider the effects of Zeeman alignment, as is demonstrated in Figure 4, for radiation tuned to the 3 -> 3 transition. In the figure it is seen that the two orthogonal linear polarizations produce population differences of opposite sign between the m = 0 sublevels of F = 3 and F = 4. This tendency is most dramatic for radiation tuned to transitions between multiplets with the same hyperfine quantum number (i.e. 3 -> 3 or 4 -> 4); in which case x-radiation cannot excite atoms out of the m = 0 sublevel, due to selection rules. The effect of alignment is still a problem, but to a lesser extent, for other transitions.

Saturation Effects

With light power levels typical of off the shelf laser diodes, it is possible to saturate the optical pumping process. When this happens, the noise associated with the fluorescence from excited atoms can be much smaller than the noise on the exciting laser. In fact it is easy to obtain a situation in which the noise on the fluorescence is totally dominated by other sources (such as detector dark noise, or Cs beam shot noise). Such conditions are only possible for transitions which produce hyperfine optical pumping. The cycling transitions (4 -> 5, 3 -> 2) are not capable of being saturated in the same sense.



Adapted From:

Feldman, Mark "The Optically Pumped Alkali Beam/Cell and Beam Tube for Utility Atomic Frequency Standards", NBS Report, May 1983.

FIGURE 3

POPULATIONS OF Cs ATOM ZEEMAN SUBLEVELS WHEN EXCITED BY LINEARLY POLARIZED LASER RADIATION TUNED TO THE 4 --> 5 TRANSITION.

$3 \rightarrow 3$ π POLARIZATION

10.64	10.17	9.17	8.01	7.33	8.01	9.17	10.17	10.64	F=4
	0	.02	.12	16.0	.12	.02	0		F=3

$3 \rightarrow 3$ σ POLARIZATION

8.08	11.22	12.03	12.31	12.40	12.31	12.03	11.22	8.08	F=4
	.09	.03	.01	.02	.01	.03	.09		F=3

Adapted From:

Feldman, Mark, "The Optically Pumped Alkali Beam/Cell and Beam Tube for Utility Atomic Frequency Standards", NBS Report, May 1983.

FIGURE 4

POPULATIONS OF Cs ATOM ZEEMAN SUBLEVELS WHEN EXCITED BY LINEARLY POLARIZED LASER RADIATION TUNED TO THE 3 -- 3 TRANSITION.

Saturation makes possible, at least in priciple, the operation of an OPCS system in which laser noise (either amplitude or frequency) is largely irrelevant.[7] This is accomplished by choosing a one or two laser pumping scheme in which each laser excites a saturable transition. The difficulty with such a system in practice arises from the rather low light levels of the Cs beam fluorescence. Because each Cs atom in the beam produces a finite number of fluorescent photons (typically ~1 photon/atom), the total fluorescent light intensity is small, and a burden is placed on the fluorescence detector capabilities. The signal to noise ratio of such a system is not necessarily better than other systems with no saturation, even though the noise level is very low.

Laser Frequency Stability

Laser diode frequency (wavelength) is very sensitive to both the forward bias current and the junction temperature. In order to assure that the wavelength of the laser remains tuned to a single optical transition it is necessary to maintain the bias current and temperature stable to a very high degree of accuracy.

The necessary stability is most easily achieved by servo-locking the laser current and heat-sink temperature via the optical transition of interest. Typically this is accomplished by locking the laser to the fluorescence signal from the Cs beam.

However, there are some practical disadvantages to locking lasers directly to the fluorescence. First, it is only possible if the laser is very narrow in linewidth and very low noise; otherwise the fluorescence signal does not contain resolved peaks to which the laser can be locked. Second, even if the fluorescence signal contains acceptably resolved peaks at low laser light levels, it does not necessarily have resolved peaks at higher light levels necessary to saturate the optical pumping. Since saturation is desirable for one or more of the lasers used, it can be necessary to lock the laser in an alternative fashion. It has been suggested, as an alternative, to lock the lasers to appropriate peaks in the saturated absorption spectrum of a Cs gas cell.[8,9,10] However, under certain circumstances it may be possible to lock the lasers to the simple absorption spectrum of a Cs gas cell. Typically the simple absorption spectrum is broadened by pressure broadening to such an extent that the hyperfine structure of the excited state is not resolved. However, the broad absorption peak can be shifted by selecting the buffer gas pressure in the gas cell, such that the peak overlaps the desired transitions in the atomic beam. Fine tuning of the lock point can be accomplished electronically by injecting a stable electronic offset.

In a gas cell, the pressure broadened absorption linewidth is roughly 0.1 to 1.0 GHz for pressures of 10 to 100 torr of typical buffer gases (such as argon, nitrogen, and helium). Such a line is routinely

"split" by a factor of 10^4, yielding a frequency locking capability of $f \lesssim 0.1$ Mlz. The linewidth of Cs atoms in an atomic beam is typically $f \sim 10$ Mlz. Hence, it should be possible to maintain the laser in a frequency range much narrower than the atomic beam absorption linewidth by locking the laser to a Cs gas cell absorption line. Furthermore, if the laser is operated in a saturation condition, as described above, then the atomic beam linewidth is broadened, and the stability requirements are correspondingly less severe.

Experimental Results

Measurements have been made on two separate Cs atomic beam tubes: OPCS-2 and OPCS-3. In both cases a number of different laser pumping sheemes have been investigated.

Results with OPCS-2

OPCS-2 is a self contained Cs beam tube, with an 8 cm. long Ramsey cavity, and two sets of 3 sapphire windows; one set at each end of the tube. Two windows at each end allow laser light to enter and exit along an axis perpendicular to both the Cs beam and a constant magnetic field (C-field) which encompasses the entire Ramsey cavity. The third window at each end of the tube is located near the focal point of an ellipsoidal mirror, whose axis of revolution is the C-field axis. A silicon photodiode detector, with a 100 mm^2 active area is placed just outside this third window, in order to detect the fluorescence from the atomic beam.

The lasers used to excite this tube have been off the shelf laser diodes manufactured by Hitachi. Most of the measurements were made using the Hitachi HLP-1400 device, a non-sealed device, with an output level of ~10 milliwatts at the D_2 wavelength, and a single mode linewidth of ~50 Mlz. Some measurements have also been made using the Hitachi HL8312A device, which is identically packaged, but capable of ~15 mW at the D_2 wavelength. The laser intensity at the atomic beam was controlled for each laser by placement of neutral density filters in the light path.

With this tube, a number of different laser configurations have been tested. We have found that the noise on the fluorescence in the probe region decreases when a laser causing transitions from the same ground state hyperfine level as the probe laser is incident in the pump region. Similarly, the probe noise increases when the pump laser is tuned to the opposite ground state hyperfine level. These observations are reasonable, since in the former case the steady state probe fluorescence is decreased by the pump laser presence, whereas in the latter case the probe fluorescence is increased by the pump laser. Accordingly, the best S/N ratio is expected to occur when the lasers in both regions excite atoms from the same hyperfine level.

We also observe empirically that saturation of the optical pumping process in the pump region produces a decrease in the noise on the fluorescence signal in the probe region. This observation also seems in line with expectations.

However, when the probe laser is tuned to a saturable transition, we are unable to detect any decrease in noise level on the fluorescence signal as the probe power is increased. In this case the noise floor is independent of laser power, but always lower than for a cycling transition. The observed noise floor is consistent with the predicted detector noise. Hence, we conclude that the very weak fluorescence signal (~50 times weaker than for the 4 -> 5 cycling transition) contains noise which is overwhelmed by the dark noise of the detector.

Based on the measured DC fluorescence in this case, and assuming that the noise associated with the saturated fluorescence is shot noise, we estimate that approximately 30 times more fluorescence must be detected in order for the shot noise to be equal to the detector noise. Conversely, a decrease in detector noise by a factor of 30 would produce the same result. Although improvements in fluorescence collection optics are possible, careful analysis indicates that a 30 times improvement is very unlikely for a commercial instrument. Similarly, it seems unlikely that a 30 times improvement in detector noise could be achieved without resorting to drastic cooling or other techniques inconsistent with a practical field instrument.

The best performance with the OPCS-2 tube has been obtained for a two laser configuration in which the pump laser is tuned to the 4 -> 4 (σ) transition and the probe laser is tuned to the 4 -> 5 (σ) cycling transition.

The S/N ratio for the clock signal (detected with a phase sensitive detector at 84 Hz) with this configuration is shown in Figure 5. The linewidth of the central Ramsey fringe was measured in this case to be 1200 Hz. The short term stability predicted from a frequency standard operating with this performance is known to be given by[11]

$$\sigma_y(\tau) = \frac{0.2}{Q(S/N)} \frac{1}{\sqrt{\tau}} \qquad (1.)$$

$$= 5 \times 10^{-11}/\sqrt{\tau}$$

We have also operated OPCS-2 in a configuration employing 2 pump lasers, in an attempt to optically pump virtually all atoms into a single m = 0 sublevel. However, the results of this effort have been disappointing compared to the predicted improvement of 6 to 8 in S/N ratio.[5] We believe this to be due to the totally ambiguous nature of the polarization of the light transmitted through the sapphire windows of this tube. This technique depends critically on one laser being tuned to a 3 -> 3 or 4 -> 4 (π) transition; while the polarization of the laser transmitted through the sapphire windows could neither be measured nor controlled due to birefringence of the window.

The OPCS-2 tube was constructed with awareness of this limitation because of the immediate availability of hard seal technology for sapphire windows, and accordingly the lack of expertise with hard-seal techniques using other, non-birefringent optical windows.

Results with OPCS-3

OPCS-3 is very similar to OPCS-2; except that the Ramsey cavity is 12 cm. long instead of 8 cm., and the windows, which are hard sealed to the tube housing, are all made from optical quality fused silica. In other respects, the two tubes are virtually identical.

S/N RATIO FOR OPCS-2

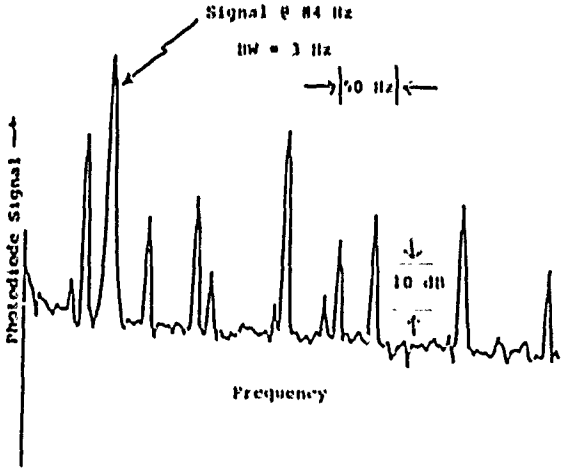


FIGURE 5

CLOCK SIGNAL FOR OPCS-2. PUMP LASER IS TUNED TO 4 --- 4 TRANSITION. PROBE LASER IS TUNED TO 4 --- 5 TRANSITION. S/N RATIO IS 45 dB/ Hz.

Results with OPCS-2, as described above, had demonstrated the importance of polarization-preserving windows, and hence considerable design effort was devoted to the hard sealed fused silica windows. Producing a vacuum tight glass-to-metal seal proved to be straightforward; however, doing so without inducing severe stresses in the glass, which in turn induced birefringence, proved to be a greater challenge.

A workable solution was eventually found, but in order to save time the windows which were installed in OPCS-3 were not anti-reflection coated on the atomic beam side as had been the case for OPCS-2. We did not expect this to be an overwhelming problem, as results with OPCS-2 had shown that scattered light was never a dominant noise source. In fact, detector noise proved a limiting factor in OPCS-2 rather than scattered light. Rough estimates indicated that a factor of two increase in scattered light in OPCS-3 compared to OPCS-2 would still result in a situation in which detector noise dominated. Coated optics, and some other design improvements were postponed to the next generation: OPCS-4.

At the time of this writing, only preliminary measurements have been made on OPCS-3. Although as expected the scattered light intensity is greater than with OPCS-2, it does not appear to dominate the noise, even when tuned to a pumping transition with a relatively low level of fluorescence.

Because the windows in OPCS-3 are polarization preserving, this tube represents the first meaningful test bed for two laser pumping techniques. We anticipate an improvement of ~6 in the S/N ratio of the clock signal due to this effect. (An 8 times improvement is predicted if one assumes that all atoms can be pumped into a single Zeeman sublevel. However, optical alignment, as discussed above, decreases that actual performance slightly.[5])

An overall improvement of ~9 in Allan Variance is expected from OPCS-3, since the increased microwave interaction length should reduce the linewidth of OPCS-3 compared to OPCS-2 by the ratio of interaction lengths (2:3).

Future Measurements

OPCS-3

As indicated above, we expect to achieve a 9 times improvement in Allan Variance with OPCS-3 compared to OPCS-2, or

$$\sigma_y(\tau) \sim 5 \times 10^{-12}/\sqrt{\tau}\,.$$

Of course this value is based on a calculation using the S/N ratio and linewidth, and hence does not include any potential drift effects due to the lasers. Therefore it is very important to perform actual measurements of the system operating as a frequency standard for extended periods of time. Such measurements are presently planned for OPCS-3.

OPCS-4

Although it is somewhat premature, the next generation tube is already on the drawing board. Modifications based on the results with OPCS-3 will be implemented. In addition, the diamond-turned aluminum mirrors used in OPCS-2 and OPCS-3 will be replaced by gold-overcoated Al mirrors, which should enhance reflectivity by at least 10%.

Summary

Development of a commercial optically pumped Cs beam frequency standard is in progress. Performance has already been achieved which corresponds to a short term stability of

$$\sigma_y(\tau) \sim 5 \times 10^{-11}/\sqrt{\tau}\,.$$

Improvements which are in proces suggest that better than an order of magnitude improvement is possible to

$$\sigma_y(\tau) \sim 3 \times 10^{-12}/\sqrt{\tau}\,.$$

This performance is possible using a compact Cs beam tube which has undergone extensive environmental, as well as performance testing in its non-optical form. Furthermore, this performance can be achieved with present, commercial, off the shelf laser diode technology. It does not require the implementation of "active" optics which might put severe constraints on the ruggedness of any instrument.

It must be emphasized, however, that such performance - although superior to competing magnetic state selection devices, would be limited by laser diode noise which is orders of magnitude above the Cs beam shot noise limit. Accordingly, improvements in laser noise hold the promise of even much better performance than predicted above.

Acknowledgements

We would like to thank R. Drullinger and L. Hollberg of the Time and Frequency Division of NBS for their countless discussions, and thoughtful suggestions.

This work is supported by the Electronics Systems Division of the Air Force Systems Command, USAF, under Contract No. F19628-85-C-0192.

References

1. G. Singh, P. Dilavore, and C.O. Alley, IEEE J. Quant. Elec. QE-7, 196 (1971).

2. M. Arditi and J.-L. Picque, J. Phys. (Paris), 41, L-379 (1980).

3. L. Lewis and M. Feldman, 35th Annual SFC, 612 (1981).

4. A. Derbyshire, et. al., 39th Annual SFC, 18 (1985).

5. M. Feldman, NBS Report to USAF, RADC, "The Optically Pumped Alkali Beam/Cell and Beam Tube for Utility Atomic Frequency Standards," (1983).

6. H.J. Gerritsen and G. Nienhuis, Appl. Phys. Lett. 26, 347 (1975).

7. V. Giordano, et. al., 1st European Freq. and Time Forum, Besançon, France, March, 1987.

8. H. Tsuchida, et. al., Japan J. Appl. Phys., 21, 561, (1982).

9. H. Hori, et. al., IEEE J. of Quantum Electron., QE-19, 169, (1983).

10. S. Ohshima, et. al., IEEE J. of Quantum Electron., QE-23, 473, (1987).

11. J. Vanier, and L.G. Bernier, IEEE Trans. Instr. and Meas., IM-30, 277 (1981).

41st Annual Frequency Control Symposium - 1987

DEVELOPMENT OF A RUBIDIUM FREQUENCY STANDARD FOR THE MILSTAR SATELLITE SYSTEM

T. McClelland, I. Pascaru and Marvin Meirs
Frequency Electronics, Inc.
Mitchel Field, NY 11553

Abstract

A radiation hardened Rubidium Frequency Standard (RFS) has been developed for the MILSTAR satellite system. Each satellite contains four Rb standards: 1 active, and 3 redundant units. The four RFS units must provide 3-year shelf life plus 10-year operation with a probability of success of 0.9996. This very severe reliability requirement is met by incorporating several key features in the new design.

Highest component reliability is obtained by using space qualified hybrid circuitry for over 80% of the electronics, and survivability is assured through a design which minimizes effects of radiation sensitive parameters such as offset voltages, and incorporates radiation hardened components where necessary.

Lamp and resonance cell reliability are assured through quantitative analysis of Rb metal content in each cell using differential scanning calorimetry. In addition, a long term aging program has been established to measure Rb consumption rates in cells, as well as random failure rates.

This program represents the first effort to determine quantitative reliability figures for Rb vacuum cells.

Introduction

A passive rubidium frequency standard (RFS) is being developed for the satellite constellation of the MILSTAR communication system. This satellite communication system is designed to have a 10-year operational life in space, and is to be hardened against well defined nuclear radiation (anti-satellite) threats.

Reliability

The reliability of the RFS subsystem on each satellite is required to be sufficient to insure a successful operational mission life of 10 years, with a probability of success, P_S, of 0.9996. This reliability must be achievable after the subsystem has been stored/tested in a ground environment for 3 years. To satisfy this requirement with a single RFS unit would require a mean-time-between failures (MTBF) of 58,000 years. We decided to develop a subsystem consisting of four RFS units; one operational and three standby redundant units. Figure 1 shows an outline drawing of this RFS subsystem.

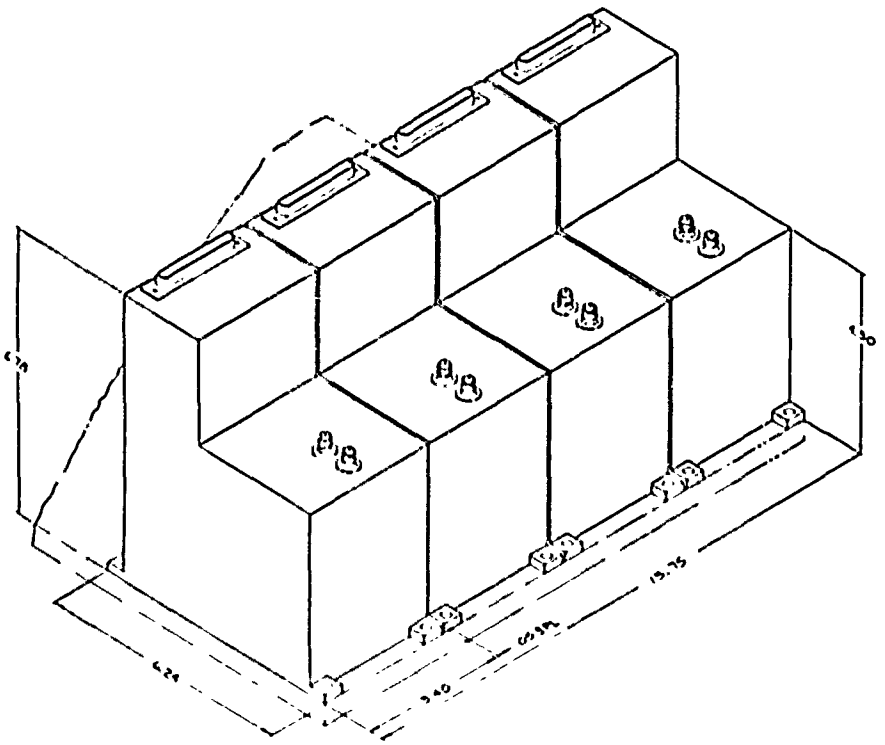


FIGURE 1

OUTLINE DRAWING SHOWING A SATELLITE SYSTEM OF FOUR RFS UNITS.

Nuclear Hardening

The RFS subsystem must survive anti-satellite nuclear threats. In principle the RFS is well suited to this requirement, since the hyperfine frequency of Rb^{87} which is detected in such a device is insensitive to nuclear radiation. However, care must be taken in the design to insure that other sensitivities are not induced by the radiation.

Design Approach

A simplified block diagram of a MILSTAR RFS unit is shown in Figure 2. The 5 MHz output of a voltage controlled crystal oscillator is internally multiplied and mixed with a synthesized frequency, also derived from the 5 MHz output. The resulting frequency (≈ 6.84 GHz) is used to excite a microwave cavity in which is located a gas cell containing Rb^{87}. Light from a Rb^{87} RF discharge lamp is passed through this cell and the transmitted light is monitored with a photodiode. A transmitted light signal is generated by frequency modulating the 6.84 GHz cavity excitation at a signal frequency of 450 Hz. This signal is detected in a phase sensitive detector (PSD), the output of which describes (as a function of the exact 5 MHz frequency) a dispersion curve. The output of the PSD is integrated and electronically fed back to the voltage control input of the VCXO, effectively locking the crystal oscillator to the Rb^{87} hyperfine frequency.

The crystal oscillator is being designed and built at Frequency Electronics. In order to stay within the lock range of the Rb system this crystal oscillator must employ a state of the art premium Q swept SC cut quartz crystal, with extremely low aging characteristics (less than 1×10^{-11}/day).

In order to meet the reliability requirements, Frequency Electronics has undertaken an RFS development effort focused on two fronts. The electronics has been designed so that most of the major circuitry can be packaged into hybrids which allow for greatly reduced size and weight, as well as increased reliability. (See Figure 3.) Table 1 shows the subassembly breakdown of a single RFS, with the corresponding hybrid circuit usage. This compact packaging also opens up space for the additional circuitry necessary for radiation hardening.

The second front consists of the development of Rb cells capable of lasting at least the required 10 year life. Cell life has been demonstrated to be the limiting failure mechanism for RFS units flown on Navstar GPS Satellites.[1] Hence we have devoted considerable effort to quantifying the aging parameters of cells, and in particular the RF discharge lamp cells, used in the RFS.

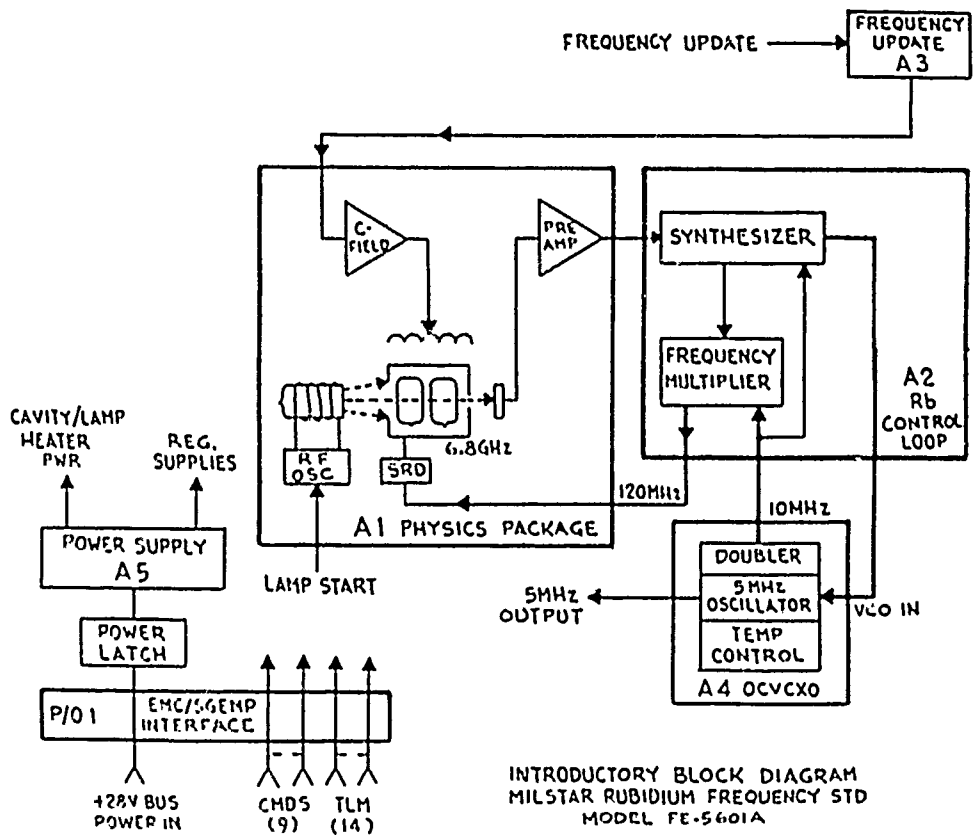


FIGURE 2

SIMPLIFIED BLOCK DIAGRAM OF RFS SYSTEM

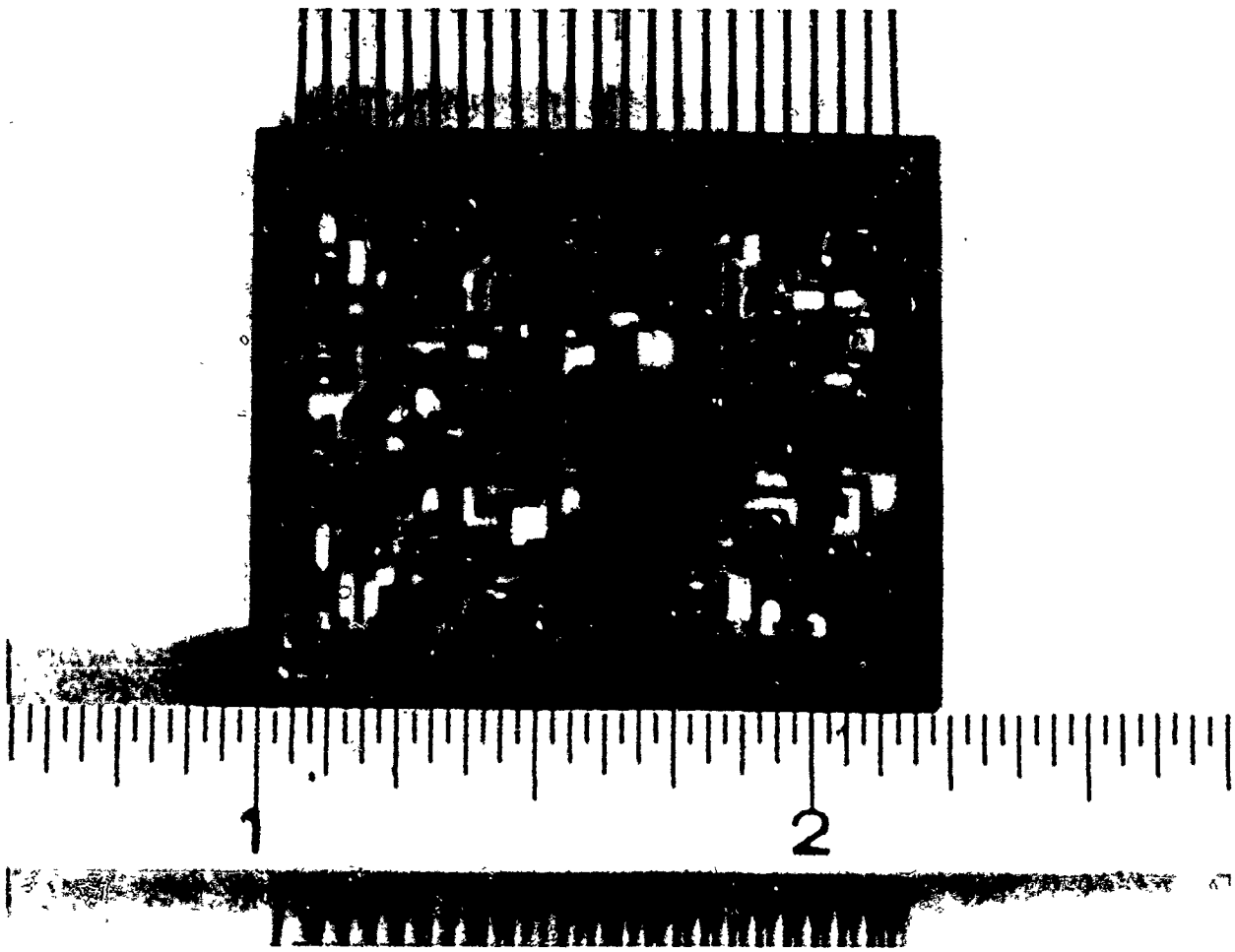


FIGURE 3

TYPICAL HYBRID CIRCUIT PACKAGE. THIS PARTICULAR HYBRID CONTAINS THE PRE-AMPLIFIER AND C-FIELD CIRCUITRY, AND IS LOCATED INSIDE THE PHYSICS PACKAGE.

Design Details

Radiation Hardening

In order to insure survival of the RFS from anti-satellite nuclear events, the conventional design concepts must be significantly modified. Even though Rb is inherently radiation insensitive, care must be taken to assure that the thermal, optical, and electrical operating conditions do not change significantly in response to the radiation. Although it is beyond the scope of this report to detail all of the corresponding design modifications, we review the important trouble spots in the following paragraphs.

Radiation Hardening in the Physics Package

The physics package contains the Rb cells and the optical detection system. (See Figures 4, 5.) Care must be taken to insure that optical elements (lenses, windows) are not adversely darkened by radiation. This could cause radiation induced frequency shifts, because of the light shift effect.[2] Optical materials have been chosen which are insensitive to radiation; such as aluminosilicate glasses and sapphire.

Also in the physics package, steps have been taken to insure that the oven control temperatures are not affected by radiation. This requires the selection of control thermistors whose resistance is insensitive to radiation.

The constant magnetic field (C-field) which is applied to the Rb atoms in the resonance cell in order to tune the lock frequency of the RFS is referenced to an oven stabilized zener diode. The zener diode has been specially fabricated and tested to insure insensitivity to radiation. Ultimately, the C-field zener represents the primary limiting to factor radiation performance.

TABLE 1

HYBRID USAGE - MILSTAR RFS

DESIGNATION	SUBASSEMBLY	HYBRID QTY	HYBRID TYPE
A1	Physics Package	2	2
A2	Rubidium Electronics	2	2
A3	Frequency Update	5	2
A4	OCVCXO	5	3
A5	Power Supply	1	1
A6	EMC/SGEMP Interface	0	0
	TOTALS	15	10

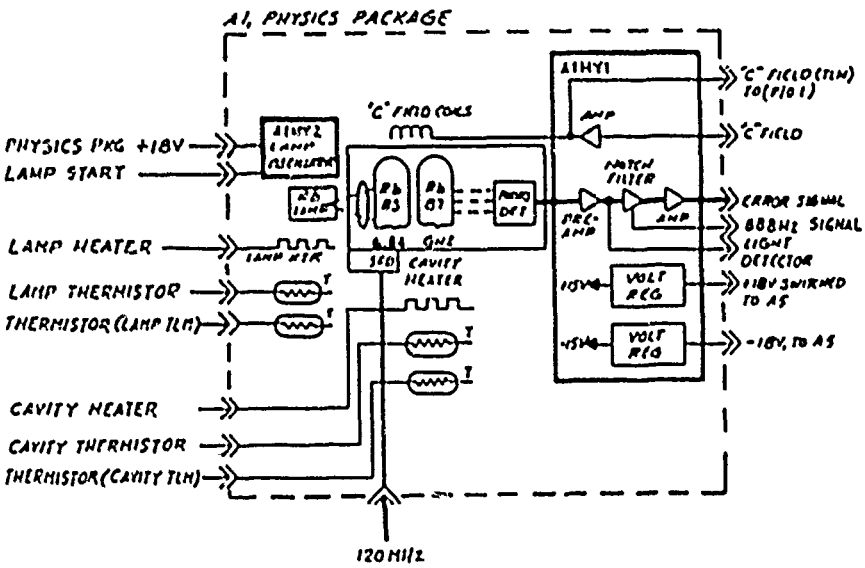


FIGURE 4

FUNCTIONAL BLOCK DIAGRAM OF THE RFS PHYSICS PACKAGE.

Finally, the photo detector, a silicon photo diode used to detect the Rb signal, must be carefully selected. Neutron radiation tends to increase the leakage current of such devices, which can increase the noise superimposed on the signal unless the detector noise is sufficiently small in the pre-irradiated state. A photo diode with sufficiently high shunt impedance, before and after radiation, has been selected.

Radiation Hardening of Servo Electronics

Several potential problems in the servo-electronics must be protected against. The performance limiting problem turns out to be the permanent offset voltage error induced on the integrator input. This appears as a shift in the Rb frequency. This problem has been minimized by choosing components with extremely small quiescent and radiation induced offsets.

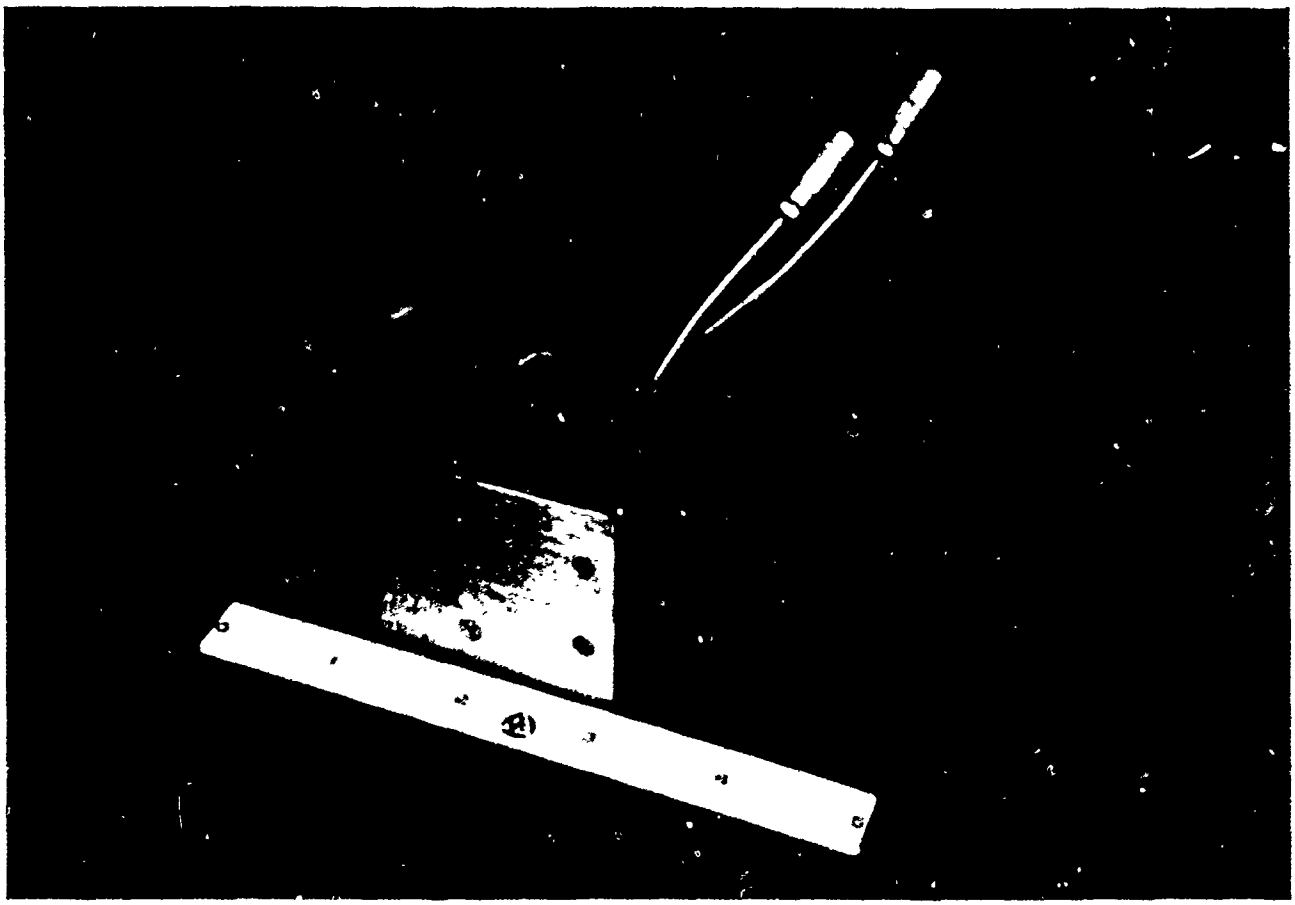


FIGURE 5

RFS PHYSICS PACKAGE, WITH COVER REMOVED.

Although lesser problems, changes in the gain and phase shift of the active elements of the servo loop must be considered. The choice of parameters must be made to minimize these effects.

Also of concern is the possibility of loss of lock during irradiation due to rapid loss of charge on the integrator capacitor. This effect has been minimized with proper choice of series resistors which act to increase the RC time constant of the effective discharge circuit during irradiation.

Rb Cell Reliability

Past studies[1] have shown that Rb metal diffuses over time into the walls of its container until eventually no free Rb is left to form a vapor. When this happens in one of the Rb cells in the RFS, the device is no longer capable of functioning properly. Fortunately, the diffusion process is very slow in the resonance and filter cells of the RFS, and does not present a major problem. However, in the RF discharge lamp, the presence of the discharge greatly enhances the diffusion process, thereby shortening the lamp lifetime. Complete diffusion of free Rb metal into the lamp cell walls has been shown to be the cause of lamp failure in RFS devices launched as part of the GPS satellite program.[1]

In order to insure adequate lifetime of Rb lamps in the MILSTAR program, Frequency Electronics has initiated an extensive measurement program. Twelve lamps are continuously monitored as part of a program to measure Rb diffusion as a function of time. (See Figure 6.) At periodic intervals, the free Rb content of each of the bulbs is measured by differential scanning calorimetry (DSC).[3] (See Figure 7.) Based on these measurements, extending over a two year period of time, an adequate characterization of the diffusion process is expected. This should allow for the definition of a reliable minimum Rb fill requirement in flight lamps. The measurement program is necessary because a gross overfill of Rb in the lamp has been shown to result in erratic behavior, which generally is sensitive to shock and vibration (due to metallic Rb moving inside the lamp). Hence a minimum fill requirement is established from the DSC measurements, and a maximum fill is established based on stability requirements.

At the time of this writing 7 months of diffusion data have been taken. A typical plot, showing a fit of the free Rb mass to a $\sqrt{t}$ diffusion expression, is shown in Figure 8. An extrapolation of this fitted curve inidcates that about 120 micrograms of Rb would diffuse into the lamp surface during a 10 year period of time. Accordingly, tests show that up to 1,000 micrograms of Rb can be placed inside the lamp without

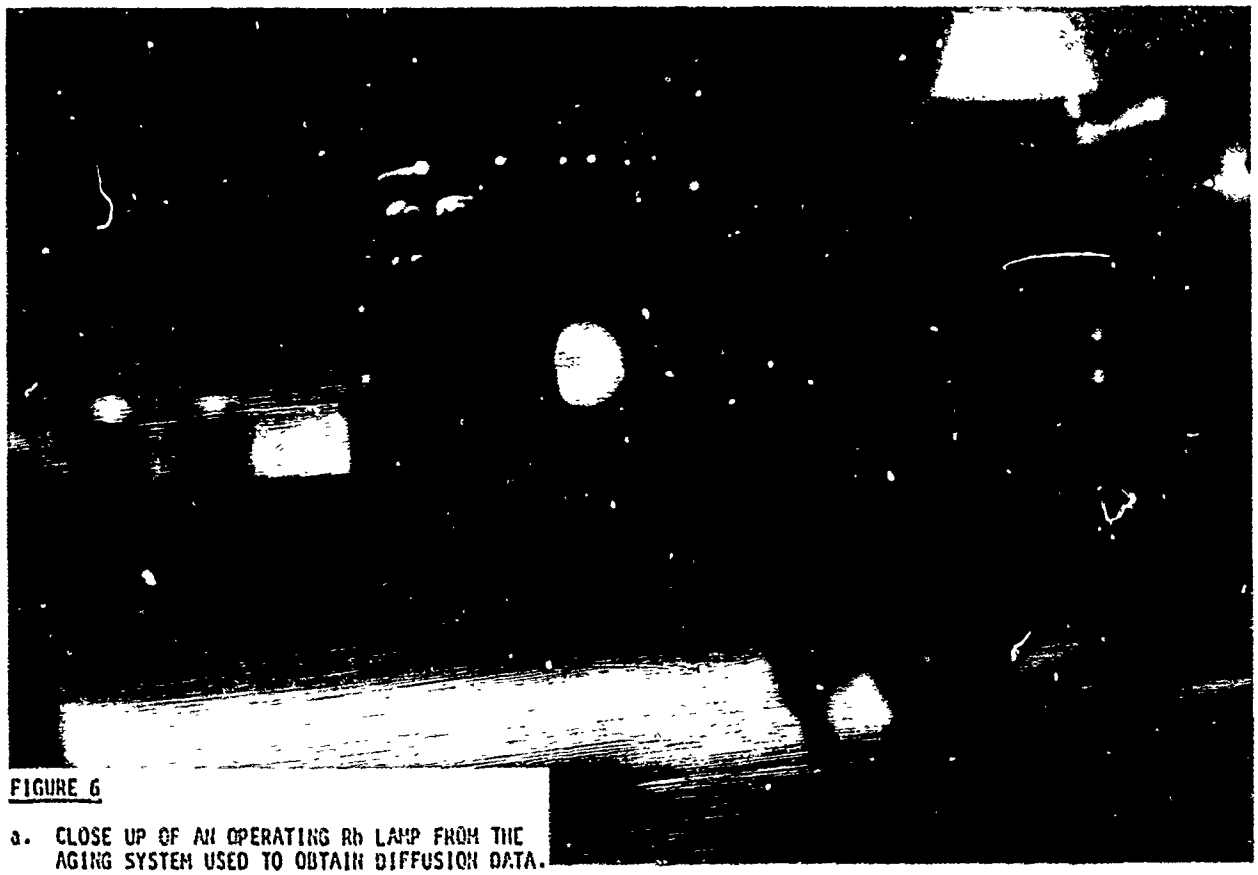


FIGURE 6

a. CLOSE UP OF AN OPERATING Rb LAMP FROM THE AGING SYSTEM USED TO OBTAIN DIFFUSION DATA.

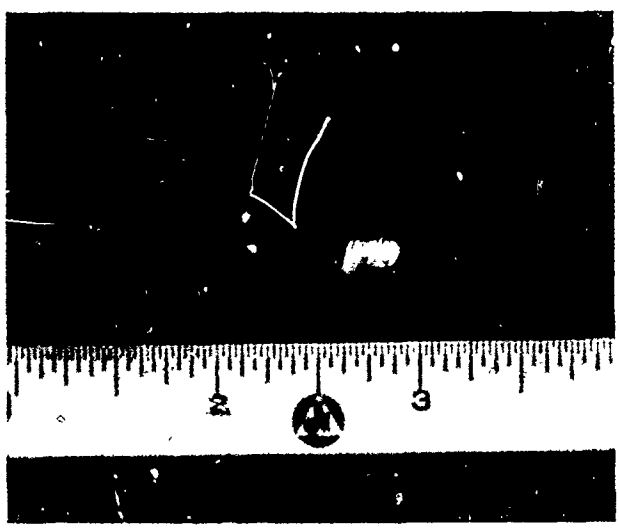
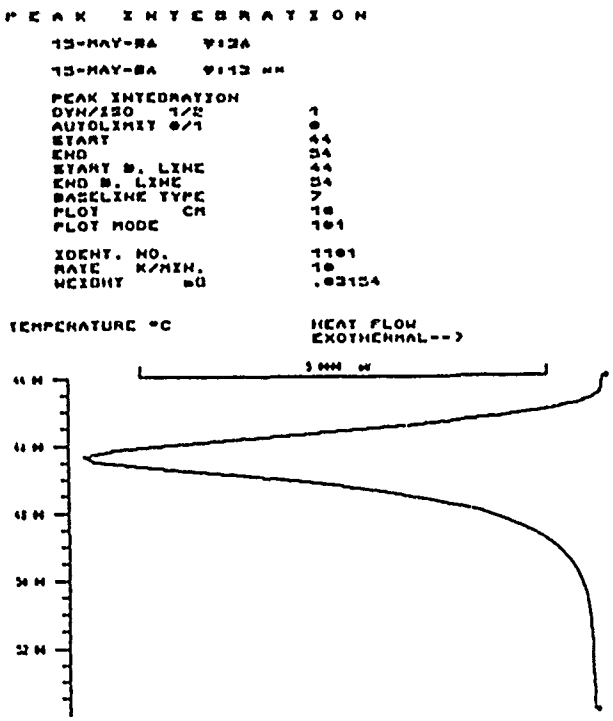


FIGURE 6

b. PHOTOGRAPH OF A Rb LAMP CELL.



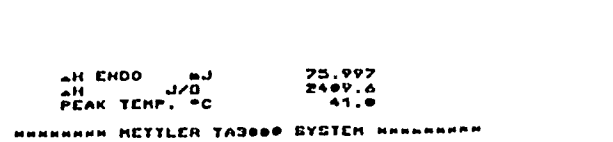


FIGURE 7

TYPICAL DSC SCAN SHOWING THE PEAK CAUSED BY THE MELTING OF Rb.

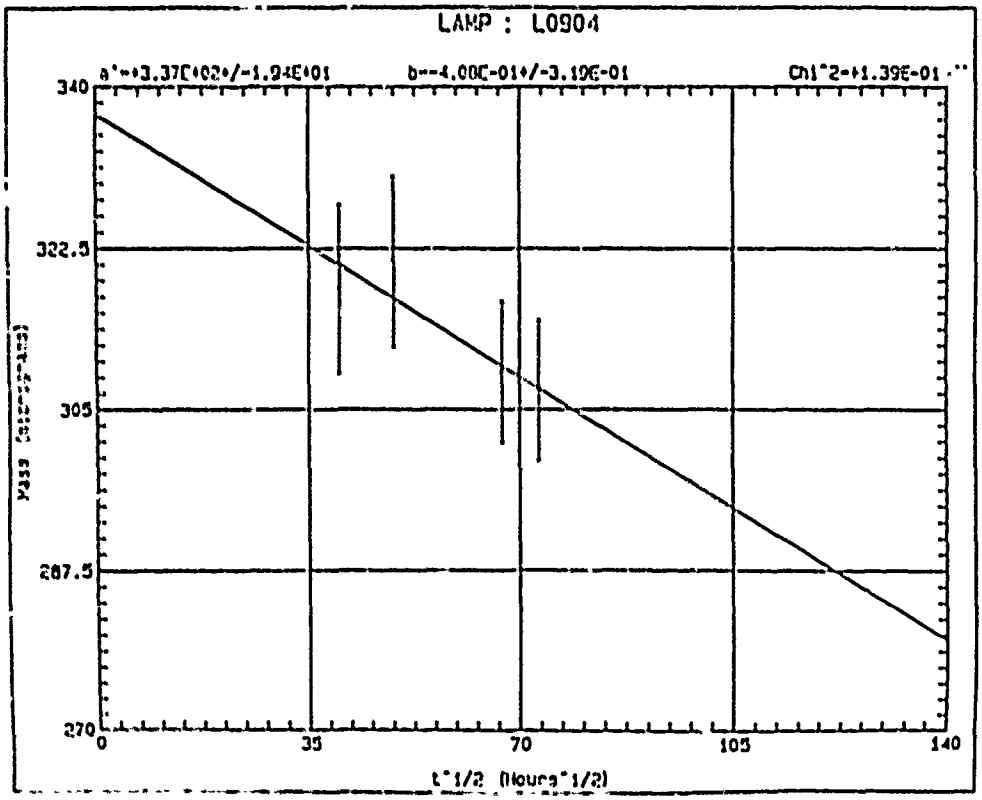


FIGURE 8

MEASURED MASS OF FREE Rb IN A LAMP CELL VS. TIME. THE LINE DRAWN THROUGH THE DATA IS A LEAST SQUARES FIT.

inducing instabilities. Hence, it appears that lamp Rb fills capable of insuring stable operation over the 10 year operational lifetime of the satellite system can be easily provided, with plenty of safety margin.

In addition to failure caused by Rb diffusion, lamps can conceivably fail due to other causes. For instance, over time cracks could develop in the glass which eventually cause a leak, with rapid loss of Rb from the cell then taking place. All such unpredictable failure mechanisms are lumped together as random failures.

It is believed that random failures of Rb lamps and Rb cells are extremely rare for properly manufactured parts. However, it is impossible to rule out such failures for most previously run RFS systems, since only very incomplete data has existed regarding the initial conditions of the various cells. Some of the GPS failures, for instance, could be due to random failures; even though the evidence suggests that Rb diffusion caused the failures.

A test fixture containing 20 carefully controlled and monitored Rb lamps has been operated since January of 1987, in an effort to establish unambiguous random failure data. Table 2 shows the upper bound random failure rate numbers which correspond to various failure free operating times. It is projected that between 60 and 80 bulb years of data will exist before the first units are flown in the MILSTAR program.

Development Tests

A breadboard system incorporating all the major design innovations has been tested, in addition to more detailed testing of all subcircuitry modules. Typical Allan variance data for the breadboard system is presented in Figure 9, and a $\bar{y}$ vs. time plot is presented in Figure 10.

Proof of design hybrids have already been built and tested for nearly 50% of the hybrids used in the RFS. (See Figure 3.) The rest of the hybrid circuits are in process and will be fully tested by the end of 1987.

Radiation testing has been performed on performance limiting components, such as the photo diode, integrator operational amplifier, and C-field zener diode. Based on the results, we are confident that the present design will achieve its goals.

Summary

A space qualified, radiation hardened rubidium frequency standard has been developed for the MILSTAR satellite communication system. Special attention has been devoted to reliability in the design of this RFS, in order to assure a ten year lifetime with very close to 100% confidence.

TABLE 2

LAMP RANDOM FAILURE RATE VS. TEST TIME

THIS TABLE TAKES INTO ACCOUNT 410,000 HRS. OF FAILURE FREE OPERATION DOCUMENTED FROM PREVIOUS RFS UNITS. FAILURE RATES ARE GIVEN IN FAILURES PER MILLION HOURS OF OPERATION.

TEST TIME (BULB YEARS)	(.9 CONFIDENCE) (0 FAILURES)	(.5 CONFIDENCE) (0 FAILURES)	(.9 CONFIDENCE) (1 FAILURES)
0	5.7	1.7	9.7
10	4.7	1.4	8.0
20	4.0	1.2	6.8
30	3.5	1.0	5.7
40	3.1	.92	5.2
50	2.7	.83	4.6
60	2.5	.75	4.2
70	2.3	.68	3.8
80	2.1	.63	3.5

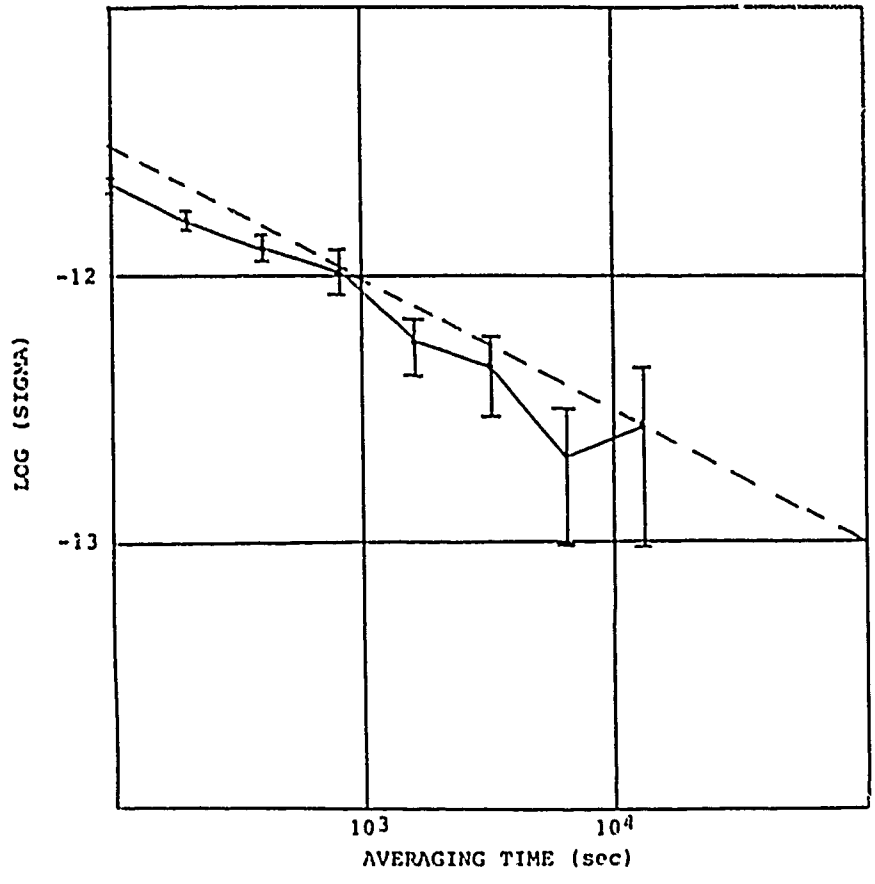


FIGURE 9

ALLAN VARIANCE OF THE BREADBOARD RFS.

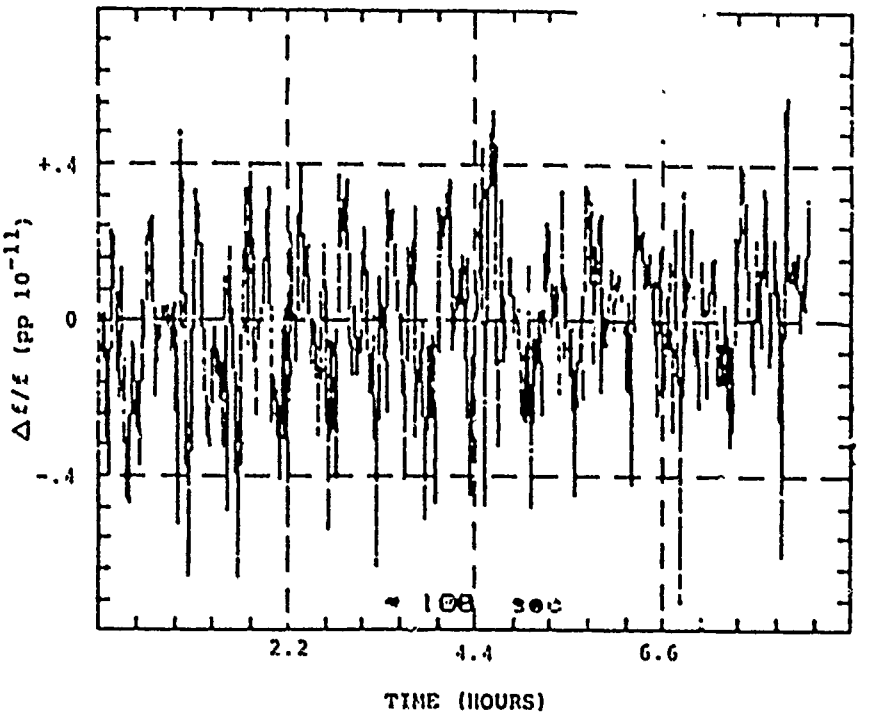


FIGURE 10

FRACTIONAL FREQUENCY ERROR (y) OF THE BREADBOARD RFS.

An extensive program to measure failures in Rb lamps has been initiated, which should result in the first dependable random failure rate statistics for such components. Furthermore, it has been demonstrated that lamps can be easily manufactured to meet the 10 year life requirement without danger of premature wear out due to Rb diffusion.

Breadboard and sub-circuit level testing have been sucessfully performed on the RFS design. An engineering model RFS, which is to be a form, fit, and function model of the flight unit, is being assembled at present, and will undergo testing before the end of 1987. Flight units are expected in early 1989.

References

1. C. Volk, et. al. "Lifetime and Reliability of Rubidium Discharge Lamps For Use in Atomic Frequency Standards" Aerospace Corporation, internal report, 1983.

2. W. Happer, and B.S. Mathur, Phys. Rev. 163, 12 (1967).

3. The use of differential scanning calorimetry techniques to measure the Rb mass non-destructively in vacuum cells was first suggested by T. Lynch at EG and G, Electronic Components Division. He also made the first measurements of this type.

41st Annual Frequency Control Symposium - 1987

HYDROGEN MASERS FOR RADIO ASTRONOMY

H. Peters, B. Owings and T. Oakley

Sigma Tau Standards Corporation
Tuscaloosa, AL

and

L. Reno

National Radio Astronomy Observatory
Socorro, NM

Summary

Ten atomic hydrogen masers are being provided by Sigma Tau Standards Corporation for the Very Large Baseline Array (VLBA), a system of ten radio telescopes under construction for deployment in locations stretching from Hawaii to the Virgin Islands. Radio astronomical observations using the method of Very Long Baseline Interferometery (VLBI) are among the most demanding applications of frequency standards, requiring the best stability and spectral purity of signals for the telescope local oscillators to maintain phase coherence at microwave frequencies over intercontinental distances.

Features which improve the stability, operating life, reliability and field utility of the masers include automatically tuned maser cavities using a technique which does not require a separate reference standard, remotely programable phase coherent receiver synthesizers giving fractional frequency resolution of 5 parts in 10 to the 17th, field replaceable vacuum pumps with extremely long life, 32 multiplexed channels of remotely addressable analog instrumental data and redundant AC and DC power supplies with battery back up. The maser is contained in one compact package measuring 46 cm (18 in) wide by 76 cm (30 in) deep by 107 cm (42 in) high, uses less than 75 watts of internal DC power and weighs 238 kg (525 lb), including batteries for 16 hours (max) of operation.

VLBA Requirements

Following is a brief summary of the specifications and performance goals which the VLBA masers must satisfy.

Frequency Stability. In radio interferometry the signals received from distant extraterrestrial sources at each remotely located antenna are heterodyned with individual stable local oscillators and the IF frequencies are recorded magnetically for later cross correlation by computer. The most important requirement for successful results is that the frequency standards driving the local oscillators contribute negligibly to the phase noise over the period of the observations. The wide band phase noise for the VLBA masers is specified to be less than .6 picoseconds rms for the maser 100 MHz outputs and the frequency stability goal is 7×10^{-14} for one second measuring intervals, decreasing as the square root of time. For longer term measuring intervals of 1,000 to 10,000 seconds the stability goal is 2×10^{-15}.

Other Requirements. The VLBA masers are to be used at isolated installations with a minimum of operational attention; all normal operations shall be automatic or be controlled remotely through a Monitor and Control (M&C) interface system. The maser standards must have two 5 MHz outputs with 90 db isolation and two 100 MHz outputs with a minimum of 70 db isolation. The outputs must be controllable in frequency with a resolution of 1×10^{-16}.

The masers should operate within specification in the temperature range of 17°C to 27°C with a temperature coefficient of less than 3 parts in 10 to the 14th per °C. The sensitivity to ambient magnetic field variations must be less than 2 parts in 10 to the 13th for a change in field of ± 1 gauss. Other effects, such as atmospheric pressure changes, humidity, vibration or powerline variations should not affect maser performance more than 1 part in 10 to the 14th. In general, the masers are to operate in a normal laboratory environment but should be rugged enough and portable enough to withstand commercial shipment. The masers and maser sub-systems must be in one self contained package. The specified power supply is either 22 to 30 volts DC or 105 to 120 VAC with less than 500 watts input.

The VLBA-112 Hydrogen Masers

General Configuration. The Sigma Tau Standards Corporation VLBA-112 atomic hydrogen maser frequency standard is an active oscillator with a natural output frequency of 1,420,405,751.xxxx Hz which is derived from quantum transitions between two of the "hyperfine" levels of the ground electronic state of atomic hydrogen. This is the famous 21 cm line of atomic hydrogen.

The maser operation, in brief, is as follows: Molecular hydrogen is supplied from a small storage bottle and passes via an electronic pressure control servo to an RF source discharge bulb where the molecules are separated into atoms. Atoms emerge from the source through a small elongated hole, the source collimator, and then pass through a magnetic "state selector" which directs a beam of atoms in the correct quantum state to a teflon coated quartz storage bulb. The bulb is located within a microwave cavity which is resonant at the hydrogen transition frequency and CW emission from the atoms is produced by maser action. Power is coupled from the cavity by a small coupling loop and is transmitted to the receiver-synthesizer system through a coaxial cable.

A low noise heterodyne receiver system is used which contains a high resolution frequency synthesizer and a voltage controlled crystal oscillator. Subsequent integral multipliers, dividers and buffer amplifiers provide several well isolated outputs at standard frequencies. To provide the proper environment for maser action to occur and to minimize systematic perturbations of the maser output frequency, the maser is maintained under high vacuum by sputter-ion pumps, which also getter the hydrogen. The cavity is also surrounded by a set of magnetic shields and isolated by a multilevel thermal control system. An axial magnetic field coil provides for control of the internal magnetic field, and a single turn coil placed

CH2427-3/87/0000/075 $1.00 C 1987 IEEE

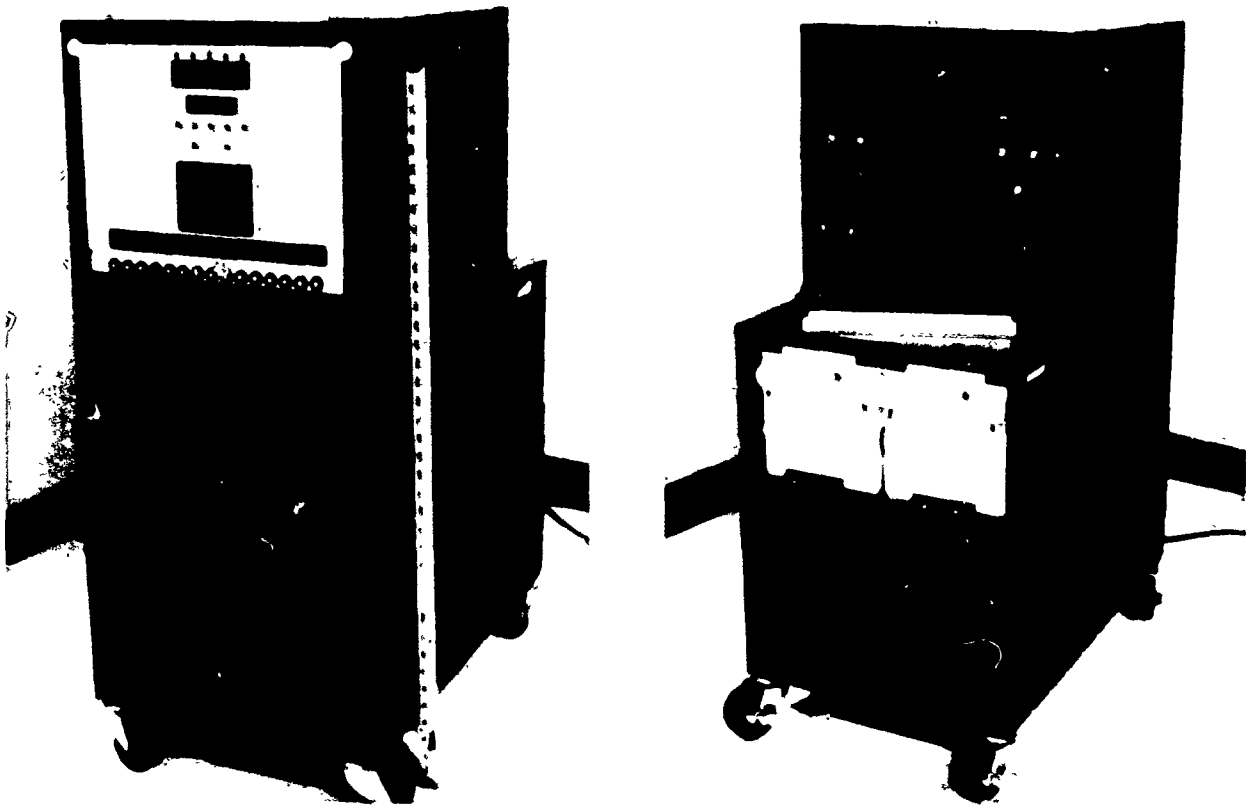


Figure 1. The VLBA-112 Hydrogen Maser.

transversley about the cavity provides a means of measurement of the field by the "Zeeman frequency" method.

One of the unique and important features of the VLBA masers is the incorporation of an automatic frequency control system to maintain the cavity at a constant frequency relative to the hydrogen emission line. This is a stand-alone system and does not require periodic source pressure changes or a separate frequency reference as used in the traditional spin-exchange method of cavity tuning. This system and many other features of the VLBA masers are discussed in more detail in References 2 and 3.

Maser Design Description. Figure 1 is a picture of the VLBA maser and Figures 2 and 3 are line drawings illustrating the arrangement of the parts and controls. These masers differ from previous developmental masers produced by Sigma Tau, described in References 2 and 3, in several ways although they are fundamentally similar. In the present design more attention has been given to electronic packaging, placement of sub-systems and controls in convenient, replaceable modules, more use of printed circuits, and other details to facilitate maintenance and repair. There are two hydrogen ion pumps, each with an estimated life of over 10 years, and a separate ion pump for regions surrounding the source and storage bulb which isolates the "inner" (hydrogen pump) system from vacuum background contaminants. Each pump may be easily changed in the field if need be.

The battery module in the present design is mounted on the maser. Two lead-acid "Maintenance Free" batteries provide up to 16 hours of operation with the masers fully operational. For partially

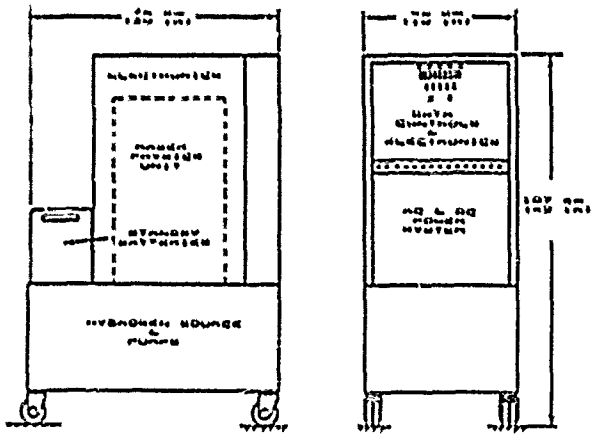


Figure 2

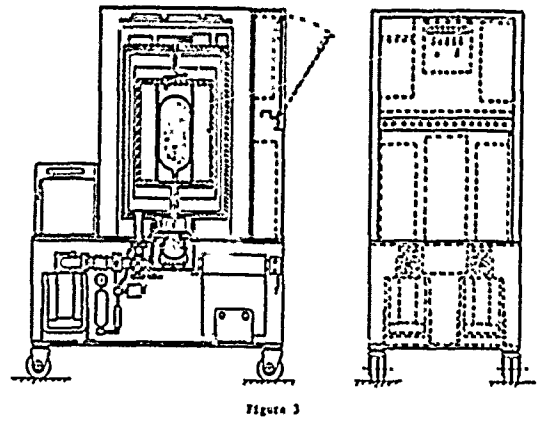


Figure 3

powered operation, such as may be used for long trips and installations, the maser will operate for periods up to a week on batteries with just the ion pumps and the cavity thermal control on; if only the ion pumps are active, about 5 watts of power is used, which will allow several weeks without AC power.

Other improvements incorporated into the VLBA masers include a very low noise receiver preamplifier having a noise figure under 1 db, a much heavier copper cavity providing better thermal homogeneity and larger thermal capacity to reduce susceptibility to ambient temperature variations. There are separate ion pump DC-DC converters for each of the three pumps. To insure phase stability, the multipliers and dividers for the 100 MHz and 5 MHz outputs are mounted on a temperature controlled plate with other potentially temperature sensitive components.

Detailed Specification. Following are the Sigma Tau Standards Corporation specifications for the VLBA Hydrogen Masers.

1. Stability, time domain: Figure 8 gives the specified frequency stability in the time domain and also the VLBA goal. The specification is a maximum which applies for realistic environmental conditions, in practice the nominal stability may be much better under ideal conditions.

2. Stability, frequency domain: The stability in the frequency domain, above the 10 Hz bandwidth of the VCO phase locked loop, is characterized by the stability of the crystal VCO. A high stability 10 MHz unit is used with phase noise specified as follows:

Phase Noise	Offset
-120 db	10 Hz
-140 db	100 Hz
-157 db	1 KHz
-160 db	10 KHz

3. Wide band phase noise: For purposes of radio astronomy, and several other applications, a useful specification is the wide band rms phase noise expressed in picoseconds such as may be observed at the output of a balanced mixer fed by signals in quadrature from two masers in a bandwidth from on Hz to up to a frequency just below the carriers. (A low pass filter which excludes the carriers must be used.) For the 100 MHz outputs, the specification is .6 ps rms and for the 5 MHz output the specification is 1.4 ps rms.

4. Drift: Included in the stability specification.

5. Settability: The output frequencies are adjustable without phase discontinuity using front panel switches or by external data system control with a resolution of 4.66×10^{-17}. The maximum continuous adjustment range is limited only by the loop hold in range of the crystal VCO to over $\pm 5 \times 10^{-9}$. The VCO frequency coarse adjustment may also be used to obtain offsets as large as $\pm 1 \times 10^{-6}$.

6. Reproducibility: Included as the maximum range of stability and environmental specifications.

7. Magnetic field sensitivity: For a ± .5 gauss external field change, the maser frequency will change by less than $\pm 1 \times 10^{-14}$.

8. Temperature sensitivity: Less than 3×10^{-14} per °C.

9. Barometric pressure sensitivity: Less than 1×10^{-14} per inch of mercury.

10. Operating life and mean time between failure: The hydrogen supply is adequate for over 50 years of operation. The ion pumps have an expected life between changes of over 10 years, and may be easily changed on site. The main failure mode is likely to be due to the expected MTBF of the electronic components and systems, which should be quite reliable.

11. Power supply sensitivity: For nominal line voltage changes or switching to batteries, the frequency will change by no more than $\pm 1 \times 10^{-14}$.

12. Power supply requirements: AC power, 115V rms ± 10%, 60 Hz, 100 watts nominal, 150 watts maximum with automatic crossover to batteries. On batteries, the typical current drain at 24 volts is 2.5 A. If external DC power is used, the requirement is 24 to 28 volts, 2.5 amperes (typical).

13. Standby battery operation: A 40 A-H, 24 volt lead-acid, "maintenance free" battery module is mounted on the maser which when fully charged will last approximately 16 hours. The batteries may be removed from the maser and located remotely if desired. Automatic recharging occurs at about 2 amperes.

14. Instrumentation: Thirty-two channels of analog data are multiplexed and selectable internally or remotely by binary addresses. Up to eight status channels are simultaneously outputed when addressed. An on-board monitor and control interface board adapts to the VLBA M&C system. Minor modification will adapt this system to other communication formats.

15. Automatic cavity tuner: An automatic cavity tuner of the cavity frequency switching type described in References 2 and 3 is continuously in operation, and no operator control is needed. The cavity spin-exchange offset frequency is set after first operation at the factory, but if it is desirable to check and reset the spin-exchange adjustment, it may be done with a digitally set resolution of 1×10^{-15} relative to the maser output frequency.

16. Size and weight: The maser is 46 cm (18 in) wide by 76 cm (30 in) deep by 107 cm (42 in) high, including casters. The overall weight is 238 kg (525 lb) including the battery module. The removable battery module weighs 34 kg (72 lb).

17. Nominal maser parameters and calibration factors:

a. Cavity loaded Q: 36,000

b. Operating line Q: 1.8×10^9

c. Cavity Q/ line Q ratio: 2×10^{-5}

d. Approximate SE frequency ratio at initial pressure settings: 5/1

e. Synthesizer factor: 4.658×10^{-17}/bit

f. Initial MPG setting: (M = 52300 approximately)

g. Cavity thermal mass: 15,600 w-s/°C

h. Cavity temperature: 50 °C

i. Other oven temperatures: 49 °C

j. Nominal oscillation frequency: 1,420,405,751.775,28 Hz assuming a magnetic field of .7 mg

Sigma Tau Hydrogen Maser Development

There are several particular design developments incorporated in the VLBA hydrogen masers which provide the performance of a "full size" oscillating atomic hydrogen maser in a relatively compact package with operational features that adapt it well to the long lived stable operation required of continuously operating basic time and frequency standards. These include the automatic cavity tuning system, the compact cavity and bulb assembly, an efficient tapered pole state selector, a compact and efficient magnetic and thermal shield configuration, an accessible arrangement of field replaceable vacuum pumps, a compact low pressure hydrogen supply, efficient and compact modularized electronics and a wide range high resolution receiver synthesizer. The cavity configuration, state selector and auto tuner are most important and only these will be detailed more fully at this time.

Cavity and Bulb Configuration. The use of metal cavities with the associated advantages of high electrical and thermal conductivity, thermal homogeneity, mechanical rigidity and ease of fabrication, ease of tuning and mounting stably within the vacuum enclosure has been one well established approach to hydrogen maser design[4,5]. To partially overcome one of the disadvantages of the large size cavity required by the relatively low frequency of the hydrogen hyperfine quantum transition and to improve the stability further, Sigma Tau Standards Corporation has analyzed the effect on the maser oscillation parameters of loading the cavity with a quartz cylinder situated about the maser storage bulb and arrived at an improved cavity and bulb design configuration[3].

Figure 4 illustrates the design. The quartz maser storage bulb is supported within a quartz cylinder which is held between the cavity end plates by spring action of a metal ring situated at the top. The thickness of the bulb and cylinder walls is such that the resultant inside diameter of the cavity is approximately 21.5 cm (8.5 in), rather than the 27 cm (10.3 in) diameter typical of a cavity loaded only with a thin-walled bulb. Analysis of the product of cavity Q x "filling factor" as calculated in Reference 3, Figure 11, gives a value approximately 28,000, which is very good and makes a very efficient design which oscillates easily.

State Selector. The service life of pumping elements and the store of hydrogen required for a maser are directly dependent upon the efficiency of the magnetic state selector which selects atoms in the proper hyperfine quantum states and focusses them into the maser storage bulb. A new quadrupole state selector with tapered pole tips was developed at Sigma Tau Standards Corporation which is much more efficient than the straight bore hexapole or quadrupole focussers widely used.

Figure 5 illustrates the state selector design. The entrance aperture is only .5 mm (.02 in) and the magnetic field gradient is very intense, approximately 2×10^5 gauss/cm at the entrance. This produces a strong acceleration to the atoms and gives a large source "capture angle". A primary practical advantage of this particular design is that the structural segments are adjustable after the magnet is assembled to provide five degrees of freedom for aligning the individual pole tips after assembly; this obviates the necessity of extremely high machining tolerances, which would make the design otherwise impractical. The design of the state selector and the analysis of focusing using the trajectory analysis of the atomic beam hydrogen masers is given in Reference 6.

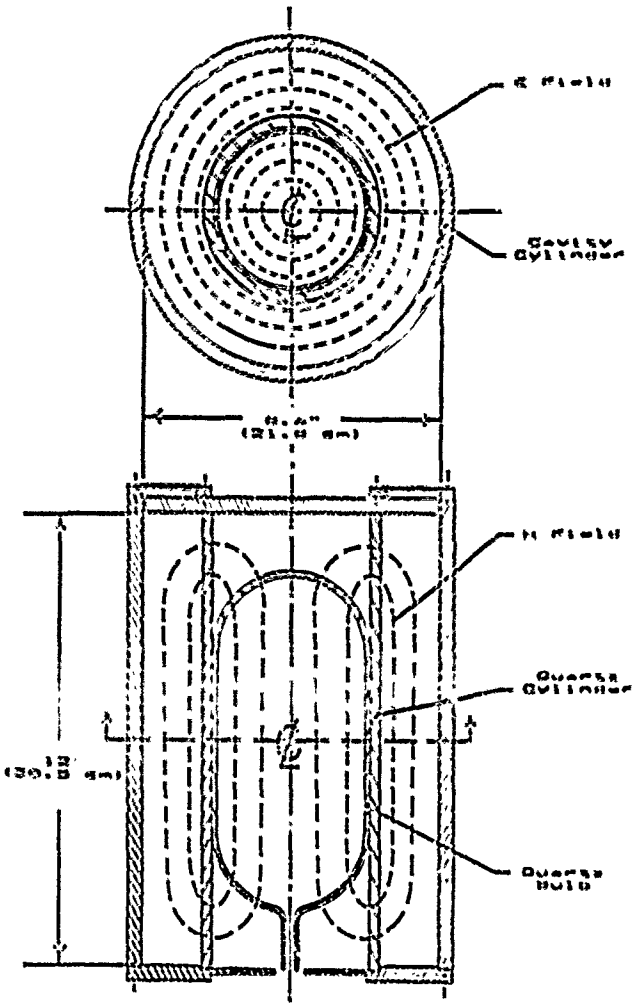


Figure 4. Cavity and Bulb Assembly

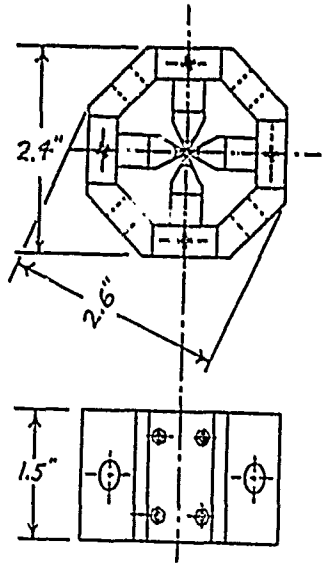


Figure 5. State Selector

Automatic Cavity Tuner. The automatic cavity tuner relates the cavity frequency to the hydrogen emission frequency through a servo system which does not require the transmission of external signals through the cavity or require a variation in beam intensity as used in spin-exchange tuning. A varactor is mounted within the cavity so that the cavity resonant frequency may be modulated electronically; as the cavity frequency changes, the coupling between the radiating atoms and the coupling loop changes, so that there is an amplitude modulation (as well as a phase modulation) of the maser output signal when the cavity average frequency differs from that of the maser signal.

In the cavity servo, a modulation period generator varies the cavity frequency periodically and a synchronous detector senses the amplitude and sign of the cavity offset and corrects the cavity average frequency. Figure 6 illustrates the amplitude variation which occurs when the cavity is detuned. Figure 7 is a block diagram showing the cavity servo system. While there is an undesirable phase variation occuring at the modulation frequency, it is cancelled out by a compensating phase shift inserted ahead of the crystal VCO so in practice the phase shift does not effect the standard output frequency. Reference 3 gives the details of the analysis of this type of cavity servo and also presents experimental result of its use in previous Sigma Tau Standards hydrogen masers.

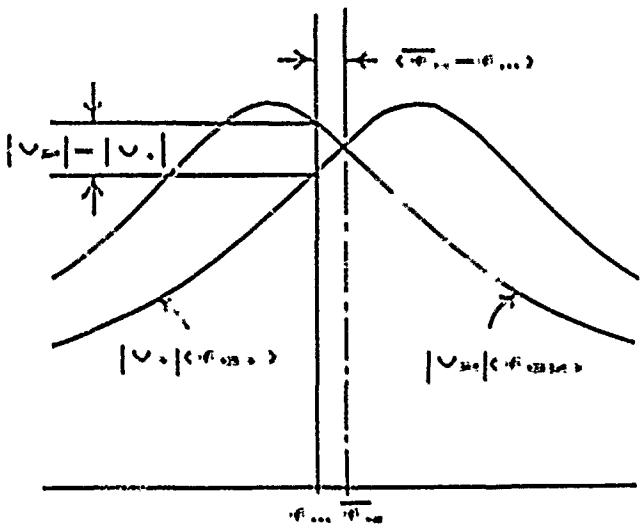


Figure 6. Cavity Frequency Modulation

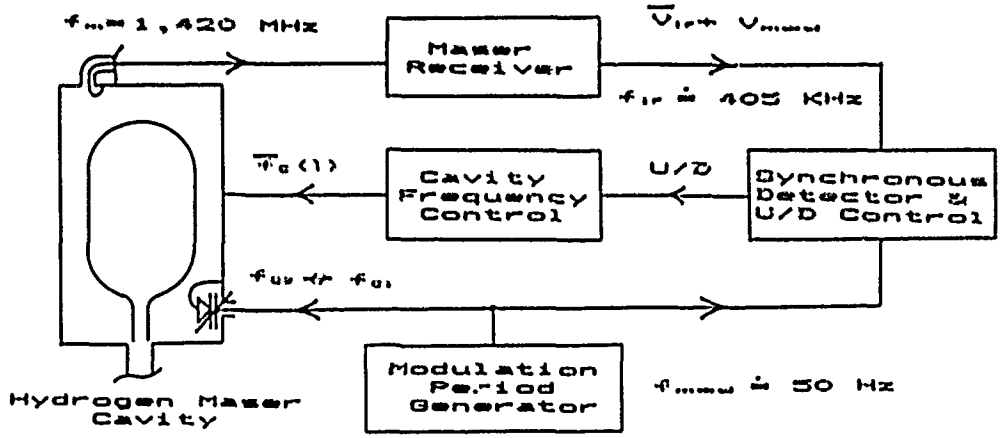


Figure 7. Cavity Frequency Servo

Performance

The first three of the ten hydrogen masers being produced for the VLBA are scheduled for delivery in July 1987. Construction is nearly completed and preliminary tests of stability have begun.

Stability. The present tests have been made under normal working laboratory conditions in an area where the temperature is held to approximately ± 1 °C by the building air conditioner, with a cycle time of approximately 20 minutes. Figure 8 shows the measured stability for time intervals between one second and 10,000 seconds. The ordinate is the two sample Allan Variance with no dead time. Also shown are the Sigma Tau Standards Corporation stability specification and the VLBA contract goal. Both masers were operating at nominal (low) beam intensity with both cavity auto tuners on.

The data for Figure 8 was obtained by connecting the 100 MHz outputs of two of the masers to a balanced mixer with a one Hertz low pass filter on the output. For measuring times between one second and 100 seconds an offset difference frequency of one Hz was set by the maser synthesizers and a period counter was used to measure the frequency differences. For 1,000 to 10,000 seconds the synthesizers were set for a slow beat frequency and the periods were measured from a strip chart recording.

Figure 9 shows stability data which was obtained, under similar conditions, using equipment supplied by the National Radio Astronomy Observatory (NRAO). In this case the curves were plotted by computer using an NRAO program, with phase measurements derived

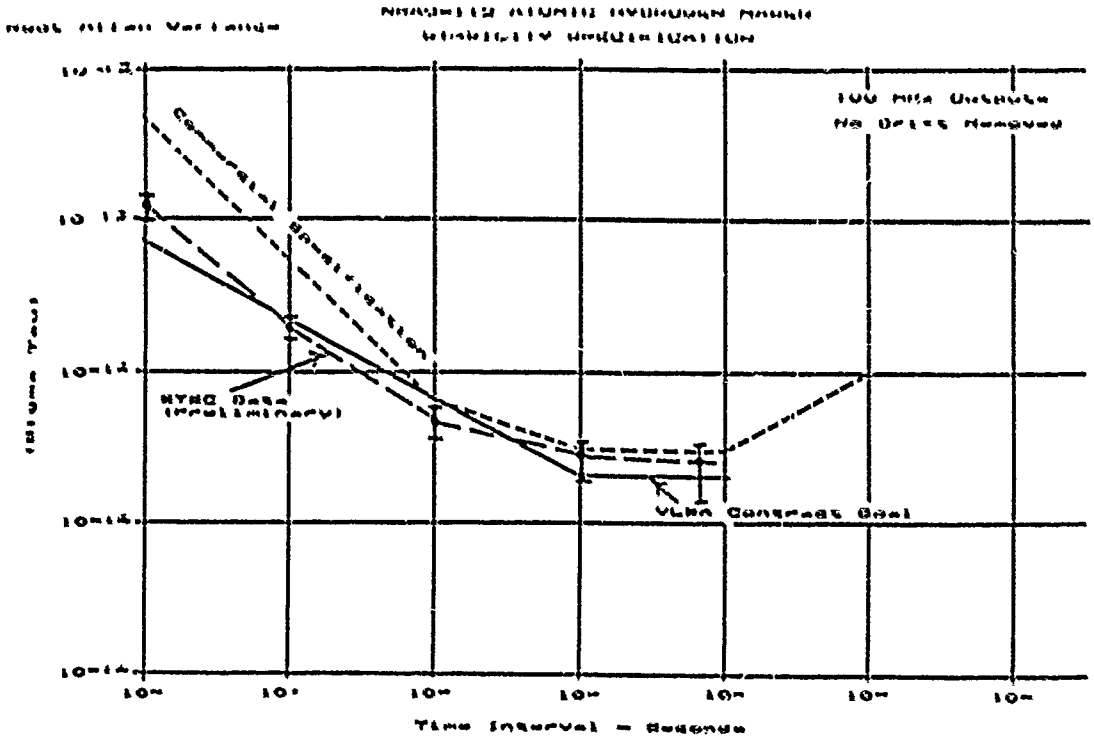


Figure 8.

FST PROGRAM OF 05/20/87
BATCH START 17:57.5 05/24/87
BATCH END 20:44.3 05/24/87
RUN START 15:09.9 05/24/87

EQUAL REFERENCE ASSUMED
TEMP=20.94C PHASE=32.6PS
TEMP=20.94C PHASE=136.9PS
TLOW=20.48C, THIGH=21.73C

TAU SEC	LAST BATCH SAMPLES	LAST BATCH SIGMA *EXP-15	CUMULATIVE SAMPLES	CUMULATIVE SIGMA *EXP-15	TIME ERROR PS
1	10009	129.9	20018	130.5	0.1
2	5004	76.1	10008	76.4	0.2
5	2001	36.8	4002	37.1	0.2
10	999	17.4	1998	18.3	0.2
20	499	12.7	998	13.6	0.3
50	199	5.8	398	7.7	0.4
100	99	4.8	198	5.2	0.5
200	49	3.1	98	3.6	0.7
500	19	2.3	38	2.2	1.1
1000	9	1.5	18	1.6	1.6
2000	4	1.5	8	1.4	2.8
5000	1	0.7	2	0.8	3.8
10000	0	0.0	0	0.0	0.0

Figure 9(a). NRAO System Stability Data

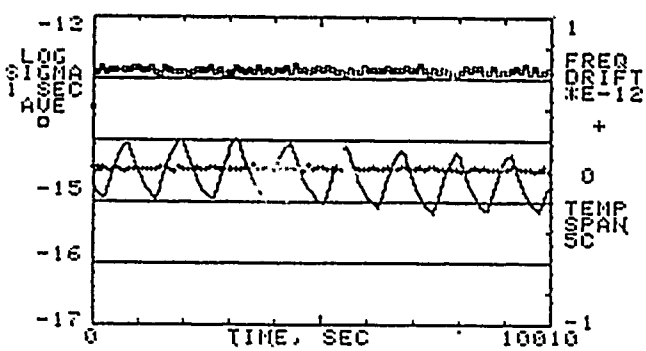


Figure 9(b). Frequency, Drift and Temperature Plot

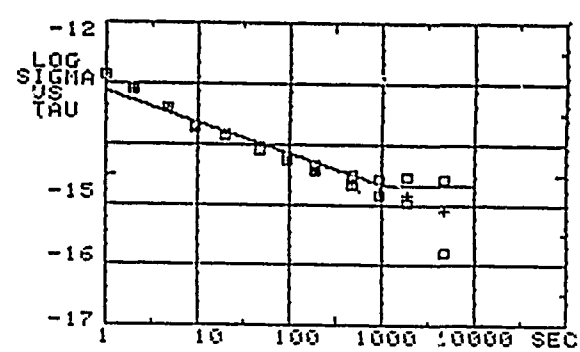


Figure 9(c). VLBA Maser Stability

from the two maser 100 MHz outputs connected to a highly stabilized balanced mixer and voltage amplifier. Figure 9(a) is a copy of the printer output of the data plotted in Figures 9(b) and 9(c). Figure 9(b), in the top trace, shows the one second stability averaged for 100 samples, the second trace gives the room temperature, which illustrates the 20 minute cycling of the temperature; the third trace, which starts at zero in the first batch, shows the relative variation of the two masers's frequencies since the measurement batch was started.

Figure 9(c) is a plot of the stability versus measuring interval in the familiar sigma tau format; also shown (the solid line) is the VLBA contract stability goal. It is noteworthy that the data taken with the NRAO computer equipment agrees very well with the stability plot obtained with counter and strip chart, shown in Figure 8.

In addition to stability in the time domain, the wide band phase noise at 100 MHz has been tested and the results are within the specifications of .6 ps rms.

Much more stability data will be taken under different conditions of operation and for longer times. The masers are still being tested and adjusted to optimize the maser temperature and cavity tuner servo gains and other parameters, however the present data is quite encouraging.

Other Performance Data. Due to the short time since the assembly of the first three masers was completed, other test data is not available at this time. There are several tests which remain to be done before delivery. These include long term stability, magnetic shielding and barometric pressure sensitivity. It is anticipated that the performance will be similar to that of the three earlier masers of similar design described in References 2 and 3, and these met or exceeded the VLBA environmental sensitivity requirements as given in the Sigma Tau Standards specifications given previously.

Acknowledgements

The authors greatfully acknowledge the support of the National Radio Astronomy Observatory under Contract VLBA-112 as well as the Associated Universities, Inc., and the National Science Foundation.

References

1. A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr., Interferometry and Synthesis in Radio Astronomy, John Wiley & Sons, Inc., 1986.

K.I. Kellerman and A.R. Thompson, "The Very Long Baseline Array," Science V229, No. 4709, 12 July 1985.

2. H.E. Peters, "Design and Performance of New Hydrogen Masers Using Cavity Frequency Switching Servos," 38th Annual Symposium on Frequency Control, June 1984.

3. H.E. Peters and P.J. Washburn, "Atomic Hydrogen Maser Active Oscillator Cavity and Bulb Design Optimization," 16th Annual PTTI Applications and Planning Meeting, November 1984.

4. H.E. Peters, T.E. McGunigal and E.H. Johnson, "Hydrogen Standard Work at Goddard Space Flight Center," 22nd Annual Symposium on Frequency Control, 1968

5. V.S. Reinhardt, et al., "NASA Atomic Hydrogen Maser Standards Program-An Update," 30th Annual Symposium on Frequency Control, 1976.

6. H.E. Peters, "Magnetic State Selectors in Atomic Frequency and Time Standards," 13th annual PTTI Applications and Planning Meeting," 1981.

THE NEW GENERATION OF HYDROGEN MASER AT SHANGHAI OBSERVATORY

Z.C. Zhai, H.X. Huang, G.X. Jiang, W.H. Luo, J.F. Lu, and C.F. Lin
Shanghai Observatory, Shanghai, China

Abstract

A new hydrogen maser for VLBI--network in China has been designed at Shanghai Observatory. The mechanical and electronic design philosophy of the new maser is described, and comparisons with previous masers made at Shanghai Observatory are detailed. The new maser is a rugged, reliable and transportable hydrogen maser. Preliminary test data and projection of stability performance are included, as well as photographs of the new maser system during the constuction phase.

Introduction

Shanghai Observatory began its hydrogen maser development at the begining of 1970 and so far has developed 6 hydrogen masers. Their design characteristics were described in Ref. 1, 2, 3. After 1984, three of Shanghai Observatory's hydrogen masers were improved again with the modifications of cavity-bulb assembly and beam optics. Figure 13 shows the frequency stability performance of the modified masers.

Meanwhile, a new hydrogen maser for VLBI network in CHINA has been developing since 1985. The new generation emphasizes the reglidness and mechanical and thermal stability, especially for the very critical cavity-bulb assembly. In the new design, we absorbed many design ideas of the masers which are developed by Dr. F.R.C. Vessot in SAO (Ref. 4, 5).

This paper described the mechanical, thermal, and electronic design of the new generation.

New Maser Design Philosophy

There are two basic aspects to the design philosophy developed at Shanghai Observatory for the new hydrogen maser. First, the maser cavity was to be as inherently stable as possible. Then a very tightly controlled thermal and mechanical environment was to be established for the cavity to permit maser operation over long periods of time without auto-tuning or other adjustment. Toward this end, the cavity was constructed of CERVIT, a material with a very small thermal-expansion coefficient (1 to 2E-7/°C) and long-term resistance to dimensional creep. To reduce the aspect of the thermal coefficient of dielectric loading, the mass of the storage bulb was reduced to 160 g (from 300 g as used in the earlier maser) without sacrificing the stiffness-to-weight ratio. A typical value for the temperature coefficient of the dielectric constant of the quartz storage bulb within the cavity is df/dT=110 Hz/°C. Figure 1 shows the cavity-bulb assembly for the new maser.

The second basic consideration was that the new maser was to be a readily tranportable, rugged, and completely integrated frequency standard. The entire system was engineered as a unit, and all electronics, including the receiver and the synthesizer, were mechanically and electrically integrated into the overall design. The end result is a relatively small instrument weighting about 300 kg and requiring a source of 28 volts DC at about 250 Watts to operate.

Fig. 2 shows a photograph of the new maser system during the construction phase.

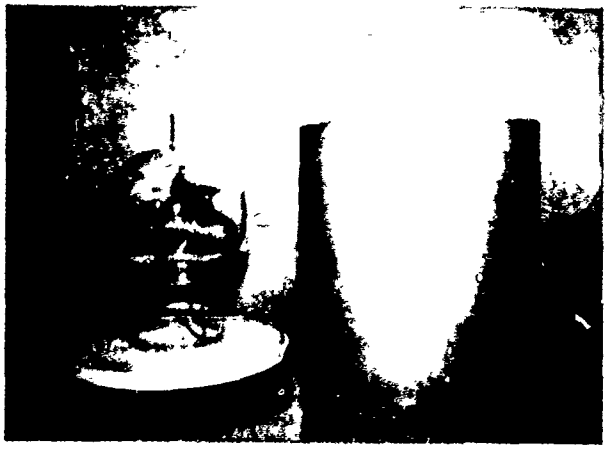


Fig. 1 - cavity-bilb assembly

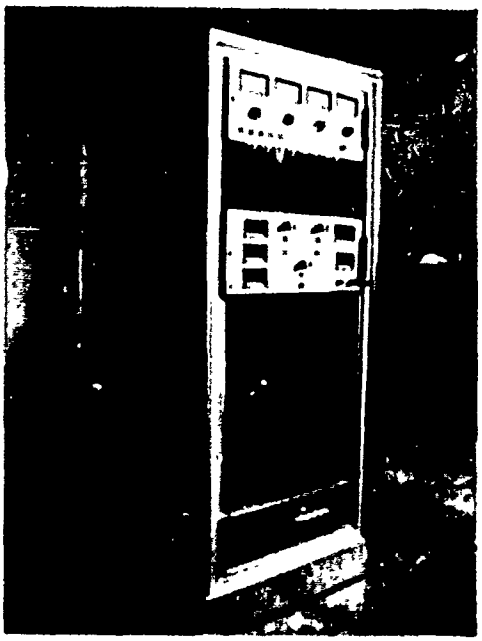


Fig. 2 - new maser system during the construction phase

Physics package designs

The differences between the new maser and any earlier masers H1 through H6 made at Shanghai Observatory are chiefly in the direction of more rugged structural and more careful thermal design and better magnetic shielding. The major departures of the new maser from either the H1 through H5 or H6 are the completely different designs, especially the cavity-bulb structure, the lightweight and the small ion pump as well as the torrispherical magnetic shields.

Figure 3 describes the cavity-bulb structure. Possible cavity-stress changes owing to the expansion

CH2427-3/87/0000-082 $1.00 C 1987 IEEE

coefficient of the external cylinder holding the cavity together are taken up by the Bellville Washer. Similar thermally induced radial stress in the base is relieved by the rollers. Strain on the bell jar owing to barometric-pressure change are isolated from the cavity by the double-base structure.

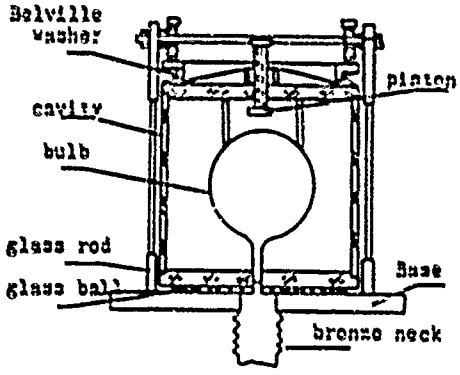


Fig. 3(a) - cavity-bulb structure for earlier maser

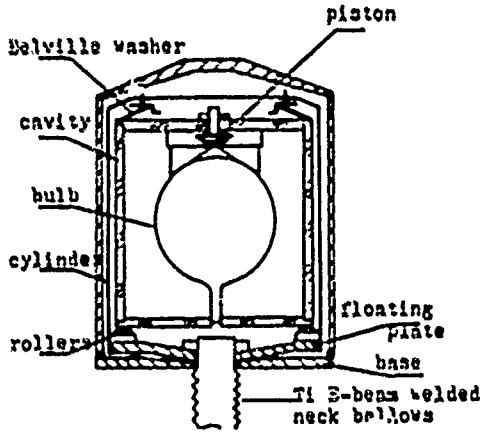


Fig. 3(b) - cavity-bulb structure for new maser

Figure 4 shows the magnetic shields. Here the new maser differ greatly from the old masers, in that torripherical shields have foam-glass insulation at their ends for supports. These shields are under far less local stress than are the flat-ended shields and have much improved lapped joints at the lower cover (the upper cover is not removable).



Fig. 4 - magnetic shields

Figure 5 shows the new maser solenoid. The multilayer printed-circuit design allows very complete cancellation of spurious magnetic fields and provides a rugged, close-fitting, and simple coil system located immediately inside the innermost magnetic shield.



Fig. 5 - new maser solenoid

The dissociator/state selector for the new maser is shown in Figure 6.

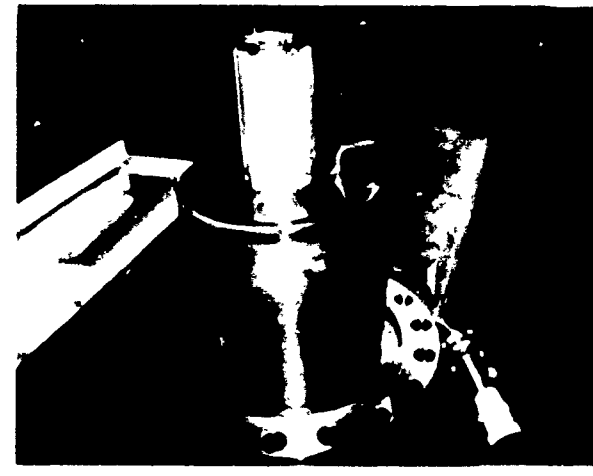


Fig. 6 - dissociator/state selector

The chief difference between this design and the initial old design include the size of the dissociator glassware and the layout of the RF coil. The dissociator size of 3.8 cm diameter x 4.0 cm long concides almost exactly with those used with excellent success by the SAO maser group. The ion-pump manifold for the new maser is shown in the lower portion in Figure 7. It can be seen that the size of the pump is much smaller than those used in the Shanghai Observatory's old masers. Fig. 8 shows a photograph of the new maser mounting on the testing rack.

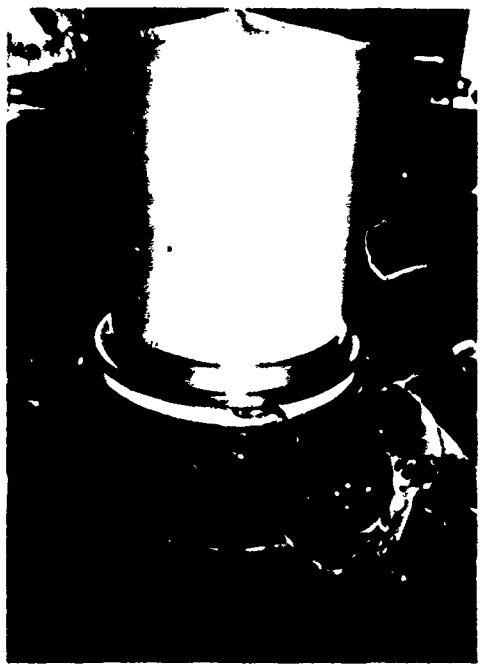


Fig. 7 - Ion pump mounting on lower portion



Fig. 8 - new maser mounting on the testing rack

Design of the New Maser Electronics

There are two basic categories of maser electronics system: (1) maser support (thermal, magnetic, presure and tuning control) and (2) maser signal processing. Unlike earlier Shanghai Observatory's designs, the new maser has all the electronics integrated into a single functional unit with all the controls, metering, and power supplies centralized. The support and processing electronics are mechanically and electrically compatible with each other and with physics package.

Maser Thermal Control

According to the experience gained at SAO (Ref. 4, 5), increased thermal gain or improved preamplifier stability alone is ineffectual in attacking the thermal gradients problem along the vacuum tank; however, division of the tank (and oven) surface into independently sensed and controlled zones has shown itself to be a powerful technique for minimizing gradient problems. Accordingly, the vacuum tank is divided into three zones -- dome, cylinder, and base -- each of these has its own sensing thermister, amplifier, and heater windings. Each zone can respond independently to external thermal loads without materially affecting the temperature of other zones. To minimize the thermal stress on vacuum-tank controller, the oven is divided into a dome-cylinder zone and a base zone. Each of these is independently controlled. The isolator-preamplifier box is a seperate thermal zone, making a total of six distinct thermal-control zones. The circulating air, which controls the temperature of the pump, the dissociator, and the upper maser electronics, is an entirely independent system with self-contained sensors and electronics.

Dissociator Oscillator

The RF dissociator oscillator in the new maser is a refinement of the one-transistor design used in earlier masers. The oscillator operates at approximately 86 MHz with 5-8 watts of DC power output. Figure 9 shows the dissociator oscillator.



Fig. 9 - dissociator oscillator

Hydrogen -- Presure Control

Hydrogen for operation of the dissociator is furnished from a two-liter gas bottle presurized to approximately 80 psi. The relatively low presure permits the use of a small, thermally agile palladium valve. Hydrogen presure within the dissociator is sensed by a thermistor pirani gauge, which is incorporated into a resistance bridge along with an identical thermistor that senses the high-vacuum side pumping system. The bridge output, which is independent of ambient temperature variations to first order, drived a servo system that controls the dissociator presure by varying the temperature of the palladium valve.

Phaselock Receiver

The phaselock receiver for the new maser is a total new design. It is a low noise, phase-stable triple-conversion phaselock loop (Ref. 6).

The first stage of low-noise preamplification is physically located within the temperature controlled

isolator box. An isolator minimizes the effects of external load changes on the cavity resonant frequency, and a low-noise amplifier, with noise figure of 1.5 dB, delivers the 1420 MHz signal to the receiver at a level of approximately -80 dBm with net isolation of 90 dB or more.

A block diagram of the RF receiver is shown in Fig. 10. Within the receiver proper, the second preamplification stage is preceded by a wide-band input filter that rejects image-frequency input noise at approximately 1080 MHz.

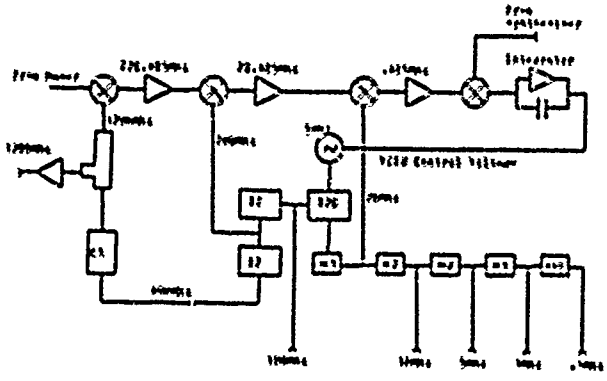


Fig. 10 The electrical schematic circuit of the receiver

The primary output is 5 MHz with auxiliary outputs provided at 100, 10, 5, 1 and 0.1 MHz.

Fig. 11 shows the phaselock receiver.

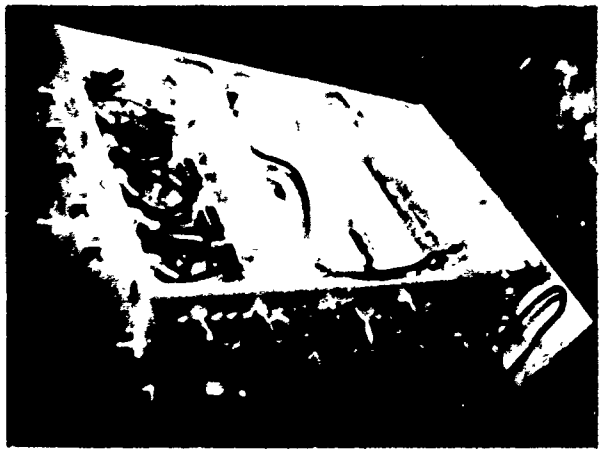


Fig. 11 - phaselock receiver

Digital Synthesizer

The digital synthesizer for the new maser is being developed, which will also be a totally new design. The new synthesizer provides a resolution of 0.001 Hz over a tuning rang of 405,750.000 to 405,753.999 Hz. The synthesizer concept combines both direct and indirect synthesis techniques which is similar to those developed in SAO maser group.

For temporary use before the new synthesizer is finished, a commercial product of H-P 3325 is put into use instead.

Monitoring System

The new maser incorporates an extensive monitoring capability. Four front-panel mounted meters, each with an eight-position selector switch are located on the monitor panel to provide quick-look monitoring of 32 functions. These include all main power-supply voltages, all heater voltages, hydrogen dissociator operating conditions, receiver/synthesizer signal level, and phaselock control voltages.

The monitor panel also includes the small four-digit counter and LED readout for displaying the output frequency of the synthesizer. All metering is visible at the top of the rack in Figure 12.

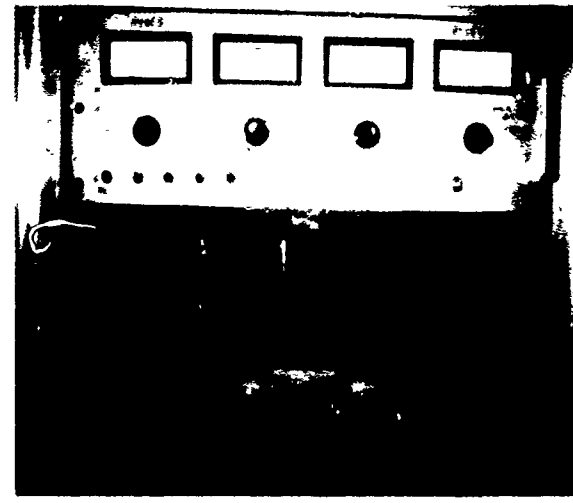
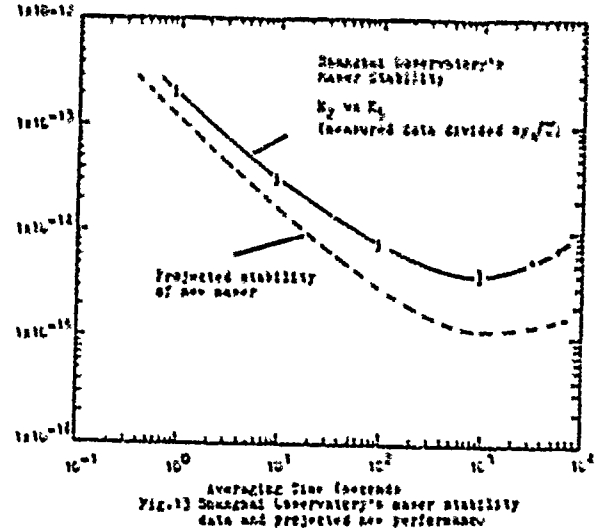


Fig. 12 - Monitoring panel

Expected Stability Data from the New Maser

Data from earlier masers

Stability data for earlier masers which are being used at Shanghai Observatory are shown in Figure 13.



Averaging time (seconds)

Fig. 13 Shanghai Observatory's maser stability data and projected new performance

Data for the new maser

From data obtained from tests on the maser during assembly, we can make the following projections about systematic effect.

A. Thermal insensitivity will be improved by a factor of at least 3, since the intrinsic thermal sensitivity of the cavity-bulb assembly had been found to be about 3 times lower, and we also expect a considerable improvement from the thermal redesign.

B. The tests on torrispherical shields indicate that axial shielding factor has an improvement by a factor of about 1.3.

C The cavity-bulb structure is more rugged and better isolated than the earlier masers. So we expect better performance from the new design.

D. Improvement of the noise figure of the maser receiver will result from by using a preamplifier with a 1.5 dB noise figure instead of the 5 dB of the earlier masers. The short-term stability in the 1/T portion of the Allan Variance will be reduced by about 2.

Figure 13 shows the anticipated performance of the new maser with all the above factors taken into account.

Acknowledgement

We would like to thank Dr. R.F.C. Vessot for providing the opportunity for the first-named author of the paper to work with the hydrogen maser group at SAO for two years.

References

[1] Zhai Zao-cheng, Shanghai Observatory's Hydrogen Maser, JIETE Vol. 27, 11, 1981

[2] Zhai Zao-cheng, Zhuang Qi-xiang, A Brief Introduction to the Atomic Time and the Frequency Standards at Shanghai Observatory, Proc. of 38th Annual Symposium on Frequency Control, 1984

[3] Zhai Zao-cheng, Zhuang Qi-xiang, Hu Jin-lun and Kou Wun-wei, Activities on Time and Fequency at Shanghai Observatory, Proc. of 16th PTTI Meeting, 1984

[4] R.F.C. Vessot, M.W. Levine, E.M. Mattison etc., Space-born Hydrogen Maser Design, Proc. of 8th PTTI Meeting, 1976

[5] E.M. Mattison and R.F.C. Vessot, Techniques Used in SAO Hydrogen Masers for Increased Frequency Stability and Reliability, Japan, 1982

[6] Operating and Maintenance Manual for the Model VIG-11H Atomic Hydrogen Maser, SAO, 1984

41st Annual Frequency Control Symposium - 1987

Performance of a Hydrogen Maser With Auto-Tuning Utilizing Cavity Q Modulation*

T. K. Tucker and G. J. Dick
California Institute of Technology
Jet Propulsion Laboratory
Pasadena, California

ABSTRACT

The hydrogen maser is the most stable frequency source generally available today for all but the shortest measuring times. For long measuring times of 10^4 seconds or greater, its performance typically deteriorates due to frequency drift of the high-Q resonant cavity. Various schemes have been devised to detect and compensate for this drift, each with its own advantages and disadvantages. The method described herein utilizes cavity Q modulation to detect cavity frequency drift from the hydrogen line by inducing a phase shift in the maser output which is proportional to the cavity tuning error. This phase shift after detection is fed back to the cavity tuning servo to compensate for detected errors. A significant advantage over other similar schemes is the lack of phase modulation of the maser output due to the feedback being around the null modulation condition.

We have implemented a cavity Q modulator, along with an incidental amplitude modulation compensator, into an active hydrogen maser and successfully demonstrated its performance against another hydrogen maser. Our tests confirm that there is no degradation of maser short term performance with the addition of cavity Q modulation. We have observed long term stability of $7X10^{-13}/\sqrt{\tau}$ with a maser operating line Q of $7X10^9$, confirming our analytical model and pointing toward an achievable stability of $3X10^{-13}/\sqrt{\tau}$. This paper will describe the Q modulation autotuner, the measurement method and equipment, and the results of the experiment.

INTRODUCTION

The Deep Space Network (DSN) has very strict requirements for precise frequency and time. These requirements arise from the need for precise spacecraft navigation as well as a variety of radio science experiments such as Very Long Baseline Interferometry (VLBI). At the present time, hydrogen maser frequency standards are used throughout the network which are stable to parts in 10^{-15} at 1000 second averaging times. This stability degrades at longer averaging intervals due to the influences of environment and cavity aging. This long term instability leads to time and frequency errors which requires frequent correction to maintain clock accuracy.

Maintenance of an accurate time scale depends upon close coordination of the time and frequency of the station clocks with respect to each other as well as the National Bureau of Standards, Boulder, Colorado. Coordination of the time and frequency of the various stations is accomplished through the use of data obtained from the Global Positioning System (GPS).[1] From these data, the drift rates of the masers are estimated and matched as closely as possible. This process requires continuous monitoring and frequent tuning adjustments to maintain the masers within an acceptable accuracy.

The requirement for frequent adjustment and tuning can be greatly reduced by continuous autotuning of the maser to compensate for long term cavity drift and/or environmental changes. The Q modulation autotuning method is capable of providing such continuous tuning without perturbing the maser short term stability.

DESCRIPTION

The Q-modulation scheme which we used in this work has been described elsewhere[2]. In summary, autotuning by fast cavity modulation operates by varying the cavity Q (or frequency) and detecting the fluctuations of phase (or amplitude) which result. If the modulation is much faster than the linewidth of the hydrogen line, the sign and magnitude of these fluctuations depend directly on the relative tuning of the cavity with respect to the radiating hydrogen atoms. A major consideration is that while one of these parameters (Q, frequency) must be substantially modulated, the other must be held constant to a very great degree, since the detected quantity (phase, amplitude) varies directly with the other parameter.

Additional problems relate to each type individually. Frequency modulation autotuning schemes have been implemented by several groups[3,4] with good success. The disadvantages of these schemes, however, include substantial phase modulation of the output signal, and sensitivity to modulation parameters including both the amplitude and duty cycle of the frequency modulation. This sensitivity is primarily a consequence of the fact that the parameter being modulated is also the parameter being controlled. Q-modulation also has its inherent disadvantages, suffering primarily from consequences of the amplitude modulation that necessarily exists in the cavity fields. These include a reduction of average output power, incidental modulation of the spin exchange tuning, and amplitude-to-phase conversion in the electronic circuits following the maser itself.

The limiting long term oscillator performance for either type of modulation depends on maser phase noise in the vicinity of the modulation frequency. Modeling this noise as white phase noise,[2] the resulting autotuned maser performance $\sigma_{at}(\tau)$ can be related in a particularly simple way to the open loop maser performance $\sigma_a(\tau)$ in terms of a crossover frequency by:

$$\sigma_{at}(\tau) = \sigma_a(\tau_c)\sqrt{\tau_c/\tau}$$

For Q-modulation, the crossover frequency τ_c is given by:

$$\tau_c = 2B\Delta_1^2 Q_h^2/(4\pi^2 F_0^2)$$

where B is the bandwidth of the measurement system, Δ_1 is the fractional Q modulation, Q_h is the hydrogen line Q, and F_0 is the operating frequency. Typical values for τ_c are .03 to .3 second.

CH2427-3/87/0000-087 $1.00 C 1987 IEEE

We have developed a PIN diode Q modulator (Fig. 1), which can be tuned for zero cavity tuning, and which also shows no cavity tuning with variation in the PIN drive amplitude. The device is mechanically tuned at the time of installation to show no cavity frequency shift even with very large Q reduction. The modulator, since it loads the cavity, does result in an incidental amplitude modulation on the cavity output signal. This amplitude modulation must be removed from the signal in order to prevent frequency tuning errors due to AM to PM conversion in subsequent receiver circuits.

IMPLEMENTATION

The hydrogen maser chosen for the installation and testing of the Q modulation auto-tuning, was an early JPL prototype unit known affectionately as P2. This maser, except for minor differences, is functionally identical to those presently used as reference masers in the JPL test facility.

The Q modulator probe was installed inside the vacuum chamber in the microwave cavity top tuning plate, and the amplitude compensation module was placed inside the RF head just outside of the magnetic shields in an area which is moderately temperature controlled.

Figure 1 is a schematic of the Q modulator, figure 2 is a simplified block diagram of the amplitude compensation module, and figure 3 is an overall block diagram of the cavity Q modulation system.

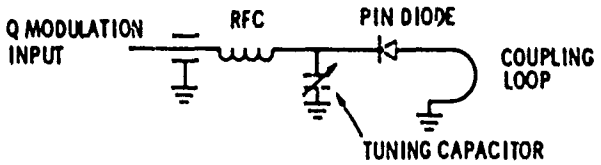


Figure 1. Q-Modulator, Schematic Diagram.

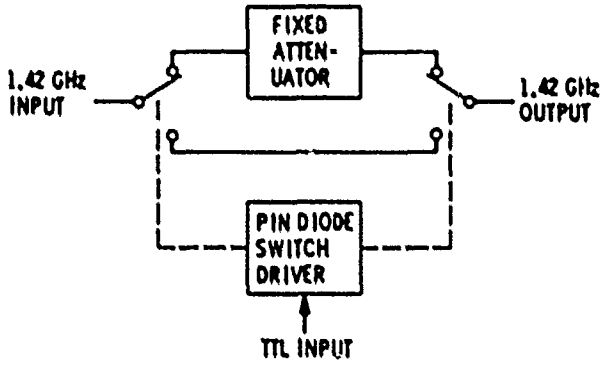


Figure 2. Amplitude Compensator, Block Diagram.

TEST METHODS & EQUIPMENT

A simplified block diagram of the test configuration is shown in figure 4. The reference used is a Smithsonian Astrophysical Laboratory (SAO), model VLG-11, hydrogen maser, S/N 21, which was installed in a temperature controlled environment. The test maser physics unit was located in an environmental chamber where its temperature could be controlled, and the electronics rack with the associated electronics equipment was located outside the test chamber.

The stability analyzer is shown in simplified block diagram form in figure 5. The 100 MHz reference signal from the reference maser is applied to the analyzer through an isolation amplifier. This reference signal is applied to the input of an offset synthesizer which offsets the reference signal by 1 Hz resulting in an output frequency of 99,999,999 Hz. This signal is then applied to an RF mixer where it is multiplied with the 100 MHz signal from the test maser, producing a difference frequency of 1 Hz. This 1 Hz signal is passed through a low pass filter, and then into a zero crossing detector. The zero crossing detector output is analyzed by a Hewlett-Packard, model, HP 1000 minicomputor which provides Allan variance data.

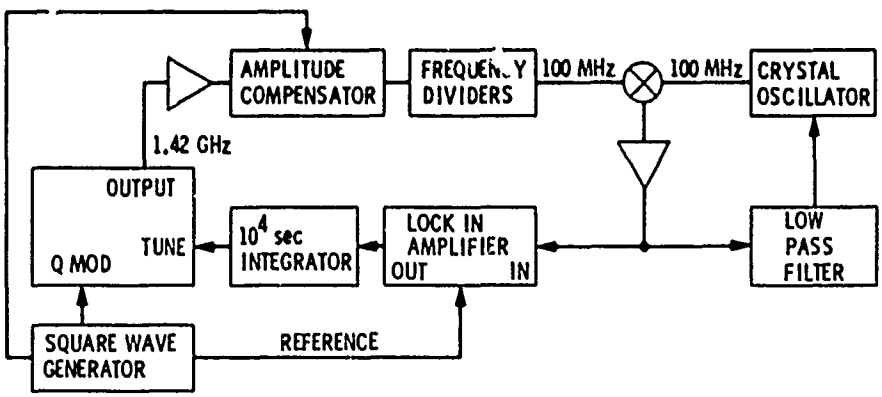


Figure 3. Cavity Q-Modulation Autotuner System.

TEST RESULTS

The first series of tests, were made to determine the operating characteristics of the maser with and without the autotuner. Allan variance data plots, taken with and without the Q modulation signal are contained in figure 6 , and clearly show that there is only minimal degradation of the maser short term stability with the introduction of the Q modulation signal.

Additional tests were accomplished by varying the temperature of the maser under test to determine the effectiveness of the autotuner to compensate for environmental influences. Figure 7 shows the temperature sensitivity of the maser without autotuning, and figure 8 after introduction of autotuning.

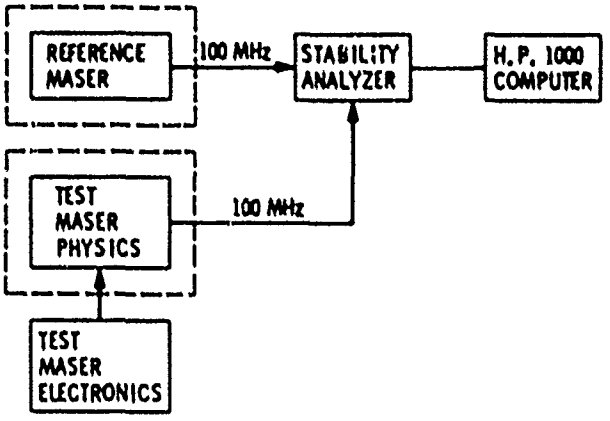


Figure 4. Test Configuration, Block Diagram.

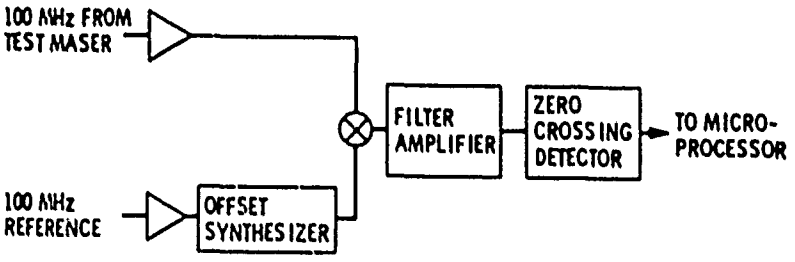


Figure 5. Stability Analyzer, Block Diagram.

From the data it can be seen that the frequency sensitivity of the non-autotuned maser to environmental temperature changes is $4 \times 10^{-13}/C^{\circ}$. With the introduction of autotuning, this sensitivity is reduced to approximately $2 \times 10^{-14}/C^{\circ}$ an improvement of 20 over the non-autotuned condition.

Figure 9 shows Allan variance data for the autotuned condition out to approximately 10^5 seconds. From the equations earlier, and for the parameters characterizing the maser and measurement; B=1Hz, Δ_f=1.33, Q=6.9×10^8, F_o=1.4214×10^9, we derive f_ϵ =.018 and $\sigma(\tau)$ =$3.0 \times 10_{-12}/\sqrt{\tau}$ for the limiting long term stability of the stabilized maser. Because the loop time constant is 5000 seconds, times longer than this value should reflect autotuning noise. It is apparent from the figure that the results are at least as good as this value, which would be 3×10^{-14} at 10,000 seconds.

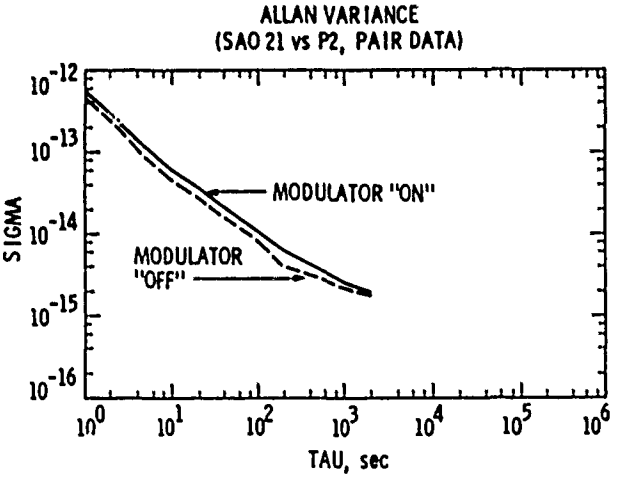


Figure 6. Allan Variance (SAO21 vs. P2).

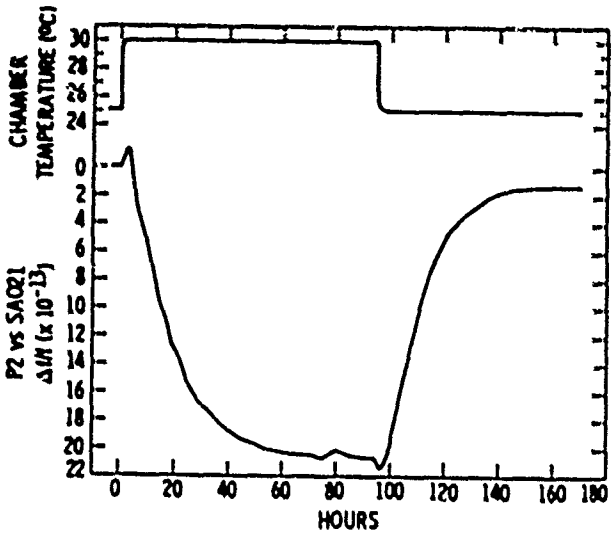


Figure 7. Open Loop Maser Temperature Response.

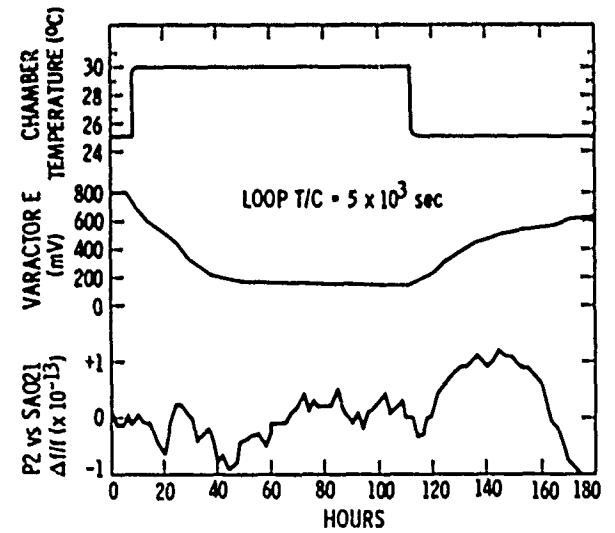


Figure 8. Autotuned Maser Temperature Response.

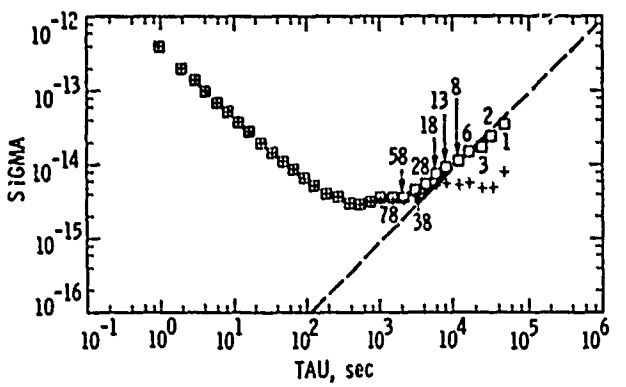


Figure 9. Allan Variance (SAO21 vs. P2), Autotuned.

CONCLUSIONS

As a result of the tests performed during the preparation of this report, a number of future improvements are brought to mind. First, the amplitude compensator should be redesigned to improve its phase vs temperature characteristics while maintaining a constant amplitude to subsequent amplifiers. Secondly, the temperature sensitivity of the maser electronics and control assembly needs improvement if the full benefits of autotuning are to be realized. The major areas of thermal sensitivity were found to be the integrator assembly and the lock-in detector amplifier unit. Replacement of the analog integrator with a digital integrator will effect a major improvement in thermal performance and permit longer integration times. And third, to increase the varactor tuning range to permit greater tuning range to compensate for a wider range of environmental factors, andalso permit longer operating times without the requirement for adjustment of the mechanical tuners.

From the above test results, we can conclude that the incorporation of an autotuner system, utilizing cavity Q modulation, substantially improves the long term stability of an active hydrogen maser and does so without sacrificing short term performance. With the introduction of continuous autotuning, we can expect a significant reduction in the requirements for field adjustments of hydrogen maser tuning.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the substantial assistance with testing and data reduction by A. Kirk, W. A. Diener, and C. A. Greenhall.

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

REFERENCES

1. Clements, P., Kirk, A., and Unglaub, R. "Results of Using the Global Positioning System to Maintainthe Time and Frequency Sychronization in the Jet Propulsion Laboratory's DeepSpace Network" Proc. 18th annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, (1986).

2. Dick, J., Tucker, T., "Fast Autotuning of a Hydrogen Maser by Cavity Q Modulation" Proc. 17th annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, (1985).

3. Peters, H. E. "Design and Performance of New Hydrogen Masers Using Cavity Switching Servos" Proc. 38th Annual Symp. Freq. Control,(1984).

4. Audouin, C., Lesage, P., Viennet, J., and Barillet, R. "Theory of Hydrogen Maser Autotuning Systems Based on the Frequency or Phase Method" IEEE Trans. Instrum. Meas., Vol. IM-29, 98-104, (1980).

41st Annual Frequency Control Symposium - 1987

HYPERFINE CONTRIBUTION TO SPIN-EXCHANGE FREQUENCY SHIFTS IN THE HYDROGEN MASER*

B. J. Verhaar, J. M. V. A. Koelman, H. T. C. Stoof, and O. J. Luiten
Department of Physics
Eindhoven Technological University
5600MB Eindhoven, The Netherlands

and

S. B. Crampton
Department of Physics and Astronomy
Williams College
Williamstown, Massachusetts

Summary

We have recalculated shifts of the ground state $\Delta m_F=0$ transition of a gas of hydrogen atoms in low magnetic field due to electron spin-exchange collisions between the atoms. We predict a new source of frequency shifts not compensated for by the usual methods of tuning the microwave cavities of oscillating hydrogen maser frequency standards. Near room temperature these additional shifts are small, but large enough to affect stability at the level of $\delta\omega/\omega \approx 10^{-15}$. At very low temperatures these shifts are large compared to the potential thermal instabilities of cryogenic hydrogen maser standards. Above 5 Kelvin they decrease rapidly and so are less severe when using neon storage surfaces near 10 K than when using liquid helium storage surfaces near 0.5 K.

Introduction

Collisions between the hydrogen atoms radiating in a hydrogen maser frequency standard affect the maser frequency in two important ways. They directly shift the transition frequency, and they broaden the radiative linewidth, which increases the frequency pulling due to cavity mistuning. The usual theoretical treatment of hydrogen atom spin-exchange collisions, which treats the energy levels during the collisions as degenerate, predicts that the direct spin-exchange frequency shifts have the same dependence on radiative linewidth as frequency shifts due to cavity mistuning.[1] Tuning the cavity so that the oscillation frequency is independent of radiative linewidth is predicted by that treatment to leave the oscillation frequency independent of collision rate.[2] Such "spin exchange tuning" methods have been important to the development of hydrogen maser standards because drifting hydrogen atom beam intensities induce changing collision rates.

Including the hyperfine structure of the energy levels during collisions to first order, but assuming undeflected classical collision trajectories and ignoring the identity of the colliding atoms, predicts additional direct frequency shifts which leave the spin exchange tuned oscillation frequency offset by an amount proportional to that part of the radiative linewidth not caused by collisions.[3] Measurements of this offset in a room temperature hydrogen maser confirmed within errors a numerical estimate of this effect.[3] The offset predicted by this calculation does not affect the stability of hydrogen maser standards unless something happens to affect that part of the linewidth not due to collisions, such as a change of relaxation by motion of the atoms through magnetic field gradients or a change of relaxation due to interactions with the storage surface.

*Supported by the Stichting voor Fundamenteel Onderzoek der Materie and NSF Grant PHY-840467.

We have recently recalculated the additional direct shifts quantum mechanically, including atom identity.[4] We find new effects which are nonlinear in the collision rate and so produce not only an offset of the spin exchange tuned oscillation frequency, but also a variation of the oscillation frequency with collision rate even after spin exchange tuning. Theoretical details can be found in the recent paper by Verhaar, Koelman, Stoof, Luiten and Crampton, hereafter referred to as (VKSLC).[4] Here we present only the results and then illustrate their implications for hydrogen maser frequency standards operating at room temperature and at cryogenic temperatures.

Spin-Exchange Frequency Shifts

Shifts of the hydrogen maser oscillation frequency due to collisions between the atoms, including direct shifts and shifts dependent on the radiative linewidth, are given by VKS.[4] Eq. (17) as

$$\delta\omega = [\Delta+\alpha\bar{\lambda}_0(1+\Delta^2)-\Omega]/\tau_2 + \Omega/\tau_0 \qquad (1)$$

with Δ the ratio of cavity mistuning to cavity width, α a constant dependent on cavity Q and filling factor, $\bar{\lambda}_0$ the cross section for frequency shifts proportional to the transition level population difference, τ_2 the radiative lifetime (inversely proportional to the linewidth $1/\pi\tau_2$), and τ_0 the partial lifetime without relaxation due to collisions. Ω is

$$\Omega = -\frac{\bar{\lambda}_1(\rho_{cc}+\rho_{aa})+\bar{\lambda}_2}{\bar{\sigma}_1(\rho_{cc}+\rho_{aa})+\bar{\sigma}_2} \qquad (2)$$

with $\bar{\lambda}_1$ and $\bar{\sigma}_1$ the cross sections for frequency shifts and broadening proportional to the sum $(\rho_{cc}+\rho_{aa})$ of level populations involved in the transition and $\bar{\lambda}_2$ and $\bar{\sigma}_2$ the cross sections for frequency shifts and broadening not proportional to level populations.[4] The cross sections are plotted in Fig. 1, and $\alpha\bar{\lambda}_0$ and Ω are plotted in Fig. 2 for particular choices of maser parameters and $(\rho_{cc}+\rho_{aa})$.

Ω generally depends on collision rate in a complicated way via the collision rate dependence of the level population sum $(\rho_{cc}+\rho_{aa})$. Careful measurements of oscillation frequencies, level populations and cavity mistunings can in principle determine the cross sections and the interesting physics which underlies them. However, the important questions for hydrogen maser standards are (1) whether these effects lead to important new sources of frequency instabilities because they couple the oscillation frequency to relaxations due to collisions between the atoms and to other relaxation mechanisms, and (2) whether there are strategies for minimizing these new sources of frequency instabilities.

CH2427-3/87/0000-091 $1.00 C 1987 IEEE

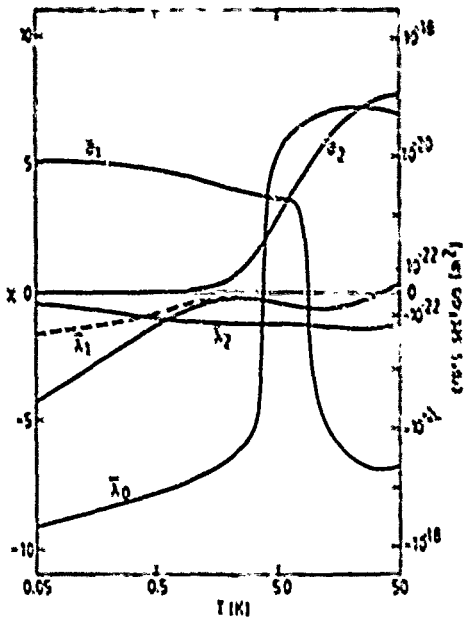


Fig. 1 Thermally averaged values of the various frequency shift and broadening cross sections as functions of absolute temperature. The left hand vertical scale is linear in $\chi = \text{arcsinh}(\bar{\sigma}_1/10^{-22}\ m^2)$ and $\chi = \text{arcsinh}(\bar{\lambda}_1/10^{-22}\ m^2)$. Solid lines, calculation to all orders; dashed line, first order calculation where it differs significantly from the calculation to all orders.

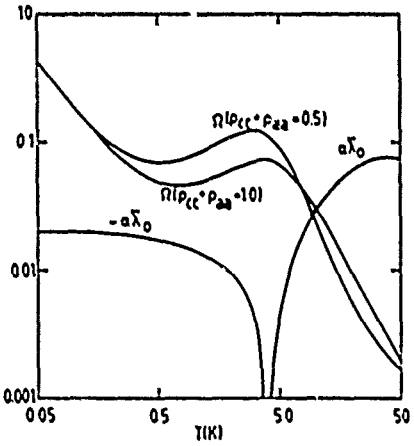


Fig. 2 The dimensionless frequency offset parameters $a\bar{\lambda}_0$ and Ω as functions of absolute temperature. Ω is given for $\rho_{cc}+\rho_{aa}$ assumed constant at 0.5 and 1.0, and $a\bar{\lambda}_0$ is given for a typical choice of cavity Q and filling factor.

Room Temperature Maser Standards

We illustrate these effects with simulations of a maser near room temperature having the operating parameters of the maser used in the experiment reported in Ref. 3 and our own preliminary values for the collision cross sections at 300 degrees Kelvin. We ignore relaxations due to magnetic field gradients and collisions with the surfaces, and we use rate equations for the evolution of level populations due to collisions assuming degenerate states during collisions.[5] Fig. 3 displays a simulation of the oscillation power level and oscillation frequency as the hydrogen atom collision rate is varied by varying the input beam intensity. $1/\tau_2$ provides a convenient measure of relaxation rates, including here only atom flow and collisions, because it can be determined from changes of $\delta\omega$ with A, as shown by Eq. (1) above.

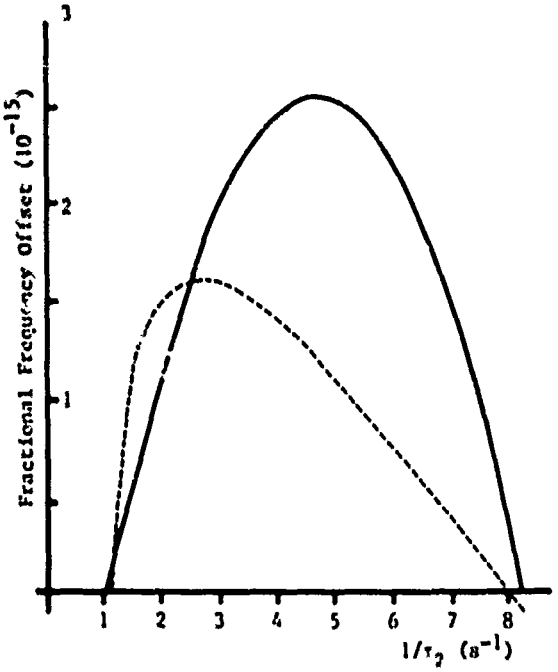


Fig. 3 Solid line: oscillation power level (arbitrary units) plotted against $1/\tau_2$. Self-sustained maser oscillation is obtained for $1/\tau_2$ ranging from 1.16 s^{-1} to 8.18 s^{-1}. Dashed curve: variations of fractional oscillation frequency offset $\delta\omega/\omega$ (from a base offset $\delta\omega/\omega = 5.5\times10^{-15}$) with $1/\tau_2$ over the full range of oscillation.

The Fig. 3 variation of oscillation frequency with $1/\tau_2$ is for a case when the cavity has been tuned so that the oscillation frequency is the same at the minimum and maximum collision rates at which self-sustained oscillation can be obtained. The fractional frequency offset at those two end points is $\delta\omega/\omega = 5.49 \times 10^{-15}$, and the offset varies nonlinearly between those two points because of the τ_2 dependence of Ω. Using a large range of τ_2 variations to set the cavity tuning does minimize the uncertainty of cavity tuning due to thermal fluctuations of the oscillation frequency at the two collision rates chosen as tuning points,[2] but it leaves a fractional variation of frequency with collision rate of order 10^{-15} per Hz of radiative linewidth at high collision rates and an even steeper variation of frequency with collision rate at collision rates just above the threshold for oscillation. In addition, the overall fractional frequency offset varies with $1/\tau_0$, so that a change of τ_0 due to a change of some relaxation rate produces a fractional frequency change at the 10^{-15} level, in addition to changes of any direct frequency shifts caused by those relaxations.[6] The effects due to $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are just large enough to be significant for hydrogen maser standards under some conditions, but they are difficult to detect because of the long

averaging times required to make measurements at the 10^{-15} to 10^{-13} level with room temperature masers. Alternatively, there may be strategies that will minimize these effects. For example, Fig. 4 displays the variations of oscillation power and frequency for the same parameters except that a higher collision rate has been chosen for the lower of the two tuning point collision rates. The precision of the cavity tuning using the tuning points displayed in Fig. 4 is less by about 1/4, but the variation of oscillation frequency with collision rate is only of order 10^{-16} if the collision rates are restricted to values between the two tuning points. Alternatively, the overall frequency offset might be reduced by tuning the cavity so that the oscillation frequency is independent of τ_s, but at the cost of greatly increasing the dependence of oscillation frequency on collision rate.

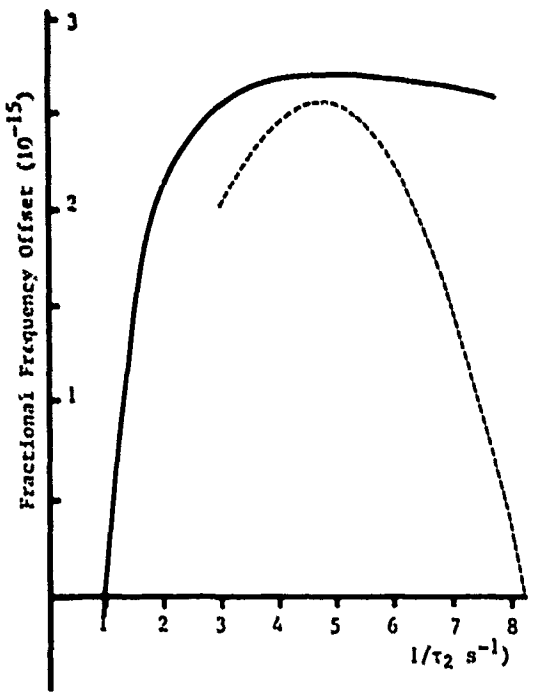


Fig. 4 Dashed curve: oscillation power level (arbitrary units) plotted against $1/\tau_2$ only for values of $1/\tau_2$ between the two collision rates used for tuning. Solid curve: variations of fractional oscillation frequency offset $\delta\omega/\omega$ (from base offset $\delta\omega/\omega = 5.5 \times 10^{-15}$) with $1/\tau_2$ over the full range of values for which there is oscillation.

Adding additional relaxations due to magnetic field gradients and collisions with the storage surfaces does not affect the results qualitatively. Overall fractional frequency offsets remain of order 10^{-14}, and deviations of fractional frequency offsets with collision rate remain of order 10^{-15}, but these can be reduced significantly by using smaller ranges of collision rates for cavity tuning.

Liquid Helium Temperature Maser Standards

Although the three liquid helium storage surface hydrogen masers developed to date[7-9] have not achieved numbers of radiating atoms as nigh as those in room temperature standards, higher numbers of atoms are essential to achieving the improvements of short term frequency stability potentially possible because of the decrease of spin exchange relaxation cross sections with decreasing temperature[10-13]. We assume that standards having atom densities high enough to produce large collision contributions to $1/\tau_2$ will be developed. Fig. 5 displays the results of a simulation assuming atom flow rate 1 s^{-1} and maximum collision rates high enough to increase $1/\tau_2$ to 8 s^{-1}.

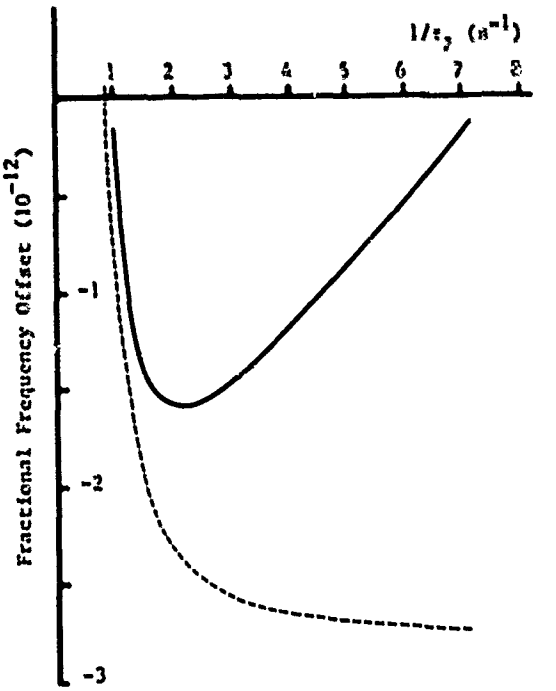


Fig. 5 Dashed curve: variations of fractional oscillation frequency offset $\delta\omega/\omega$ (from a base offset -5.5×10^{-12}) with $1/\tau_2$ when the maser has been tuned for equal oscillation frequencies at the minimum collision rate for oscillation and the maximum collision rate available. Solid curve: variations of fractional oscillation frequency offset when the maser is tuned using a higher collision rate for the lower of the two tuning point collision rates.

The Fig. 5 behavior is similar to that in Fig. 3 and Fig. 4, except that the sign of the frequency offset is inverted and the scale is three orders of magnitude larger. The problems posed by these new sources of frequency shifts are most severe if only a small variation of $1/\tau_2$ can be made by varying the collision rate. The problems are much less severe if collision rates dominate $1/\tau_2$, but they are still very large when compared to the potential thermal instabilities of cryogenic masers and even when compared to the stabilities of current room temperature standards. If liquid helium surface hydrogen masers are to be competitive as frequency standards, design studies must take into account these new sources of frequency shifts.

Middle Ground: Neon Surface Masers

Similar simulations using the parameters of neon surface hydrogen masers operating near 10 Kelvin reveal behavior that is qualitatively the same as the 0.5 K behavior, except that the scale is less by a factor 25 because of the decrease of the spin-exchange frequency shift cross sections with increasing

temperature. However, the spin-exchange relaxation cross sections are still low enough that collisions are not likely to limit the radiated power at achievable hydrogen atom beam intensities. If solid neon surfaces that are as stable as liquid helium surfaces can be developed, the neon surface hydrogen masers offer a technologically attractive alternative to the liquid helium surface masers as frequency standards.

References

1. S. B. Crampton, Phys. Rev. 158, 57 (1967).
2. 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).
3. S. B. Crampton and M. T. M. Wang, Phys. Rev. A12, 1305 (1975).
4. B. J. Verhaar, J. M. V. A. Koelman, H. T. C. Stoof, O. J. Luiten, and S. B. Crampton, Phys. Rev. A35, 3825 (1987).
5. L. C. Balling, R. J. Hanson, and F. M. Pipkin, Phys. Rev. 133, A607 (1964); 135 AB1 (1964).
6. S. B. Crampton E. C. Fleri, and H. T. M. Wang, Metrologia 13, 131 (1977).
7. H. F. Hess, G. P. Kochanski, J. M. Doyle, T. J. Greytak, and D. Kleppner, Phys. Rev. A34, 1602 (1986).
8. M. D. HUrlimann, W. N. Hardy, A. J. Berlinsky, and R. W. Cline, Phys. Rev. A34, 1605 (1986).
9. R. L. Walsworth, I. F. Silvera, H. P. Godfried, C. C. Agosta, R. F. C. Vessot, and E. M. Mattison, Phys. Rev. A34, 2550 (1986).
10. S. B. Crampton, W. D. Phillips, and D. Kleppner, Bull. Am. Phys. Soc. 23, 86 (1978).
11. A. J. Berlinsky and W. N. Hardy, Proceedings of the Thirteenth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Naval Research Laboratory, Washington, D. C. 1982 [NASA Conference Publication No. 2220, 1982 (unpublished)], p. 547.
12. R. F. C. Vessot, M. W. Levine, and E. M. Mattison, Proceedings of the Ninth Annual Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Greenbelt, Maryland, 1977 [NASA Technical Memorandum No. 78104, 1978 (unpublished)], p. 549.

SURFACE INTERACTION OF ATOMIC HYDROGEN WITH TEFLON

Edward M. Mattison and Robert F. C. Vessot
Smithsonian Astrophysical Observatory
Cambridge, Massachusetts 02138
Colin Bain, Stephen Wasserman, and George Whitesides
Harvard University Department of Chemistry
Cambridge, Massachusetts 02138

Summary

We studied the process of deterioration of a hydrogen maser's Teflon storage bulb coating, and examined the coating's physical and chemical characteristics. The deterioration, which accompanied exposure to the maser's atomic hydrogen beam, was marked by decreasing maser line Q and microwave output power. The likelihood of Teflon degradation by ultraviolet radiation from the beam source was eliminated. Teflon from the bulb, and Teflon from the same original batch, were compared with Teflon from a fresh batch. The old Teflon, which had been stored for several years, had more polarizable surface groups, more surface oxygen (of unknown chemistry), and more non-fluorinated carbon than the new Teflon. The fresh Teflon was applied to a new storage bulb and operated without deterioration. Chemical changes that occurred during storage appear to have caused the Teflon surface to deteriorate when exposed to the atomic hydrogen beam.

Introduction

The operation of the hydrogen maser, and its superior frequency stability, depend upon the ability to confine hydrogen atoms in a limited region of space for an extended length of time[1]. In present day active masers, the atoms are confined for periods on the order of a second in a quartz bulb lined with Teflon[2]. Masers produced by the Smithsonian Astrophysical Observatory (SAO) use a bulb coating of FEP-120 Teflon, a fluorinated ethylene-propylene copolymer. The Teflon coating makes the maser action possible; at the same time, however, it limits the hydrogen maser's absolute frequency accuracy, due to variability in the wall shift[3], and represents a potential source of performance degradation. We describe the process of degradation observed in a maser, and the methods used to study the nature of the coating in order to understand both the normal and abnormal operation of the maser.

Degradation of a maser's storage coating manifests itself in decreased microwave output power from the maser's resonant cavity, and in decreased line Q. The line Q, Q_ℓ, is inversely proportional to γ_2, the total rate of loss of magnetization of the oscillating ensemble of atoms, which can be written

$$\gamma_2 = \gamma_b + \gamma_r + \gamma_{se} + \gamma'$$

where γ_b is the rate of loss of atoms from the bulb aperture, γ_r is the rate of loss of H atoms on the storage surface due to recombination and other processes, γ_{se} is the relaxation rate of the magnetization due to spin exchange, and γ' includes all other relaxation processes, including magnetic relaxation and other storage wall relaxation processes. Contamination or other degradation of the wall coating increases γ_r, thus decreasing Q_ℓ for constant spin exchange relaxation. In order to compare conditions of the wall coating, it is necessary to determine the line Qs for the same value of γ_{se}; we do this by normalizing the line Q to a constant value of maser output power. Power is measured by an intermediate frequency (IF) level in the maser's microwave receiver.

Initial Maser Performance

When the maser was constructed, it had an excellent initial line Q of approximately 2×10^9. Shortly after initial maser operation the line Q and output power began to decrease. The storage bulb was removed from the maser four times and recoated with Teflon from the batch used for the original coating. Before each recoating, the bulb was tested with a drop of methanol rolled on its interior surface. The methanol adhered to a spot on the apex of the bulb, opposite the collimator tube through which the atomic beam enters. The wetted spot was crescent-shaped or circular, and approximately 10 mm in diameter, a shape consistent with the image size of the beam source. Wetting indicates that the surface has active sites capable of binding atoms and molecules, and thus possibly capable of enhancing the recombination of hydrogen atoms.

The coating process consists of rolling the Teflon dispersion in the bulb to coat the inner surface, pouring out the excess and drying the bulb in vacuum, and firing the bulb at 365°C for approximately ten minutes with oxygen flowing in the bulb to burn off any hydrocarbons in the dispersion, chiefly surfactants used to enhance wetting. The Teflon used to coat the bulb had been stored for several years and was partially agglomerated; it needed to be stirred thoroughly before being applied to the bulb in order to disperse the Teflon solids throughout the liquid.

Following each coating the line Q began with a high initial value and decreased with time. We assumed that the major source of degradation was the wetting spot, and had two initial hypotheses for its creation: either damage to the Teflon by ultraviolet light (UV) coming from the hydrogen beam source, or contamination of the surface by foreign molecules entering the bulb with the beam. Although the hydrogen beam path is blocked by a stopping disc in the state selection magnet to remove undeflected molecules and UV light, UV could conceivable reach the bulb by grazing reflection from the pole faces of the hexapole state selection magnet. To minimize this possibility we replaced the state selection magnet with one from a maser that had operated successfully, and we coated its pole faces with colloidal graphite (Aerodag), a UV absorber. We took strenous measures to eliminate the possibility of contamination. We acid etched and baked the Pyrex dissociator bulb, vacuum baked the hydrogen source system, and finally replaced the entire source with a new system whose components had been individually cleaned and vacuum baked before assembly. The result was a source system that was as clean as we could make it, and that was at least as clean as sources performing successfully in other SAO masers.

Maser Performance as a Function of Time

In November, 1986, we began a careful series of measurements of line Q and maser power as functions of time and hydrogen beam exposure. The data are shown in Figs. 1 and 2. At intervals we measured the line Q at a variety of hydrogen flux settings, and thus maser power levels, in order to be able to interpolate Q_ℓ to a constant IF level. The line Qs given in Fig. 1 are the measured values normalized to a reference IF level of 1.80 volts, corresponding to −99.2 dBm of 1.4 GHz radiation from the maser's resonant cavity.

During most of the observation period the line Q decreased with exposure to the atomic hydrogen beam, and approached a constant value for a given hydrogen flux. The saturation value was generally lower for greater flux intensity. To test the effect of the atomic hydrogen beam, we turned off the hydrogen dissociator for several periods of up to two weeks. During these periods we turned the dissociator on for intervals of a few minutes to measure the line Q. When the atomic hydrogen beam was off, the line Q either remained constant or increased. This behavior indicated that the deterioration in line Q was associated with the hydrogen beam, and

CH2427-3/87/0000-095 $1.00 C 1987 IEEE

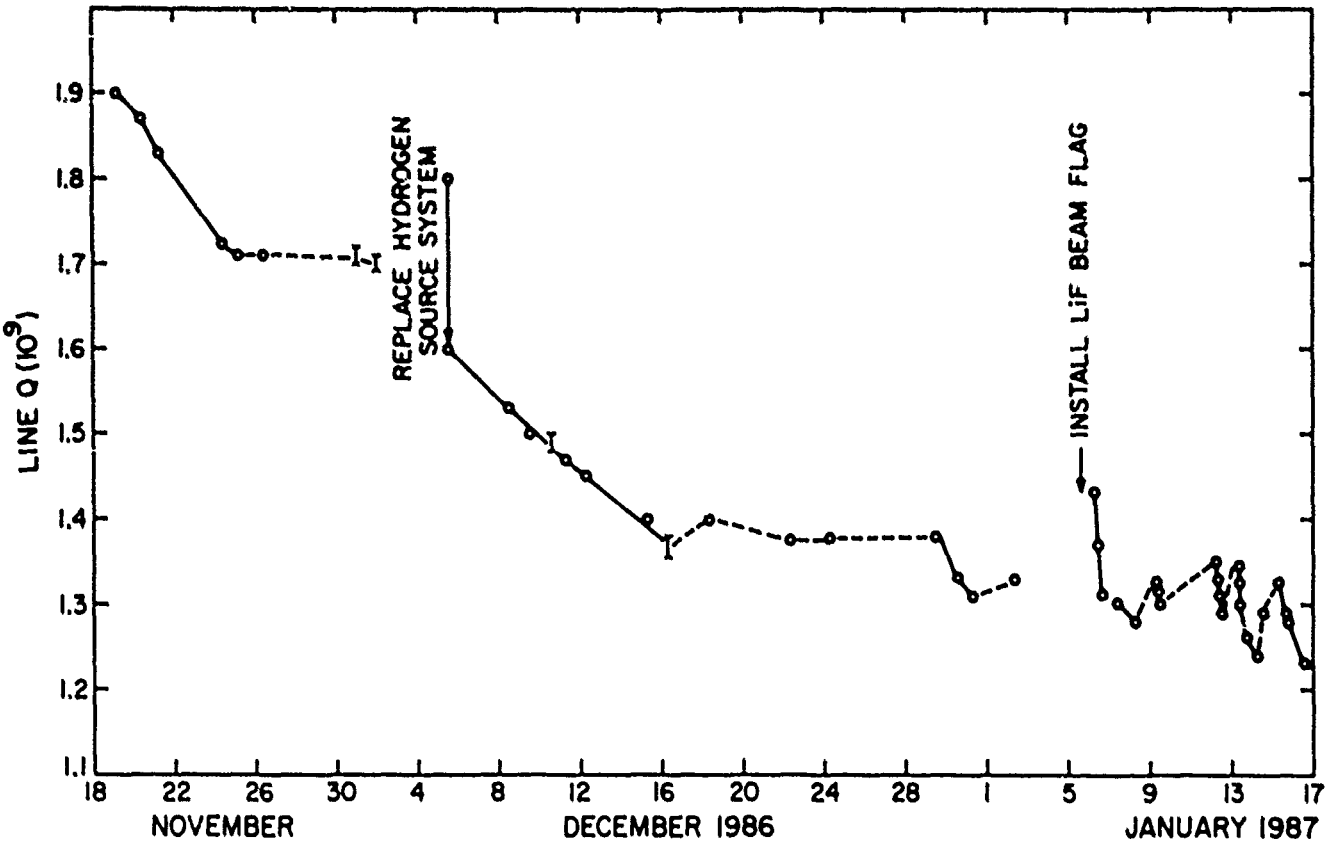


Figure 1. Initial Line Q Variation of Maser, November 1986—January 1987

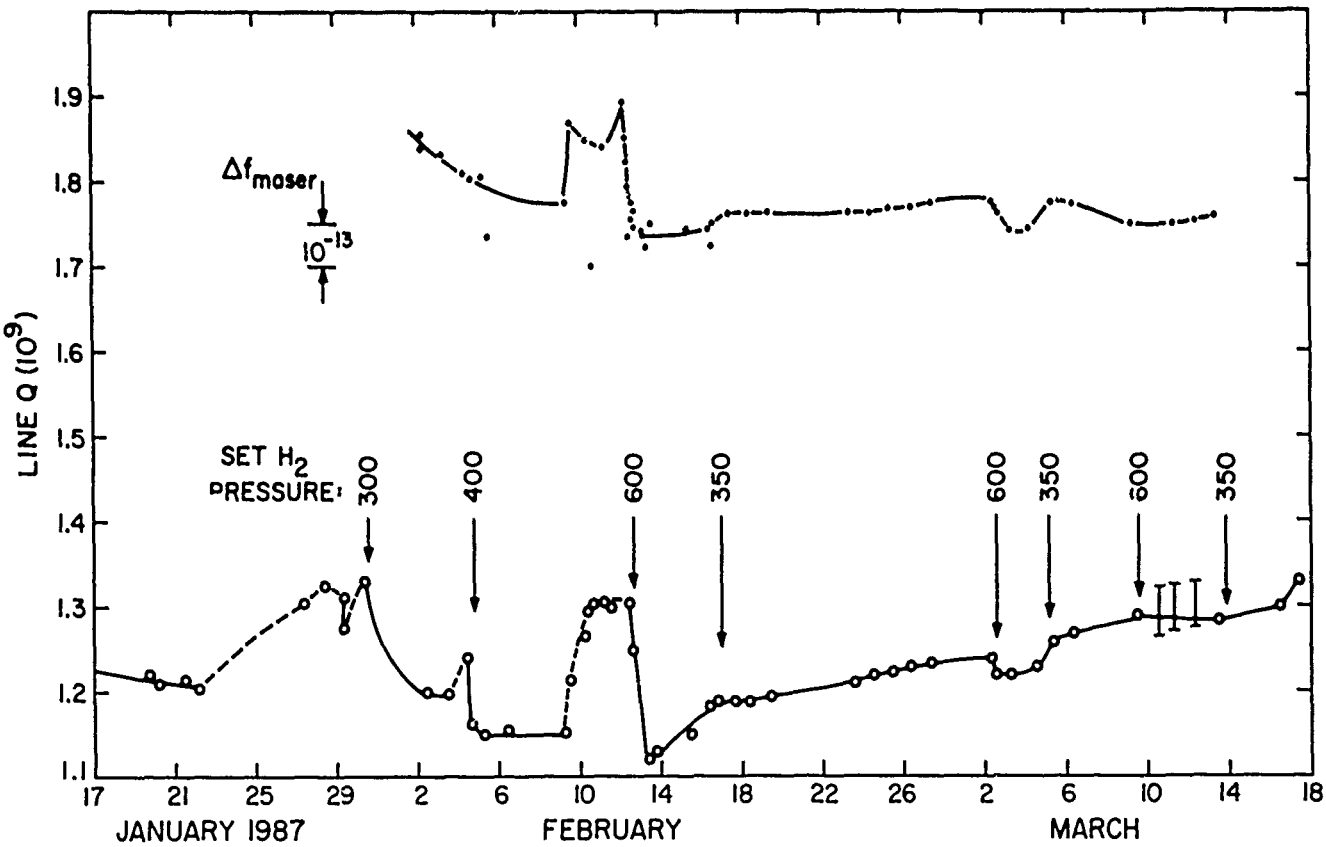


Figure 2. Line Q and Frequency Variation, January—March 1987

was not being caused by contamination due to outgassing from other parts of the maser.

In order to investigate the possible effect of UV radiation, we installed a 0.7mm thick LiF window on a movable arm in the maser, so that the window could be moved to occult the beam. The window's transmission was approximately 80% for Lyman alpha UV radiation (122 nm wavelength). The dissociator and hydrogen were left on without interruption, so that when the LiF window blocked the hydrogen beam, it did not cut off UV from the source that would otherwise have reached the storage bulb. The line Q again increased when the LiF window cut off the beam, indicating either that no significant amount of UV was present or that UV did not degrade the Teflon. (Since the Lyman photons produced in the dissociator have energies on the order of 10 eV, they have more than than enough energy to disrupt the chemical bonds in Teflon; thus this test shows that UV was not present.)

In an effort to determine how the line Q saturation level varied with hydrogen flux, we raised the flux setting to a high level on Friday, February 13, 1987. Q_ℓ first decreased to 1.12×10^9, but then, anomalously, began to increase. Four days later we decreased the flux to a setting of 350, and the line Q continued to increase. We repeated this test the next month with similar results. This behavior was inconsistent with any previous observations. During this period we measured the maser's frequency and wall shift. As shown in Fig. 2, the maser frequency followed the movement of the line Q, increasing with increasing Q_ℓ. The magnitude of the wall shift, whose sign is negative, was approximately 3.4×10^{-12} greater than that of other masers.

We terminated the tests and replaced the storage bulb with a new bulb coated with Teflon from a fresh batch. The maser's initial line Q using this bulb and coating, but the previous source structure, was 2.3×10^9 for a maser output power of −99.6 dBm. Over several months it decreased to a value of 1.5×10^9 and stabilized at that level. This is higher than the minimum value of $1.1-1.2\times10^9$ obtained with the old Teflon. The magnitude of the wall shift for the new coating is within a few parts in 10^{13} of its normal value. We coated an additional bulb with the new Teflon for use in another maser; its line Q has remained constant at approximately 2.0×10^9, and its wall shift is within 2 to 5 parts in 10^{13} of normal. These observations, particularly the difference in wall shift between the old and new Teflons, show that the coatings differ in performance. (The initial line Q drop of the new Teflon in the original maser may, however, indicate a residual effect of unidentified processes associated with the maser.)

Examination of Teflon

We cut the original storage bulb into pieces and examined the Teflon from several areas of the bulb surface by physical and chemical means. In addition, we coated extremely clean silicon wafers with Teflon from the old and new batches, using identical procedures, and tested those surfaces.

Scanning electron microscopy with a resolution of 100 angstroms showed surfaces with a low-relief "orange-peel" texture and occasional fissures. No difference in appearance between the old and new Teflon surfaces was discernible.

The polarizabilities of the Teflon surfaces were compared by measuring the contact angles between the surfaces and droplets of water, hexadecane, and ethanol. The values are given in Table 1. The results show that the old Teflon, both from the bulb and cast on the Si wafer, has more polarizable surface groups than the new Teflon. Furthermore, the Teflon coating from the bulb, both from the bulb's side and from the damaged spot at the apex, is more polarizable than the old Teflon that was not exposed to the hydrogen beam. Finally, the damaged spot is characteristic of an amorphous, unfluorinated hydrocarbon film.

Table 1. Contact Angles
(degrees)

	Water	Hexadecane	Ethanol
New Teflon cast on Si	120.5±1	58±1	42±1
Old Teflon cast on Si	118.5±1	58±1	43±1
Bulb interior (side)	116. ±2	54±1	42±3
Area of maximum damage	90	0	0

Samples of Teflon were scraped from the side of the bulb and from the new Teflon film on the silicon wafers, and were analyzed for the fractional composition of elements. The percent of carbon, fluorine, and hydrogen in the samples are shown in Table 2. The stated precision of the measurements is ±0.3% for carbon and hydrogen, and ±0.4% for fluorine. The bulk elemental analysis shows no significant difference between the old and new Teflon, within the precision of the technique.

Table 2. Elemental Analysis of Bulk Samples
(percent)

	C	F	H
Expected for Teflon	24.02	75.98	0.00
Film removed from bulb	23.08	75.80	0.06
	24.11	76.17	0.06
New Teflon cast on Si	23.97	75.76	0.03
	24.17	75.79	0.00

Samples of the Teflon surfaces were examined by x-ray photoelectron spectroscopy (XPS), sometimes referred to as ESCA. This technique samples the surface to a depth of approximately 30 angstroms, and is more sensitive to compositional differences than is the bulk elemental analysis. The results of the XPS measurements are given in Table 3. They show (a) that the surfaces of the old Teflon contain a small fraction of oxygen, whose chemical bonding could not be identified, while the new Teflon contained no observable oxygen; and (b) that the old Teflon surfaces contain more non-fluorinated carbon than the new Teflon. 80% of the carbon at the bulb's damaged spot was non-fluorinated. Fig. 3 compares plots of the XPS spectra for carbon for the new Teflon surface and the damaged spot.

Table 3. Atomic Compositions from XPS
(percent)

	F	O	C	Composition of C: CF_x	Composition of C: other C
New Teflon cast on Si	67	0	33	99	1
Old Teflon cast on Si	66	0.3	34	97	3
Side of maser bulb	64	0.3	36	96	4
Damaged ring in bulb	32	3	65	20	80

We conclude from these measurements that the old and new Teflons differ in their surface chemistry. This is particularly evident in the variations in their wall shifts. Since the old Teflon had been used successfully to coat maser bulbs several years ago, it is likely that the differences developed as a result of chemical processes during storage. While small, these differences may be

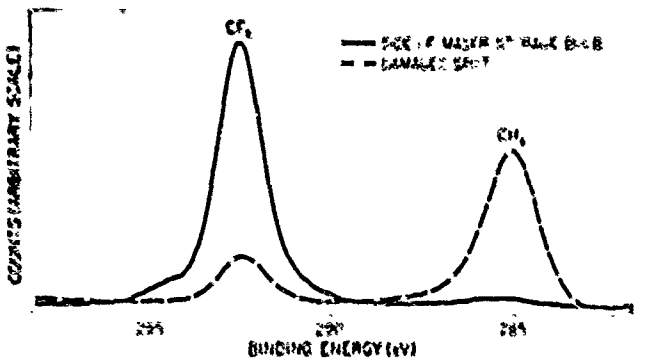


Figure 3. XPS spectra of Carbon 1S in Old and New Teflon

sufficient to degrade the maser's performance, and argue for obtaining a fresh batch of Teflon for each set of bulb coatings.

Observations on Another Storage Bulb

The significance of the damaged spot at the apex of the storage bulb was unclear. We had observed similar spots, which wet with water, methanol, and ethanol, in other masers with low or decreasing line Qs. We had the opportunity to examine the storage bulb of a maser that had operated for over 7 years and that had a high line Q, 2.3×10^9. This bulb also had a similar spot at its apex that wet with ethanol, showing that such a spot is not necessarily indicative of improper performance. Such a conclusion is reasonable. First, the wetting spot has the characteristics of a hydrocarbon film, which can be a moderately effective maser coating; prior to the use of Teflon, hydrocarbon films such as Drifilm (dimethyldichlorosilane) were used to coat storage bulbs. Second, the fact that the maser oscillated at all showed that a large number of hydrogen atoms survived their initial impact with the damaged spot. Once an atom bounces from the spot and begins its random travels about the bulb, its likelihood of hitting the spot again is equal to the fraction of the bulb area occupied by the spot, approximately 7×10^{-4}. In order to significantly relax the oscillating atoms, the damaged spot would have to have a sticking probability approaching unity, implying that the incoming beam would be quenched before oscillation could begin. We conclude that the line Q and wall shift are affected primarily by the condition of the entire bulb coating rather than by the damaged spot at the apex.

Discussion

Our investigation did not determine the mechanism for the time-varying degradation and recovery of the coating properties. The time scale for this variation is on the order of hours. This fact eliminates the possibility of hydrogen adsorption on the Teflon surface, because adsorption by means of weak van der Waals trapping would have desorption times on the order of nanoseconds, while chemical bonding would be essentially permanent, with bonds lasting on the order of years to millennia. A possible mechanism is diffusion of hydrogen atoms into the three-dimensional matrix of Teflon molecules that make up the coating.

Another open question is the cause of defluorination of the apex spot. A candidate is hydrogen atoms in the 2S state, which have a lifetime on the order of 0.14 seconds and more than enough energy to break chemical bonds. These excited atoms are quite fragile, however, and can be quenched by electric fields produced by their motion through the state-selection magnet.

Acknowledgement

We are pleased to acknowledge support from the Smithsonian Institution's Scholarly Studies Program and from the Harvard University Chemistry Department, and we thank Professor Stuart Crampton for helpful discussions.

References

1. Kleppner, D., H.C. Berg, S.B. Crampton, N.F. Ramsey, R.F.C. Vessot, H.E. Peters, and J. Vanier, Phys. Rev. 138, 972 (1965).

2. Teflon is a registered trademark of E.I. Dupont DeNemours, Inc.

3. Hellwig, H., R.F.C. Vessot, M.W. Levine, P.W. Zitzewitz, D.W. Allan, and D.J. Glaze, IEEE Trans. on Instr. and Meas., IM-19, 200 (1970).

41st Annual Frequency Control Symposium - 1987

CHARACTERISTICS AND SOURCES OF PHASE NOISE IN STABLE OSCILLATORS

T.E. Parker

Raytheon Research Division
131 Spring St.
Lexington, MA 02173
(617) 860-3054

Abstract

The frequency stability of an oscillator is a very important characteristic for many applications. Yet the causes and sources of some basic types of noise are poorly understood, particularly for close-to-carrier noise. A review is presented in this paper of the present state of knowledge about the sources and characteristics of frequency fluctuations in stable oscillators using quartz acoustic resonators (BAW and SAW) or dielectric resonators as the high Q frequency stabilizing element. A brief discussion of the various parameters used to quantify random frequency fluctuations is presented along with the relative merits of open- and closed-loop phase noise measurements. Phase noise in stable oscillators usually arises from additive voltage fluctuations and direct parameter modulation processes. Additive noise, such as Johnson noise, is reasonably well understood and will be discussed only briefly. On the other hand, modulation noise processes, such as flicker and random walk frequency noise, are only poorly understood and will be covered in more detail. The observed levels of flicker noise in amplifiers and resonators will be discussed, as well as the dependence of flicker noise level on resonator frequency, loaded and unloaded Q, and oscillator loop power level. Procedures for minimizing flicker frequency noise levels in oscillators will be presented. Finally, the known characteristics of random walk frequency noise in oscillators will be covered briefly.

Introduction

The frequency stability of an oscillator is an important characteristic for many applications. Yet the causes and sources of several basic types of noise are poorly understood, particularly for close-to-carrier noise. However, over the last few years there has been a significant increase in experimental data relating to oscillator noise and some general characteristics are beginning to become evident. This paper presents a review of the present state of knowledge about the sources and characteristics of frequency fluctuations in stable oscillators using quartz acoustic resonators (BAW and SAW) or dielectric resonators as the high Q stabilizing element. The discussion will be confined to free running, simple oscillators (no phase-locking, external filtering or buffer amplifiers) and will not include vibration effects. This will be essentially an update to Leeson's[1] discussion of oscillator noise and will be presented in the context of a simple feedback oscillator.

There are a number of parameters that are used to quantify random frequency or phase fluctuations[2] and a brief summary will be presented first. The spectral density of frequency fluctuations, $S_{\Delta F}(f)$, is the mean square frequency fluctuation in a 1 Hz bandwidth at the noise or carrier offset frequency f and is given in units of Hz^2/Hz. If $S_{\Delta F}(f)$ is divided by the square of the oscillator frequency, F_o^2, the normalized spectral density of frequency fluctuations, $S_y(f)$, is obtained.

$$S_y(f) = S_{\Delta F}(f)/F_o^2 \qquad (1)$$

This parameter has the advantage that it is invariant under frequency multiplication. Another commonly used parameter is the spectral density of phase fluctuations, $S_\phi(f)$, which is the mean square phase fluctuation in a 1 Hz bandwidth and is given in units of radians squared per Hz. $S_{\Delta F}(f)$ and $S_\phi(f)$ are related by the expression

$$S_\phi(f) = S_{\Delta F}(f)/f^2. \qquad (2)$$

Note that $S_{\Delta F}(1) = S_\phi(1)$ at f=1 Hz . If $S_\phi(f) << 1$, a small angle approximation can be made and $S_\phi(f)$ can be interpreted as the double-sideband noise-to-carrier ratio. The quantity $\mathcal{L}(f)$ is used to denote the single-sideband noise-to-carrier ratio[3] in dBc/Hz, and is expressed in terms of $S_\phi(f)$ by

$$\mathcal{L}(f) = 10\ \mathrm{Log}(S_\phi(f)/2). \qquad (3)$$

CH2427-3/87/0000-099 $1.00 C 1987 IEEE

Strictly speaking, $\mathcal{L}(f)$ should only be used if its value is less than approximately -20 dBc/Hz, but it is commonly used even for much higher values. $S_\phi(f)$ and $S_{\Delta F}(f)$ are, of course, not limited to small values, since it is only the interpretation of phase (or frequency) fluctuations as FM sidebands that is limited to small values.

Another parameter commonly used to quantify random frequency fluctuations is the two-sample, or Allan[1], variance, $\sigma_y^2(\tau)$. This parameter is the average value of one half the square of the fractional change in frequency between two adjacent frequency measurements, each made in a time interval τ. Actually, the parameter more commonly referred to for an oscillator is the square root of the Allan variance, $\sigma_y(\tau)$.

In most free running (non-phase locked) oscillators the spectral densities have a characteristic power law dependence on the noise (or offset) frequency, f. This dependence is illustrated in Fig. 1 for $S_\phi(f)$ and $S_{\Delta F}(f)$. Generally, not all of the sections are present in a single oscillator, and rarely are the f^{-2} (white frequency) and f^{-1} (flicker phase) sections present simultaneously. For each characteristic slope in Fig. 1, values of $\sigma_y(\tau)$ can be calculated[2] from either $S_\phi(f)$ or $S_{\Delta F}(f)$.

PBN-85-111

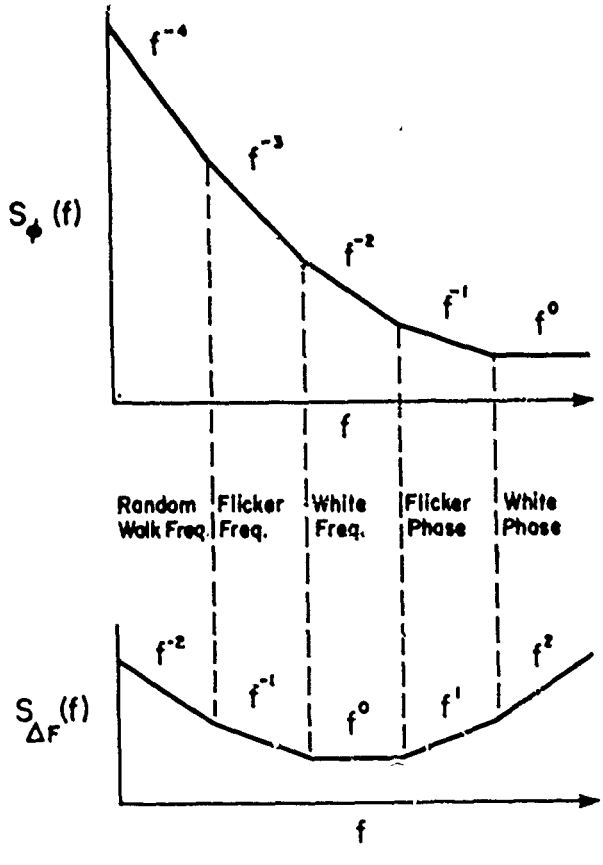


Figure 1. Power law dependence of spectral densities.

Figure 2 shows that $\sigma_y(\tau)$ has a power law dependence on τ which is analogous to that of the spectral density dependence on f. The power law dependence on f (or τ) is generic in nature and for individual oscillators the exponents may not be exactly integers (or half-integers), but may vary above or below the indicated values.

PBN-85-110A

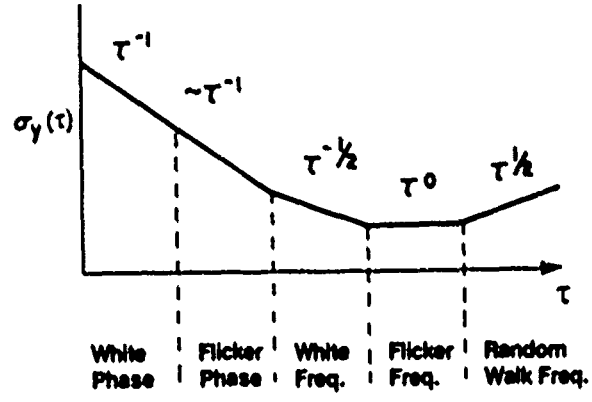


Figure 2. Power law dependence of the fractional frequency deviation.

Sources of Phase Noise

The basic feedback model of an oscillator used by Leeson is very useful in understanding the characteristics of phase noise in oscillators. Figure 3 illustrates this model in both the open- and closed-loop conditions. In the open-loop condition, the oscillator components will impart phase fluctuations to a signal which is passed through them at the intended oscillator frequency and power level. To distinguish open-loop phase fluctuations from closed-loop phase noise, the symbol, ', will be added to the open-loop spectral densities. Thus, $S'_\phi(f)$ represents the open-loop spectral density of phase fluctuations and $\mathcal{L}'(f)$ represents the open-loop single-sideband noise-to-carrier ratio. When the loop is closed, oscillation will occur if two conditions are met. First, the loop amplifier must have sufficient small signal gain to overcome the loss in the other loop components, and second, the phase shift around the loop must be an integer multiple of 2π at a frequency for which the first condition is satisfied. This approach to oscillator design has a very practical significance in that the noise characteristics of the oscillator's components as measured in an open-loop configuration have a direct bearing on the closed-loop phase noise of the oscillator.

As illustrated in Fig. 3, the noise level far from the carrier is the same in both the open- and closed-loop conditions and

PBN-79-344A

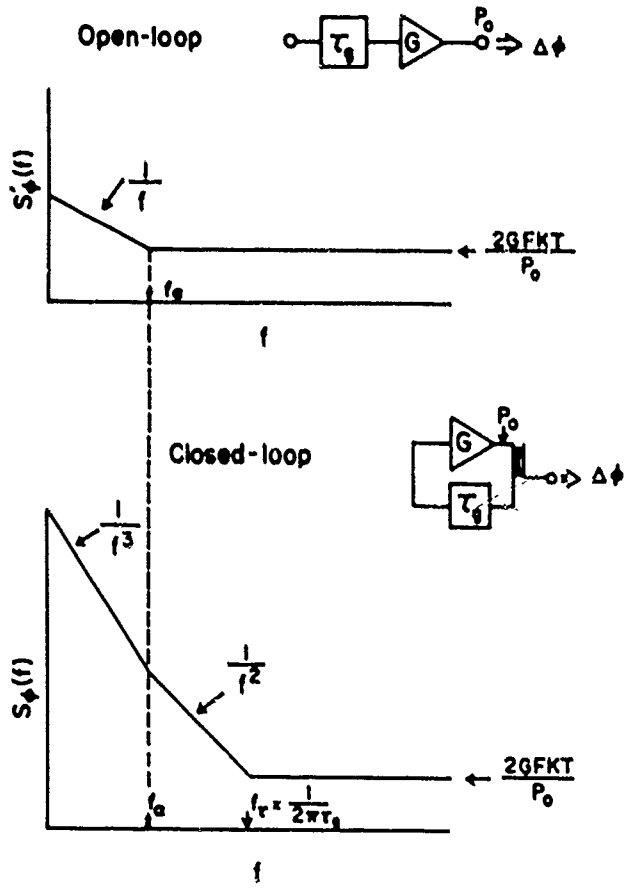


Figure 3. Open- and closed-loop configurations of a feedback oscillator.

is independent of noise frequency. However, as the noise frequency gets smaller (closer to the carrier), the closed-loop noise level, $S_\phi(f)$, begins to deviate from the open-loop level, $S'_\phi(f)$. This occurs when the noise frequency falls within the 3 dB bandwidth of the resonant device. The frequency change caused by a phase change can be calculated from the phase condition for oscillation

$$\phi_R + \phi_E = N2\pi, \qquad (4)$$

and the phase slope or group delay of the resonator

$$\tau_g = \Delta\phi/(2\pi\Delta F_o), \qquad (5)$$

giving

$$\Delta F_o = (1/(2\pi\tau_g))\Delta\phi_E. \qquad (6)$$

The oscillator frequency must change in order for the phase shift through the resonator, ϕ_R, to compensate for the change in phase shift, $\Delta\phi_E$, in the other components. τ_g can be replaced in Eq. 6 with the loaded Q, Q_L, by using the relation

$$\tau_g = Q_L/\pi F_o \qquad (7)$$

to give

$$\Delta F_o = (F_o/2Q_L)\Delta\phi_E. \qquad (8)$$

Equation 8 can be expressed in terms of the spectral density of frequency fluctuations

$$S_{\Delta F}(f) = (F_o/2Q_L)^2 S'_{\phi E}(f), \qquad (9)$$

or, by using Eq. 2, as a spectral density of phase fluctuations.

$$S_\phi(f) = (F_o/2Q_L)^2 S'_{\phi E}(f)/f^2 \qquad (10)$$

The result is that inside the 3 dB bandwidth of the resonator ($f < F_o/2Q_L = f_\tau$) the closed-loop noise begins to rise at 20 dB/decade if the open-loop noise is white. This is illustrated in Fig. 3.

The phase noise in oscillators comes from two fundamentally different sources, namely additive voltage fluctuations and direct parameter modulation. Additive noise is the familiar Johnson, or thermal noise, and will be discussed only briefly.

Additive Noise

Additive noise is generally caused by thermally generated voltage fluctuations that are added to the carrier signal and result in phase and amplitude fluctuations. As illustrated in Fig. 4, these voltage fluctuations exist at RF frequencies (F_o +/-f) and

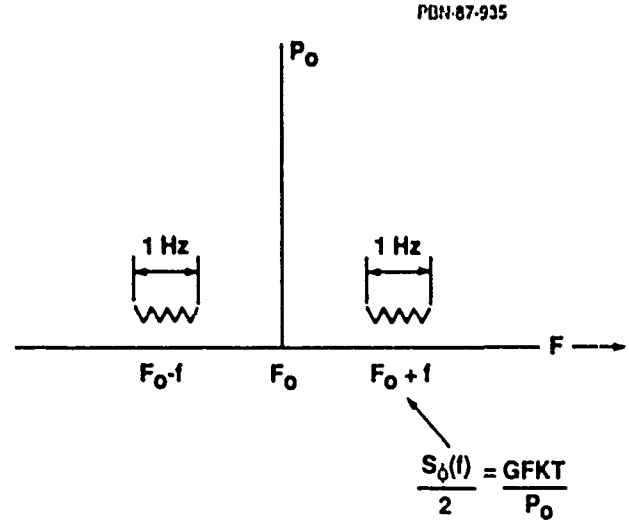


Figure 4. Additive thermal voltage fluctuations.

are simply added to the carrier frequency. This results in both AM and FM modulation[5]. However, the amplitude limiting function that is always present in an oscillator suppresses the AM noise and may cause some AM to PM conversion. Though there is some disagreement over factors of 2 in the level of the phase noise[1,6,7] the author has generally found Leeson's expression to give the best results.

$$S'_{\phi E}(f) = 2GFKT/P_o \qquad (11)$$

Here, G is the compressed power gain of the loop amplifier, F is the noise figure of the amplifier, K is Boltzmann's constant, T is the temperature in °K, and P_o is the carrier power level (in watts) at the amplifier output. For $\mathcal{L}(f)$ (in units of dBc/Hz) this becomes

$$\mathcal{L}(f) = -174 + G + F - P_o, \qquad (12)$$

where G and F are given in dB and P_o is in dBm. For offset or noise frequencies far from the carrier ($f > f_\tau$) this noise level is the same for both the open- and closed-loop conditions. Since Johnson noise is independent of noise frequency, this results in a flat (white) noise spectrum as shown in Fig. 3.

As is well known, the approach to minimizing thermal phase noise is to reduce G, F or KT as much as possible and to increase P_o. Inside the resonator bandwidth the loaded Q becomes a factor (Eq. 10) and it is desirable to have a large loaded Q. However, the resonator insertion loss and loaded Q are related, and one cannot arbitrarily increase Q_L without increasing the insertion loss. This necessitates then a larger amplifier gain G. The two competing effects result in an optimum loaded Q of approximately one half the unloaded Q and an insertion loss of about 6 dB[8]. At the present time, the state-of-the-art for the noise floor level is about -180 dBc/Hz for SAW and BAW resonators above approximately 100 MHz. For lower frequency BAW resonators the level gradually rises and is about 10 to 15 dB higher at 5 MHz due to power handling limitations of the resonators. The power level that can be used with dielectric resonators is higher than for acoustic devices, so dielectric resonator oscillators (DRO's) should be capable of noise floors better than -190 dBc/Hz.

Modulation Noise

The second basic type of noise is modulation noise, and this is the more interesting type since the sources of this noise are poorly understood. The most familiar form of modulation noise is 1/f or flicker noise. Unlike Johnson noise, flicker noise is not usually caused by additive voltage fluctuations but by direct phase or frequency fluctuations in the resonant device or by phase fluctuations in the other electronic components of the oscillator. For example, if the phase shift through the loop amplifier is fluctuating, this results in closed-loop phase noise through Eq. 10. Similarly, if the resonant frequency of the acoustic device fluctuates, the oscillator frequency will track the resonator in the same fashion it tracks the frequency-temperature characteristic of the resonator as long as the noise frequency is inside the resonator bandwidth. Figure 3 shows the effect of 1/f open-loop noise, $S'_\phi(f)$, on the closed-loop noise $S_\phi(f)$.

These phase or frequency fluctuations are undoubtedly of thermal origin, but the precise mechanism or mechanisms are not yet understood. However, the fact that they are direct phase or frequency fluctuations has some important consequences. First, increasing the oscillator power level does not result in a direct reduction in oscillator noise as it does for additive Johnson noise. This was demonstrated by Elliott and Bray in Fig. 2 of reference 9 for a case where the SAW device was the dominant source of 1/f noise. Only at very high incident power levels (>+20 dBm) has it been observed that permanent changes in flicker noise levels occur in SAW devices and this is presumably related to physical changes that take place in the resonator. For amplifiers, $S'_\phi(f)$ has generally been found to be largely independent of the amplifier's power capability, but as will be discussed later, $S'_\phi(f)$ can be increased by the level of gain compression that is present. Thus increasing the oscillator loop power generally has little or no effect on the flicker noise region of the phase noise.

A second consequence of direct modulation noise is that the open-loop phase noise levels, $S'_\phi(f)$, of each component add linearly. Unlike noise figures, where the first stage in a cascade of amplifiers dominates, there is no phase gain, and each component contributes its level of $S'_\phi(f)$ with equal weight. Thus a multistage amplifier will generally have a higher flicker noise level than a single stage amplifier. Also other loop components, such as phase shifters or power dividers, may contribute to the overall open-loop flicker noise level.

A third consequence of direct modulation noise is that the effect on oscillator noise level of increasing the loaded Q will depend on whether the resonant device or the amplifier (or other non-resonant component) is the dominant source of noise. If the amplifier is the larger source, increasing the loaded Q will reduce the oscillator flicker noise level as indicated by Eq. 10. However, if the resonant device is the dominant noise source, changing the loaded Q by impedance matching will have little to no

effect on the noise level. This observation was reported in reference 10 where SAW resonators were the dominant source of flicker noise. The resonator flicker noise is nearly independent of loaded Q for the same reason that the frequency-temperature characteristic of an oscillator is, at most, only weakly dependent on the loaded Q of the resonator.

Open-loop Flicker Noise Levels

Open-loop noise measurements are particularly useful since each oscillator component can be evaluated individually. This is especially convenient if the components are designed to operate in a 50 ohm environment. By measuring each element, the dominant source of flicker noise can be identified. Figure 5 shows the measured gain compression and flicker noise level as a function of input power for two pairs of commercial silicon bipolar transistor amplifiers. As can be seen, the flicker noise level at 1 Hz increases as the amplifiers are driven into gain compression. Of the four amplifiers measured, one was particularly noisy at low input power levels and showed little increase in noise level with increasing drive power. With 3 dB of gain compression (which is typical for many SAW oscillators that do not have separate limiters) the observed flicker noise level at 1 Hz was about -135 dBc/Hz. Since pairs of amplifiers were being measured, the level for an individual amplifier is about -138 dBc/Hz. A number of commercial silicon bipolar transistor amplifiers have been measured and flicker noise levels have been observed to fall in the range of -140 dBc/Hz to -125 dBc/Hz, with -135 dBc/Hz at 1 Hz being a typical number.

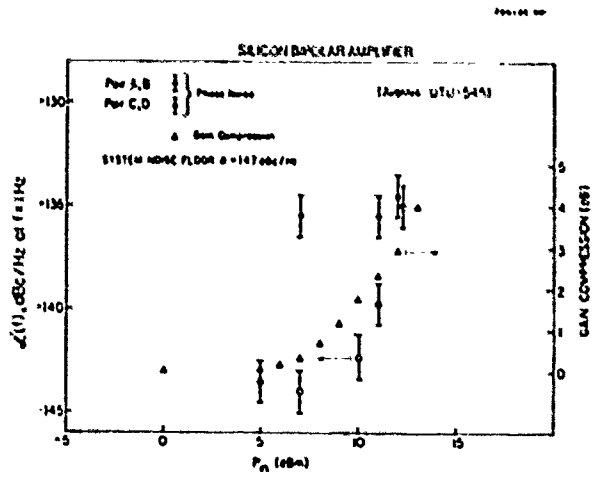


Figure 5. Open-loop flicker noise level and gain compression as a function of input power level for silicon bipolar transistor amplifiers.

Figure 6 shows the results of a similar measurement on a single commercial GaAs FET amplifier. The same characteristics are present except that the flicker noise level at 3 dB of gain compression is about 19 dB higher. GaAs FET amplifiers are well known to have higher flicker noise levels than silicon bipolar amplifiers[11] and should be avoided if at all possible.

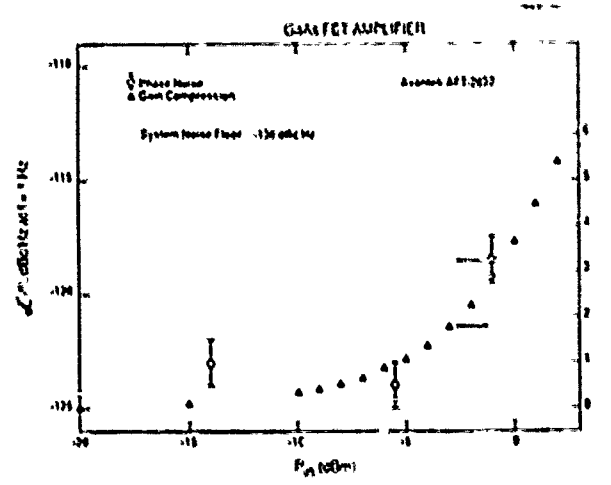


Figure 6. Open-loop flicker noise level and gain compression as a function of input power level for a GaAs FET amplifier.

The fact that both types of amplifiers showed relatively constant flicker noise levels at low input powers indicates clearly that the noise mechanism is not additive in nature. Some parameter (such as base to collector capacitance) which contributes to the phase shift through the amplifiers is fluctuating in time. Furthermore, as the amplifier is driven into gain compression, either a new mechanism appears or the low level one is modified. At this time the details of these noise mechanisms are not understood. However, it is obvious that gain compression should be avoided if the amplifiers are the dominant source of flicker noise in an oscillator. Another point worth noting is that flicker noise levels in commercial amplifiers are never specified and that clearly these levels can vary significantly between individual amplifiers that meet all of the manufacturer's specifications.

The flicker noise levels of quartz BAW and SAW resonators have been measured by a number of groups[9,10,12-18] and the typical level of $\mathcal{L}(f)$ at (or extrapolated to) 1 Hz is about -130 to -125 dBc/Hz. However there is a substantial variation from device to device and observed values of $\mathcal{L}(f)$ range from below -135 dBc/Hz to above -110 dBc/Hz. The author's group has also measured the flicker noise levels on several L-band dielectric resonators and found them to be at or less

than -135 dBc/Hz. The actual levels may be even lower since the observed levels were very close to the measurement system noise floor for the conditions under which the measurements were made. These observed levels of flicker noise indicate that BAW or SAW resonators will ususally be the dominant source of flicker noise in an oscillator if state-of-the-art amplifiers are being used, while for DRO's, the amplifiers will probably be the dominant source. There is no evidence of power dependence in resonator flicker noise levels except for permanent changes at very high incident power levels. It is interesting to note that the open-loop phase noise levels for devices as dissimilar as high Q resonators and wide-band amplifiers are as close as they are.

A characteristic of flicker noise in resonators that has not been discussed in the literature is the fact that the open-loop resonator noise may vary with the loaded Q. If flicker noise in acoustic resonators is caused by frequency fluctuations, and these frequency fluctuations are not strongly influenced by the load impedance of the oscillator circuit ($S_{\Delta F}(f)$ is constant), then from Eq. 9 we find that $S'_{\phi R}(f)$ for the resonator varies as Q_L^2.

$$S'_{\phi R}(f) = (2Q_L/F_o)^2 S_{\Delta F}(f) \qquad (13)$$

To test the validity of this assumption, the open-loop phase noise of a 500 MHz SAW resonator was measured under two different impedance matching conditions. The results are shown in Fig. 7. For Q_L equal to 5000, the observed noise level at 1 Hz was -127 dBc/Hz. For Q_L equal to 11,500, the noise level was -119 dBc/Hz. From Eq. 13 the predicted change in phase noise for the two different values of Q_L is 7 dB and an 8 dB change was actually measured. The data

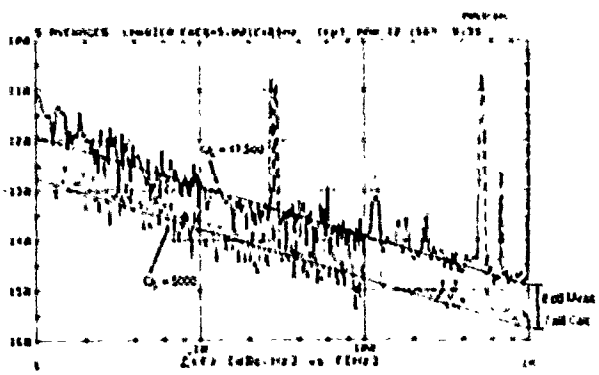


Figure 7. Open-loop flicker noise level of a 500 MHz SAW resonator for two different values of the loaded Q.

clearly shows that it is the resonator frequency that is fluctuating. Note that combining Eq. 13 and 10 gives the result that $S_\phi(f)$ (closed-loop) will be independent of Q_L as mentioned earlier. In situations where the resonator phase noise, $S'_{\phi R}(f)$, is not too different from the amplifier noise, $S'_{\phi F}(f)$, changing the loaded Q may influence which component is the dominant noise source. This may explain the tendency for the flicker noise level of the SAW oscillators in reference 10 to increase somewhat at low values of Q_L.

Dependence of $S_\phi(f)$ on F_o and Q_u

An interesting observation is that the typical value for the open-loop flicker noise level, $\mathcal{L}'(1)$, on acoustic resonators (approximately -130 to -125 dBc/Hz) is independent of the device frequency. This holds for devices ranging from 2.5 MHz to over 1 GHz. It also appears that the amplifier flicker noise is independent of the oscillator frequency (over the same frequency range) and is a constant 5 to 10 dB below the typical acoustic resonator. The fact that $S'_\phi(f)$ is constant with oscillator frequency has some interesting consequences. Starting with Eq. 10, and assuming for simplicity that $Q_L \approx 1/2Q_u$, where Q_u is the unloaded Q, we get

$$S_\phi(f) = (F_o/Q_u)^2 S'_\phi(f)/f^2. \qquad (14)$$

Furthermore, by making use of the fact that the Q F product is constant for resonators operating near the material limit for the unloaded Q

$$Q_u \times F_o = C = \text{constant}, \qquad (15)$$

we get

$$S_\phi(f) = (F_o^4/C^2) S'_\phi(f)/f^2 \qquad (16)$$

or

$$S_\phi(f) = (C^2/Q_u^4) S'_\phi(f)/f^2. \qquad (17)$$

Thus we find, since $S'_\phi(f)$ is approximately constant, that the closed-loop flicker noise level will have either a $1/Q_u^4$ or an F_o^4 dependence. This observation has been reported in references 10 and 14. Figure 8 (from reference 10) shows the observed closed-loop flicker noise level at 1 Hz as a function of unloaded Q for a number of BAW and SAW devices. The $1/Q_u^4$ dependence is clearly evident. Figure 9 shows $S_\phi(1)$ as a function of F_o for a subset of the devices in Fig. 8 which were operating near the material limit for Q_u. The F_o^4 dependence is also clearly evident. Data from three DRO's is also included.

Observations such as those shown in Figs. 8 and 9 make it possible to derive some empirical relations for estimating flicker

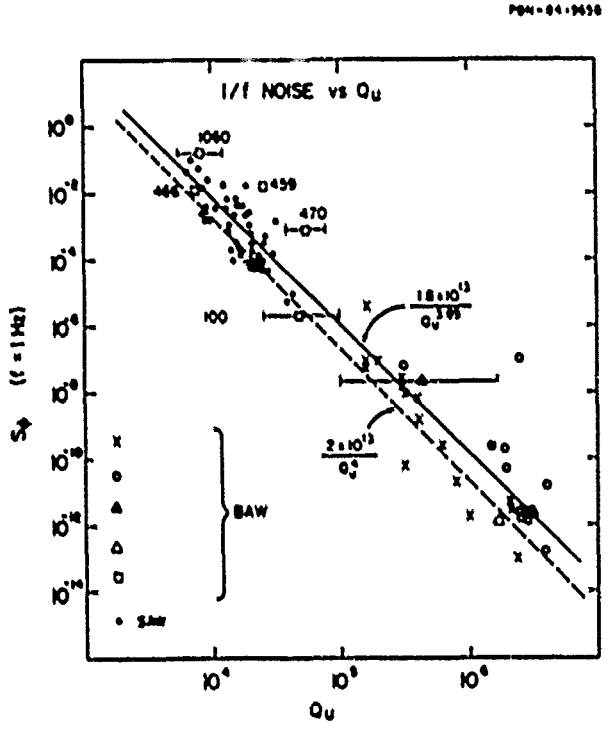


Figure 8. Oscillator flicker noise levels for a number of BAW and SAW resonators as a function of unloaded Q. From reference 10.

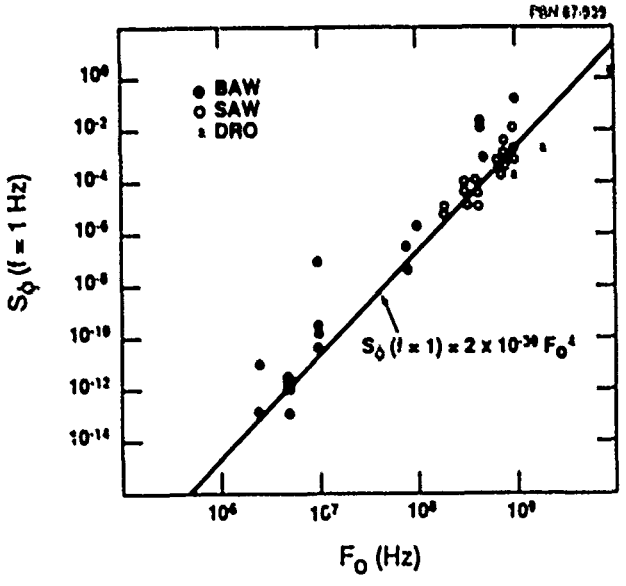


Figure 9. Oscillator flicker noise levels for a number of BAW and SAW resonators (with $Q_u \simeq Q_m$) as a function of resonator frequency.

noise levels in acoustic resonator oscillators[10,14,19] as a function of F_o or Q_u. These relations are:

$$S_\phi(f) = (2\times10^{13}/Q_u^4)(1/f^3) \quad (18)$$

$$S_\phi(f) = 2\times10^{-39}F_o^4/f^3 \qquad F_o \text{ in Hz} \quad (19)$$

$$\mathcal{L}(f) = -390 + 40\text{Log}(F_o) - 30\text{Log}(f) \quad (20)$$

$$S_{\Delta F}(f) = 2\times10^{-39}F_o^4/f \quad (21)$$

$$S_y(f) = 2\times10^{-39}F_o^2/f \quad (22)$$

$$\sigma_y(\tau) = 5\times10^{-20}F_o \quad (23)$$

These relations give an approximate lower limit to the level of flicker noise that can be achieved in oscillators using acoustic resonators that have unloaded Q's near the material limit. The noise values given in the equations generally reflect observed levels of resonator noise, but even if the loop amplifier is the dominant flicker noise source, the same functional relationship with F_o and Q_u would be maintained. The only difference would be in the proportionality constants.

For resonators that have $Q_u \leq (2/3)Q_m$, the dependence on F_o may no longer be valid and the estimates in Eqs. 19 through 23 will probably be too low. Whether Eq. 18 still holds is also open to question. If the loaded Q in a SAW resonator is reduced due to the addition of extra metal to the surface of the active acoustic area, Eq. 18 will still hold even when the loaded Q is substantially below the material limit[20]. However, if the Q is reduced by air loading there is little change in the flicker noise level[20]. Equations 20 through 23 could have been expressed in terms of Q_u, but because of the uncertainty concerning the effect of reduced values of Q_u it was decided to use F_o and limit the usage to $Q_u \simeq Q_m$.

For dielectric resonators the Q F product is higher and the inherent flicker noise in the resonator is lower than for acoustic resonators, so one would expect the oscillator flicker noise level to be lower. This is confirmed in Fig. 9, where the same F_o^4 dependence is evident for the DRO's, but the noise level is about 10 dB lower than that of the acoustic resonator oscillators.

A significant ramification of the F_o^4 dependence of $S_\phi(f)$ is that the flicker noise of an oscillator increases faster if the resonator frequency is increased as opposed to using frequency multiplication. This is particularly evident from Eq. 22, which is invariant under multiplication. This point is further illustrated in Fig. 10 where the noise levels of state-of-the-art 5 MHz (BAW) and 500 MHz (SAW) oscillators are plotted. The noise level for the 5 MHz BAW oscillator after frequency multiplication to 500 MHz

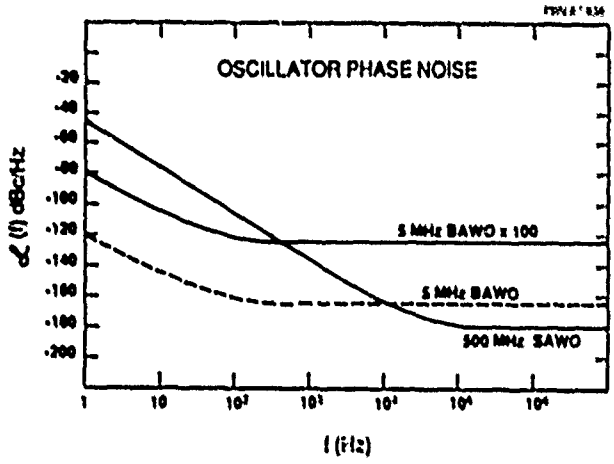


Figure 10. Phase noise spectra of a 500 MHz SAW oscillator, a 5 MHz BAW oscillator and the 5 MHz oscillator multiplied by 100.

(x100) is also shown. Note that below approximately 400 Hz the multiplied BAW oscillator is quieter than the SAW oscillator. Thus, the best close-to-carrier noise performance is obtained by using a low frequency oscillator and then multiplying. Obviously, the best overall performance could be obtained by phase-locking the SAW oscillator to the BAW oscillator with a loop-lock bandwidth of about 400 Hz.

Oscillator Noise Characteristics

Before the overall noise performance of an oscillator can be discussed, one more point regarding flicker noise needs to be made. As discussed by Kroupa[14], if a resonator is sufficiently noisy, and the additive noise floor is sufficiently low, an additional contribution to $1/f^2$ (white frequency) noise will be present. This is illustrated in Fig. 11. As Walls and Wainwright[13] observed, when the frequency of the phase noise from a resonator exceeds the bandwidth of the resonator, the open-loop resonator noise begins to fall off as $1/f^2$ as shown in the figure. If the noise floor is sufficiently low, or the resonator noise sufficiently high, there will be a region of $1/f^2$ dependence in the open-loop noise spectrum. This is equivalent to having $f_\alpha > f_\tau$. When the oscillator loop is closed, only the phase noise inside the resonator bandwidth will be modified and, in this case, it takes on a $1/f^3$ dependence. The closed-loop noise spectrum of Fig. 11 has the same shape as that of Fig. 3 except that the intercept of the $1/f^2$ section with the noise floor occurs not at f_τ as in Fig. 3, but at a larger noise frequency. Thus one cannot always use this intercept in the oscillator

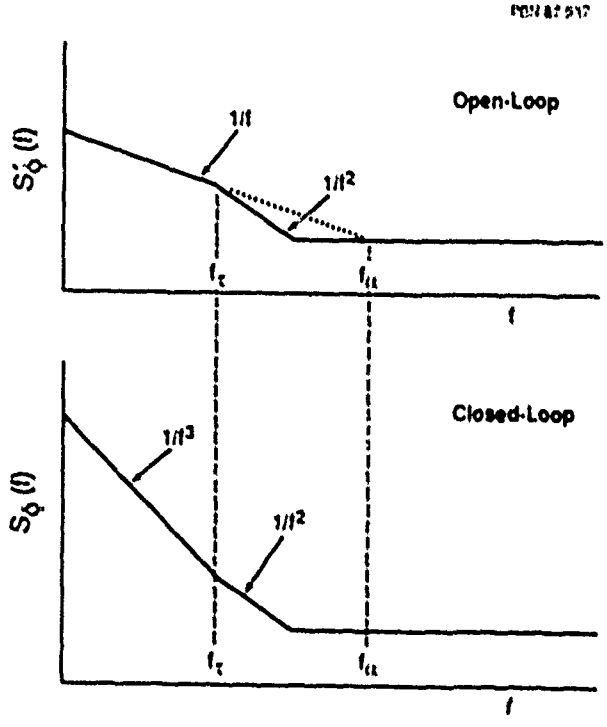


Figure 11. Open- and closed-loop configurations for a feedback oscillator in which $f_\alpha > f_\tau$.

noise as a means of estimating the loaded Q of the resonator. Having $f_\alpha > f_\tau$ is more likely to occur in low frequency (high Q) resonators where f_τ is small.

All of the various contributors to the noise spectrum of an oscillator down to a noise frequency of about 0.1 Hz have been discussed and an approximate analytic expression can be presented. By combining Eqs. 10, 11 and 19, and by using the information in Figs. 3 and 11 we get

$$S_\phi(f) = [\alpha_R F_o^4 + (F_o/(2Q_L))^2 a_E]/f^3 + [2\alpha_R Q_L F_o^3 + (2GFKT/P_o)(F_o/(2Q_L))^2]/f^2 + a_E/f + 2GFKT/P_o. \quad (24)$$

α_R is the flicker noise constant for the resonator and is approximately 2×10^{-39} $(rad/Hz)^2$ for acoustic resonators. This constant can be derived from open-loop noise measurements, $S'_\phi(f)$, on the resonator and by using Eq. 25.

$$S_{\delta F}(f) = \alpha_R F_o^4/f = (F_o/(2Q_L))^2 S'_\phi(f) \quad (25)$$

The flicker noise constant for the amplifier, a_E, is typically 6×10^{-14} $(rad)^2$, and is derived from open-loop noise measurements on the amplifier and by using Eq. 26.

$$S'_{\phi}(f) = a_E/f \qquad (26)$$

In general, the measured closed-loop oscillator noise agrees with the estimated oscillator noise based on open-loop noise measurements of the individual oscillator components to better than 3 dB. Occasionally the discrepancy may be as large as 6 or 7 dB, but on the average the agreement is within 3 dB.

Causes of Flicker Noise

Flicker noise is presently a poorly understood phenomenon and a complete discussion of the causes of flicker noise is beyond the scope of this paper. However, some comments regarding flicker noise in SAW resonators can be made. There are a number of possible sources for frequency fluctuations in a SAW resonator, and some of them are listed in Table 1. Identifying which of the many possibilities is actually causing the flicker frequency fluctuations is, however, a difficult task. The large variations (+/- 5 to 10 dB) that occur in the flicker noise level of "identical" devices makes it particularly time consuming (and expensive) to evaluate different design, fabrication or processing variables since a large number of devices must be tested. However, for SAW devices there is considerable evidence that the transducer metalization plays a major role in flicker noise. Observations have been made that the physical condition of the transducer[18,21] the type of metal used[17,21] the amount of metal used[20,21], and the type of bond wires used[21] can all influence the flicker noise level.

Table 1.

Possible sources of frequency fluctuations.

ACOUSTIC VELOCITY			EFFECTIVE CAVITY LENGTH
	TRANSDUCER		
BULK	Mechanical	Electrical	
Temperature, Temperature gradients	Temperature, Temperature gradients	Electric field strength and distribution	Temperature, Temperature gradients
Mechanical stress	Mechanical stress	Finger resistance	Mechanical strain
Defects	Defects	Interelectrode resistance	Defects
Damage	Acoustic attenuation	Capacitance	
Acoustic attenuation	Adsorption, desorption of contaminants		
Piezoelectric effect			
Adsorption, desorption of contaminants			
Surface preparation			

Another useful piece of information regarding flicker noise can be derived from the observation that $S_{\phi}(f)$ (or $S_{\Delta F}(f)$) increases as F_o^4. Some of the potential sources of flicker noise listed in Table 1 are phenomena that originate from bulk properties of the quartz, for example the resonator temperature dependence (static and dynamic) and vibration sensitivity. These parameters result in a fractional frequency change that is independent of the device frequency. For example, the fractional change in frequency with temperature for a 100 MHz SAW resonator is, to first order, the same as that for a 1 GHz resonator if they were fabricated on the same cut. Thus we have

$$\Delta F/F_o = \Delta V/V = \text{CONSTANT}$$

(no dependence on F_o). (27)

In terms of spectral densities we get

$$S_{\Delta F}(f) = F_o^2 \times \text{CONSTANT}, \qquad (28)$$

which has only an F_o^2 dependence. Therefore, these bulk properties can be ruled out as possible sources of flicker noise.

However, there is a class of effects that do have the correct dependence on F_o, and these are surface or interface phenomena. If a surface or interface property causes a fractional change in velocity, this will result in a fractional change in frequency. However, if the physical thickness of the cause of the velocity change is much less than an acoustic wavelength, λ, the perturbation to the acoustic velocity will be approximately inversely proportional to λ and will therefore increase with increasing frequency, F_o. We then have

$$\Delta F/F_o = \Delta V/V = \text{CONSTANT}/\lambda$$

$$= (F_o \times \text{CONSTANT})/V, \qquad (29)$$

which gives

$$S_{\Delta F}(f) = F_o^4 \times \text{CONSTANT}/V. \qquad (30)$$

This has the correct functional dependence on F_o, so surface or interface phenomena should be considered prime candidates for being the source, or sources, of flicker noise. This class of phenomena is also compatible with most of the observations concerning the importance of the properties of the transducer metal to flicker noise levels.

Random Walk Noise

There is another type of modulation noise that occurs at very low noise frequencies (f < 0.1 Hz) which has not been discussed yet.

This is a noise process for which $S_{\Delta F}(f)$ varies as $1/f^2$ and is therefore a random walk noise. This type of noise has been observed for both SAW[22] and BAW[23] devices, though in most BAW devices the noise can be traced to environmental factors such as temperature fluctuations[23]. However, random walk noise that is not caused by any known environmental factors has been observed in SAW devices. Figure 12 shows the measured noise levels of several SAW oscillators over the noise frequency range of 10^{-6} Hz to 10^{7} Hz. Below 10^{-1} Hz, the observed spectrum is close to being random walk in nature. Figure 13 shows the results of similar measurements on another 425 MHz SAW resonator oscillator except that the data is plotted in the time domain in terms of $\sigma_y(\tau)$. The flicker and random walk regions are also clearly evident here. The bulge in the data near f = 300 Hz was traced to temperature fluctuations but the other regions show little to no correlation to temperature variations. For SAW resonators near 400 MHz a typical level for $S_{\Delta F}(f)$ at f = 10^{-6} Hz is 3×10^{6}, though this may vary from device to device by more than an order of magnitude either way.

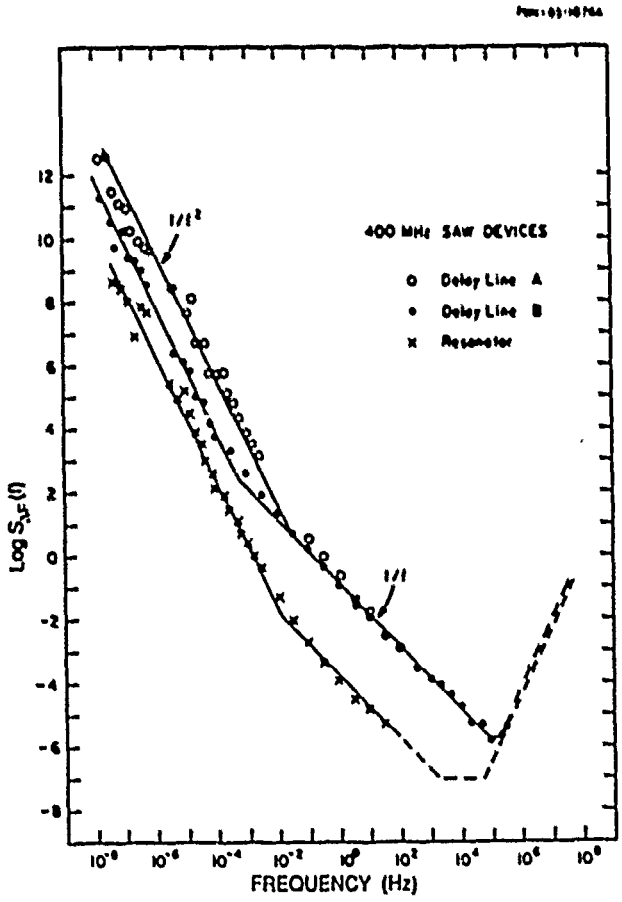


Figure 12. Spectral density of frequency fluctuations for three SAW oscillators. Random walk noise is present for f < .01 Hz.

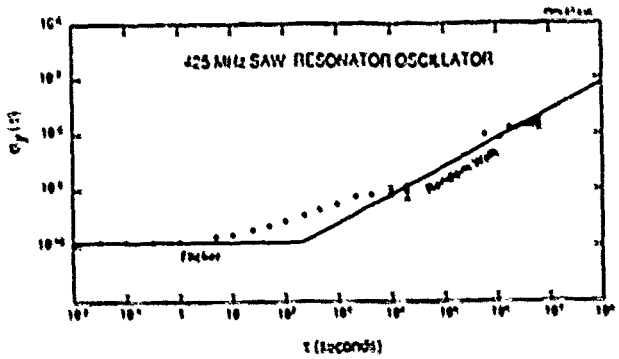


Figure 13. Flicker and random walk noise of a 425 MHz SAW resonator oscillator as seen in the time domain.

There is not enough data yet to determine precisely what the dependence on F_o is for random walk noise, but the one data point for a 5 MHz BAW device that does not appear to be caused by temperature fluctuations[23] suggests that the F_o dependence is greater than F_o^2 and may even be greater than F_o^4. If the F_o^4 dependence does in fact exist, it is then not surprising that inherent random walk noise is more noticeable in high frequency SAW devices than in low frequency BAW resonators. Even less is known about the possible causes of random walk noise than for flicker noise, but there is data for SAW devices that indicates[22] that there is a connection between the two. Furthermore, processing steps have been recently identified that reduce both flicker noise and random walk noise.

Conclusions

Considerable progress has been made over the last 10 to 15 years in the understanding of close-to-carrier noise processes in stable oscillators. Though a complete explanation is far from being in place, many major characteristics have come to light. The oscillator designer can now make informed decisions about many of the oscillator design parameters. Also, Leeson's model still provides a suitable framework for developing approximate expressions for oscillator noise characteristics.

At this time there is no way of predicting flicker noise levels in individual oscillator components, and it is therefore highly desirable that manufactures of resonators and amplifiers begin to measure and specify flicker noise levels in their products.

Acknowledgements

The author would like to acknowledge the support and assistance of Dr. J. Callerame, Dr. G. Montress, Mr. M. Loboda, Mr. J. Lang, Mr. E. Sabatino and Mr. M. Bennett.

REFERENCES

1. D. B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum," *Proceedings of the IEEE*, Vol. 54, No. 2, pp. 329-330, February 1966.

2. J. Barnes, A. Chi, L. Cutler, D. Healey, D. Leeson, T. McGunigal, J. Mullen, W. Smith, R. Sydnor, R. Vessot, and G. Winkler, "Characterization of Frequency Stability," *IEEE Transactions on Instrumentation and Measurement*, Vol. IM-20, No. 2, pp. 105-120, May 1971.

3. J. H. Shoaf, D. Halford, and A. Risley, "Frequency Stability Specifications and Measurement: High Frequency and Microwave Signals," *National Bureau of Standards (U.S.) Technical Note 632*, pp. 40-46, January 1973.

4. D. W. Allan, "Statistics of Atomic Frequency Standards," *Proceedings of the IEEE*, Vol. 54, No. 2, pp. 221-230, February 1966.

5. W. P. Robbins, Phase Noise in Signal Sources, London, U.K., Peter Peregrinus Ltd., 1984.

6. R. W. Muat, "Designing Oscillators for Spectral Purity," *Microwaves and RF*, Vol. 23, No. 7, pp. 133-160, July 1984.

7. M. F. Lewis, "The Surface Acoustic Wave Oscillator - A Natural and Timely Development of the Quartz Crystal Oscillator," *Proceedings of the 28th Annual Symposium on Frequency Control*, 1974, pp. 304-314.

8. T. E. Parker, "Current Developments in SAW Oscillator Stability," in *Proceedings of the 31st Annual Symposium on Frequency Control*, 1977, pp. 359-364.

9. S. S. Elliott and R. C. Bray, "Direct Phase Noise Measurements of SAW Resonators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1984, Vol. 1, pp. 180-185.

10. T. E. Parker, "1/f Frequency Fluctuations in Quartz Acoustic Resonators," *Applied Physics Letters*, Vol. 46, No. 3, pp. 246-248, 1985.

11. R. A. Pucel, "The GaAs FET Oscillator - It's Signal and Noise Performance," in *Proceedings of the 40th Annual Symposium on Frequency Control*, 1986, pp. 385-391.

12. J. J. Gagnepain, J. Ubersfeld, G. Goujon and P. Handel, "Relation Between 1/f Noise and Q factor in Quartz Resonators at Room and Low Temperatures, First Theoretical Interpretation," in *Proceedings of the 35th Annual Symposium on Frequency Control*, 1981, pp. 476-483; J. J. Gagnepain, M. Oliver and F. L. Walls, "Excess Noise in Quartz Crystal Resonators," in *Proceedings of the 37th Annual Symposium on Frequency Control*, 1983, pp. 218-225.

13. 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.

14. V. F. Kroupa, "Noise in Precision Oscillators," in *Proceedings of the Conference on Electronic and Piezoelectric Elements*, 1983, pp. 27a-27d.

15. G. Moulton, "Dig for the Roots of Oscillator Noise," *Microwaves and RF*, Vol. 25, No. 4, pp. 65-69, April 1986.

16. M. M. Driscoll and B. W. Kramer, "Spectral Degradation in VHF Crystal Controlled Oscillators Due to Short-Term Instability in the Quartz Crystal," in *Proceedings of the IEEE Ultrasonics Symposium*, 1985, pp. 340-345.

17. W. J. Tanski, "The Influence of a Chrome Film Bonding Layer on SAW Resonator Performance," in *Proceedings of the IEEE Ultrasonics Symposium*, 1985, Vol. 1, pp. 253-257.

18. R. C. Bray, L. L. Pendergrass, C. A. Johnsen, T. L. Bagwell, and J. L. Henderson, "Annealing Behavior and Phase Noise Performance of SAW Resonators," in *Proceedings of the IEEE Ultrasonics Symposium*, 1985, Vol. 1, pp. 247-252.

19. F. L. Walls and A. E. Wainwright, "Measurements of the Short-Term Stability of Quartz Crystal Resonators - A Window on Future Developments in Crystal Oscillators," in *Proceedings of the 6th Annual Precise Time and Time Interval Planning Meeting*, 1974, pp. 143-153.

20. T. E. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators," in *Proceedings of the 39th Annual Symposium on Frequency Control*, 1985, pp. 97-106.

21. T. E. Parker and J. Loan, "SAW Oscillators for Air Defense Systems," *Final Contract Report SLCET-TR-84-0401-F*, U.S. Army Laboratory Command, ETDL, Fort Monmouth, NJ, 1986.

22. T. E. Parker, "Random Walk Frequency Fluctuations in SAW Oscillators," in *Proceedings of the 40th Annual Symposium on Frequency Control*, 1986, pp. 241-251.

23. Y. Noguchi, T. Teramachi, and T. Musa, "1/f Frequency Fluctuations of a Quartz Crystal Oscillator and Temperature Fluctuations," in *Proceedings of the 35th Annual Symposium on Frequency Control*, 1981, pp. 484-491.

41st Annual Frequency Control Symposium - 1987

ACCURACY OF TIME TRANSFER IN SATELLITE SYSTEMS

Clifford M. Will
McDonnell Center for the Space Sciences, Department of Physics
Washington University, St. Louis MO 63130
(314)889-6244

The NAVSTAR Global Positioning System (GPS) of the U. S. Department of Defense is a military and civilian navigation and time transfer system based on a constellation of satellites carrying atomic clocks. Six satellites are currently operational (with an ultimate goal of 21 satellites), in 12-hour Earth orbits at 11,000 miles altitude. The present uncertainty in the system is 20 ns in time, or less than 10 m in position. At this level of accuracy, the effects of Einstein's theory of relativity[1,2] must be taken into account to obtain accurate and consistent measurements. These include the special relativistic time dilation and the gravitational redshift, which affect clock rates, and the Sagnac effect, which involves synchronization of orbiting clocks.

Recently, a study was carried out under the auspices of the Air Force Studies Board (Commission on Engineering and Technical Systems/National Research Council) to evaluate the adequacy of present methods for taking relativistic effects into account in GPS, and to make recommendations for future improvements. In this paper we report the conclusions and recommendations of that study[3].

(1) Current methods for treating relativistic effects are valid to well below the 2 ns level, and have been verified empirically in GPS to 5 ns.

(2) Improvement in accuracy in GPS from the current 20 ns to 2 ns is desirable and possible without new technology, by establishing timing discipline at ground monitor stations, by maintaining the high-latitude monitor station in Alaska for a few more years, and by making distance determinations to at least one satellite that are independent of the satellite clocks, in order to separate satellite ephemeris errors from clock errors.

(3) For sub-nanosecond accuracy and long-term improvements, a number of studies were recommended to determine the benefits of improved distance determinations to satellites, of improved compensation of ionospheric and tropospheric effects on the navigation signals, and of coordinated use of clocks with higher stability and reliability.

(4) Accurate time-transfer systems such as GPS offer the opportunity to study fundamental relativistic questions.

1. C. M. Will, 1981, *Theory and Experiment in Gravitational Physics*, Cambridge: Cambridge University Press.
2. C. M. Will, 1986, *Was Einstein Right?*, New York: Basic Books.
3. C. M. Will (ed.), D. B. Debra, D. Eardley, W. O. Hamilton, J.-S. Leung, B. McMillan, R. Matzner, R. F. C. Vessot, G. M. R. Winkler, and N. Yannoni, 1986, *Accuracy of Time Transfer in Satellite Systems*, Washington: National Academy Press.

CH2427-3/87/0000-111 $1.00 C 1987 IEEE

RELATING THE ALLAN VARIANCE TO THE DIFFUSION COEFFICIENTS OF A LINEAR STOCHASTIC DIFFERENTIAL EQUATION MODEL FOR PRECISION OSCILLATORS

James William Chaffee
Systems Control Technology, Inc.
Palo Alto, California 94304

Summary

Models of clock errors are of interest in understanding the NAVSTAR GPS Control Segment as well as in the design and simulation of user equipment. In these practical situations, the question of how the clock model's noise properties relate to the Allan Variance is important. In particular, the relationship between the diffusion coefficients or so-called process noise parameters for the stochastic differential equation often used in Kalman filters to model the behavior of oscillators and the Allan Variance must be understood. In this paper, the relationship is stated and derived using integrals of stochastic processes as the natural tool.

Introduction

The instantaneous time error of a clock run from a precision oscillator is sometimes modeled as a simple linear stochastic differential equation with constant coefficients. This is the case with the Global Positioning System control segment, which uses a Kalman filter to make error estimates and predictions.

In practical situations, the question of how the model's noise properties relate to the Allan variance is of importance. In this paper, the relationship is stated and derived using integrals of stochastic processes as the natural tool. The most sophisticated tool is the Wiener integral, a Stieljes-type integral of a non-random integrand with respect to a Wiener process. Appropriate references are [1-3]. References covering linear stochastic differential equations are [2, 4, and 5].

Concerns expressed by Rutman in [6] and [7] regarding the use of a power law spectral density model for the non-stationary diffusion resulting from such a model are also addressed. In fact, it is shown that some of the parameters appearing in the power law model can be precisely interpreted as parameters associated with the linear stochastic differential equation. Unfortunately, this does not apply to flicker noise.

The Stochastic Differential Equation Model

Let $x(t)$ denote the instantaneous time error of a clock run from an oscillator with instantaneous frequency $v(t) = v_0 + \Delta v(t)$, following standard terminology [7-9]. Then, formally,

$$y(t) = \frac{dx(t)}{dt}, \text{ where } y(t) = \frac{\Delta v(t)}{v_0} .$$

This notion can be given mathematical precision with the model

$$\begin{bmatrix} dx(t) \\ dy(t) \\ dz(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} dt + \begin{bmatrix} db_1(t) \\ db_2(t) \\ db_3(t) \end{bmatrix} ,$$

which will be denoted $dS(t) = AS(t)dt + dB(t)$, where the $b_i(t)$ are independent Wiener processes with constant diffusions q_i. The variable $z(t)$ is often termed the frequency drift. This model is used in the GPS Kalman filter.

This model also illustrates the concerns Rutman raises in [6] and [7]. For one thing, since the Brownian sample paths $b_i(t)$ are nowhere differentiable [10, page 9], the stochastic differentials are difficult to interpret. However, this is handled rigorously in [1-5, 10]. Of a more serious concern is that the system above does not seem to admit a stationary solution, since the coefficient matrix does not have negative eigenvalues. (See [4, Chapter 8, Section 2].) Because of this, the solution vector cannot be said to follow a model given by a power spectral law.

Nonetheless, the power law model as described in [7-9] can be applied through the fact that the solution of the above system can be Gaussian with stationary increments.

The solution to this equation is

$$S(t) = \Phi(t, t_0)S(t_0) + \int_{t_0}^{t} \Phi(t, \tau)dB(\tau)$$

where $\Phi(t, t_0) = \exp(A(t - t_0))$ and $S(t_0)$ is the initial condition [2, 4, 5]. If it is assumed that $S(t_0)$ is a non-random constant, then $S(t)$ is Gaussian with stationary increments [4, Chapter 8, Section 2].

The time error is given by

$$x(t) = x(t_0) + y(t_0)(t-t_0) + z(t_0)\frac{(t - t_0)^2}{2} +$$

$$+ \int_{t_0}^{t} db_1(\lambda) + \int_{t_0}^{t} (t-\lambda)db_2(\lambda)$$

$$+ \int_{t_0}^{t} \frac{(t - \lambda)^2}{2} db_3(\lambda) .$$

The integrals are white noise integrals and are detailed in [1-5, 10]. They can be differentiated to produce the stochastic differential equation

$$dx(t) = y(t_0)dt + z(t_0)(t - t_0)dt +$$

$$+ db_1(t) + b_2(t)dt +$$

$$+ \left(\int_{t_0}^{t} (t - \lambda)db_3(\lambda) \right) dt.$$

The Allan variance is concerned with random fluctuations over a fixed time interval τ. By assuming $x(t_0)$, $y(t_0)$, and $z(t_0)$ are non-random, they can be determined and removed, at least in theory. For this reasrn, they will be assumed zero, leaving the stochastic differential equation

CH2427-3/87/0000-112 $1.00 C 1987 IEEE

$$y(t)dt = dx(t) = db_1(t) + b_2(t)dt +$$

$$+ \left(\int_{t_0}^{t} (t - \lambda) db_3(\lambda) \right) dt.$$

This equation provides the relationship to the power law model. In order to keep the terminology consistent with the engineering literature, let t_0 be t_k and compute the integral of $y(t)dt$ between t_k and $t_k+\tau$ (see [1, 2, or 3] for details regarding these calculations):

$$\int_{t_k}^{t_k+\tau} y(t)dt = \int_{t_k}^{t_k+\tau} db_1(t) + \int_{t_k}^{t_k+\tau} b_2(t)dt +$$

$$+ \int_{t_k}^{t_k+\tau} \left(\int_{t_k}^{t} (t-\lambda) db_3(\lambda) \right) dt \quad .$$

The first integral on the right side is $b_1(t_k + \tau) - b_1(t_k)$, which has the same mean and variance as $b_1(\tau)$, and being Gaussian is thus statistically equivalent to $b_1(\tau)$, which has zero mean and variance $q_1\tau$.

The second integral has mean zero, since $b_2(t)$ is zero mean. Its variance is obtained by noting that $b_2(t)$ beginning at t_k has covariance kernal $k(s, t) = q_2((s \wedge t) - t_k)$, where $s \wedge t$ means minimum of s and t.

Now the variance of $\int_{t_k}^{t_k+\tau} b_2(t)dt$ is obtained by computing a deterministic (Riemann) integral

$$\text{Var} \left(\int_{t_k}^{t_k+\tau} b_2(t)dt \right) =$$

$$= q_2 \int_{t_k}^{t_k+\tau} \int_{t_k}^{t_k+\tau} ((s \wedge t) - t_k)\, ds dt$$

$$= q_2 \left(\int_{t_k}^{t_k+\tau} dt \left(\int_{t_k}^{t} s ds + \int_{t}^{t_k+\tau} t\, ds \right) - t_k \tau^2 \right)$$

$$= q_2 \frac{\tau^3}{3}$$

Thus, $\int_{t_k}^{t_k+\tau} b_2(t)dt$ is Gaussian with zero mean and variance $q_2 \frac{\tau^3}{3}$.

The third integral also has zero mean. If Z_t denotes $\int_{t_k}^{t} (t-\lambda) db_3(\lambda)$ for $t > t_k$, then its variance is

$$E \left[\int_{t_k}^{t_k+\tau} Z_t dt \int_{t_k}^{t_k+\tau} Z_s ds \right] =$$

$$= \int_{t_k}^{t_k+t} \int_{t_k}^{t_k+t} E[Z_t Z_s] dt ds \quad .$$

First, it is necessary to compute

$$E(Z_t Z_s) = E \left[\int_0^T (t-\lambda) I_{[t_k, t]} db_3(\lambda) \times \right.$$

$$\left. \times \int_0^T (s-\lambda) I_{[t_k, s]} db_3(\lambda) \right] \quad (T > \max(s,t))$$

$$= q_3 \int_0^T (t-\lambda)(s-\lambda) I_{[t_k, t]} I_{[t_k, s]} d\lambda$$

$$= q_3 \int_{t_k}^{s \wedge t} (t-\lambda)(s-\lambda) d\lambda$$

$$= q_3 \left(ts((s \wedge t) - t_k) - (s+t) \left(\frac{(s \wedge t)^2}{2} - \frac{t_k^2}{2} \right) + \right.$$

$$\left. + \frac{(s \wedge t)^3}{3} - \frac{t_k^3}{3} \right) .$$

If this is integrated with respect to s and t over the square with sides t_k to $t_k+\tau$, the result is $q_3 \frac{\tau^5}{20}$, with the intermediate calculations

$$\int_{t_k}^{t_k+\tau} \int_{t_k}^{t_k+\tau} ts[(s \wedge t) - t_k] ds dt =$$

$$= \frac{1}{3} t_k^2 \tau^3 + \frac{5}{12} t_k \tau^4 + \frac{2}{15} \tau^5 \quad ,$$

$$- \frac{1}{2} \int_{t_k}^{t_k+\tau} \int_{t_k}^{t_k+\tau} (s+t)[(s \wedge t)^2 - t_k^2] ds dt =$$

$$= - \frac{2}{3} t_k^2 \tau^3 - \frac{7}{12} t_k \tau^4 - \frac{7}{60} \tau^5 \quad ,$$

and

$$\frac{1}{3} \int_{t_k}^{t_k+\tau} \int_{t_k}^{t_k+\tau} [(s \wedge t)^3 - t_k^3] ds dt =$$

$$= \frac{1}{3} t_k^2 \tau^3 + \frac{1}{6} t_k \tau^4 + \frac{1}{30} \tau^5 \quad .$$

Since each of the Wiener processes $b_i(t)$ is assumed independent, this shows that $\int_{t_k}^{t_k+\tau} y(t)dt$ is Gaussian with mean zero (assuming

the "trend" has been removed) and variance

$$q_1\tau + q_2\frac{\tau^3}{3} + q_3\frac{\tau^5}{20} \quad .$$

Notice that this expression for the variance is independent of the endpoints and depends only on τ. This shows that the integrals of $y(t)$, for each t_k, form a sequence of independent, identically distributed Gaussian random variables. Of course, $t_k + \tau = t_{k+1}$ by assumption, which corresponds to there being no dead time in the measurements.

The Allan Variance

Following [7] and [9], the Allan variance is defined as an ensemble average, ignoring the questions regarding ergodicity of the processes, which relate to methods of obtaining estimates. Note that the statistics computed above are ensemble averages. The definition used here is that of the zero dead time, two-sample Allan variance

$$\sigma_y^2(\tau) = \frac{1}{2}E[(\bar{Y}_{k+1} - \bar{Y}_k)^2] \quad ,$$

where $\bar{Y}_k = \frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)dt$. By what was shown in the last section, $\bar{Y}_{k+1}$ and $\bar{Y}_k$ are independent and zero-mean. Furthermore, the variance of $\bar{Y}_k$ is independent of k for each k, and indeed the random variables $\bar{Y}_k$ are identically distributed for each k, assuming $t_{k+1} = t_k + \tau$. Therefore,

$$\mathrm{Var}(\bar{Y}_{k+1}) = \mathrm{Var}(\bar{Y}_k) = E[\bar{Y}_k^2] =$$

$$= \mathrm{Var}\left(\frac{1}{\tau}\int_{t_k}^{t_k+\tau} y(t)dt\right) =$$

$$= \frac{1}{\tau^2}\ \mathrm{Var}\left(\int_{t_k}^{t_k+\tau} y(t)dt\right) \quad .$$

which is easily calculated as

$$\frac{1}{\tau^2}\left[q_1\tau + q_2\frac{\tau^3}{3} + q_3\frac{\tau^5}{20}\right] = \frac{q_1}{\tau} + \frac{q_2\tau}{3} + \frac{q_3\tau^3}{20} \quad .$$

Then

$$\sigma_y^2(\tau) = \frac{1}{2}\left(\mathrm{Var}(\bar{Y}_{k+1}) + \mathrm{Var}(\bar{Y}_k)\right) = \frac{1}{2}\left(2\ \mathrm{Var}(\bar{Y}_k)\right),$$

because of the independence of $\bar{Y}_{k+1}$ and $\bar{Y}_k$.

Using Meditch's notation in [9], where he writes $\sigma_y^2(\tau) = \frac{N_0}{\tau} + \frac{N_2\tau}{3}$ (ignoring flicker noise, which he includes, and the drift noise, which is included in the white noise model), $q_1 = N_0$ and $q_2 = N_2$.

A Remark About Units

The units of the diffusion parameters q_i are (variance (X_i)) per unit time. Hence, if the frequency fluctuations are measured in nanoseconds over a time interval of seconds, q_1 is $(ns)^2/s$, q_2 is $(ns/s)^2/s$ or $(ns)^2/s^3$, and q_3 is $(ns)^2/s^5$.

One way to understand this is to note that the Brownian motion process b_t has variance proportional to time, so that $\mathrm{Var}(b_t) = qt$.

Maybeck [5, p.221] explains the units by noting that db_t is delta correlated and the delta function has units of $(\mathrm{time})^{-1}$, so that $\mathrm{Var}(db_t) = q\delta(t)$. In either case, q would be forced to have units of variance per time.

Note that these units are also consistent with the unitless nature of the Allan variance. If the units of q_1 are $(\mathrm{time})^2/(\mathrm{time})$, then q_1/τ would be unitless. Because of this, N_0 and N_2 will also have units of variance per time.

Conclusion

It has been shown how the two-sample, zero dead time Allan variance can be related to a white noise stochastic differential equation sometimes used to model precision oscillators in Kalman filter based estimation schemes. One particular application is the Global Positioning System's estimation scheme.

This relationship is

$$\sigma_y^2(\tau) = \frac{q_1}{\tau} + \frac{q_2\tau}{3} + \frac{q_3\tau^3}{20} \quad ,$$

where the q_i are diffusion parameters (also called process noise) related to time, frequency, and frequency drift states.

References

1. Lamperti, John, Stochastic Processes, Springer-Verlag, New York, 1977.

2. Soong, T.T., Random Differential Equations in Science and Engineering, Academic Press, New York, 1973.

3. Parzen, Emanual, Stochastic Processes, Holden-Day, Inc., 1962.

4. Arnold, Ludwig, Stochastic Differential Equations: Theory and Applications, John Wiley and Sons, New York, 1974.

5. Maybeck, Peter S., Stochastic Models, Estimation, and Control, Vol. 1, Academic Press, New York, 1979.

6. Rutman, Jacques, "Oscillator Specifications: A Review of Classical and New Ideas," Proc. 31st Annual Freq. Control Symp., Atlantic City, N.J., June 1-3, 1977.

7. Rutman, Jacques, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," Proc. IEEE, Vol. 66, pp.1048-1075, September 1978. (Reprinted in [11].)

8. Barnes, James, et al., "Characterization of Frequency Stability," IEEE Trans. Instrum. Meas., Vol. IM-20, pp.105-120, May 1971. (Reprinted in [11].)

9. Meditth, J.S., "Clock Error Models for Simulation and Estimation," Aerospace Report No. TOR-0076(6G74-01)-2, The Aerospace Corporation, 7 July 1975.

10. McKean, H.P., Stochastic Integrals, Academic Press, New York, 1969.

11. Kroupa, Venceslov F. (ed.), Frequency Stability: Fundamentals and Measurement, IEEE Press, New York, 1983.

41st Annual Frequency Control Symposium - 1987

THEORETICAL ANALYSIS OF THE MODIFIED ALLAN VARIANCE

L. G. Bernier
Laboratoire d'Electromagnétisme et d'Acoustique,
Ecole Polytechnique Fédérale, Lausanne, SWITZERLAND.

Summary

This paper presents a theoretical analysis of the modified Allan variance. It is shown that, if used under the proper conditions, the modified Allan variance can eliminate not only the temporal-spectral ambiguity it was designed to solve, but also the system bandwidth dependence of the original Allan variance.

Introduction

In recent years, much work has been done about the characterization of frequency stability in the time domain. It appears that the main requirements for a good definition of frequency stability are the following .

1) The definition should be stationary even if the underlying frequency process is not.
2) The relationship between the time and frequency domains should be bi-univocal so that the power spectral density of the frequency fluctuations could be reconstituted from the time domain measurement
3) The measurement should characterize exclusively the oscillator under test and, in particular, it should be independent of the measuring system bandwidth

Assuming that the frequency fluctuations are properly described by the usual polynomial model[1]

$$S^{+}_{yy}(f) \equiv \sum_{\alpha=-2}^{2} h_\alpha f^\alpha \qquad (1)$$

which represents the single-sided power spectral density of the normalized frequency deviation $y(t)$, it can be shown that the definition of the Allan variance meets the first requirement for all the noise processes in the model[2]. It can be shown, moreover, that the two other requirements are also met for all but two kinds of noise : the phase flicker ($\alpha=1$) and the white phase noise ($\alpha=2$) processes[1] . For both these processes, the dependence of the Allan variance upon the averaging time τ is nearly the same and the variance itself is dependent on the bandwidth of the measuring system.

The modified Allan variance was introduced by Allan and Barnes[3] in 1981 with the specific purpose of rectifying the temporal-spectral ambiguity of the original Allan variance

The temporal method used by Allan and Barnes is based on the generalized autocovariance function of the time function $x(t)$ Its actual high frequency behavior depends on the hardware bandwidth and is not tractable in the time domain. As a result, the authors assumed that the dependence on the hardware bandwidth is the same for both the original and the modified Allan variances. In a more recent work, Lesage and Ayi[4] used a spectral method and found the same harware bandwidth dependence as Allan and Barnes

This paper presents a new spectral analysis of the modified Allan variance that takes into account the exact high frequency behavior of the frequency fluctuations The principal result is the demonstration that the dependence of the modified Allan variance on the hardware bandwidth can be completely eliminated if a large number of samples is used and if a proper sampling rate is selected.

Analysis of the modified Allan variance

The original definition of the modified Allar variance was expressed in terms of the samples x_K of the time function $x(t)$[3] :

$$\text{Mod } \sigma_y^2(n\tau_0) \equiv \frac{1/2}{(n\tau_0)^2} E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}(x_{i+2n} - 2x_{i+n} + x_i)\right]^2\right\} \qquad (2)$$

where τ_0 is the averaging time and n the number of frequency samples. $E\{\ \}$ is the mean value operator. The time function $x(t)$ is defined as the integral of the normalized frequency process $y(t)$[1] . Hence, the time samples x_K may be computed from the frequency samples y_K using :

$$y_i \equiv \frac{x_i - x_{i-1}}{\tau_0} \qquad (3)$$

CH2427-3/87/0000-116 $1.00 C 1987 IEEE

Rewriting the original definition in terms of the frequency samples, one obtains

$$\text{Mod } \sigma_y^2(n\tau_0) \equiv \tag{4}$$

$$\tfrac{1}{2} E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{k=1}^{n} y_{i+k+n} - \frac{1}{n}\sum_{k=1}^{n} y_{i+k}\right)\right]^2\right\}$$

A close examination of this last form demonstrates that the calculation of the modified Allan variance covers two distinct operations as explained below

First Operation
The summations on k represent a juxtaposition of n adjacent elementary samples averaged over an interval τ_0. The result of the juxtaposition is equivalent to a single sample averaged over an interval $n\tau_0$ since the frequency counter is assumed to have no dead time

Second Operation
The summation on i, on the other hand, represents a discrete moving average (DMA) operator applied over an interval of length $n\tau_0$ using n discrete samples

The similarity between the original and modified variances can be highlighted by restating equation (4) in terms of the continuous moving average (CMA) and first increment (FI) operators as follows.

$$\text{Mod } \sigma_y^2(n\tau_0) \equiv \tag{5}$$

$$\tfrac{1}{2} E\left\{\left[\frac{1}{n}\sum_{i=1}^{n}\left(\Delta(n\tau_0)\{y(t_0+(i+2n)\tau_0,\ n\tau_0)\}\right)\right]^2\right\}$$

where t_0 is the time origin The CMA operator is defined as

$$\bar{y}(t,\tau) \equiv \frac{1}{\tau}\int_{t-\tau}^{t} y(u)\,du \tag{6}$$

and the FI operator as

$$\Delta(\tau)\{y(t)\} \equiv y(t) - y(t-\tau) \tag{7}$$

Using the same operators, on the other hand, the definition of the original Allan variance reduces to

$$\sigma_y^2(\tau) \equiv \tfrac{1}{2} E\{\zeta^2(t)\} \tag{8}$$

where $\zeta(t)$ is defined as

$$\zeta(t) \equiv \Delta(\tau)\{\bar{y}(t,\tau)\} \tag{9}$$

Comparing equation (8) with equation (5), it appears that the difference between the original and modified Allan variances reduces to a supplementary DMA operator in the case of the modified Allan variance This supplementary operator determines the "software bandwidth" discussed in Allan and Barnes' paper This is illustrated in the functional diagram of figure 1 that shows the sequence of operators that must be applied to the frequency process $y(t)$ in order to produce either the True variance, $I^2(\tau)$, or the original and modified Allan variances For n=1, it can be verified that the original and modified Allan variances are identical

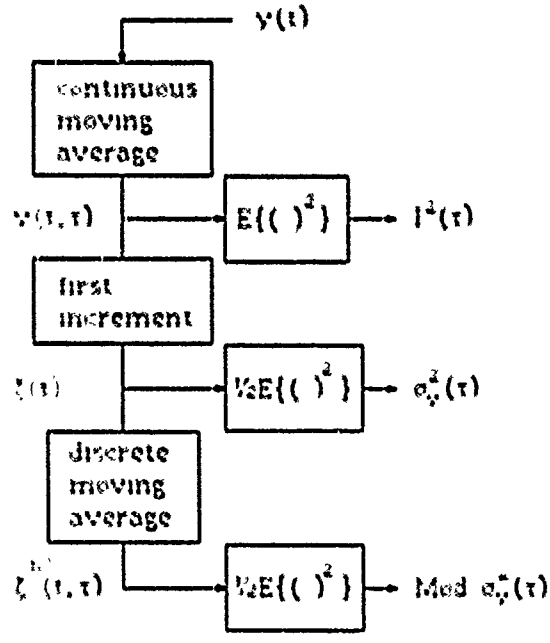


Figure 1
Operational modelization of the True variance and of the original and modified Allan variances

The asymptotic modified Allan variance

If the number n of samples goes to infinity while the averaging time

$$\tau \equiv n\tau_0 \tag{10}$$

is kept constant, the DMA operator may be replaced by its continuous equivalent. Therefore, the asymptotic modified Allan variance, i e the modified Allan variance for $n=\infty$, may be defined as

$$\text{Mod } \sigma_y^2(\infty,\tau) \equiv \tfrac{1}{2} E\left\{[\zeta(t,\tau)]^2\right\} \tag{11}$$

Hence, the asymptotic Allan variance may be calculated in the spectral domain by cascading twice the CMA transfer function and once the FI transfer function. The result is as follows

$$\text{Mod } \sigma_y^2(\infty,\tau) = 2\int_0^\infty S_{yy}^+(f)\,\frac{\sin^6(\pi f\tau)}{(\pi f\tau)^4}\,df \qquad (12)$$

This last integral was calculated for each term of the polynomial model. The result is given in table 1 below

$S^+_{yy}(f)$	Mod $\sigma^2_y(\infty,\tau)$	log-log slope
$h_{-2}\ f^{-2}$	$11/20\ \pi^2\ h_{-2}\ \tau$	1
$h_{-1}\ f^{-1}$	$1/8\ (27\ln 3 - 32\ln 2)\ h_{-1}$	0
h_0	$\frac{1}{4}\ h_0\ \tau^{-1}$	-1
$h_1\ f$	$3/(8\pi^2)\ h_1\ (8\ln 2 - 3\ln 3)\ \tau^{-2}$	-2
$h_2\ f^2$	$3/(8\pi^2)\ h_2\ \tau^{-3}$	-3

Table 1
Relationship between the asymptotic modified Allan variance and the polynomial model

From this table, it can be observed that the slope ambiguity of the original Allan variance is now resolved. The slope μ as a function of τ obeys the law

$$\text{Mod } \sigma_y^2(\infty,\tau) \propto \tau^\mu \text{ where } \mu \equiv -1-\alpha \text{ for } -2 \le \alpha < 3 \qquad (13)$$

where the condition $\alpha \ge -2$ guaranties the stationarity of $\zeta(t,\tau)$. The condition $\alpha < 3$ will be explained below. The ratios

$$R(n) \equiv \frac{\text{Mod } \sigma_y^2(n\tau_0)}{\sigma_y^2(n\tau_0)} \qquad (14)$$

were also calculated for different values of α, using table 1 and table II in 1. The result is given in table 2 below

The law (13) that determines the slope μ is in agreement with Allan and Barnes' results, as well as the ratios $R(\infty)$ computed above for α=-2, -1 and 0 (see table 2 in 3). For α=1 and 2, however, the present analysis shows that the dependence of the modified Allan variance on the hardware bandwidth can be completely eliminated. Hence, the ratios $R(\infty)$ are dependent on B.

α	$R(\infty)$	
-2	$\frac{33}{40} =$	0.825
-1	$\frac{27\ln 3 - 32\ln 2}{16\ln 2} =$	0.674
0	$\frac{1}{2} =$	0.5
1	$\frac{8\ln 2 - 3\ln 3}{2\ln(2\pi B\tau)} =$	$\frac{1.124}{\ln(2\pi B\tau)}$
2	$\frac{1}{2B\tau} =$	$\frac{1/2}{B\tau}$

Table 2
Ratios $R(\infty)$ computed for integer values of α

The difference between the high frequency behaviors of the original and modified Allan variances may be explained by a comparison of the spectral equations (12) and (16) using n=1. On one hand, there is a bandwidth dependence in the original Allan variance because the high frequency stop-band of the equivalent filter follows a f^{-2} envelope which is sufficient to limit the bandwidth of the process $\zeta(t)$ for $\alpha < 1$. On the other hand, the supplementary CMA operator associated with the asymptotic modified Allan variance yields a f^{-4} envelope which is sufficient to limit the bandwidth of the process $\zeta(t,\tau)$ for $\alpha < 3$. Hence, in the range $-2 \le \alpha \le 2$ covered by the polynomial model, the bandwidth dependence is eliminated.

Analysis with constant τ_0

In practice, the modified Allan variance is measured by varying the number n of samples while keeping constant the averaging time τ_0.

The DMA operator is defined as

$$u^{(n)}(t,n\tau_0) \equiv \frac{1}{n}\sum_{k=0}^{n-1} u\,(t - k\tau_0) \qquad (15)$$

Using the transfer functions associated with the CMA, FI and DMA operators, it can be shown from (5) that the spectral expression of the modified Allan variance is as follows.

$$\text{Mod } \sigma_y^2(n,\tau) = 2\int_0^B S_{yy}^+(f)\,\frac{\sin^6(\pi f\tau)}{(n\pi f\tau)^2\,\sin^2\left(\frac{\pi f\tau}{n}\right)}\,df \qquad (16)$$

It can be verified that expression (16) yields the asymptotic modified Allan variance for $n=\infty$ and the original Allan variance for $n=1$. A sharp cut-off frequency B is assumed.

The modified Allan variance associated with the different types of noise processes may be calculated by replacing the power spectral density $S^*_{yy}(f)$ in (16) by each term of the polynomial model. After a simple change of variable one obtains

$$\mathrm{Mod}\ \sigma_y^2(h_\alpha, n\tau_0) = \frac{2h_\alpha}{(n\pi\tau_0)^{1+\alpha}} \int_0^{\pi B n\tau_0} \frac{u^{\alpha-2}\sin^6(u)}{n^2\sin^2\left(\frac{u}{n}\right)}\,du \qquad (17)$$

An examination of this last expression leads to the following conditions of convergence toward the asymptotic modified Allan variance.

Integration Range condition
The bandwidth independence that is sought after can be achieved only if the bandwidth of the process $\zeta(t,\tau)$ is limited by the equivalent filter frequency response (i.e. the software bandwidth) and not by the measuring system frequency response (i.e. the hardware bandwidth). This condition cannot be satisfied unless the hardware bandwidth B is much larger than the center frequency $1/(\pi\tau)$ of the main lobe of the equivalent filter. This determines the following condition on the integration range in (17) which will be referred to as the integration range (IR) condition:

$$\pi B\tau_0 \gg 1/n \qquad (18)$$

Note that this condition is necessary but not sufficient for the convergence of the integral.

Asymptotic Behavior condition
On the other hand, the integrand in (17) becomes independent of n and equivalent to the asymptotic integrand in (12) on the condition that

$$\pi B\tau_0 \ll 1. \qquad (19)$$

This last condition will be referred to as the asymptotic behavior (AB) condition.

The AB condition may be interpreted as a corollary of the sampling theorem as follows. The sampling rate f_0 of the frequency counter is defined as the inverse of the sampling period

$$f_0 \equiv 1/\tau_0 \qquad (20)$$

Therefore the AB condition becomes

$$f_0 \gg \pi B \qquad (21)$$

which means that the CMA operator associated with the asymptotic Allan variance may be approximated successfully by a DMA operator only if the sampling rate f_0 is set high enough in respect to the signal bandwidth B.

The above discussion leads to the distinction of two experimental cases that depend upon the sampling rate as follows

The case of high sampling rate $f_0 \gg \pi B$
When $f_0 \gg \pi B$, the AB condition is always verified and, the IR condition is verified only for large values of n. This implies that for $\alpha=1$ and $\alpha=2$, the Allan variance (n=1) is far below the asymptotic modified Allan variance because the hardware bandwidth is too small and limits the power that gets through the equivalent filter (IR condition not satisfied). But, as the parameter n is increased simultaneously with τ, the software bandwidth becomes smaller and smaller. Ultimately, the IR condition is verified and the asymptotic modified Allan variance is reached.

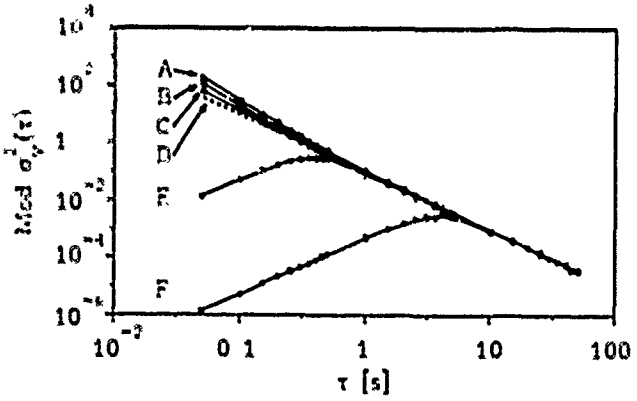


Figure 2
Theoretical calculation of the modified Allan variance for a phase flicker ($\alpha = 1$).

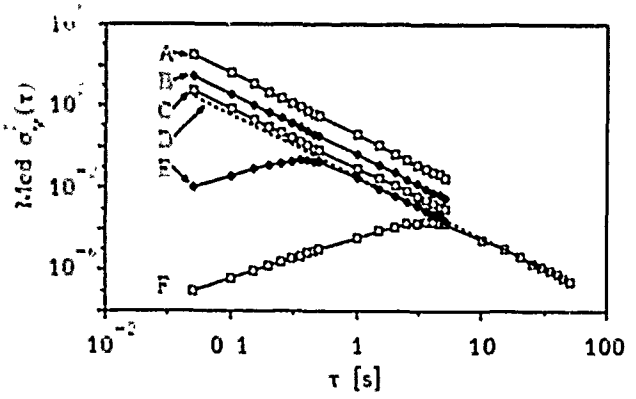


Figure 3
Theoretical calculation of the modified Allan variance for a white phase noise ($\alpha = 2$).

This case is illustrated on figure 2 and 3 by curves E and F. The parameters are given in table 3 below. The sampling period τ_0 was selected in order to match the sampling period used in the experimental verification.

Curve	*Sampling rate*	*B [Hz]*
A	high	1000
B	high	100
C	high	20
E	low	1
F	low	0.1

Figure 2 . $\alpha = 1$, $h_2 = 1$, $\tau_0 = 5.0803 \times 10^{-2}$ s
Figure 3 $\alpha = 2$, $h_2 = 1$, $\tau_0 = 5.0803 \times 10^{-2}$ s.
Curve D is the asymptotic mod Allan var.

Table 3
Parameters of figures 2 and 3.

The case of low sampling rate : $f_0 << \pi B$
When $f_0 << \pi B$, the AB condition is never satisfied, therefore the properties of equation (17) cannot be deduced from those of the asymptotic case. On the other hand, the IR condition is satisfied for all n.

The calculation of the integral in equation (17) shows that for $\alpha=2$ it is independent of n. Therefore the modified Allan variance obeys the law (13) for every n since the dependence upon τ is determined by the factor $2h_\alpha(\pi\tau)^{-1-\alpha}$ outside the integral sign. This property produces curves that are parallel to and higher than the asymptotic Allan variance. The dependence upon the hardware bandwidth is the same than in the case of the original Allan variance, since it is the same for every n including n=1. This case is illustrated on figure 3 by the theoretical curves A, B and C. Note that for B = 20 Hz, the quantity $\pi B\tau_0 = 3.19$ is not much higher than unity and that the corresponding slope is not exactly −3. For larger hardware bandwidths, on the other hand, a good parallelism with respect to the asymptotic curve is observed.

For $\alpha=1$, the integral in (17) becomes independent of n only for large values of n and the limit of the integral is the same as in the asymptotic case. Consequently there is a dependence of the variance upon the hardware bandwidth only for small values of n. The asymptotic curve is reached for large values of n. This case is illustrated on figure 2 by the theoretical curves A, B, and C.

Experimental verification

The above results were verified experimentally in the case of a band-limited white phase noise ($\alpha=2$). The frequency source was a synthesizer, phase modulated by a low-pass filtered white noise generator. The frequency samples were obtained using a pair of frequency counters controlled by a desk-top computer. The dead-time of the individual counters was eliminated by arming one counter and stopping the other using the same transition of an external clock. Hence, while one counter was counting, the other was computing a frequency sample and transmitting it to the desk-top computer through an GPIB 488 bus. The sampling rate used was the highest permitted by the counters : $f_0 = 20$ Hz ($\tau_0 = 50.803$ ms). Each measurement was obtained using 10000 frequency samples and each experimental point is the mean value of 500 samples of the modified Allan variance distributed evenly among the frequency samples. The experimental measurements are shown on figures 4 and 5 and the parameters of the different curves are given in tables 4 and 5 below.

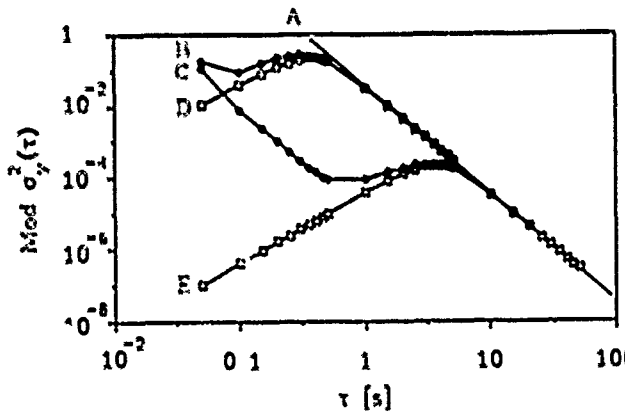


Figure 4
Experimental verification of the modified Allan variance in the case of high sampling rate for a white phase noise ($\alpha = 2$)

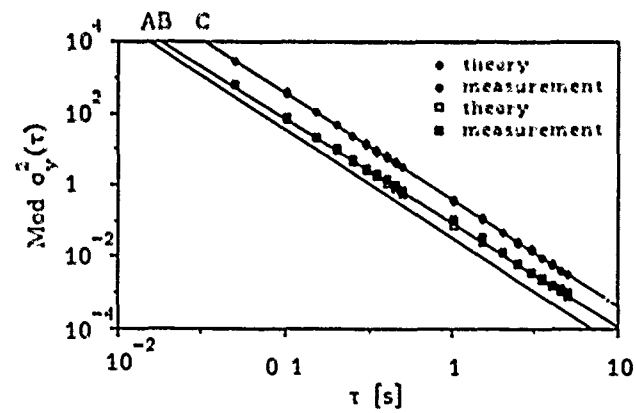


Figure 5
Experimental verification of the modified Allan variance in the case of low sampling rate for a white phase noise ($\alpha = 2$)

Curve	*B [Hz]*	*Type*
B	0 1	experimental
C	1	experimental
D	0 1	theoretical
E	1	theoretical

For all curves : $\alpha = 2$, $\tau_0 = 5\ 0803 \times 10^{-2}$
Curve A is the asymptotic mod Allan var.

Table 4
Parameters of figure 4

Curve	*B [Hz]*	*Type*
B	20	theory + measurement
C	100	theory + measurement

For all curves : $\alpha = 2$, $\tau_0 = 5.0803 \times 10^{-2}$.
Curve A is the asymptotic mod. Allan var.

Table 5
Parameters of figure 5

Verification in the case of high sampling rate

The modified Allan variance was measured for $B = 0.1$ Hz ($\pi B\tau_0 = 1.59 \times 10^{-2}$) and $B = 1$ Hz ($\pi B\tau_0 = 0.16$) using a 4th order low-pass filter. The results are shown on figure 4. Prior to normalization, the coefficient h_2 was evaluated assuming that for the largest value of n the measured variance is equal to the asymptotic Allan variance.

The experimental and theoretical curves are in very good agreement for large values of n and show that the asymptotic variance is effectively independent of the system bandwidth In the region near the maxima of the curves there is a slight discrepancy between the theoretical and experimental curves that comes from the fact that the theoretical curves were calculated for a rectangular low-pass filter while a 4th order low pass-filter was used in the experimental set-up. It can be observed, finally, that for small observation times the internal noise from the synthesizer masks the externally injected white phase noise. This effect explains why the part with a positive slope can hardly be observed on the experimental curves.

Verification in the case of low sampling rate

The modified Allan variance was measured for B=20 Hz ($\pi B\tau_0 = 3.2$) and B=100 Hz ($\pi B\tau_0 = 16$), using a low-pass Cauer filter. The results are shown on figure 5. Prior to normalization, the parameter h_2 was evaluated from the measurement of the original Allan variance (n=1) using the nominal bandwidth of the Cauer filter.

The agreement between the theoretical and experimental curves is very good Moreover, the experimental measurement confirms the lack of parallelism between the modified variance and the asymptotic modified variance when $\pi B\tau_0$ is not much larger than unity (B=20 Hz)

Conclusion

In this paper a theoretical analysis of the modified Allan variance is presented. The analysis is stated in the spectral domain and takes into account the band limiting effect of the measuring system. The principal results are :

1) A chart of the exact coefficients that relate each term of the polynomial model to the corresponding asymptotic modified Allan variance
2) Both the general and asymptotic integral expressions of the modified Allan variance as a function of the single-sided power spectral density of the frequency fluctuations
3) The definition and the analysis of the cases of low and high sampling rates that determine the two possible behaviors of the modified Allan variance when it is measured by varying n while keeping τ_0 constant

The main conclusion is that the modified Allan variance can eliminate not only the temporal-spectral ambiguity it was designed to solve, but also the system bandwidth dependence of the original Allan variance

References

1 J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources : Fifteen Years of Progress," *IEEE Proceedings,* vol. 66, no. 9, September 1978, pp 1048-1075

2 W C Lindsey, C. M. Chie, "Theory of Oscillator Instability Based upon Structure Functions," *IEEE Proceedings,* vol. 64, December 1976, pp. 1652-1666

3 D. W. Allan, J. A. Barnes, "A Modified 'Allan Variance' with Increased Oscillator Characterization Ability," *Proc. 35th Annual Frequency Control Symposium,* Philadelphia, May 1981, pp. 470-475.

4 P Lesage, T. Ayi, "Characterization of Frequency Stability. Analysis of the Modified Allan Variance and Properties of Its Estimate," *IEEE Trans. on Instrumentation and Measurement,* vol. IM-33, no. 4, December 1984, pp. 332-336.

FREQUENCY STABILITY CHARACTERIZATION OF HOPPING SOURCES

Grigorios A. Kalivas and Robert G. Harrison
Department of Electronics
Carleton University, Ottawa, K1S 5B6, Canada.

Abstract

The increasing demand for Spread Spectrum (SS) Systems utilizing Frequency Synthesizers imposes a need for development of fast Frequency Hopping (FH) Systems which constitute the most important part of FH-SS or Hybrid-SS Communication Systems. Speed limitations of other systems suggests the employment of fast-tuning Voltage Controlled Oscillators (VCOs) which are digitally controlled through a Digital-to-Analog Converter (DAC). Since these sources are not phase-locked the output noise does not obey the PLL output noise formulation. Therefore an attempt is made here to use time-domain frequency stability measures to characterize the noise of this kind of free-running Hopping Sources.

Introduction

The main problem addressed in this paper is the elucidation of a procedure leading to noise characterization of a class of frequency-hopping oscillators. Until now, various variance measures have been introduced and used, [1], [2], [3], [4], [5] to characterize the *Short Term Frequency Stability* (STFS) of free-running oscillators. An attempt is made here to apply the existing measures to frequency-hopping sources.

As pointed out in the literature, [6], [7], [8] thermal effects, such as changes in junction temperature of the active elements, power dissipation changes and instability of the bias sources are the primary factors contributing to short-term post tuning drift (PTD). All these changes are the result of tuning voltage change commands when the system is to be hopped from one frequency to another. Analog and digital noise coupled to the tuning voltage input of a Voltage Controlled Oscillator (VCO) can also effectively increase the frequency instability of the system [9]. When these effects are taken into account it becomes apparent that the STFS of a hopping system may exhibit behaviour unlike that of a free-running frequency source.

Before this section is closed the formulation of the problem is presented. Let the oscillator output signal be $g(t)$:

$$g(t) = A\cos(\omega t + \phi(t)) \tag{1}$$

and the instantaneous frequency fluctuations (IFF) be $y(t)$:

$$y(t) = \frac{d\phi(t)}{dt}\cdot\frac{1}{\omega} \tag{2}$$

A digital counter measures the number of zero level crossing points within an interval τ which is called the averaging interval. A nominal frequency F_r is measured every T_1 sec for τ sec (Figure 1).

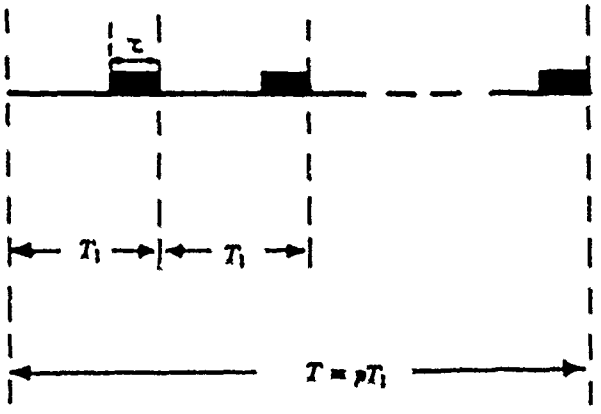


Fig. 1: Graphical Representation of Measurement Procedure

T_1 = time between two successive measurments of the same nominal frequency F_r.

p = total number of measurments needed to characterize the statistical properties of $y(t)$.

Hence the short term average $\overline{y}_\tau(t)$ represents the quantity actually measured by the counter:

$$\overline{y}_\tau(t) = \frac{1}{\tau}\int_{t-\tau}^{t} y(t)dt. \tag{3}$$

Considering $y(t)$ as a signal, equ. (3) can be regarded as a smoothing operation on $y(t)$. It can easily be shown that the following relationship exists :

$$\gamma_{\overline{y}_\tau(t)}(f) = \gamma_y(f)\left(\frac{\sin\pi f\tau}{\pi f\tau}\right)^2 \tag{4}$$

where:

f represents the frequency,

γ_y is the power spectral density (PSD) of $y(t)$, and

$\gamma_{\overline{y}_\tau(t)}$ is the PSD of $\overline{y}_\tau(t)$.

The best way of characterizing the short-term frequency fluctuations under consideration, is to use the variance of $y(t)$. This is defined as:

$$\sigma_y^2 = \int_{-\infty}^{+\infty}\gamma_y(f)df. \tag{5}$$

CH2427-3/87/0000-122 $1.00 C 1987 IEEE

Obviously this filter is not realizable and equ. (5) is approximated by:

$$\sigma_y^2 \approx \int_{-\infty}^{+\infty} |G_i(f)|^2 \gamma_y(f) df, \qquad (6)$$

where $G_i(f)$ is a filter gain term.

Description of Measurement System.

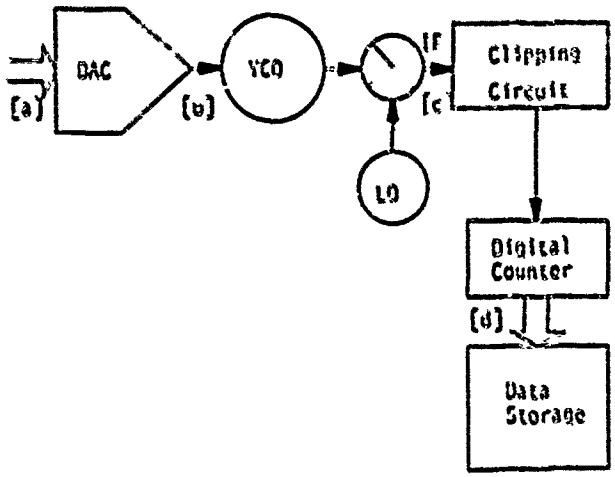


Fig. 2: System Implementation

The measurement system is illustrated in Fig. 2. At point [a] a digital tuning command is applied to the input of the DAC. A particular digital word, d_r will cause the DAC to issue a corresponding analog voltage at point [b]. This voltage is applied to the VCO which in turn issues a particular frequency F_r. This frequency is mixed with a very stable Local Oscillator (LO) to produce a downconverted replica of the VCO output which bears all the STFS information [1]. The mixing stage is needed because the frequencies are not directly counted within the VCO operating range (1 - 1.5 GHz). Centering the LO at 1.25 GHz will produce intermediate output frequencies (IF) in the 0-250 MHz range which can be easily counted.

At point [c] the IF is digitized (by clipping) and counted for a period of τ secs every T_1 secs (see Figure 1). The output of the digital counter is then stored.

Depending on the digital command at point [a], this system may be hopping between several (or even several hundred) frequencies. Hence for a sequence of digital words $d_1, d_2, \cdots, d_n$ applied successively we can have frequencies $F_1, F_2, \cdots, F_n$ respectively at the VCO output.

In our application it was assumed that the frequency noise of the hopping source at a particular frequency is mainly affected by the immediately preceding frequency.

Variance Estimates and their Equivalent Filter Representation.

The gain term $G_i(f)$ in equ. (6) depends on the way in which frequency measurments are taken and processed in order to approximate σ_y^2 [1],[4]. Specifically, the principal reason for the existence of the gain term in the variance estimator expressions is the finite number, p, of frequency measurments used to characterize the variance.

Hence the general estimate of the variance is given by the following equation:

$$\hat{\sigma}_i^2(\tau, p, T_1) = \int_{-\infty}^{+\infty} |G_i(f)|^2 \gamma_{\bar{y}_\tau(n)}(f) df. \qquad (7)$$

The different ways of calculating the noise variance are called here "variance measures" the most important of which are described below:

The first of these variance measures was introduced by Barnes [2] and Allan [3] in 1966. For each set of p samples, the following quantity is calculated:

$$\sigma_1^2(\tau, p, T_1) \equiv P_1 =$$

$$\frac{1}{p} \sum_{n=1}^{p} \left[\bar{y}_\tau(n) - \frac{1}{p} \sum_{k=1}^{p} \bar{y}_\tau(k) \right]^2 \qquad (8)$$

where

$$\bar{y}_\tau(n) \equiv \bar{y}_\tau(t_n). \qquad (9)$$

The estimate of the variance is given by:

$$\hat{\sigma}_1^2(\tau, p, T_1) \equiv E(P_1) \qquad (10)$$

where E denotes the mathematical expectation operator. Using equs. (8) and (10) it is shown [1] that the estimate of the variance is given by equ. (7) for i=1. $|G_1(f)|^2$ is as follows:

$$|G_1(f)|^2 \equiv 1 - \left(\frac{\sin p\pi f T_1}{p \sin \pi f T_1} \right)^2 \qquad (11)$$

The second variance measure was introduced by Boileau and Picinbono [4] in 1976. In this case the variance is given by the following expression:

$$\sigma_2^2(\tau, p, T_1) \equiv P_2 \equiv$$

$$\left[\bar{y}_\tau\left(\frac{p+1}{2}\right) - \frac{1}{p} \sum_{n=1}^{p} \bar{y}_\tau(n) \right]^2 \qquad (12)$$

where p is an odd number. An estimate of the variance is:

$$\hat{\sigma}_2^2(\tau, p, T_1) \equiv E(P_2). \qquad (13)$$

This is shown [4] to give $\hat{\sigma}_2^2(\tau, p, T_1)$ as in equ. (7) for i=2. In this case $|G_2(f)|^2$ is given by equ. (14):

$$|G_2(f)|^2 = \left(1 - \frac{\sin p\pi f T_1}{p \sin \pi f T_1} \right)^2 \qquad (14)$$

Hence the two different variance measures (equs. (8),(12)) produce two different filter gain expressions $|G_i(f)|^2$. If certain restrictions are imposed to the filter gain term, several other complicated measures can be derived [5]. These restrictions are meant to impose a specific shape on $|G_i(f)|^2$ in an effort to approximate equ.(5) in the best possible way. In other words this filtering of $\bar{y}_\tau(n)$ can be represented as a suitable windowing of its PSD. Various windows give different expressions for $|G_i(f)|^2$. Hence different windowing means different weighting of the existing group of measurements.

Criterion for the Choice of a Suitable Variance Estimator and the Importance of τ.

According to the above discussion the usage of equ.(7) is twofold:

(a) The first usage is as a measure to indicate the behaviour of σ_i^2 as a function of various parameters such as τ, p and T_1.

The objective of applications directly related to this aspect, is not to obtain $\gamma_y(f)$ itself but to test a model for $\hat{\sigma}_i^2$ with a finite number of measurements. Therefore no special properties are required for the filter $G_i(f)$. Clearly $G_1(f)$ and $G_2(f)$ belong to this group of filters the main advantage of which is the simplicity of the computations needed to obtain $\hat{\sigma}_i^2(\tau, p, T_1)$. Let us call these, variance measures of the *first kind*.

(b) The second usage of equ. (7) is as a way to acquire a reasonably accurate expression of the spectral density $\gamma_y(f)$ of the frequency fluctuations.

It is here that the restrictions mentioned above are imposed on $|G_i(f)|^2$. In this way a global function corresponding to the ideal variance $\sigma(T \to \infty)$ is developed [5] despite the fact that this global function will be a "finite time variance" since T will always be finite. For this group of measures complicated calculations are needed to obtain σ_i^2. Let us call this second group, variance measures of the *second kind*.

As pointed out here, for FH system applications, an estimate of the short-term noise only is needed. This is in contrast to a precise expression for the variance resulting in the estimation of $\gamma_y(f)$. Hence $\hat{\sigma}_1^2$ or $\hat{\sigma}_2^2$ are adequate for this type of characterization.

Let us now demonstrate the importance of properly selecting the averaging interval τ. It has been shown elsewhere, [1],[4], that as $T \to \infty$, the measures $\hat{\sigma}_i^2, i = 1,2$ may diverge.Simulated results in [4] indicate a behaviour as illustrated in Fig.3.

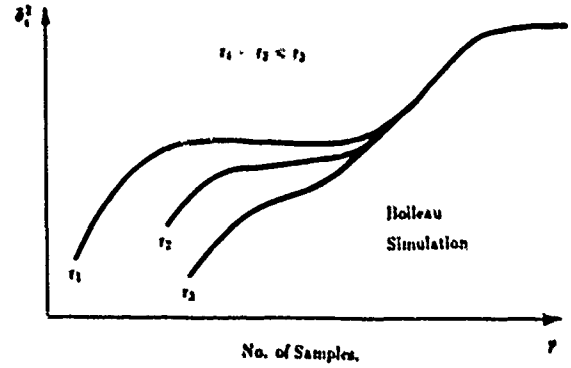


Fig. 3: Behaviour of $\hat{\sigma}_i^2$ As a Function of p

The effect of the value of τ can be clearly seen by taking into account the following considerations:

(a) $y(t)$ can be represented as follows:

$$y(t) = x(t) + n(t) \tag{15}$$

where $n(t)$ represents a noise factor with zero mean and autocorrelation $R_n(\tau')$. $R_n(\tau')$ is of short duration compared to τ (almost white noise) and its area is equal to A. $x(t)$ represents a signal containing all the information about IFF $y(t)$. An estimate of $x(t)$ is the short term average, as given by equ.(3). It can be shown that:

$$E[\bar{y}_\tau(t)] = \frac{1}{\tau}\int_{t-\tau}^{t} x(t)dt \tag{16}$$

$$\sigma_{\bar{y}_\tau}^2 \approx \frac{A}{\tau}. \tag{17}$$

From equ.(17) one can see that to reduce the variance of $\bar{y}_\tau$, τ must have a relatively large value. On the other hand τ should be small enough to ensure that $E[\bar{y}_\tau(t)]$ is close to $x(t)$ at any instant t.

(b) From equ. (4) it can be seen that τ should be small enough such that short term noise, which corresponds to the high frequency portion of the PSD, will not be altered (i.e. filtered out) in any way. Hence for small values of τ the dependance of the PSD and hence of $\hat{\sigma}_i^2$ on τ will be removed and a flat portion will appear in $\hat{\sigma}^2$-versus-p curves as indicated in Fig. 3.

A criterion for choosing between all the measures presented above can now be stated:

"For frequency hopping oscillators use a $\hat{\sigma}^2$ of the first kind. This measure should exhibit an approximately "flat" portion on the $\hat{\sigma}^2$-versus-p curve. If such a flat region exists, it can be interpreted as the short term variance of the hopping system. If there is no flat region, the value of τ should be decreased gradually until a flat part is obtained on the $\hat{\sigma}^2$-versus-p curve."

Experimental Results

Several data files were acquired using the system described in the Section 3 which was set to hop between two frequencies with τ equal to 70 μsecs. The difference between these two frequencies is conventionally called here the frequency step. Before starting to record any measurements, the system was set to hop for an adequate amount of time in order to establish thermal equilibrium. Each of the obtained data files contains approximately 15,000 data points. From the measures of the first kind, $\hat{\sigma}_1^2$ was used since it is considered to be the most widely recognized variance measure. A FORTRAN program was used to calculate $\hat{\sigma}_1^2$ according to equs. (8) and (10). Figures 4, 5, 6 illustrate $\hat{\sigma}_1^2 \times 10^7$-versus-$p$ for frequency steps F_{d1}=5 MHz, F_{d2}=10 MHz, F_{d3}=20 MHz. It can be seen from these figures that there always exists a "flat" region which may constitute a measure for the variance of the hopping frequency source. For instance the estimated variance from Fig. 4 is equal to 0.55×10^{-5}. It is also worth noting that the variance estimates are diferent for diferent frequency steps. This is expected since the short term PTD of the VCO depends on the frequency step.

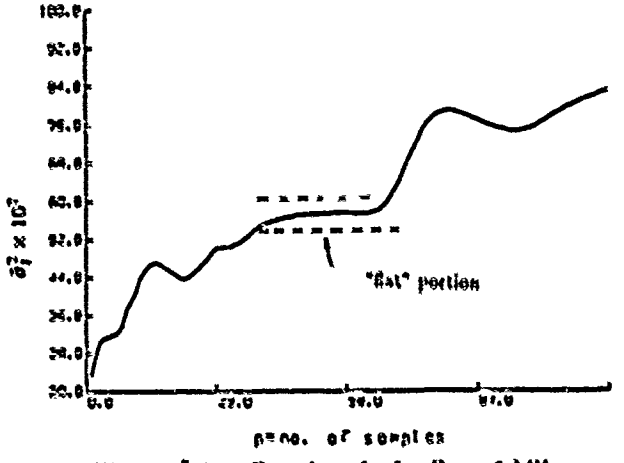


Fig. 4: ∂_1^2 As a Function of p for $F_{d1} = 5$ MHz

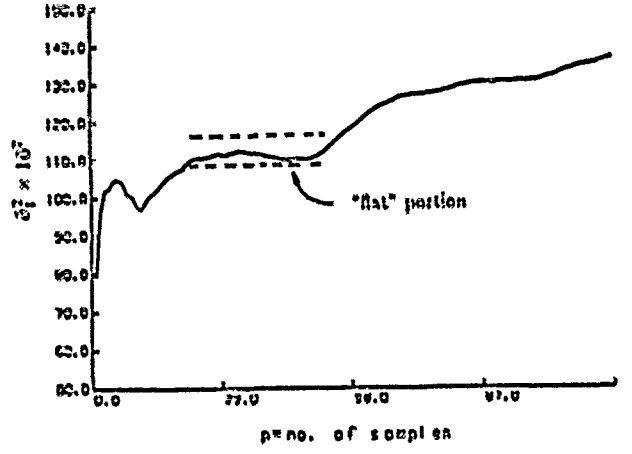


Fig. 5: ∂_1^2 As a Function of p for $F_{d2} = 10$ MHz

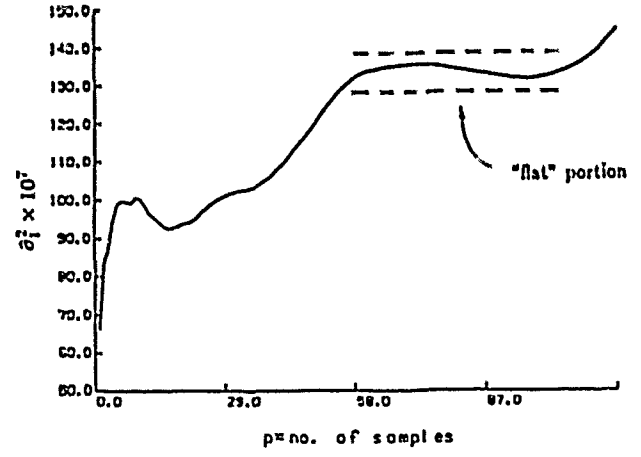


Fig. 6: ∂_1^2 As a Function of p for $F_{d3} = 20$ MHz

Conclusions

The classification of the existing variance estimators produced a criterion which can be used to choose an estimator suitable for the characterization of frequency hopping sources. It was experimentally demonstrated that ∂_1^2 can indeed be used to produce variance estimates of a hopping VCO for suitable values of r. In general it was shown that despite the fact that hopping sources exhibit orders-of-magnitude larger amounts of short term noise than stable free-running frequency sources, some of the techniques used to estimate the noise of stable sources are applicable to hopping VCO systems as well.

Acknowledgements: The authors wish to thank the Natural Sciences and Engineering Research Council of Canada for the financial support of this work.

REFERENCES

[1] J. Barnes et al., "Characterization of Frequency Stability", *IEEE Trans. Instrum. Meas.*, vol. 20, No. 2, pp. 105-120, May 1971.

[2] J. Barnes, "Atomic Timekeeping and the Statistics of Precision Signal Generators", *IEEE Proceedings*, vol. 54, No. 2, pp. 207-220, February 1966.

[3] D. Allan, "Statistics of Atomic Frequency Standards", *IEEE Proceedings*, vol. 54, No. 2, pp. 221-230, February 1966.

[4] E. Boileau, B. Picinbono, "Statistical Study of Phase Fluctuations and Oscillator Stability", *IEEE Trans. Instrum. Meas.*, vol. 25, No. 1, pp. 66-75, March 1976.

[5] E. Boileau, "Improvements of the Procedures Used to Study the Fluctuations of Oscillators", *IEEE Trans. Instrum. Meas.*, vol. 27, No. 3, pp. 210-214, September 1978.

[6] R. Buswell, "VCO's in modern ECM Systems", *Microwave Journal*, vol. 18, No. 5, pp. 43-46, May 1975.

[7] H. Schoniger, "Evaluating Tuning Response of VCO Subsystems", *Microwave Systems News*, vol. 7, No. 7, pp. 45-54, July 1977.

[8] "What You Need to Know About Microwave VCO's", *Frequency Sources Inc.* Application Note.

[9] V. Manassewitsch, "Frequency Synthesizers, Theory and Design", J. Wiley, New York, 1976.

A METHOD FOR USING A TIME INTERVAL COUNTER TO MEASURE FREQUENCY STABILITY

C. A. Greenhall

Jet Propulsion Laboratory, California Institute of Technology
Pasadena, California 91109

Summary

A commercial time interval counter can be used to measure the relative stability of two signals that are offset in frequency and mixed down to a beat note of about 1 Hz. To avoid the dead-time problem, the counter is set up to read the time interval between each beat note upcrossing and the next pulse of a 10 Hz reference pulse train. The actual upcrossing times are recovered by a simple algorithm whose outputs can be used for computing residuals and Allan variance. A noise floor test yielded a $3f/f$ Allan deviation of $1.3 \times 10^{-9}/\tau$ relative to the beat frequency.

Beat-Frequency Method

In the beat-frequency or single-mixer method of frequency stability measurement, the two sources to be compared have a small frequency offset, typically in the neighborhood of 1 Hz. The two sources at frequency f_0 are mixed down to a sinusoidal beat note at the offset frequency f_b. This sine wave is passed through a zero-crossing detector, which produces a square wave or pulse train at the same frequency, as in Fig. 1. The relative time deviation or fractional frequency deviation of the two sources is equal to f_b/f_0 times that of the square wave or, more precisely, its stream of upcrossings, which are spaced approximately 1 second apart. The improvement to be discussed below deals only with the measurement of these upcrossing times; the analog front end of the system remains the same.

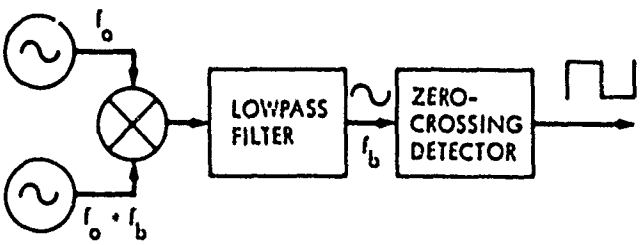


Fig. 1. Analog portion of frequency stability test set

Current Measurement System

Fig. 2 shows the current system used at the JPL Interim Frequency Standards Test Facility (IFSTF). The square wave beat note goes to a custom-built digital module that latches the readings of a free-running 1 MHz 30-bit counter at the upcrossings and writes them to a 7-track tape, which is processed offline by a mainframe computer. In this way, the upcrossing times are captured with a resolution of 1 μs. Otoshi and Franco[1] have been using a system similar to the IFSTF's, but with a 10 MHz counter, to measure Deep Space Station stability.

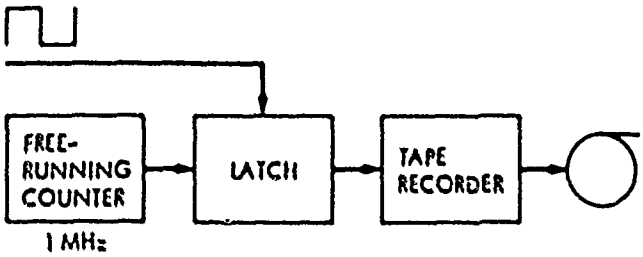


Fig. 2. Current digital portion of test set

An Improved System - The Picket Fence

Although the IFSTF system has given good service, it has several disadvantages: The counter resolution is too coarse for certain applications whose f_b/f_0 ratio is not small enough to put the quantization noise floor below the frequency instabilities to be measured. The 7-track incremental tape drives are obsolete, expensive to rent, and unreliable. The offline tape processing and graph plotting is slow and cumbersome.

Because the frequency stability testing program is soon to be moved to the new Frequency Standards Laboratory, this is a good time to develop a more accurate and more convenient test set. The laboratory already has a Hewlett Packard 1000 realtime computer and several HP 5334A universal counters, which can measure time intervals with a 1 ns resolution and interface with the computer on a IEEE-488 bus. The problem has been that these are interval counters with dead time between measurements; unaided, such a counter can measure at most every other period of the stream of beat-note upcrossings. This dead-time limitation has now been overcome by the introduction of one more element, a 10 PPS (pulses per second) reference signal, called the Picket Fence, provided by a reference signal and a divider.

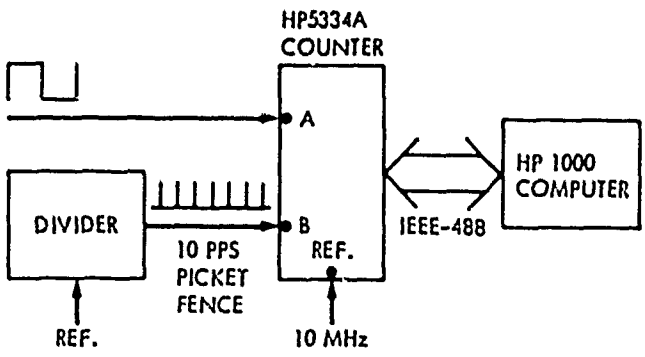


Fig. 3. Picket fence setup

CH2427-3/87/0000-126 $1.00 C 1987 IEEE

Test Setup and Procedure

The new setup is shown in Fig. 3. The beat note square wave goes into Input A of the counter, the picket fence into Input B. The same frequency standard should drive both the counter and the divider to keep the picket fence coherent with the counter. In a multichannel system, each beat note has its own counter and there is only one picket fence signal going into all the B inputs.

To carry out a test, one first uses the Period A function of the counter to make a preliminary measurement of the period p of the square wave. It is permissible to use a long gate time or the 100-reading average function when doing this measurement. Having measured the nominal period, one switches the counter to its Time Interval A to B function, and records all subsequent readings. Each reading is the time interval between an upcrossing and the next picket fence pulse. As long as the periods are not too short, the counter has time to reset itself between readings, and hence no upcrossing is missed. From these raw data, the actual upcrossing times are recovered in software by an algorithm given in detail below. Fig. 4 shows the time evolution of the measurement process.

Data Processing

Let d be the picket fence period (100 ms), p the initial period measurement, and $v_0, v_1, v_2, \ldots$ the sequence of time interval data. Let $t_0, t_1, t_2, \ldots$ be the actual upcrossing times relative to some time origin, perhaps one of the picket fence pulses. Each t_n differs from the corresponding $-v_n$ by an unknown integer multiple of d, and we would like to resolve the ambiguities.

Let Δ denote the backward difference operator, e.g., $\Delta t_n = t_n - t_{n-1}$. We make the following assumptions about the beat note:

Assumptions

1. The first period Δt_1 differs from p by less than d/2. Any two successive periods Δt_{n-1}, Δt_n differ by less than d/2. This guarantees that the 100 ms ambiguities can be resolved uniquely.

2. Each period Δt_n is greater than d - g, where g is the maximum dead time of the counter. This guarantees that no upcrossing is missed.

Since the t_n increase quickly and may contain important information in their least significant digits, they are awkward to compute, store, and use. Accordingly, the algorithm actually computes the sequence of time residuals defined by

$$x_n = t_n - t_0 - np, \quad n = 0,1,2,\ldots \quad . \qquad (1)$$

At the core of the algorithm is the signed residue function Smod(x,m), which is defined to be x minus the closest integer multiple of m to x. For example, Smod(3,5) = Smod(-7,5) = -2. If x is halfway between two integer multiples of m, then it doesn't matter whether Smod(x,m) is defined to be m/2 or -m/2.

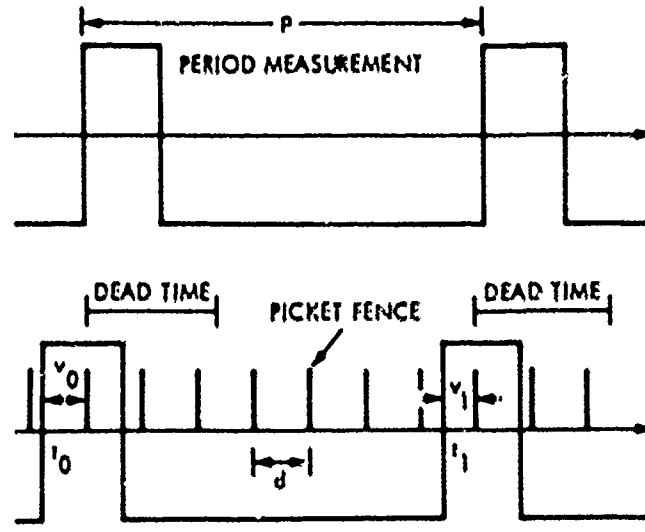


Fig. 4. Picket fence measurement process

The following algorithm incorporates a small amount of error handling to prevent one bad input from spoiling all the subsequent outputs.

Unfolding Algorithm

```
! Global parameters:  p, d
Do for n = 0, 1, 2, ...
     Read v_n
     Call Unfold (n, v_n, x_n)
     Output x_n
Enddo
```

```
Subroutine Unfold (n, v, x)

! Inputs:  n, v
! Output:  x
! Local variables
!    du:  current Δu = -Δv
!    dx:  current Δx
!    dua:  du anchor
!    dxa:  dx anchor
!    ddx:  current dx - dxa or second difference
!    vl:  previous v
!    xl:  previous x
1.   If n = 0 then
          ! Initialize
2.        dua = p
3.        dxa = 0
4.        x = 0
5.   Else
6.        du = vl - v
7.        ddx = Smod (du - dua,d)
8.        dx = dxa + ddx
9.        x = xl + dx
          ! Error handling
10.       If |ddx| < d/4 then
               ! Data OK, drag anchor along
11.            dxa = dx
12.            dua = du
13.       Endif
14.  Endif
15.  xl = x
16.  vl = v
17.  Return
```

Remarks. Let us first discuss the algorithm with the error-check lines 10 and 13 removed. (Lines 11 and 12 remain.) In the appendix, it is proved that the outputs x_n are identical to the numbers defined by Eq. (1). The algorithm accomplishes this by forcing the second difference $\Delta^2 x_n$ to be less than d/2 in absolute value (in accordance with Assumption 1) and also equivalent to $-\Delta^2 v_n$ modulo d. Nevertheless, some protection against bad data is needed. Table 1 shows what can happen if there is one error with magnitude slightly greater than d/4. The underlined value of ddx (n = 3) has been converted by Smod from -0.52d to 0.48d. As a result, the x_n start to increase linearly. This can cause computational problems for a program that analyzes the x_n. The error check with threshold d/4 anchors the current data to the last good data. If Δx_m is the anchor dxa at time n, and the condition in line 10 is satisfied at times n and n+1, then Δx_{n+1} is within d/2 of Δx_m. In the example, the anchor stays at Δx_1 for n = 2 and 3. At n = 4, it is pulled up again. As a result, the x_n go back to 0 after time 2. Time will tell whether this simple expedient is adequate in practice.

Table 1. State of the unfolding algorithm after line 9 (d = 1) with one bad input and no error handling.

n	v	vl	xl	dua	dxa	du	ddx	dx	x
0	0		line 9 not applicable						0
1	0	0	0	10	0	0	0	0	0
2	-.26	0	0	0	0	.26	.26	.26	.26
3	0	-.26	.26	.26	.26	-.26	.48	.74	1.00
4	0	0	1.00	-.26	.74	0	.26	1.00	2.00
5	0	0	2.00	0	1.00	0	0	1.00	3.00
6	0	0	3.00	0	1.00	0	0	1.00	4.00

Noise Floor Test

A single-channel hardware and software system was set up to evaluate the technique (Fig. 3). The HP 1000 collects the data in real time and stores the outputs of the unfolding algorithm on disk. Simultaneously, a user can cause any portion of the test to be processed into residuals and Allan deviation.

To measure the ultimate noise floor of the technique and to test the integrity of the measurement system, the beat note square wave was simulated by the output of a stable, low-rate pulse generator assembled from a synthesizer, a divider, and another counter. The generator's period, which could be programmed to the nearest nanosecond, was chosen to be 0.938196601 s = (10 - r)d, where d = 0.1 s, the picket fence period, and r is the Golden Ratio, (√5 - 1)/2. This period guarantees a good mix of counter readings v_n (Knuth,[2] pp. 510-11, 543). If a conveniently available 1 PPS signal had been used, the counter would have always been reading the same value. The same frequency source, a cesium standard, was used for driving the counter, the picket fence, and the pulse generator. Thus, the results include instabilities and errors of the pulse generator, picket fence divider, and counter, but not of the measurement time base that drives these components.

A test of duration 108600 s was carried out. The accumulated time residuals, with the mean frequency offset removed, remained within a 6 ns band over the entire run. The Allan deviation, shown in Fig. 5, was approximately $1.3 \times 10^{-9}/\tau$ for τ between 0.94 s and 11500 s. This shows that all the equipment maintained time coherence at the nanosecond level and that the counter met its specifications.

To see what this means for an actual test of frequency sources, recall that these numbers must be scaled down according to the source and beat frequencies. For example, if two 1 MHz sources with a 1 Hz offset were being tested, the digital part of the measurement system would contribute a $\Delta f/f$ Allan deviation of $1.3 \times 10^{-15}/\tau$.

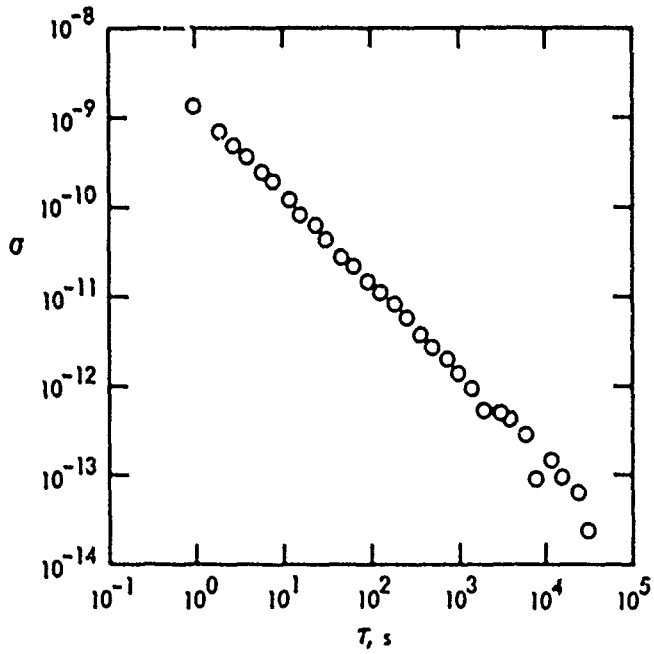


Fig. 5. Allan deviation noise floor

Conclusions

The technique given here is a method for measuring the stability of the square wave produced by mixing two offset frequency sources and passing the low-frequency sine wave through a zero-crossing detector. In contrast to the current IFSTF system, which uses custom digital hardware and has microsecond resolution, this technique uses commercial hardware of moderate cost (under $3000 per counter) with a IEEE-488 interface, and has nanosecond resolution.

The author has recently learned of the existence of a counter assembly that can latch the readings of a free-running counter with a precision of 1 ns for several input channels. The counter rolls over every 2^{24} ns, about 16.8 ms, so that in effect the counter makes its own picket fence with that period, and the same unfolding algorithm applies. A multichannel frequency stability measurement system built around this unit might be smaller and less expensive than one built around several interval counters in separate chassis.

Acknowledgment

The author would like to thank G. J. Dick and R. E. Taylor for stimulating conversations that led to the ideas used in this work.

Appendix: Proof of Correctness of the Unfolding Algorithm

To avoid confusion let the output of the algorithm be denoted by X_n, while x_n is still defined by Eq. (1). The error correction lines 10 and 13 are deleted. We shall prove by induction that $X_n = x_n$ for all n. For n = 0, both quantities are zero. For n = 1 we have

$$du = -\Delta v_1 \equiv \Delta t_1 \pmod{d},$$

$$ddx = Smod(\Delta t_1 - p, d) = \Delta t_1 - p \quad \text{(assumption 1)}$$

$$= x_1.$$

Therefore, $X_1 = x = x_1$ since dxs = xl = 0. For $n > 1$ assume the algorithm is correct up to n-1. Then

$$dus = -\Delta v_{n-1} \equiv \Delta t_{n-1} \pmod{d},$$

$$du = -\Delta v_n \equiv \Delta t_n \pmod{d},$$

$$ddx = Smod(\Delta^2 t_n, d) = \Delta^2 t_n \quad \text{(assumption 1)}$$

$$= \Delta^2 x_n,$$

$$dxs = \Delta x_{n-1},\ xl = x_{n-1},$$

$$dx = \Delta x_{n-1} + \Delta^2 x_n = \Delta x_n,$$

$$X_n = x = x_{n-1} + \Delta x_n = x_n.$$

References

1. T. Y. Otoshi and M. M. Franco, DSS frequency stability tests performed during May 1985 through March 1986, TDA Progress Report 42-86, pp. 1-14, Jet Propulsion Laboratory, 1986.

2. D. E. Knuth, The Art of Computer Programming, Vol. 3, Addison-Wesley, 1973.

POSITIONING AND TIMING STUDY OF GPS C/A CODE RECEIVERS

Qixiang Zhuang, William J. Klepczynski, and Carl F. Lukac
U. S. Naval Observatory
Washington, D. C. 20392-5100

Abstract

Using three types of differently designed L1, C/A code GPS receivers (STI, AOA and DATUM), work has been done at the U.S. Naval Observatory on investigating the actual positioning and timing abilities of the GPS C/A code receiver. Some of the results obtained during the time period from October 1985 to November 1986 are presented in this article.

A 10 m RSS accuracy in absolute position determination has been demonstrated with high reliability using an L1, C/A code GPS receiver. This is the conclusion from data measured continuously over 5 months using four different satellite configurations.

Using a delta-range measurement technique and tracking each satellite for 77 minutes, a 7 m RSS precision in position determination has been obtained.

Common-view timing data between two side-by-side GPS receivers has been compared. The corresponding measurement noise and systematic error have been analysed and are discussed.

Geodetic relative positioning has been done by making observations in the "double difference with time" interferometric mode. For baselines (up to 1500 km in length), the C/A code receivers are capable of relative positioning accuracy of the order 1 - 2 m RSS with PDOP of about 5.0.

Introduction

Concerning the absolute position accuracy of the L1, C/A code GPS receiver, some tests or analyses have been made by Parkinson and Gilbert (1983), Ashjaee (1985) and Brown and Sturza (1985).

An L1, C/A code Allen Osborne, Associates (AOA) TTRS GPS receiver was used at USNO to perform positioning tests on a long-term basis. Four satellite configurations, arbitrarily called classes 95,96,97 and 98, were chosen for daily measurement and each satellite configuration has a different geometric dilution of precision (GDOP). Tests lasted two time periods, MJD 46355 to 46443 and MJD 46452 to 46522. The satellite configurations of classes 96 and 97 changed a little, while classes 95 and 98 were kept constant during these two periods.

The Delta-range (DR) measurement technique (McCaskill, et al., 1976) was implemented, and the measurement errors and navigation solution accuracies were evaluated for pseudo-range (PR) and DR measurements, respectively. The dependence of solution accuracy of DR on the accumulated time interval also was analysed. DR measurement is useful to exploit the positioning ability and versatility of an L1, C/A code GPS receiver.

Another C/A code GPS receiver STI-5026 is mounted in the same building with the AOA receiver, where the baseline between the two antennas is about 17.7 m. Adopting a common-view approach with UTC(USNO,MC) as the common reference clock, the primary error effects are essentially canceled except for receiver delay instability and multipath distortion. According to the common-view measurements which were executed with different satellites, and different satellite elevations at different times of the day, the common-view measurement noise was found to be 2.68 ns for nighttime and 4.48 ns for daytime. A relationship between common-view systematic error and satellite elevation has been found.

In a third experiment, the AOA receiver was fixed as a stationary reference (with the coordinates as shown in Table 2) and a DATUM-9390 GPS C/A code receiver was transported to different locations as a portable one. They tracked the same satellites SV12, SV11, SV13 and SV9 in sequence over the same time interval of one hour for each satellite and were differenced to measure relative position. For the one hour time interval tracking and a PDOP of 5.0, a relative position of 1 - 2 m RSS was obtained. Over the USNO - NRAO baseline of about 247.0 km, this is about an 8 ppm geodetic survey accuracy.

Pseudo-Range Measurement

The PR model equation is given by

$$l_i = R_i + cT_0 + C_{ion} + C_{trop} \qquad (1)$$

where l_i is pseudo-range, R_i is the corresponding true range, cT_0 is receiver clock bias relative to GPS time, C_{ion} is the correction for ionospheric delay error, and C_{trop} is the correction for tropospheric delay error. In practice, the relativistic and Earth rotation corrections also should be counted in equation (1). The corresponding observation equation is given by

$$\frac{\partial l_c}{\partial x}\Delta x + \frac{\partial l_c}{\partial y}\Delta y + \frac{\partial l_c}{\partial z}\Delta z + \frac{\partial l_c}{\partial T_0}\Delta T_0 = (l_0 - l_c) + v \qquad (2)$$

where Δx, Δy, Δz are the least squares corrections to the assumed receiver coordinates, ΔT_0 is a least squares correction to the clock bias, l_0 is observed pseudo-range, l_c is computed pseudo-range, and v is the least squares residual.

Observing four satellites in sequence, the Δx, Δy, Δz and ΔT_0 can be obtained from the least squares solution of four observation equations. In practice, before doing the least squares solution to the observation equations, a sequential least squares procedure was performed to the measurement data and the fitted pseudo-ranges were obtained. The single measurement error for C/A code is on the order of microseconds. By fitting the measurement data for more than ten minutes, the error is reduced to less than 30 ns. The solution accuracy of equation (2) depends strongly on the GDOP. Based on the principle of optimum geometry distribution, four satellite configuration classes for measurement were chosen. The highly accurate UTC(USNO,MC) was used as a receiver clock, so the inconsistency of observation time for each of the four satellites in one class would not cause large measurement errors. In order to reach the stable solution quickly, a successive-substitution method was adopted, i.e. the weighted average value of the post solutions within the time period from several days to one week was used to reset the receiver coordinates. Depending upon the positioning measurements from MJD 46355 to MJD 46522, the mean values, number of data points (N), the standard deviation of residuals (RESI), the standard deviation of the mean (RMS), and position dilution of precision (PDOP) were calculated. (See Table 1.)

Figures 1, 2 and 3 are the plots of three components of antenna coordinate determination vs MJD for classes 95,96,97 and 98, respectively. The contribution of major GPS error sources to PR measurement errors for an L1, C/A code receiver can be estimated. The receiver noise can be modeled as white noise and causes about 0.23 m error after fitting. Including ionospheric and tropospheric delays, the atmospheric delay could cause errors from 1.5 m during solar activity minimum to 9 m during solar activity maximum. The satellite

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

ephemeris error and satellite clock error cause about 6 m error. Thus, the total PR measurement error could be 6.3 m to 10.9 m. For a satellite configuration with a good PDOP of 2.7 (like classes 95 and 98), the point positioning accuracy can be estimated as 16.9 m RSS to 29.3 m RSS depending on solar activity. According to the practical results of the four classes in Table 1, the joint and weighted average of point position accuracy is about 10 m RSS, which is better than the estimated value and could be explained that the observation period was around solar activity minimum and most of the measurements were made during nighttime. It also should be noted that the positioning result is based on about 5 months measurement data. Figures 1, 2 and 3 show good long term stability.

The antenna coordinate results of the four classes were averaged, using weights according to their errors, and compared with the results of Transit Satellite Doppler positioning. (See Table 2.)

Table 2. Position Results of GPS and Transit (in meters)

Coordinate component	x	y	z
GPS	1112154.48	-4842858.92	3985488.18
Transit	1112153.33	-4842855.81	3985482.74
difference	1.15	-3.11	5.44

GPS point positioning demonstrated 6.4 m RSS accuracy relative to the Transit point positioning. It is equivalent to component deviations of about 2 m in latitude, 0.6 m in longitude, and 6 m in height.

Positioning accuracy strongly relies on PDOP. It is important to choose four satellites with a good geometrical distribution. Class 95 and class 98 have the same PDOP, but the position result of class 95 is better than that of class 98. This is because positioning accuracy also relies on the measurement error of each satellite during one class. It is thought that the large measurement error for class 98 comes mainly from the greater time interval for completing measurements of all four satellites. This time interval is about 9 hours for class 98 and 3 hours for class 95. With respect to the cesium clock, the estimation of contribution to the total measurement error due to receiver clock fluctuation during this time interval could be 0.54 m for class 98 and 0.31 m for class 95, respectively. The reason for the obvious systematic drift of the longitude component for class 98 in Fig. 2 is not clear.

Delta-Ranges Measurement

C/A code GPS receivers also can measure Delta-Ranges (DR), the difference in code phase over a fixed time interval. Compared to carrier phase measurement, code phase measurement has a greater measurement noise but no ambiguity. The code phase mode equation is given by

$$\Phi_{Aj}(t_K) = \frac{f_j}{c} R_{Aj}(t_K) + \phi_A(t_K) - \phi_j(t_K) + \phi_{ion}(t_K) + \phi_{trop}(t_K) \tag{3}$$

where A refers to the receiver, J refers to the satellite, R_{Aj} is the distance between receiver and satellite, and f_j is the clock frequency of a satellite. Code phase readings can be differenced and then used as observables. If the readings are differenced with respect to time, a Doppler count is obtained by the equation

$$\Phi_{Aj}(\tau_0,\tau) = \phi_{Aj}(\tau) - \phi_{Aj}(\tau_0) = \frac{f_j}{c}[R_{Aj}(\tau) - R_{Aj}(\tau_0)] + (f_A - f_j)(\tau - \tau_0) + \phi_{ion}(\tau_0,\tau) + \phi_{trop}(\tau_0,\tau) \tag{4}$$

The DR mode equation which corresponds to equation (4) is given by

$$\rho^*_{Aj}(\tau) = R_{Aj}(\tau) - R_{Aj}(\tau_0) = \frac{c}{f_j}\phi_{Aj}(\tau_0,\tau) - c\,T_1(\tau-\tau_0) - \frac{c}{f_j}\phi_{ion}(\tau_0,\tau) - \frac{c}{f_j}\phi_{trop}(\tau_0,\tau) \tag{5}$$

where $T1=(f_A-f_j)/f_j$ is the receiver clock rate relative to the GPS clock. The DR observation equation (6) is similar to the PR equation (2), but with clock rate bias T1 instead of clock bias T0.

$$\frac{\partial\rho}{\partial x}\Delta x + \frac{\partial\rho}{\partial y}\Delta y + \frac{\partial\rho}{\partial z}\Delta z + \frac{\partial\rho}{\partial T_1}\Delta T_1 = (\rho_0 - \rho_c) + v \tag{6}$$

where ρ is Delta-Range $\rho_{Aj}(t)$, ρ_0 is the observed value, and ρ_c is the computed value. Tracking four satellites in sequence, the accumulated DR (ADR) measurements with time intervals of 15 min, 37 min and 77 min were made, respectively. The navigation messages for each of the four satellites need to be recorded. Due to the fact that there are only 7 GPS satellites in orbit at present, and good satellite distribution needs to be retained, it was difficult to increase the tracking time for each satellite.

The computer program for DR measurement is made up of 3 routines: data file preparation, ADR calculation, and the navigation solution. Figure 4 illustrates a sample of the fitted residual curve for 77 min tracking of class 98 (SV12,8,3,6) at MJD 46538. For short-term frequency stability, a rubidium clock is better than a cesium clock. Thus, SV12 and SV3 have a larger residual RMS than SV6 and SV8. Also, since the observing time for SV12 was during the day, there was greater ionospheric delay error. gure 5 shows the calculated DR value which is accumulated along with measurement time. The shape and trend of the ADR curve depends on the satellite clock frequency bias relative to the receiver clock as well as the direction and velocity of the satellite path through the sky during tracking. For comparison, class 98 measurements were performed from MJD 46537 to MJD 46541, with each satellite being tracked for 77 minutes. The position solutions of PR and ADRs were computed with accumulated times of 15 min, 37 min, and 77 min, respectively. The receiver coordinates adopted were the point position results of 5 months measurement described in a previous section.

Table 3. Comparison of Position Results between ADR and PR (unit: meter)

	ADR(15 min)			ADR(37 min)			ADR(77 min)			PR		
	dx	dy	dz	dx	dy	dz	dx	dy	dz	dx	dy	dz
mean	-13.0	1.4	-6.5	-30.6	2.4	-3.2	0.1	-2.7	-3.3	11.8	-3.2	-5.1
RMS	57.4	9.7	11.4	19.1	2.7	2.8	5.8	3.5	1.6	7.2	6.8	6.0
RSS		59.3			19.5			6.6			11.6	
DOP	33.2	5.1	7.9	14.0	2.3	2.9	7.5	1.7	1.2	1.8	1.6	1.9
PDOP		34.5			14.5			7.8			3.0	

The three-dimensional position accuracy of ADR(77 min) was about 6.6 m RSS and the systematic deviations of each antenna coordinate component were all less than 4 m with respect to the reference value. It is far better than that of ADR(15 min), ADR(37 min) and PR.

It can be seen from equation (5) that each part of the ADR measurement errors shows the form of difference corresponding to the accumulated time interval. Due to the fact that some error sources can be modeled as a function of time, the values of the errors on their differences with respect to time would be much smaller than the original values. It can be

shown that the ADR measurement error increases but the PDOP decreases as the accumulated time interval increases. Therefore, one should choose a suitable time interval for obtaining the optimum position accuracy. The main error sources and the adopted models for ADR measurement are as follows:

Receiver Noise + Quantization Noise	0.23 m after fit
Ionospheric Error	GPS correction model (TI=2hr, error< 5ns during solar activity maximum)
Tropospheric Error	$0.8(1-e^{-h/h_0})$m (h_0 =1 hr)
Sat. Ephemeris Error	0.184 T mm
Sat. Clock Error	3 $\sqrt{T}$ mm (atomic clocks)

Based on these models, the estimated total measurement error (Em), PDOP, and RSS position accuracy (Ae), for different accumulated time intervals (TI), are listed in Table 4 and are compared to the practical measurement results of point position accuracy (Am).

Table 4. Estimation of Point Positioning Accuracy by ADR Measurement

TI / solar activity	15 min max.	15 min min.	37 min max.	37 min min.	77 min max.	77 min min.	120 min max.	120 min min.
Em(m)	1.54	0.61	1.62	0.79	1.84	1.18	2.14	1.61
PDOP	34.54	34.54	14.51	14.51	7.80	7.80	5.92	5.92
Ae(m)	53.2	21.0	23.5	11.4	14.4	9.2	12.7	9.5
Am(m)		59.3		19.5		6.6		
Am-Ae(m)		+38.3		+8.1		-2.6		

Due mainly to the adopted polynomial approximation method of ADR calculation, the difference Am-Ae changes along with the increase of TI. For example, taking measurement data of SV12 on MJD 46538, the measurement noise σ and ADR measurement precision $\sigma/\sqrt{n}$ are computed for different TI. Furthermore, taking class 98 and assuming the four satellites have the same ADR measurement precision as SV12, the corresponding contributions Er to the three-dimensional position error were computed. (See Table 5.)

Table 5. Contribution of ADR Error to Position Error

TI	σ (ns)	$\sigma/\sqrt{n}$ (ns)	Er (m)
15 sec	25	---	---
10 min	14.5	2.3	---
15 min	15	1.9	<26.7
37 min	16	1.3	< 7.7
77 min	18.1	1.0	< 3.2

After the first fit of 15 seconds, the measurement noise of the C/A code could be reduced to below 25 ns during solar minimum. After the second fit, σ increases and $\sigma/\sqrt{n}$ decreases along with the increase of TI, but not obviously. After solving the observation equation (6), the elements of the line of sight (LOS) matrix also increase along with the increase of TI. Thus, the position error Er decreases and the measurement values get better than the estimated values.

Timing in Common-View

The code phase reading $\phi_{Aj}(t_k)$ in equation (3) can also be differenced with respect to two receivers at locations A and B at a common local epoch t, thereby obtaining an interferometric single difference phase reading. The model equation is given by

$$\Phi_{ABj}(t) = \phi_{Aj}(t) - \phi_{Bj}(t) = \frac{f_j}{c}[R_{Aj}(t) - R_{Bj}(t)]$$

$$+(f_B - f_A)(t - t_0) + \phi_{Am}^{AB} + \phi_{trop}^{AB} \qquad (7)$$

which can be transformed to the time difference equation

$$t_{ABj}(t) = t_{Aj}(t) - t_{Bj}(t) = \frac{1}{c}[R_{Aj}(t) - R_{Bj}(t)]$$

$$+\left(\frac{f_B - f_j}{f_j} - \frac{f_A - f_j}{f_j}\right)(t - t_0) + t_{Am}^{AB} + t_{trop}^{AB} \qquad (8)$$

where $f_B - f_A$ is the frequency offset between the two receivers at A and B, t_0 is the initial time, t is the observation time, and f_j is the satellite clock frequency. Equation (7) is the interferometric mode of relative positioning and equation (8) is the so-called common-view mode of time transfer. Two GPS receivers with different design, AOA and STI, were placed in the same building with a common time and frequency reference UTC(USNO),MC), and their antenna were seperated by about 10.7 meters; equation (8) becomes

$$t_{ABj}(t) = \frac{1}{c}[R_{Aj}(t) - R_{Bj}(t)] + t_{Am}^{AB} + t_{trop}^{AB} \qquad (9)$$

From MJD 46557 to MJD 46576, 14 common-view paths for each day were made. The results are given in Table 6. A refers to AOA receiver, B refers to STI receiver, t is local observation time, and EL is satellite elevation.

Table 6. Common-View Results from MJD 46557 to 46576

SV_j	average t (hr)	EL(deg)	t_{ABj}(ns)	RMS_j(ns)	N
3	19.95	71	45.38	3.67	16
6	20.70	33	39.33	2.76	18
8	20.97	80	45.94	2.18	18
11	21.67	47	39.13	1.88	15
8	21.93	60	46.61	2.91	18
9	22.47	34	39.33	3.31	18
6	22.73	64	43.39	2.30	18
11	23.40	61	46.87	2.53	15
9	23.67	49	42.89	2.63	18
13	0.73	40	42.56	3.45	18
13	1.83	58	39.78	2.86	18
6	7.63	24	41.35	5.53	17
9	9.07	37	39.53	4.21	17
12	10.33	58	46.00	4.49	16

RMS_j, the standard deviation of the daily measurement $t_{ABj}(t)$, can be considered as the measurement noise estimation for each of the common-view paths. It varied from 1.9 ns to 5.5 ns level. Figure 6 is the plot of the common-view measurement noise vs local observation time. Apart from the contributions of receiver delay instability and multipath distortion instability, it can be clearly seen that there exists the non-cancelling effect of the atmospheric delay error. The average of common-view measurement noise was about 2.68 ns for nighttime and 4.48 ns for daytime during the period MJD 46557 to 46576. The corresponding fractional frequency stability limitation due to common-view measurement noise for our side-by-side comparison was about 3.1×10^{-14} to 5.2×10^{-14} per day, respectively.

It should be mentioned that there exists about a 42.65 ns systematic error, the mean value of $t_{AB_j}(\tau)$ measurements for each of the common-view paths. Common-view systematic errors could be caused by incorrect determination of receiver system bias, the error of the adopted receiver coordinates, and systematic errors of the satellite ephemeris. In general, one can say that the incorrect receiver system bias should cause the same size of systematic error for each of the common-view paths, while errors of receiver coordinates and satellite ephemeris could cause a different systematic error for each of the common-view paths depending on the geometry between satellite and receivers. In our experiment, the $t_{AB_j}(\tau)$ measurements only have a standard deviation of 3.1 ns. Hence, this large systematic error of 42.65 ns is thought to come from the offset of system bias determination between the two receivers.

After substracting 42.65 ns from each measurement of the common-view paths, the obvious relationship between the remaining systematic error and satellite elevation was found. Applying a linear fit to the data points in Figure 6, the intercept (0 deg) -7.12 ns, the slope 0.14 ns/deg, the standard deviation of residual 1.84 ns, and correlation coefficient 0.786 were computed. One also can see that the total systematic error is within 10 ns as satellite elevation varies from 24 deg to 80 deg, which could be mainly from the systematic errors of the adopted coordinates between the two receivers. In our case, the adopted values are the average results of four classes of GPS positioning for the AOA receiver and the results of Transit Satellite Doppler positioning for the STI receiver. From Table 2 it is seen that systematic errors obviously exist between GPS positioning and Transit Doppler positioning, especially in the height component. A large deviation in the height component also is the reason that the common-view systematic errors relate to satellite elevation. Thus, if one requires 10 ns or better timing accuracy, it is important to determine the receiver coordinates accurately in the same positioning system for all stations. If the systematic errors depending on satellite elevation were corrected further, common-view accuracy could approach 1.8 ns in our comparison.

Relative Positioning

Geodetic relative positioning using GPS C/A code receivers has been tested by making observations in the "double difference with time" interferometric mode, i.e., one kind of differential operations. The "double difference with time" code phase observable can be obtained by differencing the single phase difference with respect to time, or the Doppler count with respect to two receivers at locations A and B over a common time interval. The mode equation is given by

$$\begin{aligned}\Phi_{AB_j}(\tau_1,\tau_2) &= \Delta f_{AB}(\tau_2-\tau_1)+\Phi^{AB}_{com}(\tau_1,\tau_2)+\Phi^{AB}_{env}(\tau_1,\tau_2)\\ &\quad+\frac{f_j}{c}\left[R_{Bj}(\tau_2)-R_{Bj}(\tau_1)-R_{Aj}(\tau_2)+R_{Aj}(\tau_1)\right]\\ &= \Phi_{AB_j}(\tau_2)-\Phi_{AB_j}(\tau_1)\\ &= \Phi_{B_j}(\tau_1,\tau_2)-\Phi_{A_j}(\tau_1,\tau_2)\end{aligned} \tag{10}$$

where,

$$\rho_{AB_j}(\tau_1,\tau_2) = R_{Bj}(\tau_2)-R_{Bj}(\tau_1)-R_{Aj}(\tau_2)+R_{Aj}(\tau_1) \tag{11}$$

The observation equation corresponding to equation (11) is given by

$$\begin{aligned}&\frac{\partial\rho}{\partial x_A}\Delta x_A+\frac{\partial\rho}{\partial y_A}\Delta y_A+\frac{\partial\rho}{\partial z_A}\Delta z_A+\frac{\partial\rho}{\partial x_B}\Delta x_B+\frac{\partial\rho}{\partial y_B}\Delta y_B\\ &+\frac{\partial\rho}{\partial z_B}\Delta z_B+\frac{\partial\rho}{\partial f_{AB}}\Delta f_{AB}\\ &=(\rho_o-\rho_c)+v\end{aligned} \tag{12}$$

By holding the coordinates of receiver A fixed and setting the corresponding unknowns to zero, the remaining unknowns in equation (12) are reduced to four and a navigation solution can be done by tracking four satellites.

In "double difference with time" mode, most of the error sources will be essentially canceled. For a local area, relative positioning measurement errors will be the receiver noise and quantization noise only, which can be estimated to approximately 0.32 m. With a PDOP of 5.75, the estimate of RSS relative positioning accuracy is about 1.8 m.

Due to the fact that there exists systematic error between two receivers, and also asymmetry of the signal path from the satellite to each of the two receivers in interferometric (common-view) mode, it is necessary to determine the systematic error for each of the common-view paths. For calibration the DATUM receiver was placed side-by-side with the AOA receiver. There was a common time and frequency reference, the antenna coordinates were well known, and the baseline length between the two antennas was about 5.9 meters.

In this calibration mode, user clock error and user position error will be canceled further. It is useful to determine separately the remaining systematic error for each common-view path which includes the systematic error between the two receiver systems and the systematic error between the different common-view paths. Due to the non-cancellation of satellite ephemeris error and atmospheric error in the case of large baseline common-view, the systematic errors of the "double difference with time" observable, which were determined at the side-by-side location, might change while the portable receiver is moved to a remote site.

The effects of this kind of non-cancellation are discussed as follows:

1. Atmospheric error

The tropospheric error only becomes significant if the satellite elevation angles are very low. In our common-view paths, the satellite elevation angles are: SV12=28.5 deg, SV09=36.5 deg, SV13=67.5 deg, and SV11=28.0 deg. It can be assumed that no significant error is introduced from tropospheric delay.

If we choose a separation between sites of less than about 1500 km in baseline and site latitudes above 30 deg, the ionospheric error in "double difference with time" observable will reduce to about 1-2 ns.

2. Satellite ephemeris error

The relation between "double difference with time" observable and satellite position error is shown in Figure 8. The user and satellite position vectors are $\vec{R}$ and $\vec{S}$, respectively. The user positions at A and B are well known and the range errors due to satellite position error are $\Delta\rho A$ and $\Delta\rho B$, respectively. We have

$$\rho = |S - R|$$
$$\rho' = |S' - R|$$
$$\Delta\rho = \rho' - \rho$$

The single difference observable is

$$\rho_B' - \rho_A' = (\rho_B - \rho_A) + (\Delta\rho_B - \Delta\rho_A)$$

where $(\Delta\rho_B - \Delta\rho_A)$ is determined during calibration. For a remote site B1, the equation is

$$\rho_{B1}' - \rho_A' = (\rho_{B1} - \rho_A) + (\Delta\rho_{B1} - \Delta\rho_A)$$

It can be seen that the change of systematic error is $(\Delta\rho_{B1} - \Delta\rho_B)$. For the "double difference with time" observable, the change of systematic error is

$$(\Delta\rho_{B1} - \Delta\rho_B)t2 - (\Delta\rho_{B1} - \Delta\rho_B)t1$$

The value of the change can be estimated in our measurement for given satellite position errors and relies on the change of geometry between satellite and receiver. Taking SV09 for example, Figure 9 shows typical curves of systematic error changes (in meters) vs the log of baseline length (in kilometers) for different satellite position RSS error s.e.. It might be different for different remote site locations. We can see from the curves that the changes of systematic error are far below 1 m in most cases and for baselines less than 1500 km and 10.4 m of satellite position RSS error, the systematic error change will be less than 0.3 m.

The AOA, DATUM(1) and DATUM(2) receivers are denoted RA, RB and RB', respectively. The systematic errors of the "double difference with time" observable $\rho_{AB_j}(t_1, t_2)$ were determined between different receiver pairs and their possible change quantities were also calculated. (See Table 7.)

Table 7. Systematic error of observable $\rho_{AB_j}(t_1, t_2)$ (unit: meter)

Receiver pair	Baseline length	MJD	common-view path j SV12	SV11	SV13	SV09
RA-RB	A-B=5.9m	46650 to 46657	-1.8134 (±1.0556)	+0.8889 (±0.6636)	-0.7234 (±0.4758)	+1.0430 (±0.8090)
RA-RB'	A-B=5.9m	46742 to 46746	-1.9318 (±1.3100)	+0.5876 (±0.9539)	+0.1464 (±1.2467)	+2.4541 (±1.1113)
change of sys.error	A-B= 247818m	sat.error 10.4m RSS	+0.026	+0.016	-0.081	-0.033

Table 7 shows that the systematic errors of $\rho_{AB_j}(t_1, t_2)$ differ from the different common-view paths and receiver pairs in this calibration mode. After correcting the systematic errors, the determined baseline lengths are compared with the actual baseline lengths and the initial baseline lengths which are caculated from the adopted antenna coordinates of two receivers. (See Table 8.) Even though there exists large errors in the adopted antenna coordinates of the portable receiver, this method still can determine baseline length with mean value 6.65 m (1.52 m r.m.s.) in comparison with the actual length 5.9 m.

Table 8. Calibration of baseline length estimates (actual baseline length = 5.9 m)

MJD	Initial error of antenna coordinates LAT (arc sec)	LON (arc sec)	HEI (m)	Initial baseline length(m)	Determined baseline length(m)	PDOP
46650	-0.150	-0.102	-09.73	10.44	6.09	5.75
46651	-0.006	-0.492	-12.63	17.70	6.51	5.75
46652	-0.006	-0.492	-12.63	17.70	7.34	5.75
46655	-0.234	-0.252	-08.73	10.57	5.40	5.75
46656	-0.012	-0.462	-25.63	28.46	9.33	5.75
46657	-0.228	-0.192	-04.53	6.24	5.22	5.75
mean					6.65	
r.m.s.					1.52	

Results of geodetic relative positioning made from different baselines and receiver pairs are shown in Table 9.

Table 9. GPS baseline estimate

Receiver from	to	Number of pair	Length	r.m.s.	observing sessions	PDOP
BLDG78(USNO)	BLDG16(USNO)	RA-RB'	165.36m	1.47m	5	5.76
BLDG78(USNO)	BLDG16(USNO)	RA-RB	166.30m	1.22m	4	5.76
BLDG78(USNO)	ICB* (NRAO)	RA-RB	247818.95m	2.07m	5	5.90

* ICB: Interferometer Control Building of the National Radio Astronomy Observatory (NRAO), Green Bank, WV

Conclusions:

1. After calibration of the systematic error, different receiver pairs RA-RB' and RA-RB have a credible result of relative positioning, with the baseline length estimates within the error ranges of each other.

2. It is expected and has been discussed above that the longer baseline length has a greater error. In addition, the increase of PDOP from 5.76 to 5.90 also is a factor. It can be overcome by a small adjustment of the satellite tracking schedule once a month.

3. Along with the increase of the number of operational GPS satellites, the satellite configuration could be improved and PDOP could decrease further. It would be possible to determine relative positions to decimeters with PDOP of 2.

4. In the differential mode of operation of a C/A code receiver, a major error source is the measurement noise of the code phase. If the carrier phase could be measured for a C/A code receiver, the relative positioning accuracy would be greatly increased.

Acknowledgment

The authors wish to thank our colleagues for supporting this work. Among those, Mithran Miranian of the USNO Time Service Department who set up the data transfer link between the AOA receiver and HP A900 computer for data collection, Marie R. Lukac of the USNO Nautical Almanac Office who put relative positioning measurement data in machine-readable format, and Len Howell and his colleagues of Interferometer Control, National Radio Astronomy Observatory, who assisted in the experiment when the portable GPS receiver was at NRAO.

References

1. B. W. Parkinson and S. W. Gilbert, "Navstar: Global

Positioning System-Ten Years Later," Proc. of IEEE, Vol.71, No.10, October 1983.

2. Javad Ashjaee, "New Results on The Accuracy of the C/A Code GPS Receiver," Proc. of First International Symposium on Precise Positioning with the GPS, April 1985.

3. Alison K. Brown and Mark A. Sturze, "Static Point Positioning with an L1, C/A Code GPS Receiver," Proc. of First International Symposium on Precise Positioning with the GPS, April 1985.

4. J. A. Klobuchar, "Ionospheric Correction for the Single Frequency Users of the Global Positioning System," IEEE Trans. on National Telesystems Conference, 1982.

5. V. Ashkenazi, L. G. Agrotis and J. Yan, "GPS Interferometric Phase Algorithms," Proc. of First International Symposium on Precise Positioning with the GPS, April 1985.

6. A. J. Van Dierendonck, et al., "The GPS Navigation Message," Journal of The Institute of Navigation, Vol.25, No.2, Summer 1978.

7. A. J. Van Dierendonck and W. C. Melton, "Applications of Time Transfer Using Navstar GPS," Journal of The Institute of Navigation, Vol.30, No.2, Summer 1983.

8. D. W. Allan, et al., "Accuracy of International Time and Frequency Comparison via Global Positioning System Satellites in Common-View," IEEE Trans. Vol.IM-34, No.2, June 1985.

9. T. McCaskill, J. Buisson and A. Buonaguro, "A Sequential Range Navigation Algorithm for a Medium Altitude Navigation Satellite," Journal of The Institute of Navigation, Vol. 23, No. 2, Summer 1976, pp. 164-176.

Table 1. Point Position Results of Pseudo-Range Measurement

CLASS **		LATITUDE (Arcseconds) *			LONGITUDE (Arcseconds)			HEIGHT (METRES)		
		MJD1	MJD2	MJD3	MJD1	MJD2	MJD3	MJD1	MJD2	MJD3
95	MEAN	13.7692	13.7257	13.7598	01.0435	00.9128	00.9920	58.8836	53.0270	56.7260
	N	62	28	89	60	30	90	62	30	91
	RESI	0.1233	0.0764	0.1067	0.0970	0.1514	0.1273	5.9603	7.7472	6.2206
	RMS	0.1280	0.0789	0.1109	0.1277	0.1582	0.1476	6.2979	8.5584	7.2442
	DOP	1.7	1.7		0.9	0.9		1.9	1.9	
96	MEAN	13.9148	13.8167	13.8653	01.1026	01.0577	01.0869	62.3679	50.4125	58.2295
	N	67	36	101	67	36	103	68	36	104
	RESI	0.2209	0.2700	0.2248	0.1347	0.1356	0.1390	9.0040	8.1580	9.5380
	RMS	0.2322	0.2962	0.2364	0.1380	0.1412	0.1401	9.2120	8.6824	10.6527
	DOP	2.8	2.8		1.0	1.3		2.5	2.9	
97	MEAN	13.7615	13.8075	13.7903	01.2338	01.0507	01.1431	58.1760	61.9003	59.4344
	N	51	37	90	51	37	88	52	36	89
	RESI	0.0835	0.1773	0.1457	0.1773	0.2068	0.1888	5.5734	7.5785	7.0563
	RMS	0.0963	0.1812	0.1516	0.1987	0.2166	0.2095	5.7531	8.2869	7.4904
	DOP	1.6	1.7		1.4	1.5		1.6	1.8	
98	MEAN	13.6354	13.7149	13.6640	01.2093	00.8063	01.0363	54.7280	55.0998	54.8928
	N	56	46	104	59	46	104	59	47	106
	RESI	0.0852	0.1108	0.1141	0.1887	0.1888	0.2088	7.6099	6.9993	7.4384
	RMS	0.0886	0.1326	0.1271	0.2227	0.2639	0.3127	7.6424	7.2857	7.4533
	DOP	1.6	1.6		1.4	1.4		1.6	1.6	
Weighted Average	MEAN	13.7526	13.7599	13.7578	01.1320	00.9518	01.0550	58.1274	55.6217	57.1536
	RMS	0.1247	0.1583	0.1474	0.1641	0.1893	0.1946	7.0632	8.1180	7.9474

* MJD1=46355 - 46443 MJD2=46452 - 46522 MJD3=46355 - 46522

** CLASS	MJD1		MJD2	
	SV	PDOP	SV	PDOP
95	12,9,11,13	2.7	12,9,11,13	2.7
96	9,6,13,8	3.9	11,13,8,12	4.2
97	6,11,8,3	2.7	11,8,3,9	2.9
98	12,8,3,6	2.7	12,8,3,6	2.7

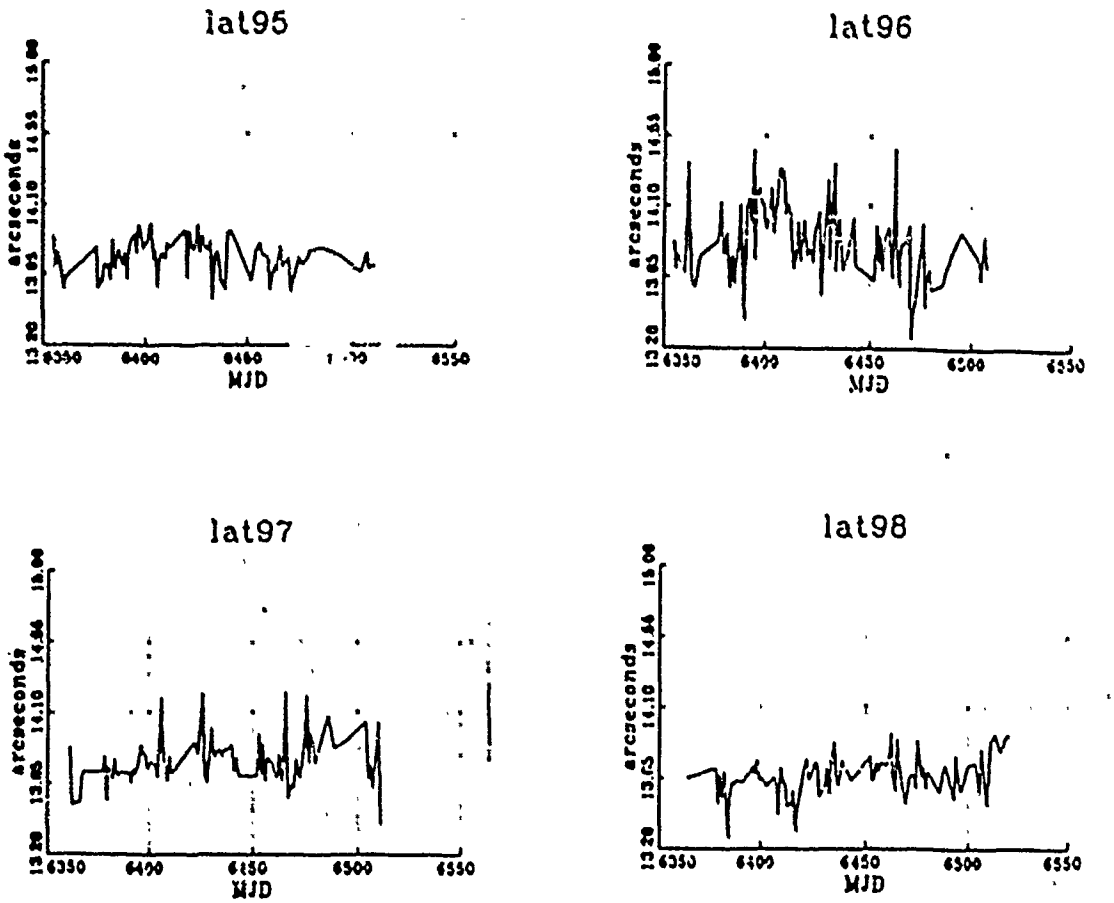


Figure 1. Plots of Antenna Latitude for Different Satellte Confígurations of Class 95,96,97 and 98

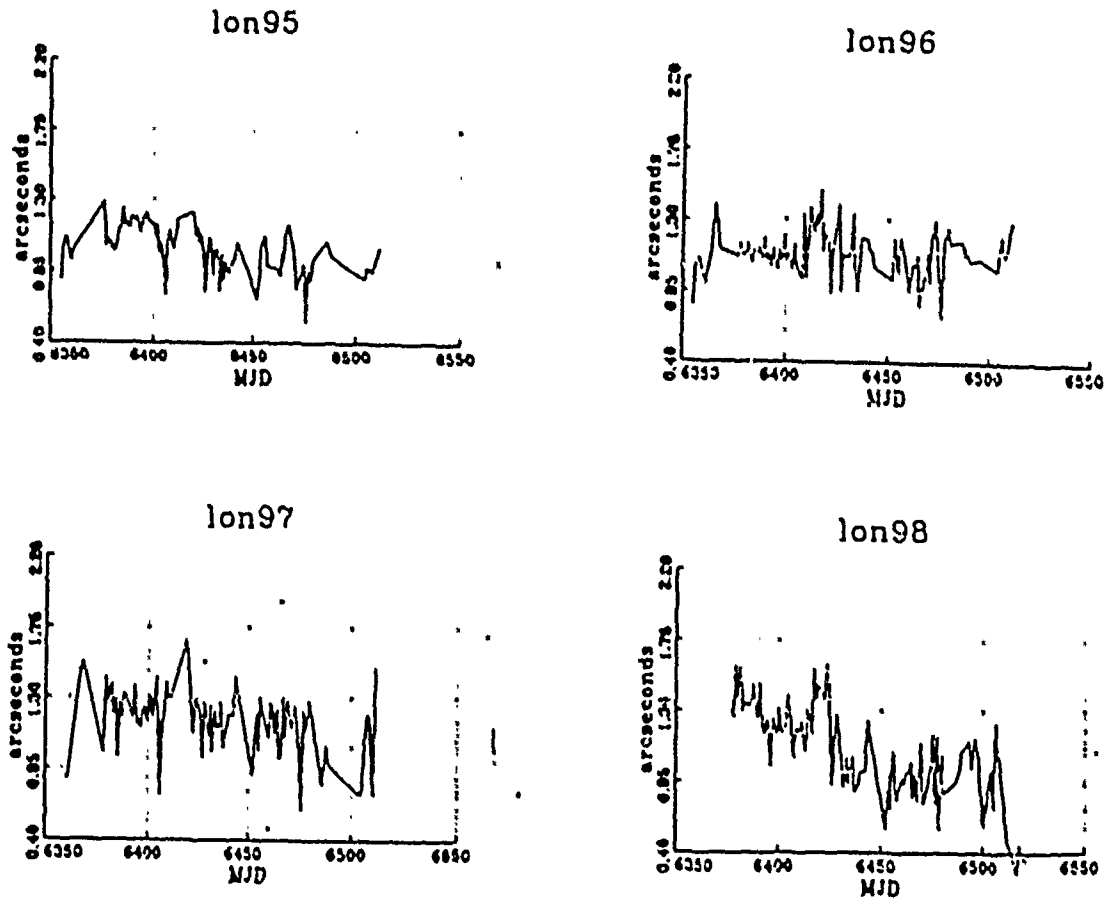


Figure 2. Plots of Antenna Longitude for Different Satellite Configurations of Class 95,96,97 and 98

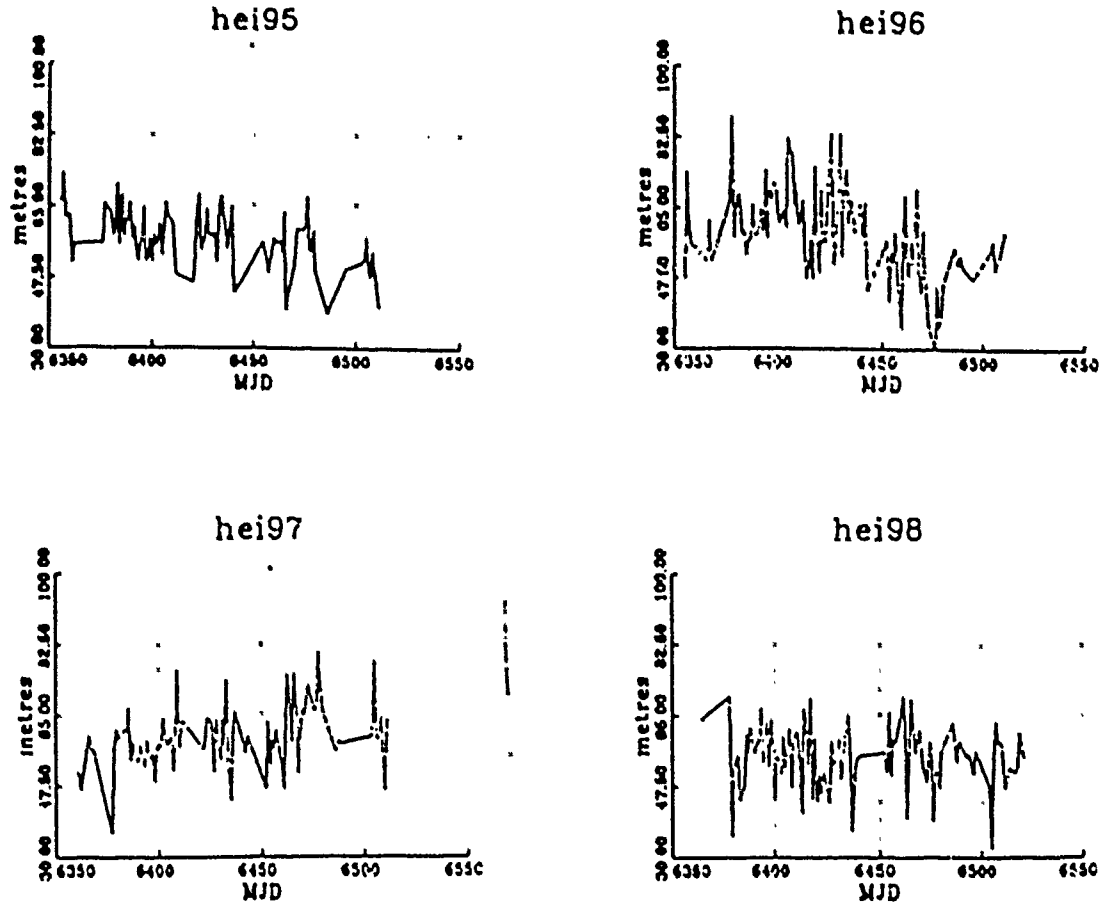


Figure 3. Plots of Antenna Height for Different Satellte Configurations of Class 95,96,97 and 98

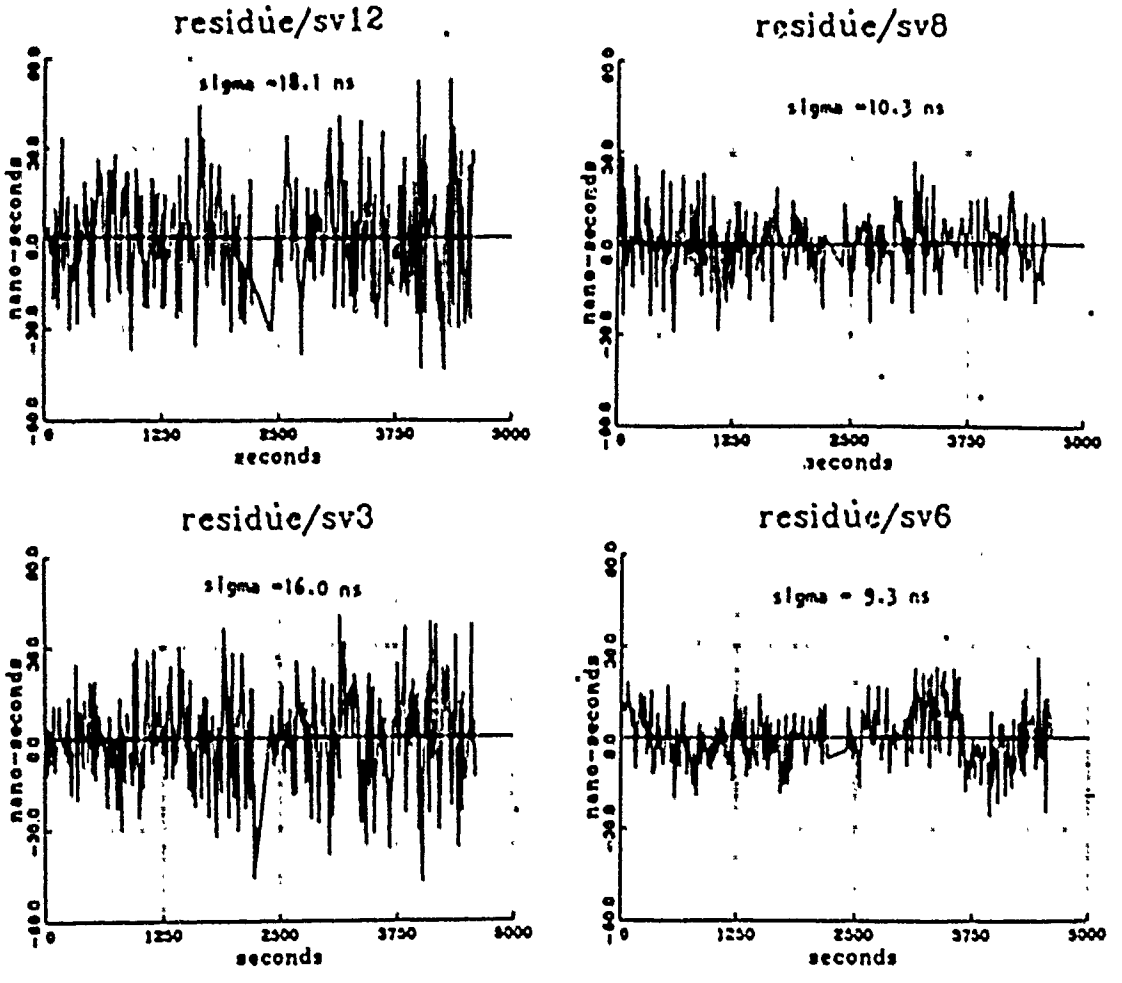


Figure 4. Fitted Residual Curves of 77 min Measurement for SV 12, 8, 3, 6 on MJD46538

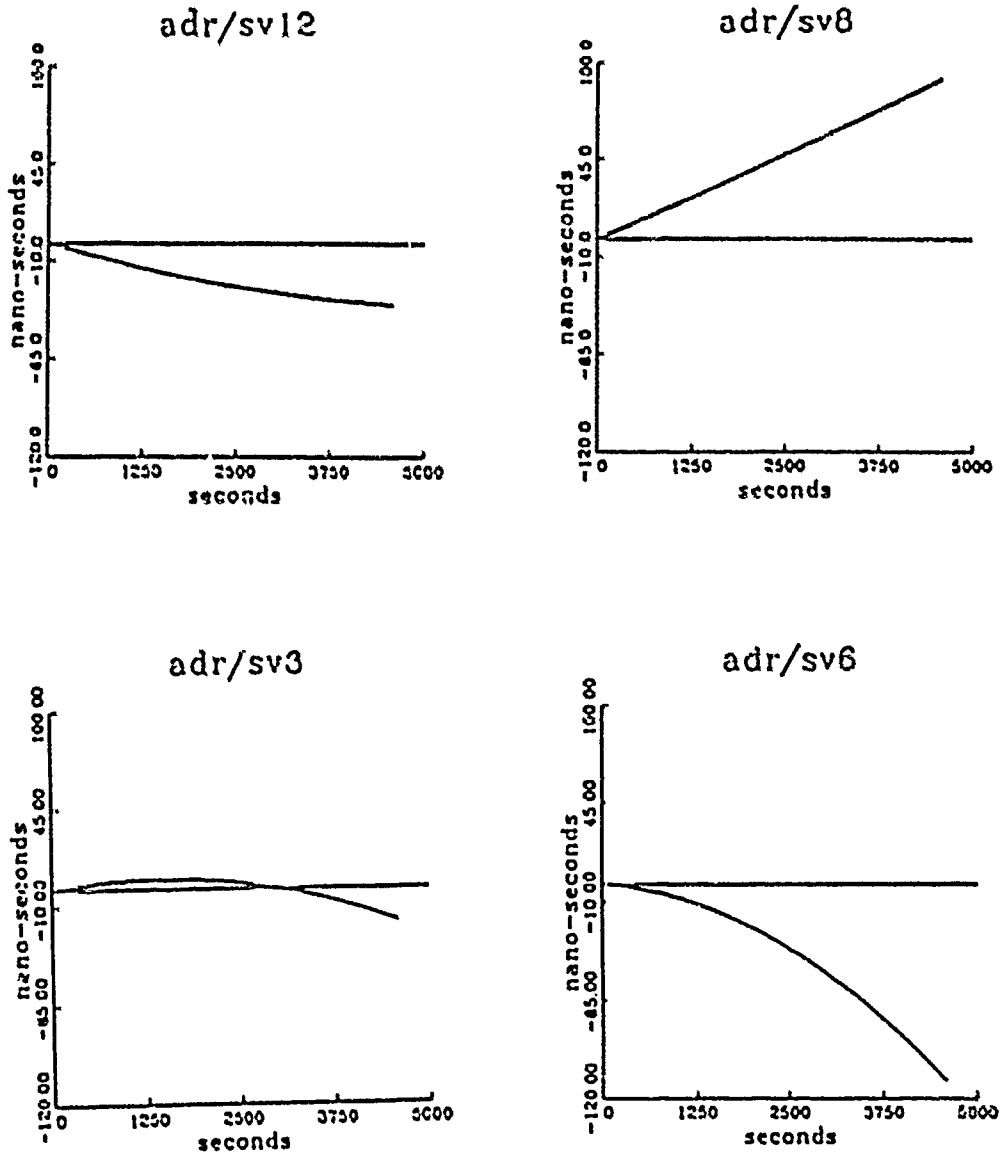


Figure 5. Polynomial Fits to Accumulated Delta-Range Measurement for SV 12, 8, 3, 6 on MJD 46538

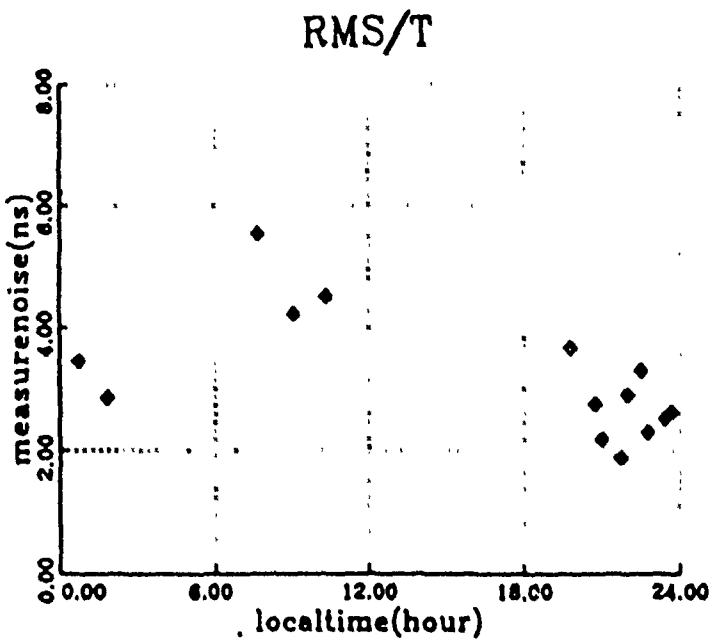


Figure 6. Common-view measurement Noise vers Local Observation Time for Two Side-by-Side GPS Receivers at USNO from MJD 46557 to 46576

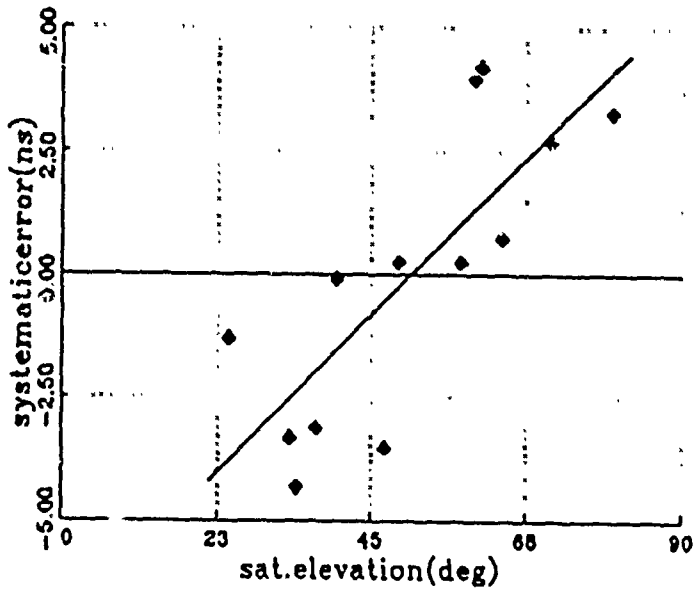
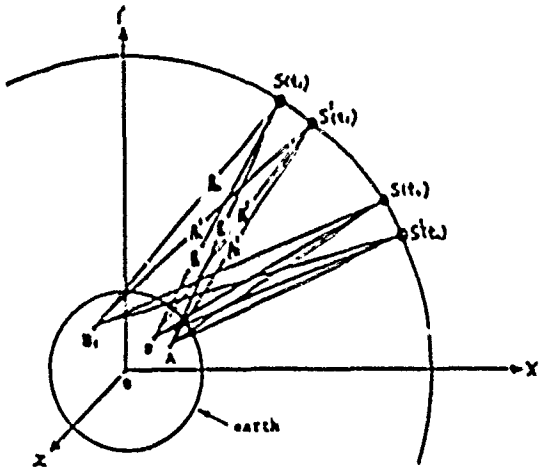


Figure 7. Common-View Systematic Error vers Satellte Elevation for Two Side-by-Side GPS Receivers at USNO from MJD



satellite position S — satellite position with error S'

reference GPS receiver position A — remote GPS receiver position B1
side by side GPS receiver position B
range ρ — t2 - t1 = 1 hour

Figure 8. Relation between 'double difference with time' observable and satellite position error

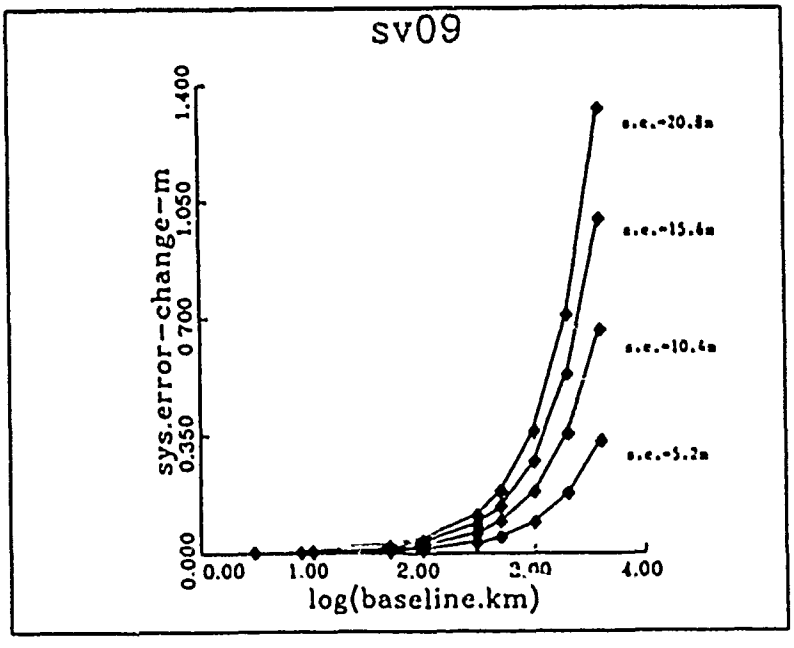


Figure 9. Systematic Error Change VS Baseline Length

41st Annual Frequency Control Symposium - 1987

ESTIMATING COMBINED ERRORS DUE TO PROPAGATION AND EPHEMERIS AND THEIR EFFECT ON TIME AND FREQUENCY TRANSFER

by

David W. Allan* and Lin Ping-ping†

*National Bureau of Standards
Boulder, Colorado 80303

and

†Beijing Institute of Radio Measurement
Beijing, China

ABSTRACT

We now have a global network of timing centers with frequency standards having stabilities of a few parts in 10^{14} which are monitoring the GPS. It has been shown that by taking differences of the common-view time differences between two timing centers and between pairs of satellites, one can arrive at a statistically optimum estimate for a weighting factor for each common-view path. With this approach, GPS common-view measurement noise of a few parts in 10^{14} is achievable for an integration time of 1 day.

Using the above weighting factors, this paper developes an algorithm for estimating a weighted linear error of the differential ephemeris plus propagation errors for each satellite. This can be done between any pair of timing centers which have receivers and clocks with adequate stability. Since most of the time transfer receivers operate at the L1 frequency (1.575 GHz), this technique reveals information regarding the accuracy of the ionospheric models broadcast at this frequency as part of the GPS data word.

Once the individual satellite's differential propagation and ephemeris errors are estimated, the statistical properties of each can be combined to obtain a statistically weighted estimate of the common-view measurement variations limiting the comparison of the clocks between the two remote sites. Optimum statistical weighting yields a significantly better measurement noise than can be obtained from a simple average. For example, between NRC in Ottawa, Canada and NBS in Boulder, Colorado demonstrated stabilities of $\bar{\sigma}_y(\tau) = 1 \times 10^{-14}\ \tau^{-3/2}$, where τ is in day's, have been achieved. This is equivalent to a white time modulation noise of less than 1 ns sampled once per day. On the other hand, without proper care in the data processing errors can accumulate to several tens of nanoseconds in common-view time comparisons.

INTRODUCTION

The goal of this paper is to investigate the limitations caused by the propagation and ephemeris errors associated with time and frequency comparisons between remote clocks via GPS satellites in common view. The common-view technique capitalizes on cancellation of some errors as a given satellite is viewed at the highest angle -- limited by other scheduling constraints -- between the two remote locations.[1,2] Typically L1 (1.575 GHz) frequency receivers are employed. If the tracking schedules for comparing the clocks at the remote sites are identical, then the GPS clock error cancels perfectly. Because the vectors are not far from parallel, a significant amount of the ephemeris error is cancelled. The correlation distances for the ionosphere extend only to about 1000 kilometers[3]; hence, for global time comparisons, cancellation of ionospheric errors seems not to be significant except as they occur through the modeling. The broadcast model for the ionospheric delay is used in the common-view calculations, and it appears that these modeling errors are the largest contributors to the inaccuracy of time and frequency transfer via this technique.

In the context of this paper there are four concepts that need to be defined: time stability, time accuracy, frequency stability, and frequency accuracy. Specifically, in this paper we address only stability and accuracy of the GPS common-view measurement technique and not that of the remote clocks. Our goal again is to see how well we can compare the time and frequency of the remote clocks using this technique as limited by the propagation and ephemeris errors. Time stability is a measure of the change in the measurement system time delay from one time to a time τ later. Time accuracy can be conceptualized in terms of a perfect portable clock which is used to measure the absolute time difference between the two remote clocks. Frequency stability is usually specified in terms of $\sigma_y(\tau)$ and/or modified $\sigma_y(\tau)$, (denoted $\bar{\sigma}_y(\tau)$).[4-6] Frequency accuracy of the measurement system is a measure of how well one can determine the absolute frequency difference between the two remote standards.

METHOD OF ANALYSIS

References [2&7] showed that over a given day's tracking schedule of GPS satellites, two remote clocks having a set of common-view values can be compared with a day-to-day time stability of a few nanoseconds. On a given day if the difference of the time difference between these two remote clocks is calculated from data obtained through two different GPS satellite vehicles (SV), then the times of these remote clocks drop out of the equations and we are left only with the difference in the common-view propagation and ephemeris errors between these two

Work of the U.S. Government; not subject to U.S. copyright.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

tracks. For most of the international timing centers the effect of the remote clocks on this assumption amounts to only a few nanoseconds.

We will denote the two remote clock sites as A and B. At sites A and B we measure, respectively, time differences at a time t_i of clock A and B minus GPS:

$$x_{AG}(t_i) = x_A(t_i) - x_G(t_i) - x_{D_A}(t_i) \qquad (1)$$

and

$$x_{BG}(t_i) = x_B(t_i) - x_G(t_i) - x_{D_B}(t_i), \qquad (2)$$

where x_A, x_B, and x_G are the true time deviations for clocks A, B, and GPS via the i^{th} SV respectively, and the x_Ds are the errors at A and B between these measures and the truth. Subtracting (2) from (1) gives us the common-view estimate equation for the time difference between clocks A and B:

$$x_{AB}(t_i) = x_A(t_i) - x_B(t_i) - x_{D_{AB_i}}(t_i), \qquad (3)$$

where $x_{D_{AB_i}}(t)$ is now the differential delay error via SV_i. It has been shown that the differential delay errors are significantly smaller than either of the error terms in equations (1) and (2), hence, the value of common view for time and frequency comparisons between remote clocks [2].

We have a similar measurement at a time t_j of the common-view estimate of the time difference $x_{AB}(t_j)$ via SV_j:

$$x_{AB}(t_j) = x_A(t_j) - x_B(t_j) - x_{D_{AB_j}}(t_j). \qquad (4)$$

Subtracting (4) from (3) we obtain

$$x_{AB}(t_{ij}) = x_A(t_i) - x_A(t_j) + x_B(t_i) - x_B(t_j) - x_{D_{AB_{ij}}}(t_{ij}). \qquad (5)$$

If $t_i = t_j$ then, of course, the first four terms on the right of (5) cancel in pairs. In a typical pass of the GPS constellation the maximum value for $|t_i - t_j|$ is about 6 hours or less. For typical high-performance cesium-beam frequency standards employed at international timing centers the rms contribution of the first four terms is about 2 ns or less for a set of such passes taken as a time series.

If clocks A and B have a frequency difference, this will cause a bias in the value given by (5), which will have no effect on the statistical analysis to follow, but can have an effect on the linear estimate. If the frequency difference changes outside of the normal noise of the clocks, then that change will have an effect on the statistics.

Given that each path i and j is nominally independent of the others, and given the above conditions, let us simplify (5) to

$$x_{AB}(t_{ij}) \approx x_{ji}(\bar{t}_{ij}), \qquad (6)$$

where $\bar{t}_{ij}$ is the average of the track times t_i and t_j. As shown in reference [4] we can do N-corner statistics on (6) to determine the optimum weight, w_i, for each common-view track.

Let us next assume that the variance of the deviations of $x_{D_{AB_i}}(t)$ is proportional to the linear deviation. If this is true then we can write as an estimate of the jth linear deviation,

$$\dot{x}_{jo} = \sum_{i=1}^{n} w_i (x_{ji} - x_{io}), \qquad (7)$$

where there are n values from the SVs on a given day and x_{IO} is the true deviation of the differential-delay common-view error. The above assumption yields

$$\sum_{i=1}^{n} w_i x_{io} = 0; \qquad (8)$$

hence, even though x_{IO} is not known, if our assumptions are valid, we can estimate the linear deviation as in (7).

Multipath and coordinate errors in the GPS receivers at A and B can also bias the value calculated in (5). Since these are nominally constant they will not affect the variances--only the linear estimates.

Since the largest error in the linear deviation for common-view time and frequency transfer is believed to be in the ionospheric modelling on the GPS L1 frequency, we have performed a global analysis of the estimate given by equation (7). The stations used in the analysis were picked from around the world. These stations include the following:

TABLE 1

	TIME CENTER	WEST LONGITUDE	LATITUDE
CSIRO	Commonwealth Scientific and Industrial Research Organization Australia	204°.8	33°.8S
NBS	National Bureau of Standards Boulder, Colorado	105°.3	40°.0N
NRC	National Research Council Ottawa, Canada	75°.9	45° 4N
PTB	Physikalisch-Technische Bundesanstalt Braunschweig, West Germany	307°.7	52° 3N
RRL	Radio Research Laboratory Tokyo, Japa	220°.5	35°.7N
USNO	US Naval Observatory Washington, D.C.	77°.1	38° 9N

Figure 1 is a plot of the day-by-day estimate of the combined propagation and ephemeris common-view errors between USNO and NBS via SV 12 (NAVSTAR 10) given by equation (7). The Fourier frequency spectrum is characterized by white-noise phase modulation (PM); hence, filtering is appropriate in order to see the low-frequency characteristics of the data. A 30-day exponential filter was chosen through which to plot the day-by-day estimates given by equation 7. Figure 2 shows a plot of data obtained through both NAVSTAR 10 and the other vehicles available between Washington, D.C. and Boulder, Colorado using the USNO

and NBS L1 receivers. If the residuals are white noise PM, the measurement noise is given by [6]

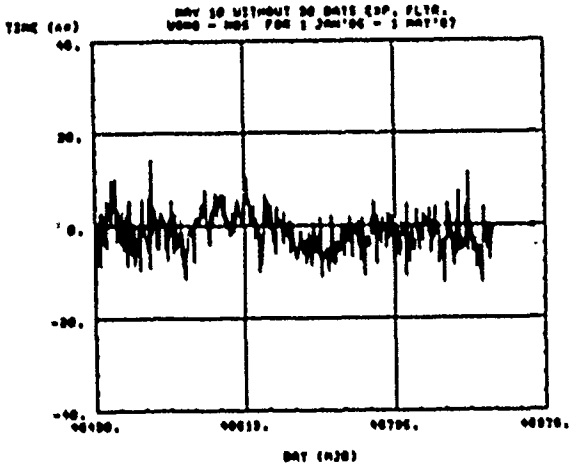


FIGURE 1 A plot of the daily estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Boulder, Colorado and Washington D.C. via NAVSTAR 10.

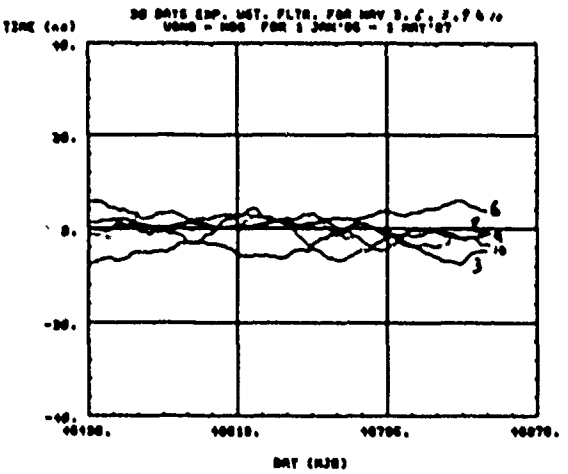


FIGURE 2 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Boulder, Colorado and Washington D.C. via the different NAVSTAR satellites indicated.

$$x_{rms}(\tau_0) = \frac{1}{\sqrt{3}}\,\tau\,\sigma_y(\tau) = \frac{\tau^{3/2}}{(3\,\tau_0)^{1/2}}\,\bar{\sigma}_y(\tau)\,, \qquad (9)$$

where τ_0 is the data spacing i.e., 1 sidereal day. The composite measurement noise is given by

$$x_{rms}(\tau_0) \quad \left[\sum_{i=1}^{n}\frac{1}{\sigma_i^2}\right]^{-1/2}, \qquad (10)$$

where the σ_i^2 come from the N-cornered-hat statistical analysis using (6). The average standard deviation of the residuals across the different SVs is listed in Table 2. Comparing it to the composite measurement noise, which is also listed, one obtains a feel for the benefit of performing a combined optimal weighted estimate to obtain improved time stability. This factor ranges between 3 and 14 for the data in this paper. Equation (9) gives the relationship between the time stability and the frequency stability. The time accuracy is probably more closely related to the standard deviation of the residuals as listed in Table 2. The frequency accuracy of the measurement, on the other hand, will be given by the magnitude of the integration time chosen in measuring the frequency difference between clocks A and B. The value of $\bar{\sigma}_y(\tau)$, once a weighted set of the common-view tracks is taken, will be an estimate of that accuracy. This measurement accuracy has been demonstrated to be significantly better than the accuracy of the best primary frequency standards in the world.

Figure 3 shows the smoothed deviations of the residuals for the path from Australia to Tokyo. We

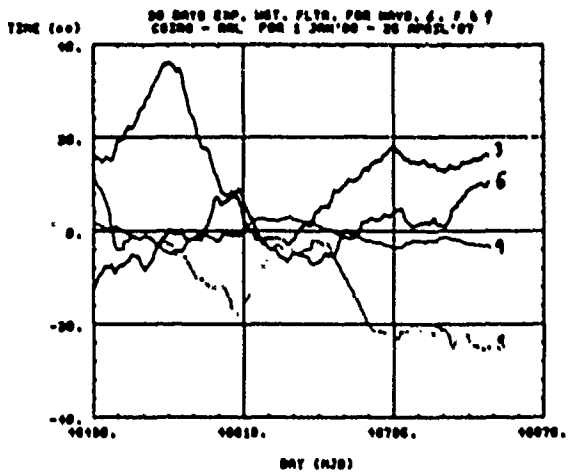


FIGURE 3 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Australia and Japan via the different NAVSTAR satellites indicated.

see a lot more low-frequency variations and some evidence for systematic errors in these data. The next common-view path shown in Figure 4 is from Tokyo to Boulder, Colorado between RRL and NBS. Notice an apparent annual term for the residuals with NAVSTAR 10, as well as with NAVSTAR 6. Figure 5 shows data for the common-view path across Asia from RRL to PTB. Figure 6 shows comparisons between PTB and NBS on a

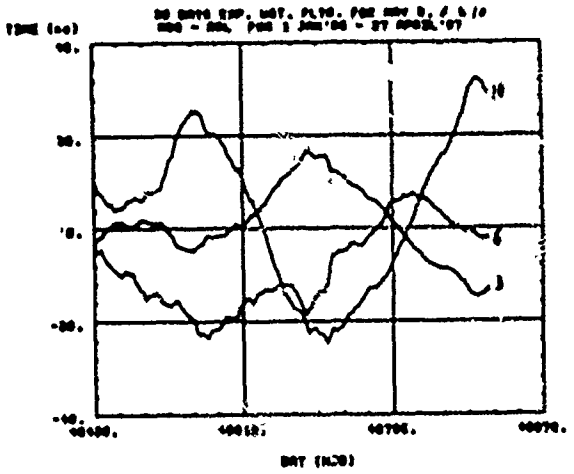


FIGURE 4 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Colorado and Japan via the different NAVSTAR satellites indicated.

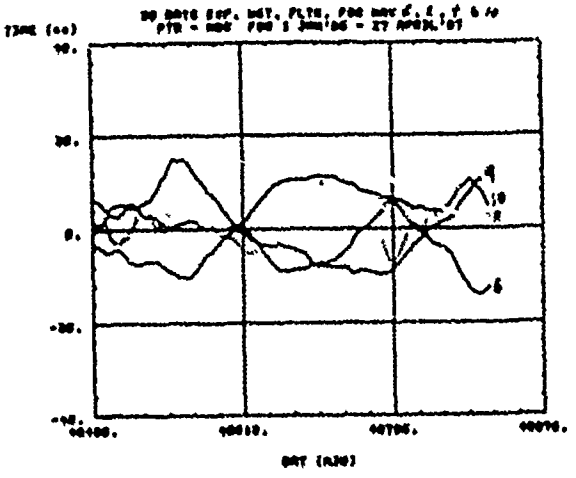


FIGURE 6 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between West Germany and Colorado via the different NAVSTAR satellites indicated.

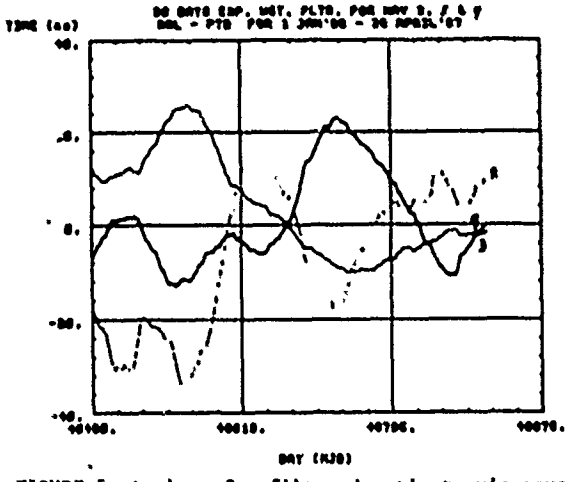


FIGURE 5 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between West Germany and Japan via the different NAVSTAR satellites indicated.

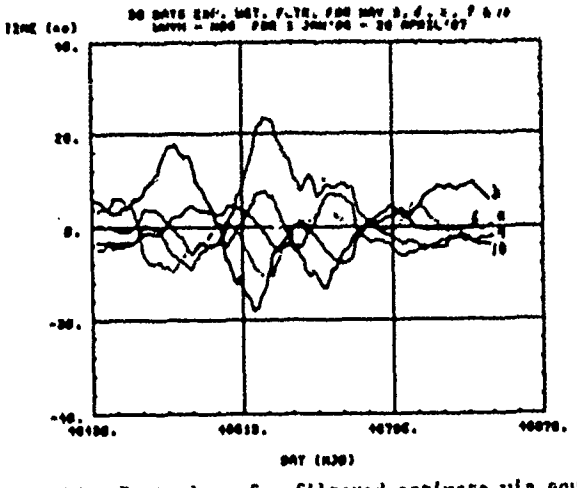


FIGURE 7 A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Hawaii and Colorado via the different NAVSTAR satellites indicated.

path across the Atlantic plus the continental US. These data also show an apparent annual variation with NAVSTAR 6 and 10 of about 20 ns.

The path between Hawaii and Boulder was chosen because of the proximity of the clock at WWVH to the equator -- placing greater demands on the ionospheric modeling. The peak-to-peak scatter shown in Figure 7 seems to be a bit larger and the variability seems to be higher. No annual term is evident. In contrast the comparison across the continental US between Ottawa, Canada and Boulder, Colorado involves a very high-latitude station. These data, shown in Figure 8, are quite different and show very small variability and an indication of significant biases which could be due to multipath or coordinate problems at one or both of the sites.

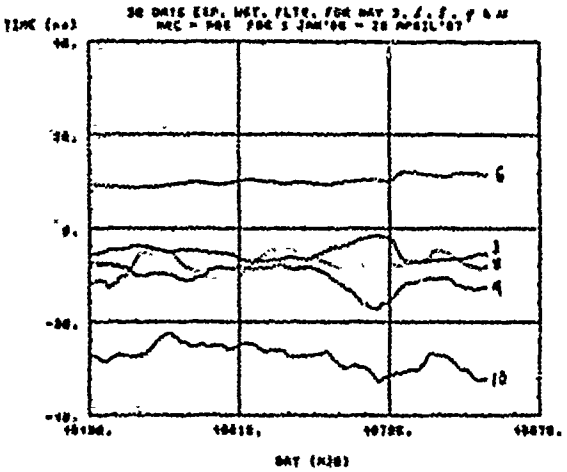


FIGURE 8- A plot of a filtered estimate via equation 7 of the differential ephemeris plus propagation errors for GPS common-view measurements between Ottawa, Canada and Boulder, Colorado via the different NAVSTAR satellites indicated.

Conclusion

Table 2 lists the paths i.e., the end locations of clocks A and B, the standard deviation of the filtered data over the last 100 days and the weighted measurement noise for each of the common-view paths studied. Global time comparison accuracies of about 20 nanoseconds or less are available from a weighted set of GPS satellites used in common view. Time stabilities are typically only a few nanoseconds for global time comparisons. Frequency stabilities may be characterized by the square root of the modified Allan variance, $\sigma_y(\tau)$, equal to a few parts in 10^{14} times $\tau^{-3/2}$ (τ in days) for integration times from one day to a couple of weeks. The full accuracy of state-of-the-art primary frequency standards is available at a remote site through a weighted average of the GPS satellites used in common view i.e., values less than 1 part in 10^{14} are achievable for integration times of a few days.

TABLE 2

PATH	AVERAGE STANDARD DEVIATION (ns)	DAY-TO-DAY MEASUREMENT NOISE (ns)
USNO - NBS	4.2	1.4
CSIRO - RRL	18.6	3.4
NBS - RRL	20.5	5.2
RRL - PTB	12.1	4.8
PTB - NBS	8.5	2.0
WWVH - NBS	5.9	2.3
NRC - NBS	19.0	1.4

REFERENCES

1. Allan, D.W., Weiss, M.A., "Accurate Time and Frequency Transfer During Common-View of a GPS Satellite", Proc. 34th Ann. Freq. Control Symp., USAERADCOM, Ft. Monmouth, N.J. 07703, May 1980.

2. D.W. Allan, D.D. Davis, M.A. Weiss, A. Clements, Bernard Guinot, Michel Granveaud, K. Dorenwendt, B. Fischer, P. Hetzel, Shinko Aoki, Masa-Katsu Fujimoto, L. Charron, and N. Ashby, "Accuracy of International Time and Frequency Comparisons Via Global Positioning System Satellites in Common-View", IEEE Transactions on Instrumentation and Measurement, Vol. IM-34, No. 2, June 1985.

3. Klobuchar, J.A. (1978), "Ionospheric Effects on Satellite Navigation and Air Traffic Control Systems," NATO AGARD proceedings, Lecture Series No. 93, Recent Advances in Radio and Optical Propagation for Modern Communication, Navigation, and Detection Systems.

4. Barnes, J.A., Chi, A.R., Cutler, L.S., Healey, D.J., Leeson B., McGunigal, T.E., Mullen, Jr., J.A., Smith, W.L., Sydnor L., Vessot, R.F.C., Winkler, G.M., "Characterization of Frequency Stability," Proc. IEEE Tran. on Instrumentation and Measurement", Vol. IM-20, 105, NBS Technical Note 394, 1971.

5. Lesage, P., Ayi, T., "Characterization of Frequency Stability: Analysis of the Modified Allan Variance and Properties of Its Estimate," IEEE Transactions on Instrumentation and Measurement", Vol. IM-33, No. 4, December 1984, pp. 332-336.

6. Allan, D.W., "Time and Frequency Characterization, Estimation Prediction of Precision Clocks and Oscillators," IEEE, Ultrasonics Ferroelectrics and Frequency Control, Special Issue on Frequency Control, Nov. 1987.

7. Weiss, M.A., and Allan, D.W., "An NBS Calibration Procedure for Providing Time and Frequency at a Remote Site by Weighting and Smoothing of GPS Common View Data," Proceedings of the CPEM Conference, Gaithersburg, MD, June 1986.

Ku-BAND SATELLITE TWO-WAY TIMING USING A VERY SMALL APERTURE TERMINAL (VSAT)*

David A. Howe
Time and Frequency Division
National Bureau of Standards
Boulder, CO 80303

Summary

The NBS Boulder Laboratory has recently completed installation of a 6.1 meter Ku-band (14/12 GHz) satellite earth station and has acquired two 1.8 meter portable earth terminals and commercial spread-spectrum modems. This equipment was procured for the purpose of doing two-way time transfer experiments in collaboration with other timekeeping laboratories. Each portable earth terminal, often called VSAT for "very small aperture terminal" when the dish diameter is 1.8 meters or less, is a complete Ku-band earth station. The VSAT can provide full duplex capability to transmit to and receive from a geostationary satellite of the Fixed Satellite Service (FSS). Thus, two-way time comparisons can be done directly between NBS and a portable VSAT through a Ku-band satellite. Phase measurements have been performed of earth station and VSAT facilities in various loop-around schemes using a satellite simulator. Frequency stability $\sigma_y(\tau)$ at various carrier-to-noise density (C/N_0) ratios will be reported. Stability plots have shown that performance is $4 \times 10^{-10}\tau^{-1}$ for a C/N_0 ratio of 65 dB-Hz and that this white noise behavior continues to a few days where σ_y is 3×10^{-15}. Absolute phase delay measurements show reproducibility to better than 1 ns over a 16 day sample time. This suggests the potential for accuracy to this level given an appropriate calibration. Preliminary results of loop tests through satellite will be reported.

This paper describes a measurement method which determines the principle time delay constant affecting the accuracy of time synchronization using the two-way technique. The method does not rely on traditional injection and timing of an RF pulse but is an in situ determination having fewer measurement uncertainties.

Introduction

Continuing growth in satellite telecommunications opens up new opportunities for doing accurate transfers of time between suitably equipped laboratories. Wideband Ku-band satellite transponders in the 14/12 GHz band permit the use of small, low-cost earth stations with minimal terrestrial interference and relatively straightforward licensing. The short wavelength of the 12 GHz downlink frequency along with the high power of the satellite transponders yields adequate signal levels even when using only a small 1.8m dish on location. These small earth stations are portable and can be conveniently placed for connection to time standard outputs. When the dish diameter is 1.8m or less, the earth station is referred to as a "very small aperture terminal" or VSAT. The VSAT can provide full duplex capability to transmit to and receive from a geostationary satellite of the fixed satellite service, or FSS.

Recent time transfer experiments using the two-way technique through geostationary satellites have been demonstrated to a level of several nanoseconds [1,2]. The development of a modem specifically designed for this task creates additional opportunity for doing research on the two-way time transfer technique via satellites of the Fixed Satellite Service (FSS) by simplifying the procedure [3]. The modem uses phase-shift keying modulation in roughly 4 MHz of bandwidth and a pseudo-random noise (PN) sequence as the signal modulation. Therefore, the signal is indistinguishable from the normal communications link noise. There is, of course, an increase in the noise density over the 4 MHz which the signal occupies, but as discussed later this increase can be small enough that no interference is caused to other transponder users.

This paper discusses tests conducted using a Ku-band portable VSAT having a dish diameter of 1.8m with the spread-spectrum modem. This equipment is part of a system directed toward two-way time transfer experiments with various timekeeping laboratories. In addition, tests have been conducted with a satellite earth station facility having a 6.1m dish at the National Bureau of Standards, Boulder Laboratories. Finally, actual satellite loop tests were performed using the VSAT and the NBS earth station.

Facilities

The central component of the present two-way satellite work is a Ku-band earth station having a 6.1m dish located on top of the radio building at NBS, Boulder. It is referred to as the "hub" in the sense that other involved earth stations (VSAT's) have less capability and the use of the hub during experiments helps establish initial satellite connections. A picture of the antenna is shown in the background of Figure 1 and a picture of the earth station equipment which is located about 30 meters away is shown in Figure 2. Figure 3 is a block diagram of the principal parts of the earth station. Pointing of the antenna is by motorized positioners controlled by a microprocessor-based driver. The 6.1m dish has a receive gain of 56 dB and has a low noise amplifier (LNA) with a 2.5 dB noise figure. The transmitted output power is +26 dBm maximum (approximately 0.4 watts) measured near the feed. A complete description of the hub is given in reference 4.

The VSAT consists of a complete earth station RF package with 70 MHz input and output intermediate frequency and attached to a 1.8m dish on a free-

*This work has been partially supported by The Rome Air Development Center under contract F30602-85-0055. Contribution of the National Bureau of Standards, not subject to copyright.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

standing pedestal. It is shown in the foreground of Figure 1. The dish uses a J-hook prime-focus-feed system. The mount is a simple elevation-over-azimuth assembly using galvanized steel braces arranged in a triangular geometry designed for resting directly on the ground or, in this case, on the roof. Pointing can be done by one person by physically moving the whole terminal to position and locating the satellite by observing the received signal. Since several satellites might be picked up, a unique identifying signal such as a video (television) transmission is very helpful during this pointing exercise. The 6.1m antenna of the hub can automatically point to a given satellite so the signals from it can be used as a reference. This operation is described later.

In-Cabinet and Free-Space Loop Tests

In order to test the various subsystems comprising the ground equipment, several loop-around schemes were employed, which can be divided into two basic categories: (1) in-cabinet tests involving the use of a satellite simulator connected to the transmit and receive (14 and 12 GHz) waveguide ports of the VSAT or earth station, and (2) free-space tests using the satellite simulator connected to horn antennas which are located away from and directed at the VSAT with its dish fully operational. Figure 3 shows the placement of the simulator for in-cabinet tests for the hub earth station. The transmit signal from the 1 watt power amplifier is switched to a power meter with a sample of the signal routed to the input of the simulator. The simulator shifts the frequency by 2.3 GHz and this output is fed to the LNA (low noise amplifier). Figure 4 is a block diagram of the in-cabinet test using the VSAT. The VSAT RF transceiver contains upconverter, power amplifier, LNA, and downconverter all in one package. Connection of the simulator is straightforward. Figure 5 is a block diagram of the free-space loop test used with the VSAT. In this test, the VSAT 1.8m dish is deployed and horn antennas are used with the simulator along with a LNA for gain ahead of the simulator.

With the modem sending and receiving PN spread-spectrum modulation in loop tests, one can analyze the phase stability of the 1 pps sent round trip through the ground equipment. This sets an upper limit on the expected stability using this equipment. Stability measurements are performed using the two-sample Allan variance of the phase readings from a time-interval counter (TIC). Frequency stability (σ_y vs. τ) from one second to a few thousand seconds is shown in figure 6 for three carrier-to-noise-density (C/No) ratios using the hub equipment with an in-cabinet loop test. Carrier-to-noise-density ratio is a figure of merit parameter for communication links and the measurement technique is described in Appendix I. For a C/No of 65 dB-Hz, σ_y is about $4 \times 10^{-10} \tau^{-1}$. Also shown is σ_y with the modem in its internal (70 MHz) loop test mode which presumably is a best case condition representing a high C/No ratio. Also plotted in figure 6 is the calculated two-sample variance from the white phase jitter published with the modem. Daily measurements at 65 dB-Hz show fairly good agreement with extrapolated short-term measurements with actual data taken to 4 days and stability of a few x 10^{-15}. Thus long-term stability is good. For reference purposes, one voice-grade satellite channel typically represents a C/No of about 55 dB-Hz and the modem works reasonably at this level as shown by the stability measurements of figure 6.

Figure 7 shows frequency stability results of in-cabinet loop tests performed on the VSAT transceiver. C/No was 75 dB-Hz taken at the upper signal limit of the modem in order to uncover any noise degradation due to the transceiver. Identical results were obtained in the internal loop tests and in-cabinet transceiver loop tests indicating that no degradation had taken place. As an additional test, three modems were connected in series so the loop around involved transmission by one modem, reception by a second with retransmission by the second, reception by a third with retransmission by the third through the VSAT transceiver and simulator and finally reception by the first modem. The frequency stability measurement results in figure 7 show virtually no stability degradation compared to single modems, so the apparatus does well even in this situation.

The VSAT was set up with the 1.8m antenna assembled on its free-standing mount for making range measurements to the translator located up to 26 meters away. Horn antennas and a LNA were added to the translator as shown in the block diagram of figure 5. The distance r is taken from the open end of the feedhorns to the open end (phase center) of the J-hook prime focus feed horn of the 1.8m VSAT dish. This range measurement yields the delay through the VSAT extrapolated to zero distance after removing the translator delay by taking measurements as shown in figure 8. The slope, that is the round trip delay as a function of distance, is determined by the speed of light, and is 14.95 cm/ns. Three measurements were taken and one can see the agreement with this slope in figure 8.

Day-to-day loop-around time delay variations were measured with the VSAT in-cabinet and free-space. These results are plotted in figure 9 and show reproducibility to the 1 ns level. These results are excellent and show the potential for satellite synchronization at this level. There is a discernible increase in the noise level of the free-space data. Its cause is unknown but may be due to translator environmental sensitivity. Nevertheless, the results are encouraging since the measurement is "round trip" and represents accumulated phase noise.

Satellite Loop Tests

After testing this equipment (the ground segment), we proceeded to loop tests involving actual satellites operating in the domestic Ku-band with assigned uplink between 14.0-14.5 GHz and downlink between 11.7-12.2 GHz. The configuration for doing these satellite loop tests is essentially identical to the free-space tests involving the translator except that the dish antenna is directed at the satellite transponder. Before this operation is done, the first consideration is deciding which satellites make suitable candidates for tests. This decision is dependent on the receiver G/T and transmit EIRP of the ground facility, the satellites in view and their G/T and EIRP for a given connection, the minimum acceptable C/No, and (in this case) the available bandwidth [5]. Tests were done on three satellites: SATCOM K2 (81 W), SBS 3 (95 W), and SPACENET II (69 W). These transponders are primarily used for video transmissions, and are available on a pre-emptible, partial-transponder, occasional-use basis as long as

no interference is caused to other normal traffic. The spread-spectrum modem requires a minimum C/No of 50-55 dB-Hz, the equivalent of about one voice grade channel, spread across a bandwidth of 2 to 4 MHz. In practice, the modem requires more than 60 dB-Hz in order for the demodulator to acquire initial lock because of the effects of any unknown frequency offset errors in the transponder and because of occasional spurious interference in the roughly 4 MHz bandwidth.

Figure 10a shows a block diagram of key elements of the loop test using the hub and its 6.1m dish. The intermediate frequency used into and out of the upconverter and downconverter equipment is 70 MHz and connections are made to the corresponding transmit and receive ports of the modem. A 1 pps and a 10 MHz reference were provided by a rubidium standard and the transmitted 1 pps identification point used in the modulated PN sequence is approximately synchronized with the external 1 pps reference. The transmitted and received 1 pps as determined by the modem are used as start and stop pulses and are measured using a time interval counter having a resolution of 35 picoseconds (rms) per second. The negative going transitions provide triggering. Figure 10b shows a similar setup but instead using the VSAT with its 1.8m dish. In this case, the upconverter and downconverter are part of the transceiver which is attached to the backside of the dish. About 15 meters of coax cable separates the modem's 70 MHz ports from the input/output of the VSAT.

First loop tests were conducted using the hub earth station and SATCOM K2 located at 81 W. A 4 MHz segment at the high frequency edge of transponder number six was used for the test. The advantage to this kind of segment is its availability and low-cost since it is useful mainly for in-house voice communications by satellite operators and its marketability is not great due to potential interference problems described here. Figure 11 shows spectrum analyzer displays of signals received from SATCOM K2(6B) using the modem and hub earth station. The center of each display marks the frequency of the loop test signal as a clean carrier (figure 11a) and with PN modulation (figure 11b). Figure 11c represents a wider frequency span. Of the normal video signal which may occupy the upper half of transponder number six (designated 6B), there is usually only a small amount of power that extends to the last 4 MHz. However, video signals are frequency modulated and may extend to 24 MHz wide. A full transponder on SATCOM K2 is 54 MHz wide and can carry two video signals. Allowing for a guard band between two video signals, one sees the possibility of interference even at the transponder edges. An adjacent video signal is clearly seen in the spectrum analyzer displays of figure 11. Another form of interference comes from a second transponder which is operating at nearly the same frequency but has orthogonal antenna polarization. This configuration is called the dual polarization mode. Commonly, horizontal and vertically polarized signals share the same satellite permitting greater utilization of the frequency spectrum. If signals are present on the other polarized transponder, then the interference depends on the signals' frequencies, amplitudes, and the degree to which the orthogonal component can be nulled by the earth station antenna. Cross-polarization interference is evident in figure 11 from video transmissions from the other transponder. Another potential problem with use of the edge of a transponder is frequency-response roll-off which creates (1) amplitude-to-phase conversion and (2) more difficult decoding by the modem. This roll-off, as it turned out, was not significant in this test and hence was not a problem.

There are other forms of interference than those stated above, but they were negligible in these tests. Interference analysis is described in reference 5. There is, however, a type of interference that is subtle yet significant when a low C/No signal is shared with video transmissions; its effect is called "loading". The transponder output power amplifier is not linear when operating near saturation; that is, its overall gain becomes a function of its input signal level. When there are no other signals on the transponder, a given signal level is retransmitted back based on the satellite's specifications thus yielding a predictable received C/No. Multiple carrier systems (usually non-video) generally require that the received satellite flux density be several dB less than that required for saturated ouput in order to reduce intermodulation interference. However, when one or two video signals share the transponder, the system is usually operated near saturation. If only one video signal occupies the entire transponder, it may be operated at or even above saturation. This can cause the shared PN signal to drop by as much as 6 to 8 dB causing a net reduction in the received C/No by this amount. Since σ_y is increased as C/No is decreased, unexpected characteristics in σ_y may occur if transponder loading is changing during sampled time intervals. A few transponders have an automatic gain step capability that alleviates this problem by adjusting the gain of the transponder so low level data channels are less affected by loading.

Satellite loop tests were performed using the VSAT and SBS-3 located at 95 W. Setting up the VSAT and actually doing the loop around proved to be a formidable task that is greatly simplified with the availability of the hub earth station. The VSAT with its small dish, low power, and fixed 2.3 GHz transmit/receive offset introduces a new set of problems compared to the hub earth station. In addition to making sure that the link calculation with a given satellite (in this case SBS-3) yields at least 60 dB-Hz, one has to consider the problem of pointing at the correct satellite, assuring that the correct frequency and power level is actually transmitted, and verifying that the spread spectrum signal is being received.

Figure 12 shows spectrum analyzer displays of received signals from the VSAT in a loop test with SBS-3. Clean carrier is shown in figure 12a with a 10 MHz scan and 10 kHz analysis bandwidth, figure 12b is the same display but with PN modulation on, and figure 12c is with PN modulation and with an expanded scan of 50 MHz and 1 MHz analysis bandwidth. Scan time is 10 seconds for figures 12a and 12b, and one can see the difficulty in discerning an identifiable spread spectrum signal in the noise. These displays are typical of the kinds of signals obtained using the modem. The modem usually took several minutes to acquire lock of the return signal from the transponder. A difficulty with using the particular brand of VSAT used here is that the frequency offset between transmitting and receiving is preset at 2.3

GHz and the transponder has an allowable error in its re-transmit offset of ±24 kHz. The signal to the modem should be within ±1 kHz for its best accuracy, but no provision exists for compensating for any frequency error. For this test, the frequency error back to the modem was about 12 kHz which yielded an off-scale condition of a frequency discriminator meter located on the panel of the modem. A convenient method of remotely incrementing the offset in the VSAT in less than 1 kHz steps would be ideal. As it turns out, two-way time transfers will usually occur between a VSAT and the hub earth station, and frequency synthesizers are used at the hub for transmit and receive independently so all frequency errors can be compensated in this configuration.

Time delay through the satellite is about 250 ms in these loop tests. Analysis of the round-trip time (phase measurement) was done again using the two sample Allan variance and its square-root σ_y, and typical results are shown in figure 13. These results are consistent with in-cabinet and free-space loop tests done on the ground equipment and shown in figure 6. In addition to loop tests described thus far, plots are shown in figure 13 for hub earth station to SBS-3 and SPACENET II (69 W), transponder 22. The stability plots follow τ^{-1} behavior as expected for white noise except for the case of VSAT and SBS-3 with loading. As described earlier, the effect of satellite loading is to change or "modulate" the C/No causing a change in σ_y dependent on when and how much loading occurs during the data sample.

Two-Way Time Transfer Using A Common Reference Standard

In making estimates of the time transfer accuracy using the two-way technique, signal delays everywhere in the link are of concern. Such delays come from cables, amplifiers, filters, converters, and, of course, the transmission link itself. In doing a time comparison using the two-way time transfer technique, the absolute values of the signal delay are not directly involved; instead, the difference between the transmit path and the receive path is the parameter of interest (the differential delay). Realizing that two transmit/receive facilities are needed for the transfer, it is ultimately the difference of the two differential delays of the involved facilities that is essential (the "offset of the differences"). Appendix II derives this offset of the differences term. The accuracy and stability of this term gives an upper limit on the accuracy and stability of the time comparison.

Most measurements of ground equipment delays involve the timing of an injection RF pulse and the detection of the same pulse at a point before the antenna for the transmission portion. The delay measurement for the receive portion is done in the same way but with the injection pulse usually at the LNA and the envelope detector after the appropriate receive chain equipment. There are two common difficulties with this approach. The first is that the pulse injection and detection scheme itself introduces a measurement uncertainty since this is not the way the equipment normally operates. A more favorable measurement would be done in situ. Second, the measurement does not include the antenna and its associated orthomode transducer and feed system. One cannot assume that the antenna's differential delays are zero.

Figure 14 is a diagram showing the basic scheme in which the hub earth station and VSAT simultaneously use a common transponder with two separate spread spectrum sequences (near orthogonal [3]) timed by a common 1 pps reference. A modem is used at the hub and another is used at the VSAT and the transmit sequence (Tx) is indicated as "0" for the hub and "1" for the VSAT. The receive sequence (Rx) is "1" for the hub and "0" for the VSAT so each earth station receives the sequence from the other one and not itself. This is the basic configuration used between separate locations doing a simultaneous two-way time transfer. In the case here, the two earth stations are colocated with a common 1 pps reference thus allowing a direct measurement of differential timing errors. Furthermore, this approach allows a direct measurement of system accuracy and stability [1]. Although no data has been analyzed at this time, the concept has been demonstrated using the hub and VSAT equipment operating simultaneously through SBS-3 using two separate spreading sequences. Time interval counter measurements need to be taken and the results compared as described in Appendix II. Figure 15 shows a method of directly measuring the offset of the differences with one time interval counter (TIC) and hence establishing a calibration with the use of these earth stations.

Conclusions

A number of tests have been performed on commercially available Ku-band satellite telecommunications equipment and, in particular, on a small, self-contained earth terminal (VSAT). The VSAT is used in conjunction with precise time transfers via satellite to a central hub facility at NBS, Boulder, with the commensurate differential time offset between VSAT and hub having been directly measured. Tests have also been conducted using a specialized spread spectrum modem designed for two-way time transfer. Short term stability data with this equipment in three different loop-around tests (in cabinet, free space, and via satellite) have been performed under various conditions and various C/No ratios. Satellite loop tests used SATCOM K2 (81 W), SBS 3 (95 W), and SPACENET II (69 W). Loop tests with the hub and with the VSAT were also described. Stability performance is on the order of $4 \times 10^{-10}\tau^{-1}$ for a C/N_0 ratio of 65 dB-Hz. Long term stability data in a satellite loop test has not yet been obtained, but in-cabinet and free-space loop tests of the ground equipment using a satellite simulator show that this white noise behavior continues to a few days where σ_y is 3×10^{-15}. Absolute phase delay measurements show reproducibility to better than 1 ns over a 16 day sample time. Future tests will be done to gather long term stability measurements and long term reproducibility in loop tests involving actual satellites.

This paper described a direct measurement technique for determining the differential offset term for two satellite earth stations (one, a portable VSAT) in close proximity to each other using a common satellite and a common reference clock. Usually, the differential offset constant has been determined in satellite time comparisons by measuring delays of injected RF pulses, a method which has certain inaccuracies. The method used here provides a direct measurement of the delay for two earth stations under actual operating conditions. The direct measurement

makes the fewest assumptions regarding the uncertainty in the determination of the differential offset constant.

Future work will be directed toward analysis of the accuracy and stability of the common-view/common-clock scheme as outlined here. At the present time, the plan is to incorporate more extensive instrument control and automated data analysis to facilitate data reduction and documentation of results. The ground segment and satellite loop tests of frequency stability point to an accuracy capability in the range of a few nanoseconds or better and stability of several x $10^{-10}\tau^{-1}$ for a C/N_0 ratio of 65 dB-Hz.

Acknowledgements

The author is grateful to D. Wayne Hanson of NBS for valuable guidance and discussions in all aspects of this work. He also wishes to acknowledge the help of Ray Conover of Conus Communications for extensive partial transponder availability and Bob Trotter and Terry Maraia of the NBS Radio Frequency Management Office for expediting the VSAT authorization.

Appendix I

Measurement Methods

Of interest in this paper is a relationship between standard frequency stability measurement techniques and standard satellite, signal-to-noise ratio parameters. In the case of frequency stability the measurement performed is the two-sample Allan variance of the phase noise of the ground segment in both in-cabinet and free-space, loop-around schemes. Frequency stability measurements from one second to a few thousand seconds were performed for various carrier-to-noise-density (C/N_0) ratios. Carrier-to-noise density ratio is a general figure of merit parameter for a satellite communications link. Frequency stability data also was taken at one day intervals to look at long term stability and its agreement with extrapolated short term stability results. The actual C/N_0 measurement is made using a spectrum analyzer sampling the unmodulated pure RF carrier of the modem compared to the density of noise in a 1 Hz bandwidth. The spectrum analyzer had a minimum resolution bandwidth of 10 Hz but the noise component was white and allowed straightforward calculation to 1 Hz bandwidth. The analyzer incorporated a correction for doing noise bandwidth measurements at various analyzer settings. These corrections were found to be accurate to the rated specification of 1 dB by scanning the shape of the response curve and scaling the bandwidth as high as 100 kHz and (again assuming white noise) seeing correct closure of the equivalent noise bandwidth at 1 Hz. At low (C + N)/N ratios, it is difficult to accurately measure the signal level because of the presence of the noise component. To overcome this difficulty the carrier signal was introduced at a higher level, accurately measured by the analyzer, and then a precision attenuator was applied to only the carrier in order to reduce its level to a known value.

Some data which is presented in the documentation which accompanies the modem is useful for the analysis of frequency stability presented here. Figure 6 shows a plot from data of the white noise phase jitter versus carrier-to-noise density ratio [2]. A value for σ_y can be computed and the value for 75 dB-Hz is included with the data. If we assume that modem data is the classical variance about the mean of the phase jitter (σ_x^2) then the relationship to the Allan variance is [6,7]

$$\sigma_x^2(\tau_0) = \frac{\tau^2\sigma_y^2(\tau)}{3}.$$

Appendix II

"Offset of the Differences" Measurement

For measurement of the differential delay terms which show up in the two-way transfer scheme, one can use the portable dish in conjunction with a fixed ground station. With a common clock, one can calibrate out these differential delay terms. This is shown in the following analysis in which TI(1) and TI(2) are the time interval counter readings at locations 1 and 2 respectively in a two-way time transfer involving locations 1 and 2.

TI(1) = ΔT + u/c(2) + sat.path(2 to 1) + d/c(1)

and TI(2) = -ΔT + u/c(1) + sat.path(1 to 2) + d/c(2).

ΔT is the time difference of the clocks at 1 and 2, u/c denotes time delay through the up-conversion at locations 1 or 2, and d/c denotes the down-conversions. Sat.path represents the total signal path delays up to and through the satellite and down for signals going from location 1 to 2 and vice-versa. Sat.path includes any delays due to the earth's rotation. ΔT can be calculated as

$$\Delta T = \tfrac{1}{2}(([TI(1)-TI(2)] + [u/c(1)-d/c(1)] - [u/c(2)-d/c(2)] + sat.path(1 \text{ to } 2) - sat.path(2 \text{ to } 1)).$$

If we assume sat.path time delays are reciprocal, except for the time difference term due to the earth's rotation, then sat.path(1 to 2) = sat.path(2 to 1) + δT(rotation) and we have

$$\Delta T = \tfrac{1}{2}(([TI(1)-TI(2)] + [u/c(1)-d/c(1)] - [u/c(2)-d/c(2)] + \delta T(rotation)).$$

Now with the two earth stations co-located and using a common clock, δT(rotation) = 0 and ΔT = 0 and the difference in the u/c's and d/c's is explicitly the difference in the time interval counters. We have

$$TI(2)-TI(1) = [u/c(1)-d/c(1)] - [u/c(2)-d/c(2)] = \text{constant}.$$

This constant which is the offset of the differences can be used for subsequent two-way time transfers using the earth stations, one of which is a portable VSAT which can be located with another earth station to yield a "calibration" of that earth station.

In doing two-way time transfer experiments through geostationary satellites, there exists a limit on the

knowledge of the time delay difference between the outgoing signal and the received signal. This non-reciprocity is due to the difference of paths and the difference of equipment between the uplink and the downlink. Using spread spectrum modulation with different pseudorandom codes for the two directions of time transfer, it is common to assume the difference in the transmission paths to and from the satellite as well as through the satellite transponder to be zero and the earth's rotation correction can be computed [8]. Certainly the ionospheric dispersion and the effects of water vapor dispersion are small (below 100 ps) [9]. The most significant time delay difference error enters in the ground segment. The use of cables, interconnects, conversions, and test points for instruments creates the most significant absolute delays and thus the opportunity for significant differential delays.

References

[1] "Time Comparison Experiments with Small K-Band Antennas and SSRA Equipments via a Domestic Geostationary Satellite," M. Imae, et. al., IEEE Transactions on I & M, Vol. IM-32, No. 1, March 1983, pp. 199-203.

[2] "Timing by Satellite: Methods, Recent Developments, and Future Experiments," D. Kirchner and W. Riedler, Proc. Int. Symp. on Satellite Transmissions, Graz, Austria, 25-27 September, ESA SP-245 (No. 1985).

[3] Mitrex 25000 Documentation, Ph. Hartl, et al., Institute for Navigation, University Stuttgart, Germany, January 1985.

[4] "Stability Measurements of Ku-Band Spread Spectrum Equipment Used for Two-Way Time Transfer," D. A. Howe, Proc. 18th Annual Precise Time and Time Interval (PTTI) Applications Planning Meeting, December 1986. Also "Progress Toward One Nanosecond Two-Way Transfer Accuracy Using Ku-Band Geostationary Satellites," D. A. Howe, IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control, Special Publication on time and frequency (1987) to be published.

[5] "Digital Communications, Satellite/Earth Station Engineering," K. Feher, Prentice-Hall Inc., Englewood Cliffs, N. J., 1981. Also "Digital Satellite Communications," T. T. Ha, Macmillan Publishing Company, New York, N. Y., 1986.

[6] "Characterization of Frequency Stability," J. A. Barnes, et. al, IEEE Transactions on I & M, Vol. I-20, May 1971, pp. 105-120.

[7] Private communication, D. W. Allan, National Bureau of Standards, 325 Broadway, Boulder, CO. 80303.

[8] "Practical Implications of Relativity for a Global Coordinate Time Scale," N. Ashby and D. W. Allan, Radio Science, Vol. 14, No. 4, 1979, pp. 649-669.

[9] "Atmospheric Absorption and Dispersion of Microwave Signals in the Centimeter and Millimeter Region," H. A. M. Al-Ahmad, H. Smith, E. Vilor, Portsmouth Polytechnic Internal Report No. 84/5, January 1984.



Fig. 1. 6.1 meter antenna used as a hub (shown in background) and 1.8 meter VSAT (shown in foreground).



Fig. 2. Earth station equipment.

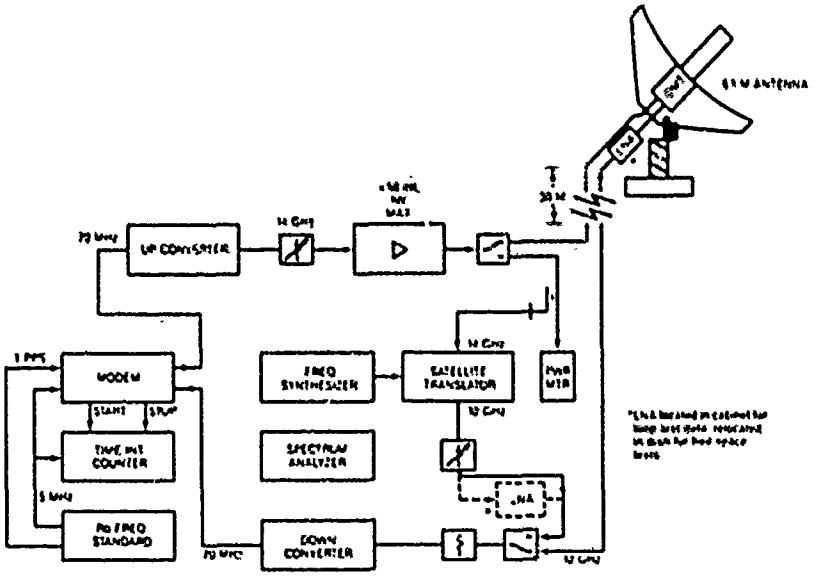


Fig. 3. Block diagram of principal parts of the hub earth station.

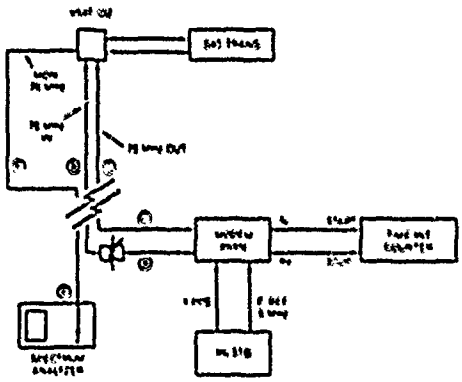


Fig. 4. Block diagram of VSAT in-cabinet loop-around test.

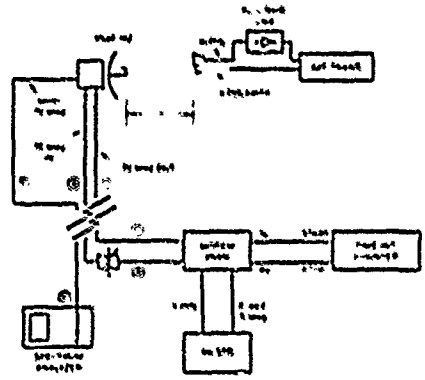


Fig. 5. Block diagram of VSAT free-space loop-around test.

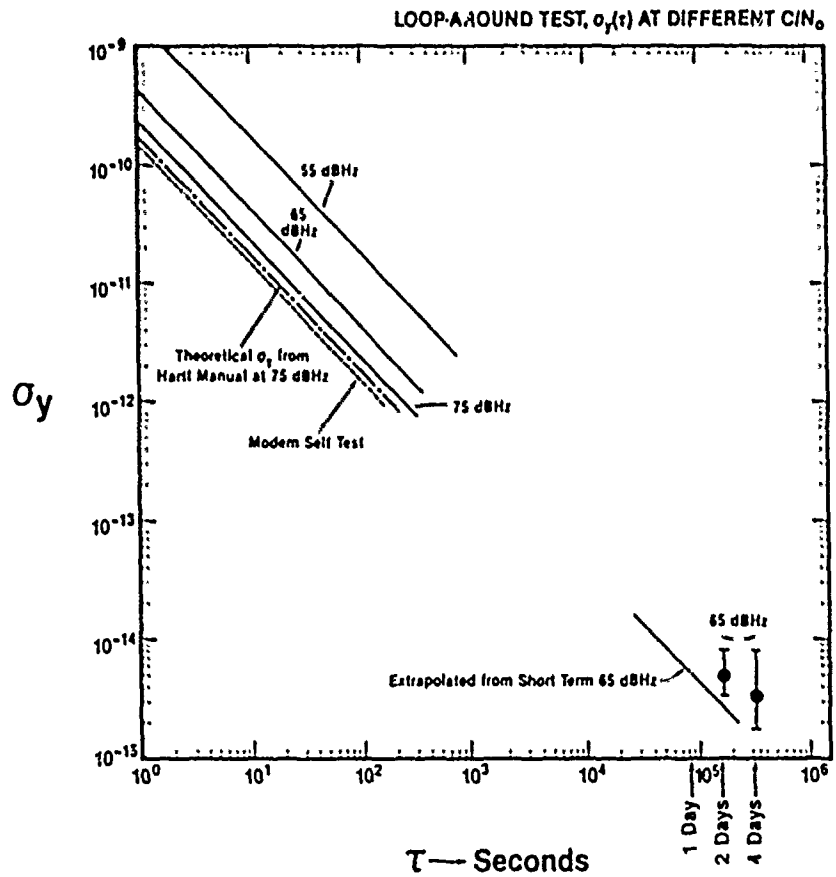


Fig. 6. Frequency stability measurements of hub earth station equipment in an in-cabinet loop test.

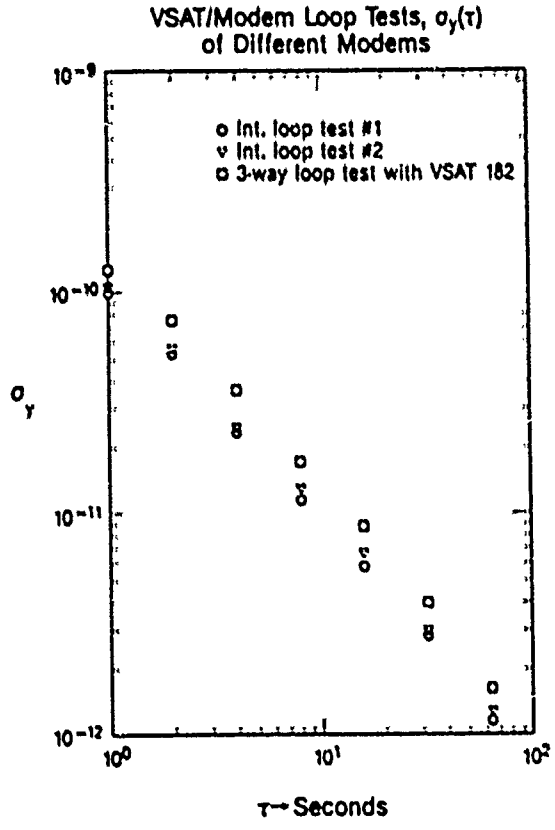


Fig. 7. Frequency stability measurements of VSAT equipment in an in-cabinet loop test and using three different modems for comparison.

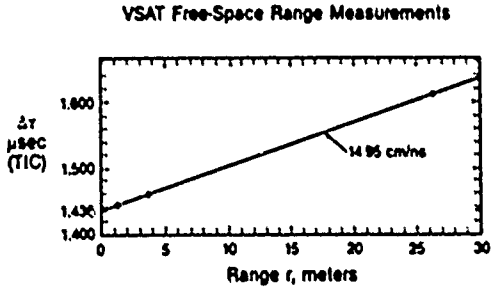


Fig. 8. Three VSAT free-space range measurements can be extrapolated to zero distance to determine delay with the dish. Time interval counter (TIC) readings are a function of range r. Slope is exactly expected result and extrapolation to zero is straightforward.

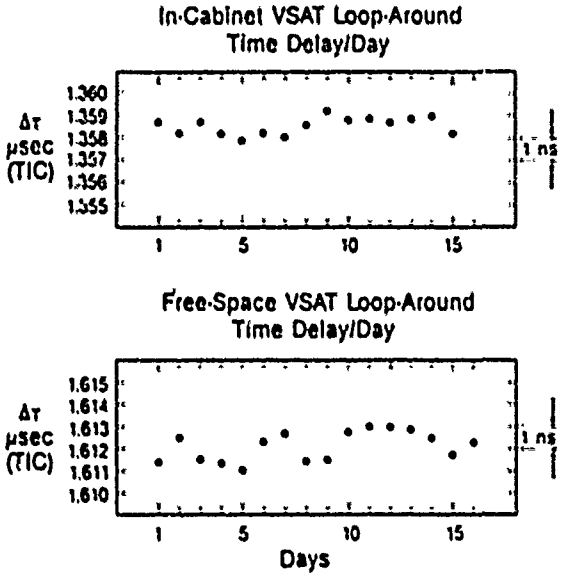


Fig. 9. Day-to-day loop-around time delay using VSAT in-cabinet and free-space showing total in-cabinet and free-space loop-around time delay reproducibility using VSAT.

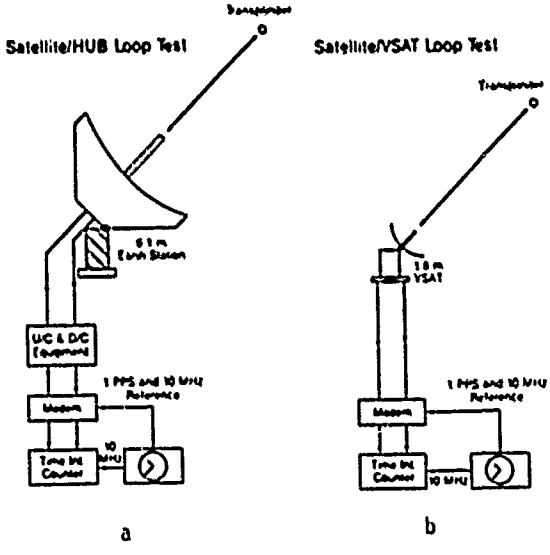


Fig. 10a. Block diagram of principal part of satellite loop tests with hub.

Fig. 10b. Block diagram of principal part of satellite loop tests with VSAT.

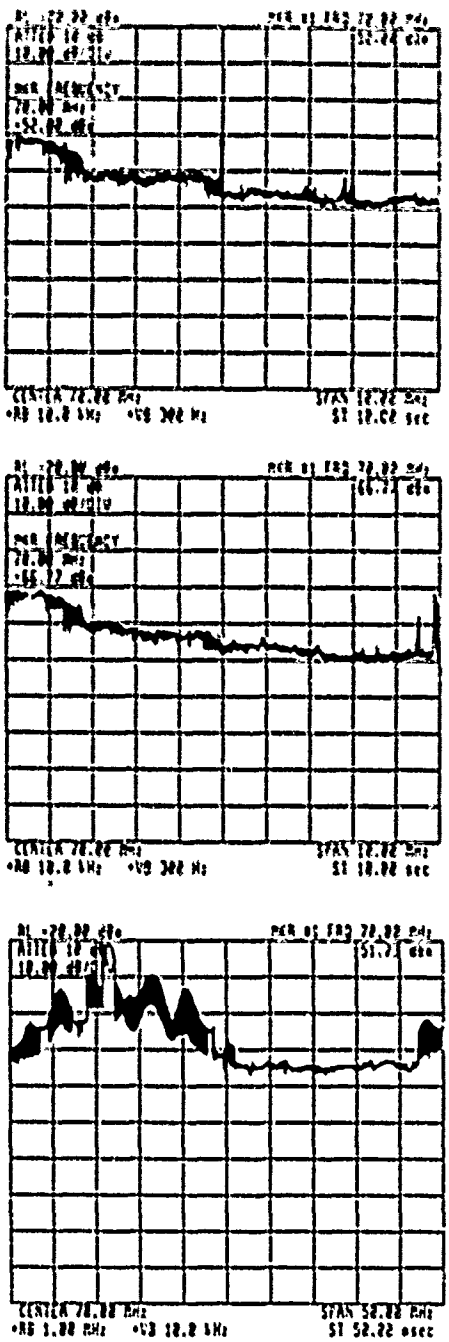


Fig. 11. Spectrum analyzer displays of loop tests with hub earth station and SATCOM K2(6B). Adjacent video signal and cross-polarization interference are easily identified.

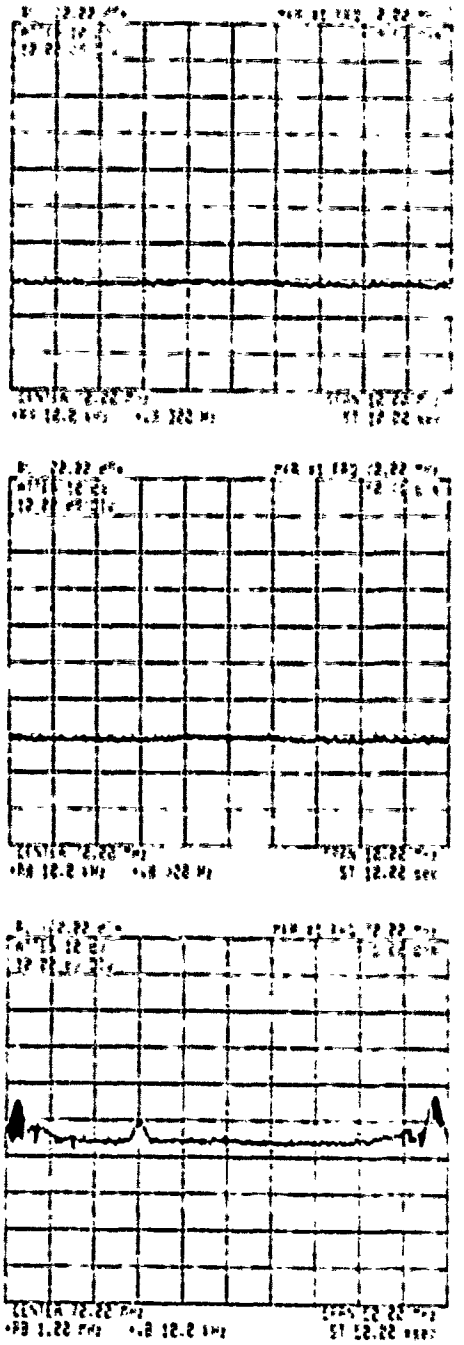


Fig. 12. Spectrum analyzer displays of loop tests with 1.8 m VSAT and SBS-3(7A).

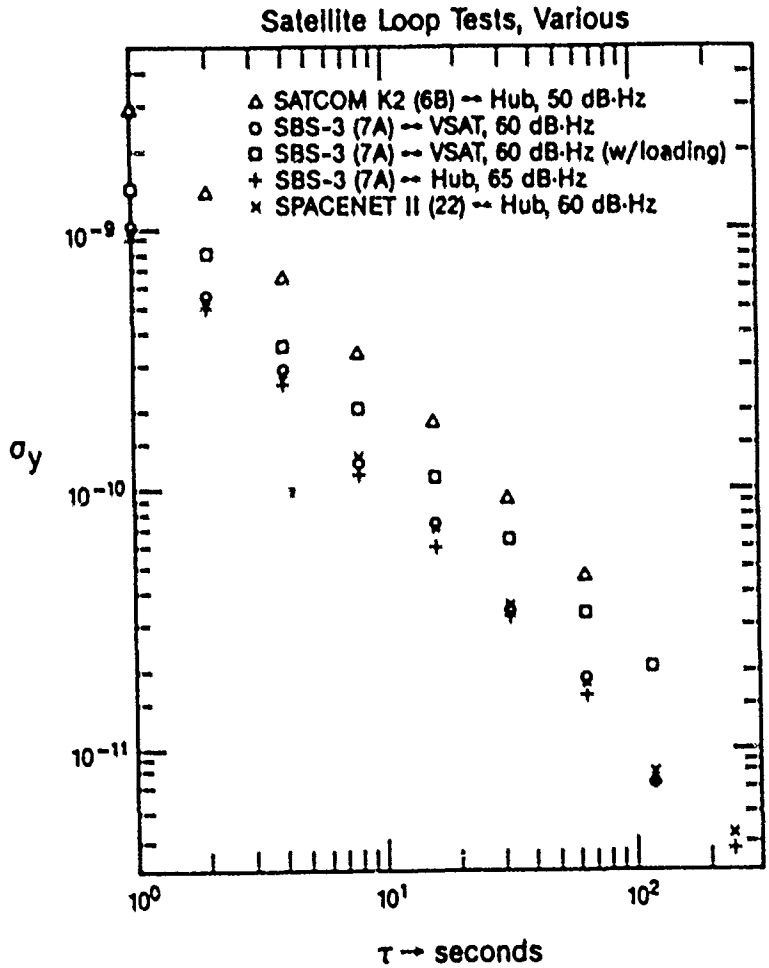


Fig. 13. Frequency stability measurements of loop tests involving hub and VSAT with three satellites.

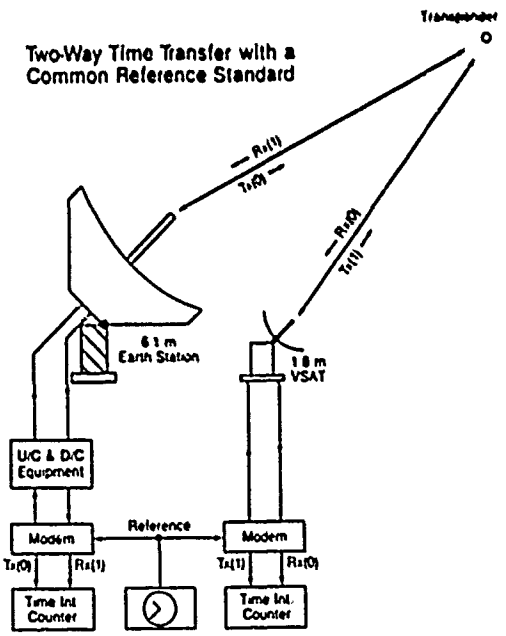


Fig. 14. Common view/common clock scheme for measurement of ground-segment differential delay constant.

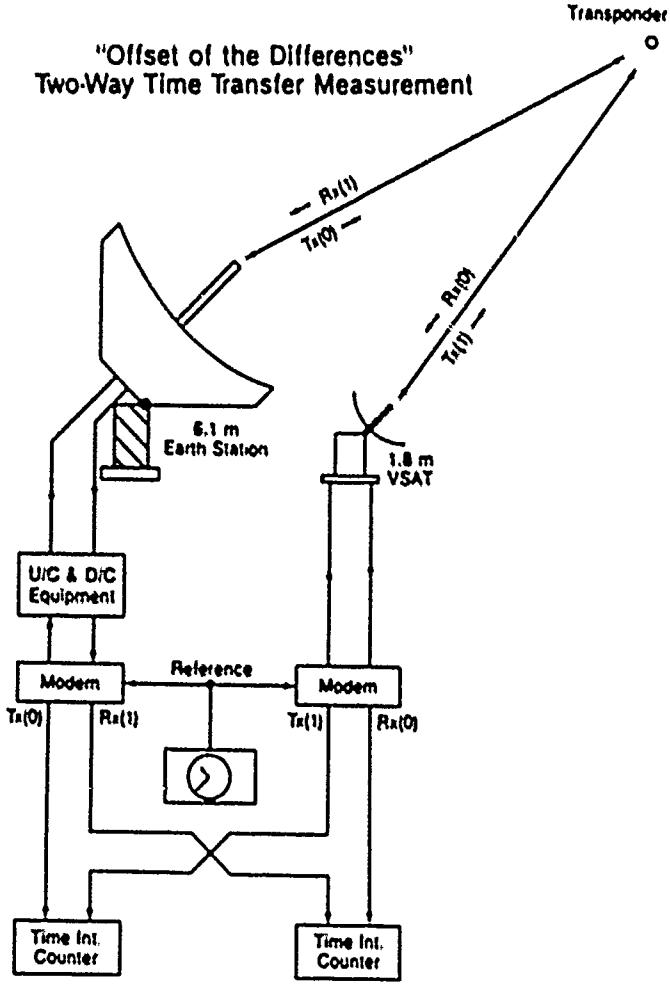


Fig. 15. Technique for direct measurement of offset of the differences. If the left time interval counter (TIC) is zero (representing identical on-time transmission of 1 pps from hub and VSAT), then right TIC is exactly the offset term.

41st Annual Frequency Control Symposium - 1987

REFERENCE FREQUENCY DISTRIBUTION OVER OPTICAL FIBERS:

A PROGRESS REPORT

George Lutes
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

ABSTRACT

The results were recently reported of an experiment in which a 100-MHz reference frequency was transmitted over a 14-km long fiber-optic link. A differential stability of 1.5×10^{-15} for 1000 seconds averaging times was measured for this link. However, there were several factors that limited the stability of the link. A phase change due to bending the cable resulted in a sensitivity to microphonics and a degradation to the Allan variance. Also, variations of phase delay due to temperature changes caused degradation of the Allan variance as well.

This paper considers fundamental and practical stability limits on reference frequency distribution over fiber-optic links. It also reports on recent progress in reducing the sources of instability, including the use of optical isolators to greatly reduce sensitivity to cable bending.

I. INTRODUCTION

Modern frequency standards, such as hydrogen masers and mercury ion standards, generate reference frequencies that are very stable, better than 1×10^{-15} for 1000 seconds averaging times.[1] The distribution of these signals to remote users without significantly degrading their stability has become a difficult technical challenge.

The ability to distribute precise reference frequencies over distances of tens of kilometers will result in considerable cost savings and improved reliability in the NASA/JPL Deep Space Network (DSN). To this end, fiber-optic reference frequency distribution system development is an ongoing task at JPL. The goal is to achieve a transmission stability of 1×10^{-17} for 1000 seconds averaging interval over a distance of 22 kilometers.

A transmission system degrades the frequency stability of a signal by adding frequency variations to the signal and degrading its signal-to-noise (S/N) ratio.

The transmission medium is an important element of a transmission system. The medium's loss, delay stability, bandwidth, susceptibility to interference from outside sources, reliability, and cost are all important considerations in the design of precise frequency distribution systems.

As a result of studies and numerous experiments made at JPL,[2-5] it has been concluded that optical fiber is the most promising medium that can be used to distribute precise reference frequencies.

II. FREQUENCY STABILITY

The accepted measure of frequency stability is the Allan variance,[6] which can be given by

$$\langle \sigma_{\Delta f/f(2,T,\tau)} \rangle = \frac{1}{f_0} \sqrt{\frac{1}{2N} \sum_{i=1}^{N} (f_i - f_{i-1})^2} \qquad (1)$$

where $f_i - f_{(i-1)}$ is individual successive measurements of the frequency being measured, N is the number of measurements, f_0 is the nominal mean frequency, T is the time between measurements, and τ is the averaging time of each frequency measurement. The measurement of Allan variance is dependent on the averaging time of each frequency measurement. This implies that the noise-causing instabilities have components of many frequencies. Therefore, the averaging time must be specified.

III. FREQUENCY STABILITY DEGRADATION

As mentioned earlier, frequency distribution systems degrade the frequency stability of signals that are transmitted through them. Basically, there are two mechanisms responsible for this degradation, frequency variations that result from group delay changes in the signal path and a decrease in S/N ratio.

Frequency variations generated by group delay changes in the signal path can be a result of a changing environment around the signal path, such as changes in temperature or pressure, or they can result from changes in the material that makes up the transmission medium itself.

If the group delay through a signal path changes when a signal is transmitted through it, the frequency of the signal at the receiver is offset relative to the frequency of the signal at the transmitter by

$$f_\Delta = f_0 \frac{dD}{dt} \qquad (2)$$

where f_Δ is the offset frequency at the receiver, f_0 is the input frequency at the transmitter, dD/dt is the rate of change of group delay through a transmission medium with respect to time, D is the delay through the transmission medium in seconds, and t is the time in seconds.

A constant rate of change of group delay through a signal path adds a constant frequency offset to a signal passing through it. In this case, the frequency stability of the transmitted signal is not degraded relative to the frequency stability of the reference signal. However, if the rate of change of group delay through the signal path is not constant, which is usually the case, the frequency of the transmitted signal varies with the rate of change of group delay. This variation in frequency is given by

$$\frac{df_\Delta}{dt} = \frac{d^2D}{dt^2} . \qquad (3)$$

CH2427-3/87/0000-151 $1.00 C 1987 IEEE

These frequency variations represent a degradation in frequency stability of the transmitted signal relative to the frequency stability of the reference signal.

One of the desirable features of most transmission systems is the ability to transmit a signal with minimal degradation to the S/N ratio. This is particularly true for transmission systems designed to transmit precise reference frequencies. Frequency stability can be directly related to S/N ratio. For white phase noise this relationship is given by[7]

$$\sigma(\tau) = \frac{0.276\sqrt{S_\phi}\sqrt{f_h}}{f_0\,\tau} \qquad (4)$$

where τ is the averaging time, S_ϕ is the spectral density of phase noise, f_h is the cutoff frequency of the measurement equipment, and f_0 is the mean frequency being measured.

Reduction of the S/N ratio in a transmission system can occur because of poor system design such as incorrect gain distribution in the system. Excess signal loss in the signal path, poor circuit design, or faulty circuitry can also contribute to degradation of the S/N ratio.

IV. REDUCTION OF THE EFFECT OF GROUP DELAY CHANGES

Three methods can be used to reduce frequency stability degradation of a signal as a result of group delay changes in the signal path. The magnitude of the group delay changes can be reduced, the rate of change of group delay can be slowed down, and an electronic feedback system can be used to reduce the effect of group delay changes.

The first method can be implemented by actively or passively controlling the environment around the transmission medium and other system components. However, there are practical limits to the extent that the environment may be controlled. For instance, the transmission medium may be spread out over a great distance, making it impractical to actively control the environment around it. In this case, the only practical means of protecting the transmission medium from environmental changes may be insulation. Usually there is a practical limit on the amount of material that can be used for this purpose.

Earth is an excellent material to use for insulation. It is a good thermal insulator and can provide a lot of mass at low cost. Since temperature changes are usually the greatest contributor to group delay variation in a transmission medium, the rate of change in group delay can be reduced substantially by adding mass. A combination of mass and insulation can be very effective in reducing temperature effects on group delay.

At the frequency stability levels required in the DSN, on the order of a part in 10^{17}, it is not sufficient to use only the first two methods. Electronic systems are being developed to further reduce the effect of group delay changes in the transmission medium.

Two types of electronic systems are used to reduce the effect of group delay changes on a transmitted signal. In both systems the transmitted frequency is separated into two signals at the remote end of the signal path. One of the signals is used as a remote frequency reference. The other signal is transmitted back to the reference end of the signal path to be used to determine the phase at the remote end of the signal path.

One type of system[8] uses a variable delay device in series with the signal path. The delay of this device is controlled by an electronic feedback circuit so that it virtually cancels any group delay change in the signal path. If the group delay in the signal path changes, the delay in the variable delay device is forced to change by the same magnitude in the opposite direction. The net change in delay is near zero.

The second type of system adjusts the phase of the transmitted signal in such a way that group delay changes in the signal path have no effect on the frequency stability of the received signal.

Figure 1(a)-(c) is used to explain the operation of this type of system. In Figure 1(a), a signal is transmitted through a transmission line with a length of 2 D. The input phase of the signal is θ_1 and the phase of the output signal is θ_2. The phase of the signal at the midpoint is

$$\theta_m = \theta_1 + \frac{\theta_2 - \theta_1}{2}. \qquad (5)$$

The phase at the midpoint can be determined if the phase at the output of the transmission path is known relative to the input signal. However, the output end of the transmission line may be a long distance away. In that case, the phase of the output signal cannot be measured relative to the input signal. To circumvent this problem, the transmission line may be cut in the middle and

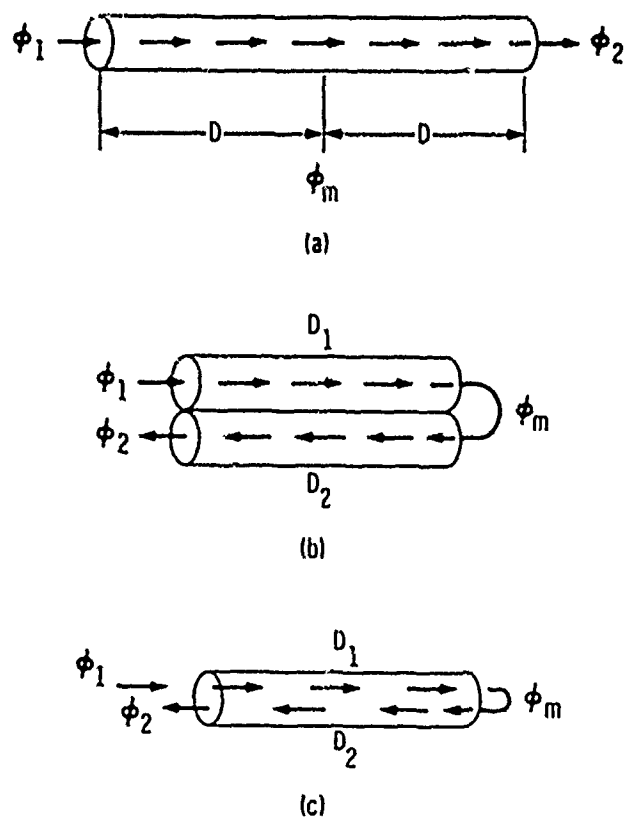


Figure 1 - Evolution of a bidirectional method to determine phase at the far end of a signal path.

folded back on itself, as shown in Figure 1(b). The input and output signals are then at the same location. Now the phase of the output signal can easily be measured relative to the input signal phase, and the phase at the midpoint (remote end) of the folded transmission line can be determined.

There is a problem with this configuration. In a real system, the group delays in the two sections of transmission line will not change equally when subjected to environmental changes. One transmission line will become longer or shorter than the other transmission line. By referring to Figure 1(b), one can see that the phase at the midpoint will be

$$\theta_m = \theta_1 + (\theta_2 - \theta_1)\frac{D_1}{D_1 + D_2}. \qquad (6)$$

The phase of the signal at the midpoint cannot now be determined unless the delay is known in both the forward and return paths, D_1 and D_2, respectively.

In order to guarantee that the signal paths are equal in each direction, the signal must be transmitted through the same transmission line in both directions, as shown in Figure 1(c). This puts the additional requirement on the system that high isolation be maintained between the signals transmitted in opposite directions. Cross talk between the input and return signals will change the apparent phase of the return signal. This will result in an error in the phase measurement at the remote end of the signal path. The magnitude of this phase error is found by adding the vectors of the cross talk and the output signal and is

$$\alpha = \arctan\frac{B \sin \beta}{A + B \cos \beta} \qquad (7)$$

where A is the magnitude of the return signal, B is the magnitude of the cross talk, and β is the phase angle of the cross talk relative to the phase of the return signal.

So far it has been shown that the phase of a signal at the remote end of a signal path can be determined. However, it is not only required that the phase be determined, but also that it be held constant so that no frequency offset will be generated. This is accomplished as shown in Figure 2. The input signal phase is constantly adjusted so that a fixed relationship exists between the phases of the input and return signals relative to a reference phase, θ_0.

A conjugate phase relationship is used in this example, which gives the result

$$\theta_m = (\theta_0 - \theta_1) + \frac{(\theta_0 + \theta_1) - (\theta_0 - \theta_1)}{2} = \theta_0. \qquad (8)$$

This result shows that when a conjugate phase relationship is maintained between the input and return signals, the phase of the remote signal is the same as the phase of the reference signal, θ_0. This conjugate phase relationship is not required to maintain the frequency stability of the transmitted signal. It is only required, in this system, that the phase relationship between the input and return signals be fixed.

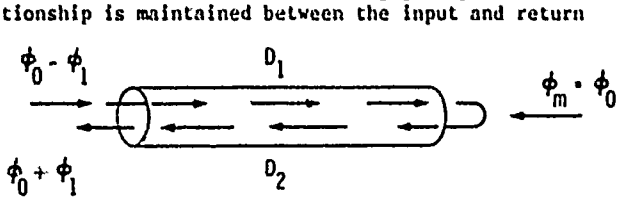


Figure 2 - Depiction of the conjugation method to maintain constant phase at the far end of a signal path.

The portion of the signal path in the input and remote terminals of a constant phase relationship system is outside of the two-way signal path through the transmission medium. The terminal delays may be such that the two-way signal paths are not exactly the same. The results of this situation can be analyzed by assigning different delays, D_1 and D_2, to the forward and reverse signal paths in Figure 2 and applying them to (8) to obtain

$$\theta_m = (\theta_0 - \theta_1) + \frac{2\theta_1 D_1}{(D_1 + D_2)}. \qquad (9)$$

From equation (8) it can be seen that θ_m is no longer equal to θ_0. However, θ_m is still not affected by changes in group delay through the signal path as long as the forward and reverse signal paths change proportionately.

This can be analyzed by adding proportional delays to both the forward and reverse signal paths. Adding delays KD_1 and KD_2 to (9) gives

$$\theta_m = (\theta_0 - \theta_1) + \frac{2\theta_1(D_1 + KD_1)}{(D_1 + KD_1) + (D_2 + KD_2)}. \qquad (10)$$

This reduces to the original equality in (9). Thus there is no phase variation at the remote end of the signal path due to changes in the group delay through the signal path.

However, if the group delay through the signal path changes disproportionately in one direction, the phase of the signal at the remote end of the signal path will also change. This can be shown by adding a delay, D_1, to the forward signal path only in (9) to get

$$\theta_m = (\theta_0 - \theta_1) + \frac{2\theta_1(D_1 + KD_1)}{(D_1 + KD_1) + D_2},$$

$$\theta_m = (\theta_0 - \theta_1) + \frac{2\theta_1 D_1(K + 1)}{D_1(K + 1) + D_2}. \qquad (11)$$

From this it can be seen that if the group delay changes in one direction, as does D_1 in (11), then the phase of the output signal also changes. This creates an offset frequency.

The precise frequency distribution system being developed at JPL uses a fixed conjugate relationship between the input signal and the return signal at the reference end of the signal path. An electronic negative-feedback system maintains this conjugate relationship. A simplified block diagram of this system is shown in Figure 3.

The input reference signal of the phase conjugation system is split by an RF power splitter. One signal drives the L port of a phase detector, and the other signal drives the R port of another phase detector. Each of the I ports of the two phase detectors is connected to one of the inputs to a low noise differential amplifier (op-amp).

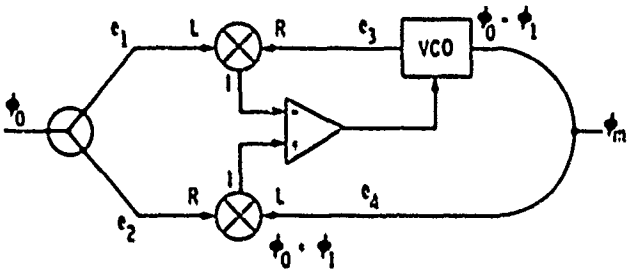


Figure 3 - Block diagram of a system to maintain phase conjugation between a transmitted signal and a return signal.

The output of the amplifier is zero only when the remaining R and L ports are driven by an input signal and a return signal that are conjugate to one another around the phase of the reference signal. Any error signal at the output of the differential amplifier will drive the phase of the signal at the output of the VCO until a conjugate relationship is reached. At this point, the feedback loop will maintain the conjugate relationship between the two signals.

Two methods that can be used to transmit precise reference frequencies have been discussed. Both methods depend on proportional change in group delay in both directions through the signal path and on the ability to transmit signals in both directions in the same signal path with high isolation between them. How well these criteria can be met is largely dependent on the transmission medium used.

V. OPTICAL FIBERS

There are two general classifications of optical fibers, multimode and single-mode. Multimode fiber supports the propagation of many modes. It is available with bandwidths as high as 1 GHz•km and loss as low as 0.5 dB/km at 1300 nm wavelength.

The delay through multimode fiber is dependent on the average velocity of all of the modes being propagated through it. The number of modes and the mode mix can vary depending on the source and conditions. This makes the group delay through multimode fiber susceptible to differences for signals propagating in opposite directions. This effect can be minimized by filling all of the modes in the fiber. This can be done by using an LED light source and the right launch conditions. However, it is difficult to assure this condition.

A form of noise called modal noise is also characteristic of multimode fiber. It is caused by changes in the mode mix resulting from modulation of the light source and movement of the cable or connectors.

Although multimode fiber can be and has been used to transmit precise reference frequencies, it is not the best choice. When it is used in this application, an LED source should be used and the cable should be protected from movement.

Single-mode fiber of the type most commonly used today propagates only the lowest order bound mode, which consists of a pair of orthogonally polarized fields. Special fiber that propagates only one polarization is available, but it is a relatively new development and is quite expensive at this time.

Because it propagates only one mode, single-mode fiber is less susceptible to different delays for signals propagating in opposite directions. However, this condition can still occur under some circumstances because the two orthogonally polarized fields associated with the optical mode propagate at different velocities.

The loss in single-mode fiber is less than 0.5 dB/km at 1300 nm wavelength, and the bandwidth at this wavelength can be greater than 100 GHz•km.

Single-mode fiber designed to operate at 1300 nm wavelength is an excellent medium for the transmission of precise reference frequencies and is used at JPL for this purpose. A wavelength of 1300 nm is used because the bandwidth and loss of optical fiber designed for use at this wavelength are adequate, the technology is mature, and the components are readily available.

VI. FIBER-OPTIC REFERENCE FREQUENCY DISTRIBUTION SYSTEM

A block diagram of a single-mode fiber-optic reference frequency distribution system being developed at JPL is shown in Figure 4. The system uses a 1300 nm single-mode optical fiber with an optical directional coupler at each end to provide a means to send signals both ways through the fiber.

An optical transmitter and receiver pair are attached to the two fibers. The optical transmitters are commercial units that emit a 1300 nm optical carrier that is directly amplitude modulated. The 3 dB bandwidth of the modulation frequency response is 3 GHz. A high isolation fiber-optic isolator is in series with each optical transmitter to isolate it from reflections on the optical fiber. These fiber-optic isolators were recently developed at JPL.[9] They provide greater

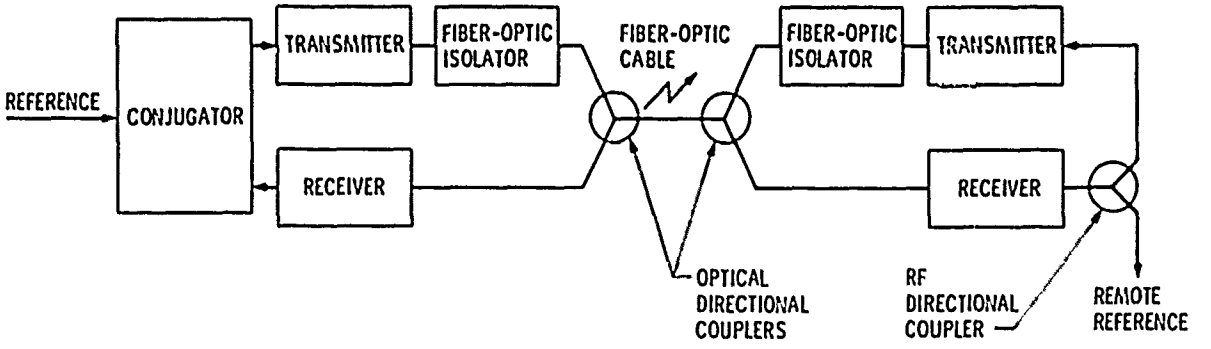


Figure 4 - Block diagram of fiber-optic reference frequency transmission system.

than 70 dB isolation and have a forward loss of only 2.6 dB.

Without these isolators, phase instabilities in the RF modulation frequency are generated in the laser diodes. This instability is a result of changes in the amplitude of extraneous modulated light entering the laser diode.[10] The sources of this extraneous light are the return signal and reflections on the optical fiber. The changes in amplitude are caused by moving or bending the optical fiber. The result of this effect is a pronounced sensitivity to microphonics if isolators are not used.

The optical receivers consist of commercial PIN photodiode detectors followed by wideband, low noise amplifiers. The receivers have a 3 dB bandwidth of 3 GHz and, like many wideband devices, have a small temperature coefficient of delay.

A phase conjugation circuit like the one previously described establishes and maintains a conjugate relationship between the transmitted and return signals.

Phase locked loops are used for two purposes, to provide a constant amplitude RF signal to the modulation input of the laser transmitters and to improve the short-term phase noise at the remote end of the transmission system.

A phase locked clean-up loop is needed at the output of the frequency distribution system to eliminate the wideband noise from the fiber-optic system. This clean-up loop must use an oscillator that is as good as that used in the frequency standard if the frequency stability of the transmitted signal is to be comparable to the frequency stability of the standard.

VII. THEORETICAL AND PRACTICAL LIMITS

The short-term Allan variance is largely dependent on the S/N ratio of the transmitted signal and for white noise has the relationship given in (4).[7] Typical analog fiber-optic systems in use today have a S/N ratio of 120 to 130 dB/Hz[11-12] for transmission distances less than 30 km. For these distances, the S/N ratio is determined primarily by the laser noise. The phase noise of several laser diodes has been measured at JPL and, for offset frequencies from 1 Hz to 10 kHz from the carrier, was in agreement with the manufacturer's reported value.

The Allan variance corresponding to the S/N ratio of today's laser diodes and laser diodes under development is shown in Figure 5. The frequency stability limit due to S/N ratio can be lowered by raising the frequency since the stability as given in (4) is inversely proportional to frequency. For this reason, JPL is planning to develop 1 GHz systems in the future.

Much of the short-term noise in the present system results from microphonics. As stated earlier, the sensitivity to microphonics should be reduced considerably with optical isolators.

Long-term instabilities due to changes in the group delay through the signal path can be reduced by electronic feedback circuits. These circuits are subject to errors that result from nonreciprocal changes in delay in the optical fiber or terminal equipment. They are also subject to drift or noise

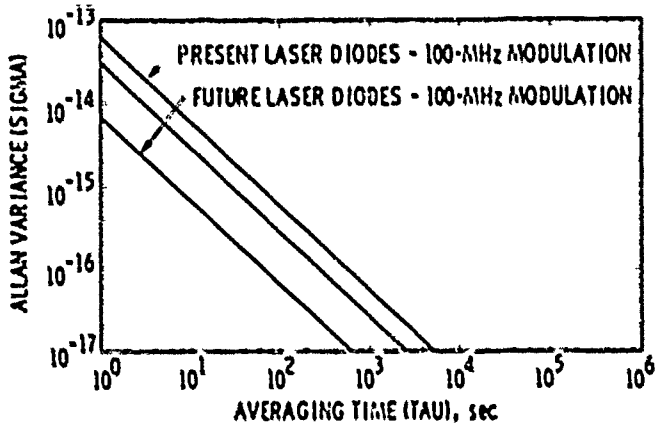


Figure 5 - Frequency stability limits imposed by the S/N ratio of present and future laser diodes.

in the error detection circuitry. Fundamental limits due to these effects are not yet fully understood. It seems likely that the stability of a transmission system could approach the best stability that can be measured, at least until the S/N ratio limit is reached.

VIII. STATUS OF FIBER-OPTIC FREQUENCY TRANSMISSION DEVELOPMENT

A stability of 1.5×10^{-15} for 1000 seconds averaging interval for a 14-km long fiber-optic frequency transmission system was reported in December of 1986.[13] This system was unstabilized; in other words, no electronic feedback system was used. The stability data for this link is shown in Figure 6.

This system was very sensitive to microphonics as a result of light reflected back into the laser from the optical fiber. The previously described optical isolator was developed to reduce this effect.

Tests using the optical isolator are in progress. However, the optical fiber pigtail between the laser and the isolator is still subject to microphonics. A new isolated laser package that should eliminate this problem is being developed.

An electronic stabilizer is also under development. Several prototype versions have been tested and several needed improvements have been

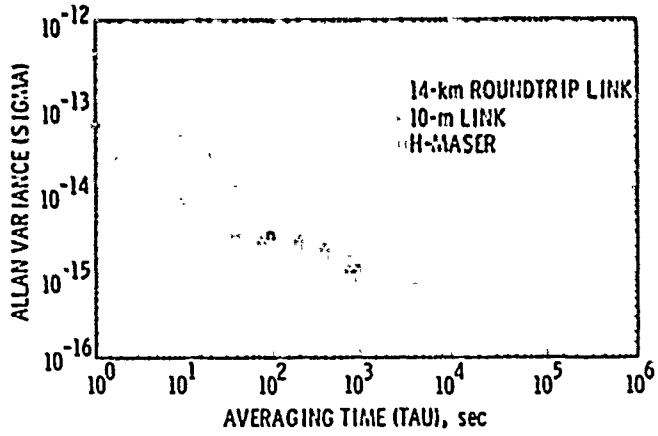


Figure 6 - Allan variance vs. averaging time for the 14-km link, 10-m link, and hydrogen maser.

identified. These improvements are currently being implemented and test results will be reported soon.

IX. FUTURE DEVELOPMENT

Tests will be performed on a fiber-optic transmission system consisting of an optical transmitter, an optical isolator, a cable, and a receiver, to determine

- the sensitivity of system components to temperature variations,
- phase change vs bending of the cable,
- sensitivity to microphonics,
- the extent of nonreciprocity of delay in the fiber,
- the magnitude of phase noise and the Allan variance.

The knowledge gained from these tests will be used to improve the existing system performance and to design a short-distance frequency distribution system.

The primary reference frequency to be used in the DSN will be raised from 100 MHz to 1 GHz, and a fiber-optic frequency transmission system will be designed for this frequency.

X. CONCLUSION

Frequency references generated with today's frequency standards can be distributed over tens of kilometers distances with fiber-optic systems with little degradation to the frequency stability of the signal.

Frequency standards could be improved considerably in the future, and reference frequency transmission systems will be challenged to distribute the reference frequencies they generate.

Some fundamental and practical limits on the stability of fiber-optic reference transmission systems are not yet fully established, but work is being performed to determine the limits.

ACKNOWLEDGMENT

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

The author wishes to thank Dr. L. Maleki, Dr. R. Sydnor, Dr. C. Greenhall, Dr. G. Dick, A. Kirk, L. Primas, P. Tu, and C. Corrigan for their suggestions and help in the preparation of this paper.

REFERENCES

[1] A. Kirk, P. Kuhnle, and R. Sydnor, "Evaluation of modern hydrogen masers," Proc. 14th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 359-392, NASA Conference Publication 2265, Nov./Dec. 1982.

[2] L. A. Bergman, S. T. Eng, A. R. Johnston, and C. F. Lutes, "Temperature dependence of phase for a single-mode fiber cable," Proc. Third International Conference on Integrated Optics and Optical Fiber Communications, p. 60, USA-IEEE, April 27-29, 1981, San Francisco, CA.

[3] C. Lutes, "Optical fibers for the distribution of frequency and timing references," Proc. 12th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 663-680, NASA Conference Publication 2175, Goddard Space Flight Center, Dec. 1980.

[4] C. Lutes, "Development of optical fiber frequency and time distribution systems," Proc. 13th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, pp. 243-262, NASA Conference Publication 2220, Naval Research Laboratory, Dec. 1981.

[5] K. Y. Lau, "Propagation path length variations due to bending of optical fibers," The Telecommunications and Data Acquisition Progress Report 42-63, pp. 26-32, Jet Propulsion Laboratory, March-April 1981.

[6] J. A. Barnes, A. R. Chi, L. S. Cutler, D. J. Healey, D. B. Leeson, T. E. McGunigal, J. A. Mullen, W. L. Smith, R. Sydnor, R. F. C. Vessot, and G. M. R. Winkler, "Characterization of frequency stability," Technical Note 394, National Bureau of Standards, Oct. 1970.

[7] L. S. Cutler and C. L. Searle, "Some aspects of the theory and measurement of frequency fluctuations in frequency standards," Proc. IEEE, special issue on Frequency Stability, Vol. 54, p. 136, Feb. 1966.

[8] K. Y. Lau, "A voltage-controlled optical radio frequency-phase shifter," The Deep Space Network Progress Report 42-53, pp. 24-32, Jet Propulsion Laboratory, July-Aug. 1979.

[9] C. Lutes, "A low loss, high isolation, single-mode fiber-optics isolator," submitted to Optics Letters, April 1987.

[10] K. Lau, Ortel Corp., Alhambra, CA., private communication.

[11] K. Lau, "Signal-to-noise calculations for fiber-optics links," The Telecommunications and Data Acquisition Progress Report 42-58, pp. 41-48, Jet Propulsion Laboratory, May-June 1980.

[12] A. Yariv, Introduction to Optical Electronics, Holt, Rinehart and Winston, 1971.

[13] C. Lutes and A. Kirk, "Referen. frequency transmission over optical fiber. Proc. 18th Ann. Precise Time and Time Interval (PTTI) Applications and Planning Meeting, Dec. 1986, to be published in the proceedings.

41st Annual Frequency Control Symposium - 1987

ELECTRODIFFUSION OR SWEEPING OF IONS IN QUARTZ

J. J. Martin
Department of Physics
Oklahoma State University
Stillwater, OK 74078-0444

SUMMARY

Electrodiffusion or sweeping is a high temperature process which selectively exchanges monovalent ions in alpha-quartz. Sweeping is commercially employed to replace the alkalis present in as-grown quartz with hydrogen. Electrodiffusion has been shown to improve the radiation hardness of quartz oscillator crystals and to significantly lower the production of etch tunnels. The first effect is important in critical aerospace and defense applications. While the second directly impacts the production of devices by photolithographic techniques and of very high frequency bulk wave quartz oscillator crystals.

Sweeping affects two types of defects: point defects such as the substitutional aluminum with it's associated interstitial alkali and the extended dislocation networks with their precipitated impurities that form the etch tunnels. The process associated with the point defects is, at least, qualitatively understood. At high temperatures, the interstitials are thermally liberated from their trapping sites and can then migrate along the relatively open Z-axis channels. Therefore, if the sample is heated with an electric field applied along the Z-axis, these interstitials can be swept out at the negative electrode provided replacement ions are brought in at the positive electrode. In the commercial process, the replacement ions are protons from the water vapor in the surrounding air atmosphere. Thus, the aluminum-alkali centers are converted into Al-OH centers. The precipitates in the dislocation networks are probably modified so that they are less reactive to the NH_4HF_2 or HF etchants.

Sweeping is usually carried out on a fully lumbered bar such as the Y-bars used for AT-cut crystals. Evaporated or sputtered metal film electrodes are applied to the Z-faces. The electroded bar is placed in the furnace and then slowly brought up to the desired temperature. The electric field can be first applied at the operating temperature (usually near 500°C) or at room temperature. The field is established such that the negative electrode is the seed-side of the bar; therefore, the impurities are swept back toward the seed. If the field is first applied at the operating temperature a large initial current is observed; this current then slowly decays until it reaches a steady value which is caused by the movement of protons through the sample. If instead, the field is applied at room temperature; a peak or plateau in the sample current is observed near 250-300°C. This peak probably has the same origin as the large initial current which is observed when the field is applied at high temperatures. Once the current becomes steady, the sweeping is thought to be complete. The bar is then slowly returned to room temperature with the field still on. An effective DC ionic conductivity can be measured during this cool down part of the cycle. Plots of the product of this conductivity times the absolute temperature versus the reciprocal temperature show that the process is thermally activated with activation energies of 1.5 to 2 eV.

A number of tests are available for the determination of the completeness of the sweeping run. For example, the observation of a steady current towards the end of the run is usually assumed to indicate that the sweep is successful. If the run is repeated with the electric field applied at room temperature the current peak or plateau near 250°C will be missing. Both of the above tests are very qualitative. A reliable test for the completeness of the reconversion of aluminum-alkali centers into Al-OH centers can be performed by comparing the concentration of Al-hole centers after an initial irradiation at 77 K with the concentration found after subsequent irradiations at 300 K and then at 77 K. If the initial and final concentrations are the same then the conversion was complete. This test is made on small samples removed from the bar. The presence of the Al-OH center can be determined by low temperature infrared absorption measurements. High temperature (300-400°C) measurements of the equivalent series resistance of a finished crystal can also be used to determine if the material has been well swept. The alkali ions present in unswept quartz cause a large exponentially increasing acoustic loss or resistance at higher temperatures. Since sweeping removes these alkalis, swept crystals will show only small increases in resistance. All of the above tests deal only with the point defects in the crystal. At the present time no test other than direct etching is available to test sweeping effectiveness on the etch pipe defects.

INTRODUCTION

Electrodiffusion or sweeping is a high temperature process which selectively exchanges monovalent ions in alpha quartz. At the present time, sweeping is commercially employed to replace the alkalis present in as-grown quartz with hydrogen. Sweeping affects two types of defects: 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. The replacement of the alkali associated with the aluminum by hydrogen gives rise to the improved radiation hardness of swept oscillator crystals. Sweeping also alters the extended dislocation networks so that the production of etch channels is greatly reduced. The first effect is important in critical aerospace and defense applications while the second directly impacts the production of quartz devices by deep-etch-processing. Verhoogen[1] in his study of ionic diffusion in natural quartz apparently carried out the first deliberate alkali sweeping experiments. King[2] was the first to apply a sweeping process to quartz later used for resonator studies. By carrying out the process in a vacuum, he introduced color centers into the natural quartz bar which caused low temperature acoustic loss peaks in resonators made from the bar. Kats[3], in his extensive infrared study of natural and cultured quartz, demonstrated the sweeping of specific alkalis and protons into and out of quartz. Fraser[4] described the basic process for sweeping specific alkalis and clearly demonstrated that the 53 K acoustic loss peak is caused by the Al-Na center. Kreft[5] showed that holes could be swept into quartz if the process is carried out in vacuum at temperatures above the phase transition. King[6] suggested that vacuum swept quartz should be free of radiation induced transient frequency shift. Recent studies by Lipson[7] and by Antnes, Garcia, and Koehler[8] show that the holes can also be swept in at temperatures below the 573°C phase transition. It is the purpose of this paper to review current sweeping technology and related properties of quartz.

PROPERTIES OF SWEPT QUARTZ

Sweeping affects both point and extended defects. The OH^- growth defects and substitutional aluminum are probably the most important point defects in as-grown quartz. The room temperature broadened OH^- absorption bands are used to determine the optical Q of the material.[9,10] The Al-Na and the radiation-produced Al-hole center have acoustic loss peaks which directly affect oscillator preformance[4,19]. Other point defects present in as-grown quartz include substitutional germanium and both silicon and oxygen vacancies. A recent study by Euler and Kahan[11] suggests that Ge may play a role in low-dose radiation effects. There have been a number of papers reviewing the general properties of point defects in quartz.[12-16] Johnson and Irvine[17] review etch channels and extended defects in another paper in this proceedings.

The replacement of the interstitial alkali trapped adjacent to the substitutional aluminum by hydrogen is the most direct consequence of air or hydrogen sweeping quartz. Fig. 1 compares both the room temperature and the liquid nitrogen temperature IR absorbance spectra of a 5 MHz 5th overtone AT-cut resonator blank in the as-received condition, after a strong (10 Mrad) irradiation, and after sweeping. The low temperature as-received spectrum shows the four growth-defect bands at 3350, 3400, 3437, and 3581 cm^{-1}; the three lower frequency bands are completely broadened in the room temperature curve. Low temperature acoustic loss measurements show that this material contains no Al-Na centers. Irradiation creates Al-hole centers and the $Al\text{-}OH^-$ center from the Al-Li center. The spectrum for the irradiated sample shows that hydrogen in the growth-defects has been completely removed and transferred to the aluminum site forming the $Al\text{-}OH^-$ center which is responsible for the bands at 3367 and 3306cm^{-1}. It is interesting to note that the 3367cm^{-1} is still observable at room temperature. In this particular sample the OH^- growth-defects contain almost exactly enough hydrogen to compensate the 10-15 ppm aluminum present in the sample after irradiation. Sweeping does not cause any changes in the as-grown OH^- bands, but does replace the alkali at the aluminum and form the $Al\text{-}OH^-$ center. The $Al\text{-}OH^-$ concentration is nearly the same for the irradiated condition and after sweeping; thus, very few Al-hole centers were produced when the as-received sample was irradiated. We subsequenlty irradiated the swept sample and observed only a very small decrease in the $Al\text{-}OH^-$ band. A comparison of the production of Al-hole centers by ionizing radiation in swept versus unswept samples taken from the same bar shows that fewer centers are produced in swept material[20]. The production of paramagnetic oxygen vacancy centers is also greatly reduced in swept quartz. The reader should remember that sweeping does not remove point defects but only modifies them. For example, sweeping converts the Al-Li center present in as-grown quartz into $Al\text{-}OH^-$ centers.

The modification of the point defects discussed above gives rise to a number of the improvements in crystal performance attributed to sweeping. Comparison studies by Capone et al.[20] Poll and Ridgway,[21] Young, Koehler, and Adams[22] show that swept quartz usually exhibits lower steady state frequency offsets when exposed to ionizing radiation than unswept quartz. Figure 2 which is taken from Capone et al. shows the frequency offset versus radiation dose for a number of crystals. High Q material contains Al-Na centers and the radiation induced reduction of their loss peak causes the positive shift. The conversion of the Al-Li centers present in the other unswept material into Al-hole centers is probably the cause of their negative offset. On this scale the swept electronic grade material shows nearly zero offset. More recent

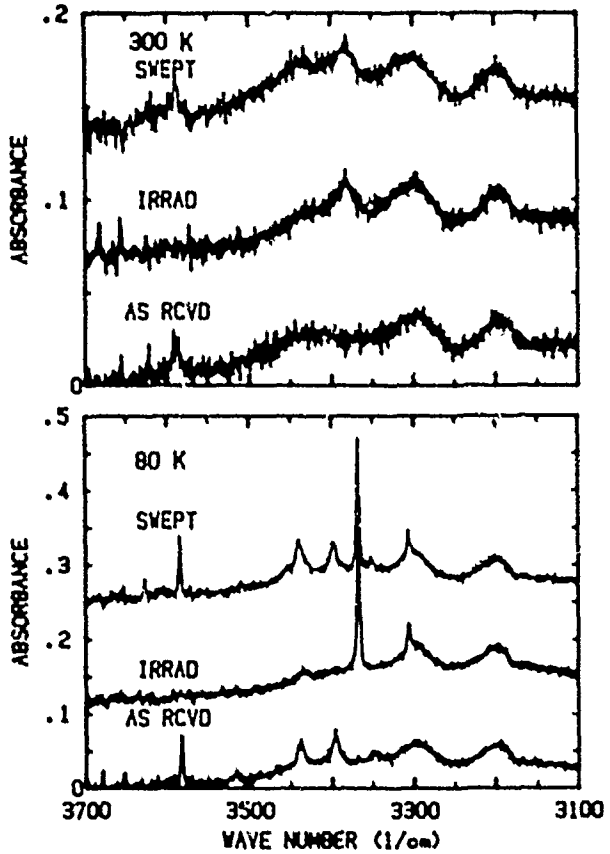


Figure 1. The IR spectra of a 5 MHz 5th overtone AT-cut resonator blank containing 10-15 ppm Al is shown in the as-received condition, after a strong room temperature irradiation, and after sweeping at both room temperature and 80 K. The scale has been expanded for the room temperature curves; the irradiated and swept curves have been offset for clarity.

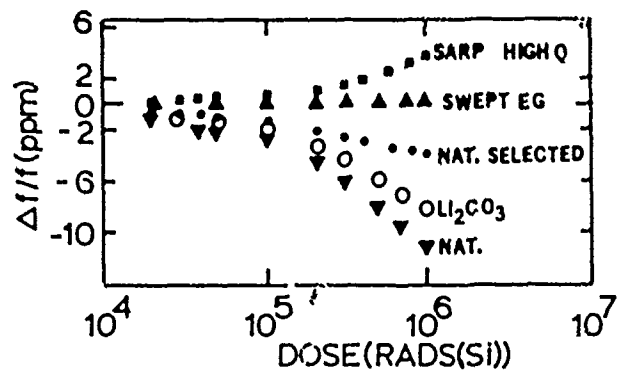


Figure 2. The fractional frequency offsets of a a number of unswept crystals and for a swept electronic grade crystal are show. After Capone et al.[20]

measurements on modern material often show frequency offsets in the 10-100 ppb range for swept material[23,24]. When exposed to pulses of ionizing radiation air swept quartz crystals do not show the transient resistance increases usually seen in unwept material.[25-28] They may, however, show transient frequency offsets because of the temporary production of Al-hole centers. King[6] has proposed that vacuum

swept quartz is free of both transient frequency and resistance offsets. The published frequency offset results for vacuum swept quartz are somewhat contradictory: Pelligrini et al.[24] report offsets off -22 and -7 ppb for vacuum swept Premium Q crystals while Kahan et al.[23] report much larger offsets for vacuum swept material.

Sweeping also improves the room temperture mechanical Q of quartz crystals. Capone et al.[20] report that sweeping raised room temperature the Q of Electronic Grade material from $5x10^5$ to over $1x10^6$. Young, Koehler, and Adams[22] have report that sweeping raised the Q of optical grade quartz from 2 million to 2.6 million. Since the room temperature IR spectrum at 3500cm^{-1} is unchanged by sweeping, there is no comparable increase in optical Q. This mechanical Q improvement comes about because alkalis are removed from the sample. At higher temperatures the improvement is more dramatic; the measurements of Lipson et al.[7] of the crystal resistance of 5MHz 5th overtone AT-cut crystals as a function of temperature is shown in Fig. 3. The resistance of the unswept crystal is over 1000 ohms at 350°C while that of the swept and vacuum swept crystals stays under 100 ohms. This result is important for pressure transducer[29] and other high temperature applications of quartz crystals.

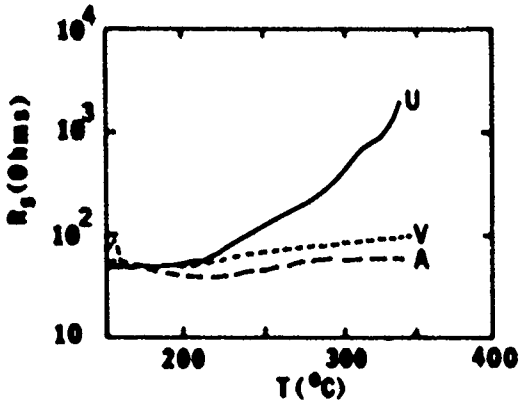


Figure 3. The crystal resistance versus temperature curves for an unswept crystal and for air and vacuum swept crystals are shown. After Lipson et al.[7]

While the properties discussed above all result from the modification of point defects by sweeping, sweeping also alters the extended dislocation networks that cause etch channels. There have been several pulished studies showing that the etch channel density in swept quartz is much less than in comparable bars of unswept material.[30-34] Figure 4 presents the results for several of these studies. The pairs of bars A through G show the data obtained by Balascio and Armington[32] on seven stones of cultured quartz. Half of each stone was left unswept while the other half was swept; then AT slices from each pair were etched. The low unswept etch channel density in bar E came about because it was grown on an X-growth seed. Bars H and I represent the published results of Gaultieri[33] and of Hunt and Smythe.[34] X-ray topographs show that after sweeping the basic dislocation networks remain in the quartz[17]. Hanson[35] has used synchrotron radiation to make X-ray topographs of a number of crystals. His results show a direct correlation between the disiocation networks seen in the topographs and with etch channels which were optically observed after etching. Dislocation networks are high stress areas where impurities are likely to aggregate; it seems possible that sweeping somehow modifies these impurities so that the region is less reactive. Gaultieri and Lukart[36] have observed gold from the cathode electrode diffusing back into the bar along at least some extended defect channels.

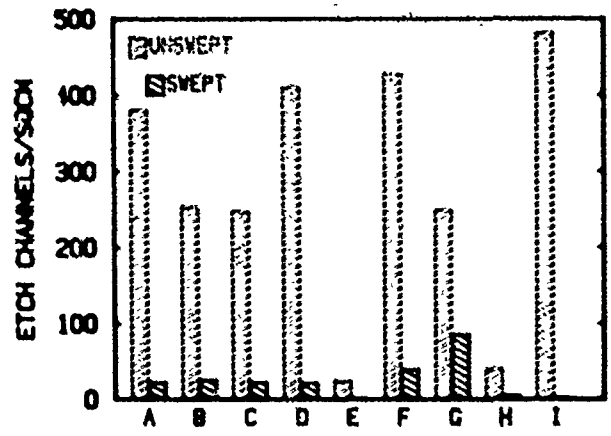


Figure 4. Pair bars A through G represent a comparison of the etch channel densities in unswept and unswept samples taken from the same quartz stones by Balascio and Armington.[32] Bars H and I show the published comparison between unswept and unswept quartz made by Gaultieri[33] and by Hunt and Smythe.[34].

THE ELECTRODIFFUSION PROCESS

As mentioned above, sweeping affects both point and extended defects. The process that is associated with the point defects is qualitatively understood. Most as-grown quartz contains substitutional aluminum which behaves as an acceptor; an associated interstitial alkali, usually lithium, provides the charge compensation[19,20]. Additional unidentified point defects are know to trap protons forming the OH^- related growth defects[3,37] responsible for several IR absorption bands. When thermally liberated from their traps these interstitial ions can drift along the relatively large Z-axis channels present in the alpha-quartz structure. Therefore, if an electric field with a component parallel to the Z-axis is applied to the sample at high temperatures the mobile positive interstitials will be forced toward the cathode (negative electrode) setting up a space charge region near the anode. If a source of positive ions is present at the anode, such as the NaCl film used for sodium sweeping[38], new positive ions will be forced into the sample. At the same time, an equal quantity of the old interstitials leave the sample and move into the negative electrode. Commercially, sweeping is done in air and the original alkalis in the sample are replaced by protons. The protons come from water vapor present in the atmosphere which is probably catalyzed by the metal film positive electrode (anode) with the proton entering the film and then moving into the sample[39]. Figure 5 which is taken from Lipson et al.[7] shows that the replacement proceeds as a wave front moving through the sample. After three hours the Al-OH center has been formed in the first2mm of the 15.2 mm thick sample; as time progresses to 12 hours the "front" moves about half way through the sample. The upper curve in Fig. 5 shows the variation in the as-grown OH^{-1} though the sample and that the growth-defects do not change with sweeping. Lipson et al. and Anthes, Garcia, and Koehler[8] have also shown that a similar "front" of Al-hole centers moves through the sample from the anode to the cathode during vacuum sweeping. Once the ionic replacement process is complete one might assume that only one species of ion is

moving through the quartz sample. Sweeping is a one-dimensional process: the author's students have made several attempts to sweep both alkali and hydrogen in the X and Y directions with no success. The process does work whenever there is a significant component of the applied field parallel to the Z-axis such as in AT, BT, and SC crystal blanks.

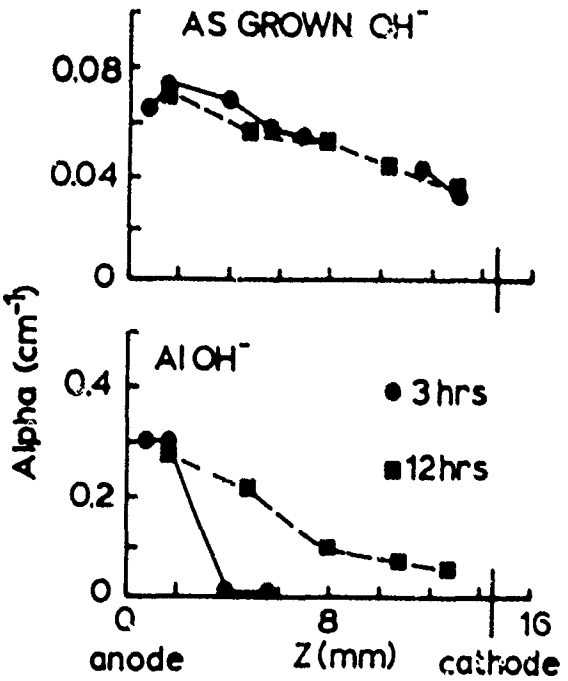


Figure 5. The upper curve shows a nearly 2 to 1 variation in as-grown OH^- through the sample. The lower curve shows the progression protons forming Al-OH^- centers after 3 and 12 hours of air-sweeping. After Lipson et al.[7].

The motion of the ions through the quartz sample is directly related to the ionic conductivity with the number of mobile ionic charge carriers governed by the number of traps and the depth of their trapping wells. If we assume that only one kind of ion is mobile, the conductivity, σ, can be written as $\sigma = ne\mu$ where n is the density of ions, e is the electronic charge, and is μ their mobility. The mobility which is related to the diffusion coefficient is given by

$$\mu = (e/kT)\nu a^2 \exp(-E_m/kT) \quad (1)$$

where ν is the attack frequency, a is the jump distance, and E_m is the activation energy for interstitial migration[40]. If there is only one kind of trap present, aluminum for example, the number of mobile ions per unit volume can be found from the law of mass action to be

$$n = (c/2)^{1/2} N_o \exp(-E_a/2kT) \quad (2)$$

where c is the mole fraction of traps, N_o is the number of SiO_2's per unit volume, and E_a is the association energy between the ion and the trap.[41] The factor of 1/2 in the exponential term comes about because two defects are created: the mobile ion and the fixed ionized acceptor trap. Thus, the conductivity can be expressed as

$$\sigma T = A\exp(-E/kT) \quad (3)$$

where $E = E_m + E_a/2$ and

$$A = \nu a^2 (c/2)^{1/2}/k \quad (4)$$

Equations 3 and 4 show that the ionic conductivity of quartz is thermally activated. The assumption of only one kind of trap is not justified: cultured quartz has the OH^- growth defects (as-grown OH^-) in addition to the substitutional aluminum, as-received natural quartz has both Al-OH^- and Al-Li as well as OH^- related defects which also contain alkalis[3]. Consequently, the conductivity should not be expected to go as the square root of the trap concentration. Nevertheless, the conductivity should increase with increasing defect concentration. Figure 6 shows the apparent DC ionic conductivity measured during cool-down at the end of Li- and H-sweeping runs plotted as log(σT) vs 1000/T for samples with 70(HA-A), 10(PQ-J), and 5(TD-A) ppm aluminum. The lithium data shows some curvature giving activation energies near 1 eV at the higher temperatures while the hydrogen data show activation energies ranging from 1.6 to 2.1 eV[16]. At the lower temperatures the results shown in Fig. 2 are in reasonable agreement with the AC-conductivity results of Green et al.[43] for similar samples. When applied to the sweeping process these conductivity results show that sweeping is thermally activated and that the sweeping current should be larger for the lower purity samples.

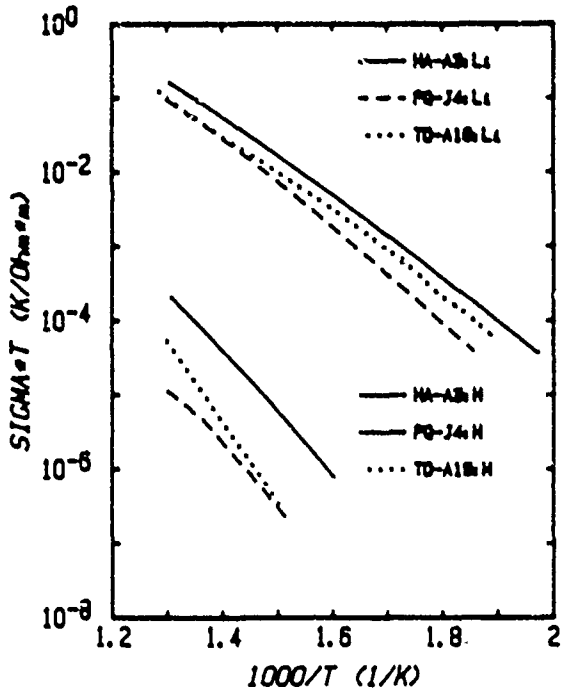


Figure 6. The apparent DC-ionic conductivity taken at the end of Li- and H-sweeping runs is shown for samples with 70(HA-A), 10(PQ-J), and 5(TD-A) ppm Al.

Figure 7 compares the current versus time curves taken during H-sweeping runs on two AT-cut 5 MHz 5th overtone 15mm diameter resonator blanks, one with 10-15 ppm Al and the other with 0.2 ppm Al. Both samples were taken from cultured stones. Both samples were swept with a field of 2000 V/cm applied at room temperature and then taken through the temperature versus time cycle shown as the heavy solid curve in Fig. 7. Both samples show current peaks or plateaus during the warmup portion of the cycle that are similar to those first reported by Hanson[44]. The origin of these peaks is not understood; Hanson, and Lopez, West, and Martin[42] have shown that they disappear when the sample

has been completely swept. When the temperature cycle reaches the 500°C operating temperature the sample current shows an initial fast decay followed by a slowerdecrease until it reaches asteady value. The steady current is usually taken as an indication that the ionic exchange is complete. The currents during the 500°C portion of the cycle roughly scale with the aluminum content. Figure 8 shows the initial current versus time behavior for a 10-15 ppm aluminum blank taken from the same bar with the field first applied when the 500°C sweeping temperature was reached. The log(i) versus log(t) plot shows that the current decreases as $t^{-0.5}$ which is characteristic of one-dimensional diffusion. The large initial current observed when the field is first applied at the full sweeping temperature must be closely related to the current peaks observed during warmup.

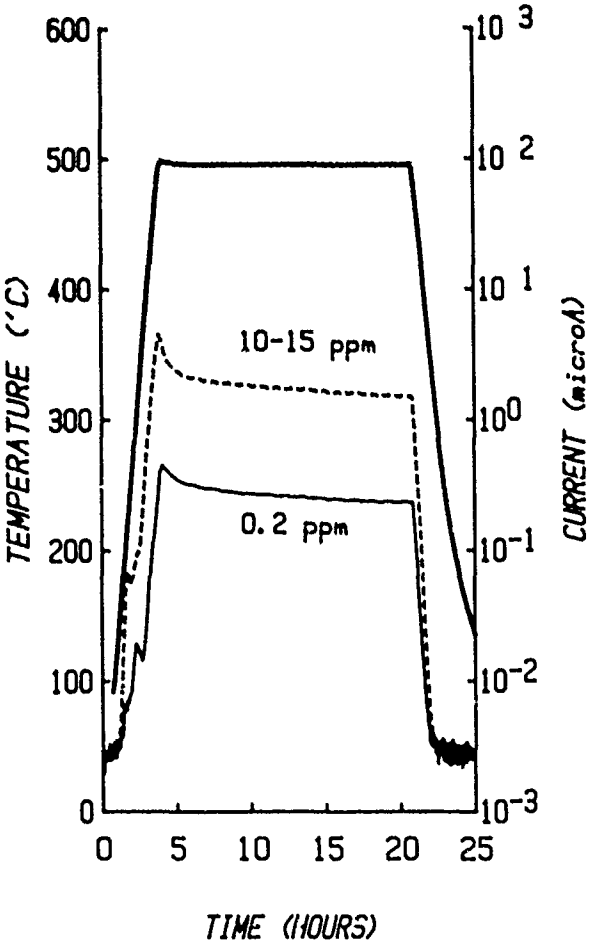


Figure 7. The sweeping current versus time curves for samples with 0.2 and 10-15 ppm Al are shown for the case where the electric field is applied at room temperature and the system is taken through the temperature-time cycle given by the heavy solid curve. The sweeping was done in an H_2 atmosphere.

SWEEPING TECHNOLOGY

Figure 9 shows a block diagram of a hypothetical electrodiffusion system similar to the one described by Brown, O'Connor, and Armington.[45] The system consists of an electric furnace with a programmable temperature controller. To avoid thermally shocking the quartz bar the temperature should be slowly raised to the desired operating temperature and then slowly cooled at the end of the run. The system includes a separate monitor for

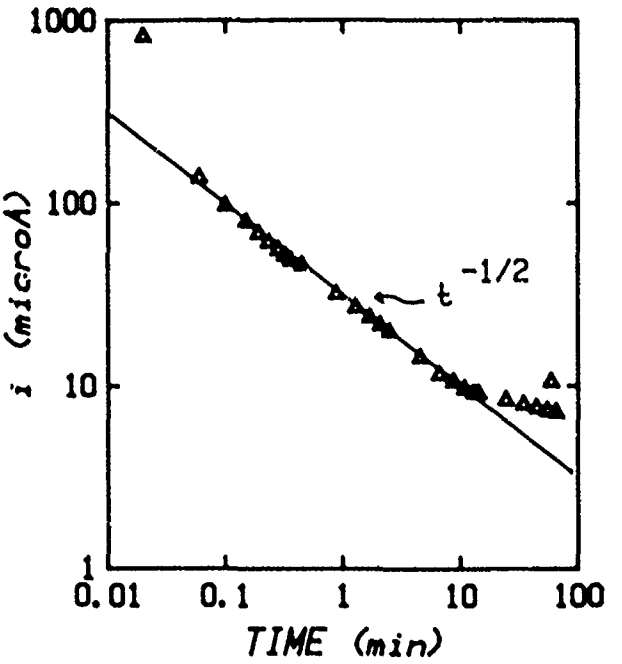
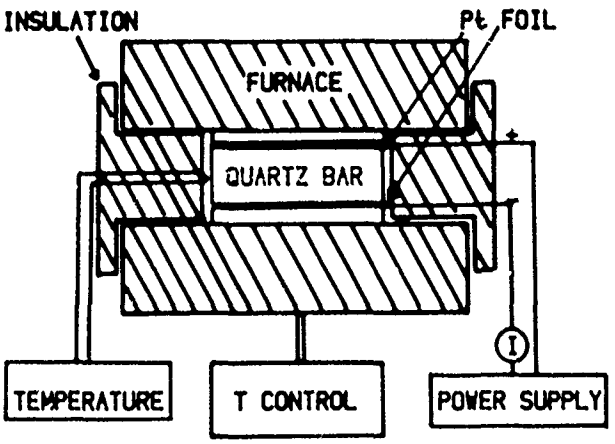


Figure 8. The initial sweeping current versus time for a 10-15 ppm Al sample is shown with the field first applied when the sweeping temperature was reached. The current decays as $t^{-0.5}$.



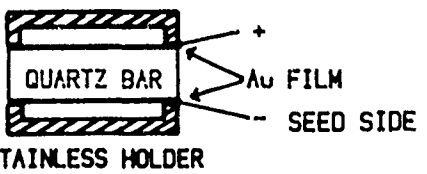


Figure 9. A block diagram for a hypothetical air sweeping system is shown.

an independent temperature check. Air sweeping which exchanges hydrogen for the alkalis requires electric fields of 1000-1500 V/cm; so a 2000 V power supply is used. Fields on 10-20 V/cm are adequate for sweeping alkali ions such as sodium.[31] Larger fields result in larger currents and faster electrodiffusion. Kahan et al.[23] report the use of 1000 V/cm for the electrodiffusion of holes in a vacuum sweeping experiment. Gaultieri[33] has compared with reduction of etch channel density with applied electric fields for fixed sweeping times and found that higher fields give lower density. If is not clear if this result is caused just by the faster sweeping or if there is a direct field effect. A current meter or recorder is needed to monitor the process. With most modern digital current meters it is necessary to place the meter in the negative (ground) lead. Figure 10 shows a block diagram of the computer assisted measurement system that we use at OSU with an extension for the simultaneous sweeping of two samples. The power supply is voltage programmed by the D to A which is controlled by the computer. The sweeping currents are found by measuring reading the voltage drop across the 1000 ohm precision resistors with a HP3478A digital multimeter. The input of the DMM is scanned between the thermocouple and the current measuring resistors by the computer controlled relay box. The relay box also triggers a motor driven independent set point on the temperture controller. For each sweeping run the temperature and current versus time data like that shown in Fig. 7 as well as data on the quartz bar are recorded and stored on disc. The computer assisted system greatly simplifies data analysis; for example, one can readily plot out the apparent DC conductivity versus reciprocal temperature as shown in Fig. 6. Hanson[44] has described a similar system. To prevent back-diffusion, the electric field is usually left on during the cool-down phase of the run.

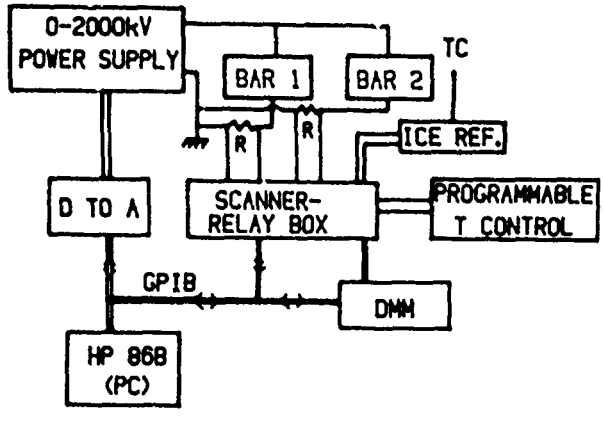


Figure 10. A computer controlled electrodiffusion system is illustrated. Sweeping currents are measured by reading the voltage drop across the resistors.

Sweeping is usually carried out on a fully lumbered bar such as the Y-bars used for AT-cut crystals. The field is established such that the cathode (negative electrode) is the seed side of the bar; thus, the ions are swept towards the dirtier region of the bar. After sweeping the cathode side is ground off. Originally, platinum foil electrodes were pressed against the sample. Since the hydrogen (or the holes in vacuum electrodiffusion) enters the quartz bar through the anode; the positive electrode should have a good intimate contact with the surface. The modern technique is to use evaporated or sputtered metal film electrodes deposited on polished surfaces. In a series of papers, Gaultieri and co-workers[33,36,39] have reported on the effects of surface finish and electrode type on the efficiency of the electrodiffusion process. They found that pressed foil electrodes do not provide sufficiently uniform electrical contact with the sample. They also found evidence of diffusion of the cathode electrode material into extended defects. This back-diffusion was reduced by using well polished surfaces.

The quartz used for electrodiffusion should be selected based upon the final application. Brice[46], in a review paper, discussed the selection of quartz for various applications. He reported that higher quality quartz gave higher yields of finished devices. The same results must hold for swept material. As discussed above, electrodiffusion affects both point and extended defects. The performance of quartz crystal devices for operation in radiation environments or at high temperatures is most strongly influenced by the presence of point defects such as the substitutional aluminum. Therefore, low aluminum, low OH^- content quartz should be chosen for those applications and then swept. The yield of devices produced by deep etching is influenced by the presence of the dislocation networks. Consequently, low dislocation quartz is desired for deep etch processing.

EVALUATION OF SWEPT QUARTZ

A simple direct test for the question, "Is it done?" is needed when carrying out an electrodiffusion run. The most common on-line test seems to be the appearance of a steady current; by this criteria the current versus time traces shown in Fig. 7 show that the sweeeping is not quite complete. Other tests include checking to see if the current is "Ohmic;" such a test was performed periodically during the sweeping run shown by the upper current trace in Fig. 7. Fig. 11 shows current density versus electric field traces taken at four different times after the system reached the 500°C operating temperature. The slope of the curves seems to be converging towards a value slighlty lower than the one shown for 1095 minutes. The current observed for zero field is not understood. Again, the run appears to be not quite finished. Other tests included the appearance of a high activation energy characteristic of hydrogen for the ionic conductivity data that one can obtain on "cool-down."[33] If the sweeping run is repeated with the electric field applied at room temperature, the "warm-up" current peaks will be missing[44,42]. Both samples shown in Fig. 7 passed these two tests.

While no indirect test of the effectiveness of sweeping run for the reduction of the tendency to form etch channels exists a number of tests for point defects are possible. One simple test is to irradiate the bar; if the bar colors then it is assumed that it is not fully swept[29]. However, this test is ineffective on low aluminum content cultured quartz. This material shows essentially no coloration even in the unswept condition. Room temperature infrared absorption measurements such as those shown in the upper part of Fig. 1 may be useful. A cultured quartz AT-cut 5 MHz 5th overtone 15 mm diameter resonator blank containing 10-15 ppm alumninum was used for this data. The Al-OH^- band at 3370 cm^{-1} for the room temperature results is clearly seen for both the swept and irradiated cases. Quartz growers routinely make similar room temperature scans for the evaluation of their as-grown quartz; the Al-OH^- band appears to be as strong as the broadend growth-defect OH^- bands used for the determination of optical Q. It also seems possible to develop a calibration giving the aluminum content of a swept bar. A similar test for an irradiated unswept bar would be unreliable because the amount of Al-OH^- produced depends upon the amount of OH^- growth defects

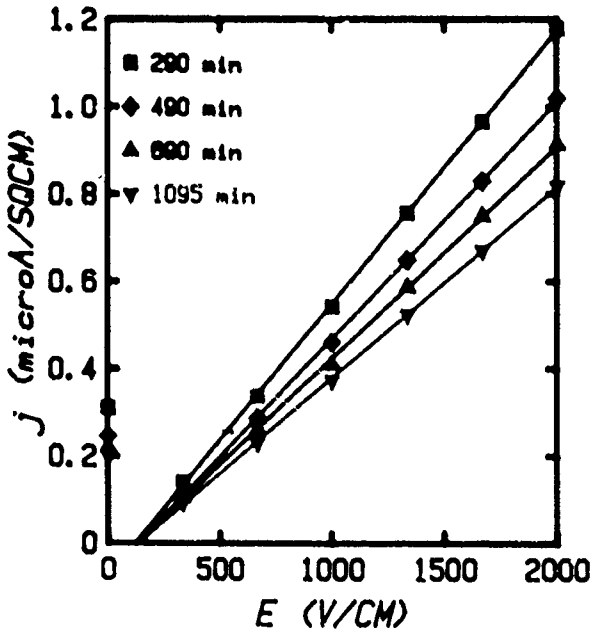


Figure 11. The current density versus applied electric field is shown for four different times after the sample reached the 500°C sweeping temperature. The slopes of the curves are converging to a value slightly lower than the 1095 min. curve. The zero field current is not understood.

present as well as the amount of aluminum. The author believes that such data would not be useful for aluminum concentrations below about 5 ppm. However, IR measurementsat liguid nitrogen temperture as shown in the lower portion of Fig. 1 would be useful down to about 0.5 ppm of aluminum; the Al-OH^- band has shifted to 3367 cm^{-1} and the OH^- growth-defect bands are fully resolved. Lipson et al.[7] have shown that low temperature FTIR spectroscopy can be used to scan probe the homogeniety of the OH^- growth defects and of the Al-OH^- produced either by irradiation or by sweeping. They and Anthes, Garcia, and Koehler[8] have also shown that a He-Ne laser can be used to probe the Al-hole center in vacuum swept samples.

Halliburton et al.[16] have developed a reliable but complicated test for the effectiveness of air or hydrogen sweeping. The test involves direct elecron spin resonance (ESR) measurement of the Al-hole centers produced by irradiation. Suppose that a sample containing concentration C3 of substitutionalaluminum is partially swept so that it contains a concentration C1 < C3 of Al-OH^- centers and C2 of aluminum alkali centers. We then have C1 + C2 = C3. An initial irradiation at liquid nitrogen temperatures will convert the C1 Al-OH^- centers into C1 Al-holes and leave the C2 aluminum alkali centers unchanged. C1 is readily mesured by ESR techniques down to levels of ppb's. If the sample is subsequently irradiated at room temperature the C2 aluminum alkali centers are converted into Al-OH^- centers. A third irradiation, this time at 80 K, then produces C3 Al-hole centers which can again be measured by ESR techniques. The sweeping efficiency is given by:

$$\text{Sweeping Eff.} = (C1/C3)\times 100\%. \qquad (5)$$

Figure 12 shows that concentrations C1 and C3 taken after the initial and final low temperature irradiations on 8 bars of quartz. Bars 1 through 4 were unswept, bars 5 and 6 were 65% and 92% swept, while bars 7 and 8 were fully swept. Since only a ratio (C1/C3) needs to be determined, the ESR test is relatively painless; the aluminum content can also be found by comparing the strength of the ESR signal to that of a standard sample. The ESR test is typically performed on a small sample with X,Y,Z dimensions of 7 mm x 2 mm x 3 mm. The test described above is applicable to air or hydrogen swept quartz.

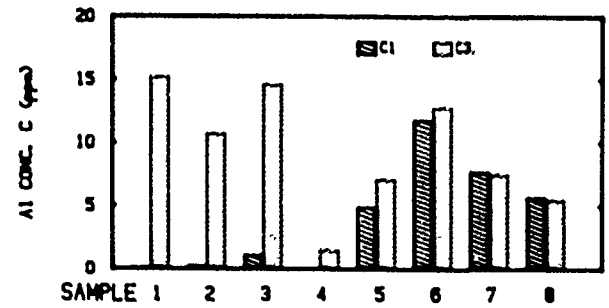


Figure 12. The first bar for each sample gives the Al-hole center concentration C1 as measured by ESR after the initial low temperaure irradiation. The second bar gives the Al-hole concentration C3 as measured after the room temperature and subsequennt low temperature irradiations. The sweeping efficiency is (C1/C3)100%. C3 is also the total aluminum content.

We have measured the Al-hole center concentration in a number of vacuum swept samples in the as-vacuum swept condition and then applied the test above. The samples usually show an initial as-vacuum-swept Al-hole concentration ranging from 3 to 40% of the total aluminum content and that the samples are 100% swept according to the above standard.[47] These results and frequency offset data have led Kahan et al.[23] to propose that the required charge compensation is located at a site some distance away from the substitutional aluminum.

Koehler[48] has developed two tests of sweeping effectiveness. The first is based upon the relatively long lived radiation induced electrical conductivity present in unswept quartz. The second test is a measurement of the crystal resistance at high temperatures. The test can be performed either on finished blanks in a gap holder or on finished crystals. He defines the index ratio as

$$\text{Ratio} = (R_{400} - R_{25})/100 \qquad (6)$$

where R_{400} is the crystal resistance at 400°C and R_{25} is the resistance at 25 °C. Figure 13 shows the results presented by Koehler for tests on a number of crystal units. Those crystals which were well swept showed small resistance changes and transient radiation induced frequency offsets which could be explained in termsof the radiation induced temperature shift. The unswept or poorly swept crystals showed large resistance changes and transient resistance changes when exposed to a radiation pulse. This test should be attractive to crystal manufacturers since the series resistance of is readily measured.

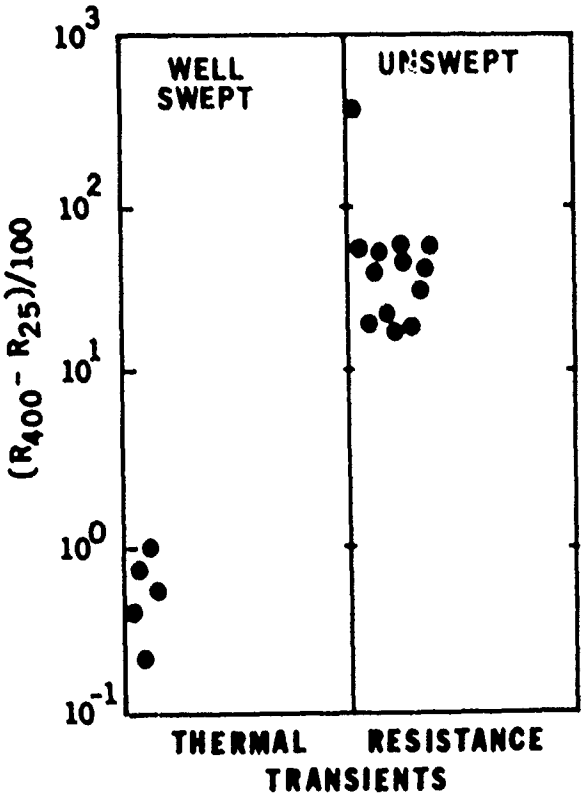


Figure 13. Crystals which have only a thermal transient frequency offset when exposed to a pulse of ionizing radiation show only small increases in series resistance at 400°C as shown on the left. Crystals which are poorly swept show large resistance increases at high temperatures and large transient series resistance when exposed to radiation pulses. After Koehler.[48]

REFERENCES

1. J. Verhoogen, Am. Mineral. 37, 637 (1952).
2. J. C. King, Bell System Tech. J. 38, 573 (1959).
3. A. Kats, Phillips Res. Rpts. 17, 133 (1962).
4. D. B. Fraser, J. Appl. Phys. 35, 2913 (1964).
5. G. Kreft, Rad. Effects. 26, 249 (1975).
6. J. C. King, U. S. Patent 3,932,777 (1976).
7. H. G. Lipson, A. Kahan, R. N. Brown, and F. Euler, Proc. 35th Ann. Freq. Control Symposium, EIA, p322, May 1981.
8. J. P. Anthes, P. Garcia, and D. R. Koehler, IEEE/OSA Topical Meeting on Lasers in Material Diagnostics in Conjunction with the Southwest Optics '87 Conf. Feb. 9-13, 1987, Albuquerque, NM.
9. B. Sawyer, Proc. 42st Ann. Freq. Control Symposium, IEEE, May 1987 (in press).
10. J. F.Balascio and N. C. Lias, Proc. 37th Ann. Freq. Control Proceedings, IEEE, p157, May 1984.
11. F. Euler and A. Kahan, Phys. Rev. B., 35, 4351 (1987).
12. D. B. Fraser, Physical Acoustics, W. P. Mason, ed., (Academic, New York, 1968), Vol.V, Chap. 2.
13. J. A. Weil, Radiat. Eff. 26, 261 (1975).
14. A. Kahan, Proc. Natl. Bur. Stand. Seminar on Time and Frequency: Standards Measurement Useage, Boulder, CO 10.1-10.31 (1977).
15. D. Griscom, Proc. 33rd Ann. Freq. Control Symposium, EIA, p98, May 1979.
16. L. E. Halliburton, M. E. Markes, and J. J. Martin, Proc. 34th Ann. Freq. Control Symposium, EIA, p1, May 1980.
17. G. Johnson and R. Irvine, Proc. 41st Ann. Freq. Control Symposium, IEEE, May 1987 (in press).
19. J. J. Martin, J. Appl. Phys. 56, 2536 (1984).
20. L. E. Halliburton, N. Koumvakalis, M. E. Markes, and J. J. Martin, J. Appl. Phys. 57, 3565 (1981).
20. B. R. Capone, A. Kahan, R. N. Brown, and J. R. Buckmelter, IEEE Trans. Nucl. Sci., NS-17, 217 (1970).
21. R. A. Poll and S. L. Ridgway, IEEE Trans. Nucl. Sci., NS-13, 130 (1966).
22. T. J. Young, D. R. Koehler, and R. A. Adams, Proc. 32nd Ann Freq. Control Symposium, EIA, p34, May 1978.
23. A. Kahan, F. Euler, H. Lipson, C. Chen, and L. E. Halliburton, Proc. 41st Ann. Freq. Control Symposium, IEEE, May 1987 (in press).
24. P. Pelligrini, F. Euler, A. Kahan, T. M. Flanagan, and T. F. Wrobel, IEEE Trans. Nucl. Sci. NS-25, 1267 (1978).
25. J. C. King and H. H. Sander, Radiat. Eff. 26, 203 (1975).
26. J. C. King and H. H. Sander, IEEE Trans. Nucl. Sci. NS20, 117 (1973).
27. J. C. King and H. H. Sander, IEEE Trans. Nucl. Sci. NS-19, 23 (1972).
28. D. R. Koehler and J. J. Martin, J. Appl. Phys. 57, 5205 (1985).
29. G. S. Galtz, Proc. 38th Ann. Symposium on Freq. Control, IEEE, p245 May 1985.
30. J. R. Vig, J. W. LeBus, and R. L. Filler, Proc. 31st Ann. Freq. Control Symposium, NTIS, p113, May 1977.
31. J. J. Martin, R. B. Bossoli, L. E. Halliburton, B. Subramaniam, and J. D. West, Proc. 37th Ann. Freq. Control Symposium, IEEE, p164, June 1983.
32. J. F. Balascio and A. F. Armington, Proc. 40th Ann. Freq. Control Symposium, IEEE, p70, May 1986.
33. J. G. Gaultieri, Proc. 39th Ann. Freq. Control Symposium, IEEE, p247, May 1985.
34. J. R. Hunt and R. C. Smythe, Proc. 39th Ann. Freq. Control Symposium, IEEE, p292, May 1985.
35. W. Hanson, Proc. 42st Ann. Freq. Control Symposium, IEEE, May 1987 (in press).
36. John G. Gaultieri and Donald W. Eckart, Proc. 40th Ann. Freq. Control Symposium, IEEE, p63, May 1986.
37. J. J. Martin and A. F. Armington, J. Crystal Growth, 62, 203 (1983).
38. S. P. Doherty, J. J. Martin, A. F. Armington, and R. N. Brown, J. Appl. Phys. 51, 4164 (1980).
39. J. Gaultieri, 41st Ann. Freq. Control Symposium, IEEE, May 1987 (in press).
40. A. B. Lidiard, Hand. of Physics, 20, 246 (1957).
41. A. D. Franklin, in Point Defects in Solids, J. H. Crawford and L. M. Slifkin, eds. Plenum, NY, Vol 1, Ch. 1 (1972).
42. Augusto R. Lopez, J. D. West, and J. J. Martin, Mat. Res. Soc. Symp. Proc. Vol. 60, p451 (1986).
43. E. R. Green, J. Toulouse, J. Wacks, and A. S. Nowick, Proc. 38th Ann. Freq. Control Symposium, IEEE, p32, May 1984.
44. W. P. Hanson, Proc. 38th Ann. Freq. Control Symposium, IEEE, p39, May 1984.
45. R. N. Brown, J. J. O'Connor, and A. F. Armington, Mat. Res. Bull. 15, 1063 (1980).
46. J. C. Brice, Rev. Mod. Phys. 57, 105 (1985).
47. L. E. Halliburton, Private Communication, May 1987.
48. D. R. Koehler, Proc. 35th Ann. Freq. Control Symposium, EIA, p.322, May 1981.

ACKNOWLEDGEMENTS

The author wishes to thank the following for a number of helpful discussion on sweeping and the general properties of quartz: A. F. Armington, J. Balascio, F. Euler, J. Gaultieri, L. E. Halliburton, W. Hansen, A. Kahan, D. R. Koehler, J. C. King, H. Lipson and A. S. Nowick. The work at OSU was supported by the USAF.

41st Annual Frequency Control Symposium - 1987

ETCH-CHANNELS IN SINGLE CRYSTAL CULTURED QUARTZ

Gary R. Johnson and Robert A. Irvine

Sawyer Research Products, Inc.
35400 Lakeland Blvd.
Eastlake, Ohio 44094

Abstract

Previous work to identify, characterize and reduce etch channels in single crystal cultured quartz is reviewed. Experiments in this study confirm that etch channels result from selective etching near edge dislocations. The dominant sources of the dislocations in the grown crystal are determined to be latent seed dislocations and the seed crystal interface. Using x-ray topography, optical and scanning electron microscopy the composition of the interface is seen to consist of several solid phase inclusion species coexisting with bands of dislocations identified as misfit dislocations by Burgers vector invisibility criteria. It is proposed that the inclusions nucleate from a reaction of elements present in the system.

Introduction

Device Considerations

Applied as a frequency control or timing device, components fabricated from single crystal cultured quartz are widely used in military, industrial and consumer electronic equipment. New designs continually offer higher performance, miniaturization and lower costs. To answer the need for device improvements, quartz component manufacturers implemented processes requiring the exposure of quartz to strong etchants. In some cases, this was an effort to adapt the batch photolithographic techniques of the semiconductor industry to quartz resonator production, offering lower costs and realization of innovative designs. In other instances, the design makers were trying to reduce the subsurface damage caused by previous lapping operations to improve performance. On still other occasions, the fabricators were trying to achieve very thin sections for high frequency operation. Each of these efforts was constrained by the formation of micron sized holes (often referred to as etch channels) in cultured substrates on exposure to etchants.

Figure 1 is typical of the surface condition following prolonged etching in ammonium bifluoride (ABF).



FIGURE 1

Note the appearance of pits with well defined inverted apices which sometimes serve as the starting point for long narrow channels. Higher magnification by scanning electron microscopy (SEM) is shown in Figures 2 and 3.



FIGURE 2



FIGURE 3

Etch channels limit device makers by:

1. Creating electrical faults upon application of conductive electrodes.

2. Acting as stress concentrators to reduce the ultimate strength of the device.

3. Trapping contamination resulting in increased aging.

The first widespread application of photolithographic processing techniques occurred in the manufacture of tuning fork resonators. An example of a device with size about 1 mm x 5 mm is shown in Figure 4.



FIGURE 4

Thicksten (1983) catalogued various etch related defects encountered in fabricating tuning forks. Figure 5 illustrates the presence of etch pits and channels in photolithographic processed parts.



FIGURE 5

In some cases, electrical discontinuities are caused by voids in the metalization attributable to etch defects. Additionally, the shock resistance of the device is severely impaired by the presence of stress concentrators. As devices undergo further miniaturization and more complex shapes are required, the need to reduce the concentrations of pits and channels increases.

Defect Characterization and Origin

The well known method of etching to detect the presence of dislocations in metals and crystals was applied to cultured quartz early in its development and remains a standard characterization technique. The formation of etch pits in the presence of dislocations is predicted by free energy considerations. Cabrera, Levine and Plaskett (1954) describe the conditions necessary to open a dislocation into an etch pit when the crystal is placed in an undersaturated medium. The basic consideration is an energy balance between the dissolution of a volume of crystal into the undersaturated medium and the release of potential energy stored as strain near the dislocation balanced against formation of additional surface area free energy. Two of these terms are nearly the same throughout the bulk of the crystal. However, the increased strain energy near the dislocation results in much higher etch rates in those areas.

The qualitative similarity between patterns formed by Lang topographs and etching is readily seen in Figure 6.



A) X-RAY TOPOGRAPH B) PHOTOGRAPH AFTER ABF ETCHING
FIGURE 6

However, in quartz it was noticed that in addition to the etch pits which existed in other materials, small areas, beginning at the apices of some of the pits, etched much faster than the surface, creating channels extending well below the surface. Arnold (1957) observed that very large (over 250 microns) inclusions related to the formation of etch channels. At about the same time Augustine (1957) and Augustine, Hale and Berry (1957) were conducting similar etching experiments and reporting similar observations.

The composition of channels in natural and cultured quartz was investigated in Buerger (1954) and Cohen (1960). The possibility of stuffing quartz with beta-eucryptite and beta- spodumene to explain the presence of large quantities of certain impurities was mentioned.

Nielsen and Foster (1960) conducted further etching experiments with similar results. They noted the presence of fluid inclusions spanning the thickness [0003] of the seed plate and a one-to-one correspondence between these inclusions and etch channels in the grown crystal. They further discussed possible causes of these channels, mentioning dislocations, nonuniform incorporation of impurities and a combination of these two possibilities. They ultimately decided that they could come to no conclusion.

Further important etching studies were carried out by Tsinzerling and Mieonua (1963) and Hanyu (1964). This work concluded that the faces of the surface pits matched naturally occurring faces of quartz and attributed the pits to surface intercepts of dislocations.

Spencer and Haruta (1966) applied the transmission x-ray diffraction topographic methods developed by Lang to single crystal cultured quartz. This study established that these linear defects had some characteristics of dislocations and if they were dislocations, they were most likely screw or mixed types.

Line drawings illustrating the general idea of screw and edge dislocations are shown in Figures 7 and 8.

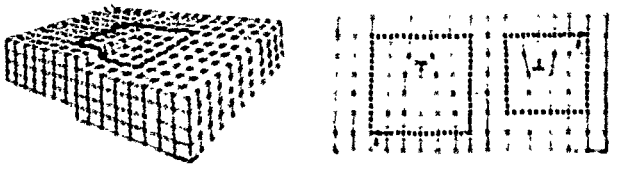


FIGURE 7 SCREW DISLOCATION FIGURE 8 EDGE DISLOCATION

On a unit cell scale, the screw dislocation appears as an offset creating a spiral staircase effect. An edge dislocation is visualized as an extra plane of atoms. The Burgers vector is labeled b.

If g is the reciprocal lattice vector of a set of planes and b is the Burgers vector of the dislocation, the contrast effect of a dislocation image will disappear when $g \cdot b = 0$. Conversely, the contrast effect will maximize when $g \cdot b = 1$. This implies that the image at dislocation lines exhibits the maximum when the reflecting planes are oriented perpendicular to the Burgers vector and vanish when they are parallel to b. In the case of a pure edge dislocation, b and the dislocation line both lie in the same plane, the axis of the edge is obtained from the criteria that $g \times L = 0$, where L is the direction of the dislocation line.

An extensive study of dislocations by Lang and Miuscov (1967) confirmed the link between etch pits and dislocations. Specifically, they concluded that with rare exception, pits formed at dislocations and that no pits occurred not at dislocations. They further decided that based on contrast effects from various planes that only about 15% had C axis component [0003] associated with screw dislocations. Presumably the remaining 85% exhibited strong edge dislocation characteristics.

Lang and Miuscov further observe that dislocations are randomly oriented early in the growth cycle, but that dislocations soon organize into a polygonal cell boundary network (Figure 6 illustrates), the final stage of which appears as growth hillocks on the rough final surface with orientation approximately (0001). Fault surfaces were observed in the topographs which were interpreted as indicative of impurity segregation in the cell walls.

Further confirmation of the relationship between etch pits, etch channels and dislocations was presented by McLaren, Osborne and Saunders (1971). Once again it was concluded that most of the dislocations are of the edge variety, some mixed and rarely finding a pure screw dislocation. As in other studies, the dislocations were confined to a narrow (10°-25°) cone, centered about [0003]. The topographs produced illustrated dislocations originating in the seed or from the seed surface, propagating in the general direction of growth. A suggested model for a pure edge dislocation is presented in Figure 9.

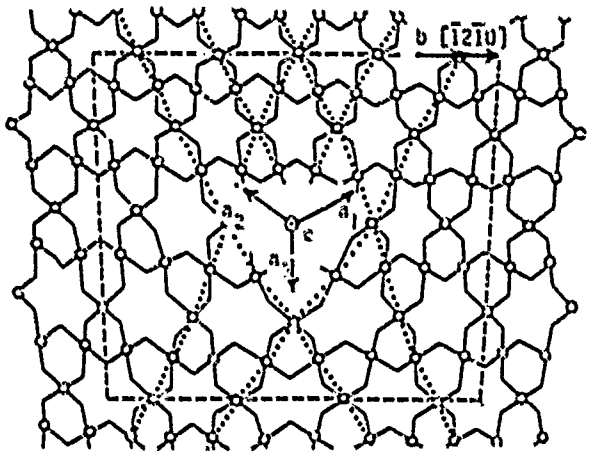


FIGURE 9

An important feature of this model is a hollow core with a series of unbonded oxygens. It was hypothesized that this was an obvious place for the gathering of impurities, especially hydrogen and substitutional aluminum (Al^{+3}).

Iwasaki (1977) draws the same association between edge dislocations, etch pits and etch tunnels as previous researchers. Another structural model is proposed with the key feature of unbonded oxygen surrounding a hollow core as presented by McLaren et al. (1971). However, the problem of transporting etchant to the reaction interface down the micron sized channel is discussed. The mechanism posed is to assume vapor etching by the outgassing of an acid with higher mobility along the channel than in the case of liquid.

Barns et al. (1978) provided additional confirmation that etch channels and pits are related to dislocations. They also observed certain shallow flat bottom pits which they associated with flaws created during grinding or lapping. However, the majority of their report concerning pits described sharply defined apices and etch channels; the same as mentioned by earlier observers. This research failed to uniquely determine the Burgers vector, thus leaving open the question of the character of the dislocation.

Etch channels were separated into two categories:

1. Those propagating from fluid inclusions in the seed and

2. Those emanating from particulate inclusions near the seed.

Homma and Iwata (1973) provided another link between observations with etching, Lang topography and polygonal cell structure. They also applied ion probe methods to samples taken within tens of microns of the as-grown Z-surface and detected a large concentration of aluminum and iron in a 10 micron area of high dislocation density between cell walls. The boundary of high dislocation density observed by Lang and Miuscov (1967) was not seen in this work.

Saha, Annamalai and Bandyopadhyay (1979) conclude that about 80 to 85% of the dislocations in cultured quartz are pure edge dislocations constrained to a cone of about 25° around the C axis. These dislocations congregate along cell wall boundaries which are also a region of high impurity segregation. Additionally, the remainder of the dislocations are predominantly of screw character, each of them associated with a single terraced cobble apex.

The various relationships shown by the cumulative work is briefly illustrated in Figure 10.

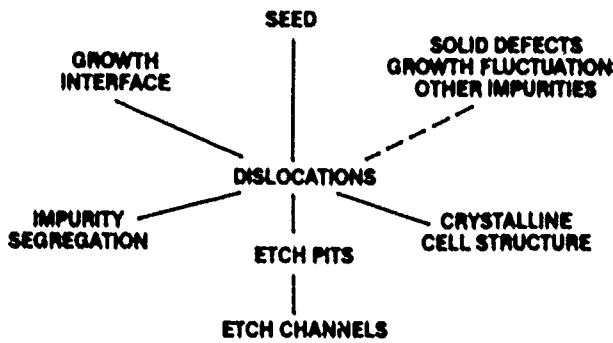


FIGURE 10

Efforts to Reduce Dislocations

The main thrust of the Barns et al. (1978) work was to develop a reproducible technique for the growth of low dislocation density cultured quartz. Based on their analysis of inclusions, showing that iron was a critical component in inclusions which created dislocations, combined with knowledge that dislocations present in the seed would propagate, they found their best results using low dislocation seeds in a noble metal liner to isolate the growth of the crystals from steel autoclaves. Standard Brazilian nutrient was employed as were typical growing conditions. Analysis indicated that all dislocations formed etch pits. Though no analysis was reported, they indicated the dislocations decorated with iron formed channels.

They reported low dislocation material and concluded that it was possible to create dislocation free material under normal growth conditions in noble metal lined vessels if nearly perfect seeds are used. They further stated that across a broad growth rate variation there was no relationship to dislocation density. Crystals grown on (01$\bar{1}$1) seed plates were found more prone to dislocations than those grown on (0001) seed plates in this study.

The problem of obtaining suitable seeds was addressed by Zarka, Lin and Buisson (1981). Previous research consistently illustrated propagation of dislocations nearly parallel to the direction of growth. Zarka et al. chose seeds from the greater X-region with growth direction [2$\bar{1}$10] to minimize dislocations in the

[0003] direction of intended growth. This lowered the dislocations propagating from the seed, although the topographs show that considerable dislocations nucleated from inclusions in close proximity to the seed.

In a series of reports prepared by Armington et al. (1981, 1982, 1984, 1985, and 1986) the topics of lowering impurities, especially aluminum, and reduction in dislocation density are simultaneously addressed. The summary of the result of this activity is shown below in Figure 11.

Experimental Variable	Effect on Dislocations
Growth Environment	none
Noble liner	
Unlined	
Mineralizer	
Seed Source	strong
Low defect natural	
Greater X cultured	
"Normal" cultured	
Low strain	
Supply Source	none*
High purity cultured	
High purity glass	
High purity sand	
Seed Preparation	none
Acid etch	
Hydrothermal etch	
Swept	
Growth Rate	none*

*Showed strong effect on aluminum uptake.

FIGURE 11

In contrast to prior investigations, the 1985 report indicated that most dislocations originated within seeds supplemented by a small number nucleating at inclusions. The 1986 report indicated that typical etch channel densities of under 50 cm^{-2} were possible using greater X seeds and pure Z cultured quartz nutrient. Suitable results were realized across a variety of vessel surface materials, mineralizers, growth rates, and seed preparation methods.

Croxall et al. (1982) created quartz with dislocation density of under 10 cm^{-2}. The method was based on employing a noble metal vessel surface, high purity nutrient and growth solution, low dislocation seeds, and standard growing conditions appropriate to sodium hydroxide. They reported the appearance of dislocations in the seed crystal interface, but reported no dislocation generation at inclusions.

Post Culturing Processing

Although the complete elimination of dislocations is desirable, most devices can tolerate the presence of dislocation densities of 10^3 cm^{-2} or greater and many photolithographic uses are feasible with densities of 10^2 cm^{-2}. However, in the case of certain devices with relatively large dimensions and a low tolerance for dislocations (especially etch channels), processes designed to reduce the formation of etch pits and channels in the presence of dislocations have been developed. Electrolytic sweeping as a method to inhibit etch channel formation has been reported by Vig et al. (1977), Martin et al. (1983), Gualtieri (1985), and Armington and Balascio (1985). Large reductions are consistently reported with Balascio and Armington (1986) reporting a decrease in mated half bars from 328 cm^{-2} to 36 cm^{-2}. They further reported an incidence of a reduction from 24 cm^{-2} to 1 cm^{-2} in crystal grown on a greater X seed.

Another technique reported by Bernot (1985) concerns the addition of various fluoride salts to the normally used acid etchants. This work showed that the addition of KF along with unnamed secondary additives to either HF acid or ABF greatly reduced the occurance of etch pits and etch channels. Both of these results were reported using unswept commercially available quartz.

The motivation for developing these further steps was the desire to fabricate a precise altimeter requiring an etch channel free active area and a chemically polished surface. This combination is a severe requirement for presently available material and requires swept high quality material and special etchant to achieve successful fabrication. This device is shown in Figure 12 (patent 4,479,070).

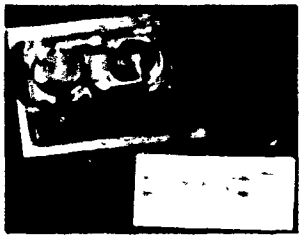


FIGURE 12

Experimental

With the long term objective to make available commercially produced quartz crystal with fewer than 10 etch channels cm^{-2} and 50 etch pits cm^{-2}, a reliable and convenient dislocation characterization procedure was designed. Other experimental activities involved identifying various dislocation types by Lang topography and exploring the relationship between dislocations and inclusions.

Characterization

The first step of the investigation was to develop a standard technique for characterizing dislocation defects. The two available methods were Lang topography and etching. The etch method was chosen due to its relatively low cost and convenient availability. To be effective, the characterization procedure had to be consistent, repeatible, representative and transferrable. After investigation, the following procedure was developed:

AT sample orientation.

Lapped surface.

ABF solution saturated at 75°.

Solution agitated. Duration approximately 2.5 hours.

0.1 mm minimum material removal.

To obtain valid results, it is especially important to assure adequate stock removal. Figure 13 illustrates.

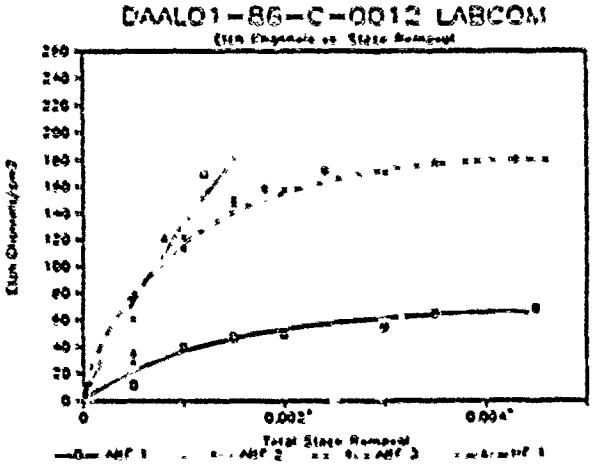


FIGURE 13

Bulk Inclusion Study

This effort was to investigate the relationship between etch channels and inclusions in the bulk of the crystal. An historical data base was established from 66 production runs of a standard crystal type grown to an average Z dimension of 26 mm on a Z plate seed with X 37 mm at a growth rate of about 0.35 mm day^{-1} under standard conditions appropriate for the sodium carbonate process. The seed population was from normal production, indicating a large variance in dislocation density was expected. The process was modified through two series of changes. The first involved decreasing dissolved and particulate impurities in the growth solution. The second stage added further particulate filtration.

The effect on etch channel density is shown in Figure 14.

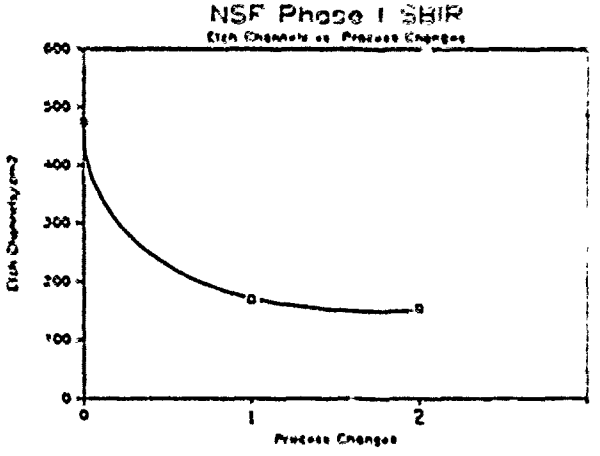


FIGURE 14

The inclusion density showed no significant changes throughout the same variations.

From beginning to end the etch channel density decreased from mean 474 (standard deviation 316) to mean 153 (standard deviation 164). Although not quantified, it is worth noting that an important part of this variation is the result of test variations. This is especially true in the case of high densities, where accurate counting is difficult. Using an anlysis of variance technique described in Box, Hunter and Hunter (1978), no significance was discovered between the initial and final process regarding inclusions. However, significance was detected regarding etch channels. Figure 15, a scatter diagram between etch channel and inclusion densities, illustrates the data and visually confirmed a lack of correlation between a concentration of the various species of bulk inclusions and etch channels.

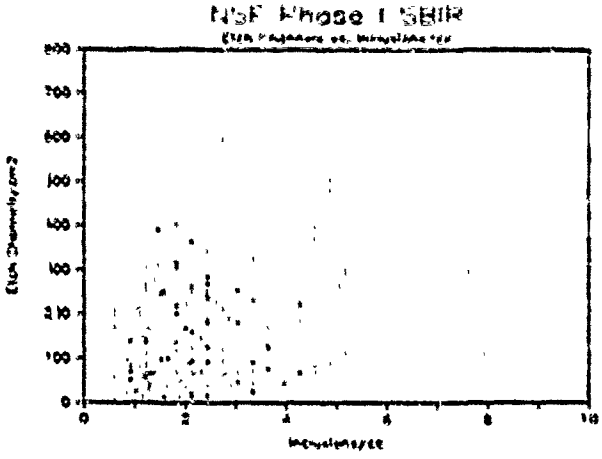


FIGURE 15

Tentatively, the reduction in etch channels across the processes has been attributed to either a decrease in fine particles directly deposited on the seed or serving as nucleation sites for plates, boxes, etc. which form a debris plane or to direct improvements in the chemical purity of the solution.

Identify and Catlogue Inclusion Species

Since little effect was attributable to bulk inclusions, a microscopic study of inclusions was initiated, with special emphasis on the seed crystal interface.

Starting at the centerline of the seed (between the Z faces) and working outward in the direction of growth, Figure 16 summarizes the results.

SPECIES	LOCATION	SIZE (MM)	FREQUENCY
BUBBLE	SEED		
PLATE	BOUNDARY	[illegible]	HIGH
BOX	BOUNDARY	[illegible]	HIGH
DIRT	BOUNDARY-BULK	[illegible]	HIGH
PIN CUSHION	BOUNDARY-BULK	[illegible]	LOW
CLUMP	BULK	[illegible]	<10 cm^{-2}
PATCH	BULK	[illegible]	LOW
LOG JAM	BULK	[illegible]	LOW

FIGURE 16

Figures 17 through 28 illustrate the various inclusion species.



FIGURE 17 BUBBLE



FIGURE 18 PLATE



FIGURE 19 PLATE SEM



FIGURE 20 BOX

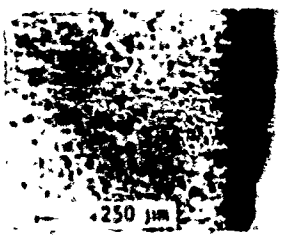


FIGURE 21 DIRT

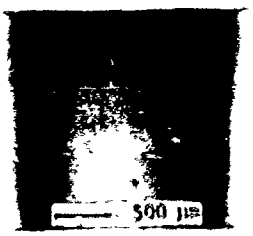


FIGURE 22 SEED VEIL

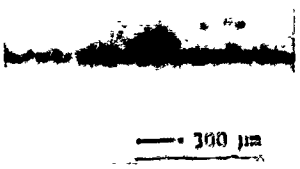


FIGURE 23 PIN CUSHION



FIGURE 24 P.C. SEM

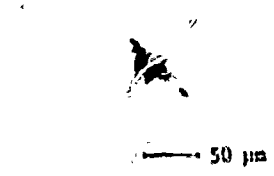


FIGURE 25 CLUMP



FIGURE 26 CLUMP SEM

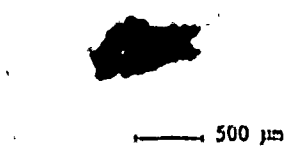


FIGURE 27 PATCH



FIGURE 28 LOG JAM

The scale which forms on the interior surface of the steel autoclave wall was examined. The results of an SEM photograph are shown in Figure 29.



FIGURE 29

This has long been suspected as a source of inclusions in cultured quartz. However, as a result of this investigation, it is concluded that based on dissimilar morphologies that wall scale is not an important source of inclusions in cultured quartz. Tentatively, it is proposed that inclusions in cultured quartz largely result from a reaction of elements present in the growing solution.

Identify and Catalogue Dislocation Species

Lattice defects in Y and Z cut cultured quartz crystals were investigated by x-ray topography. Burgers vector was determined and structural inferences were made by examination of the contrast effect viewed in topographic images reflected from a variety of planes.

Topographs on a Y cut crystal grown at an average growth rate of about 0.4 mm day^{-1} under standard commercial growth conditions for the sodium carbonate process were prepared. The results, using several values of g, are shown in Figures 30 to 36.



FIGURE 30 $[\bar{1}2\bar{1}0]$



FIGURE 31 $[01\bar{1}0]$



FIGURE 32 [0003]



FIGURE 33 $[\bar{1}011]$



FIGURE 34 $[10\bar{1}1]$



FIGURE 35 $[10\bar{1}2]$



FIGURE 36 [10$\bar{1}$0]

Examining the figures, it is seen that the maximum contrast was obtained for [$1\bar{2}10$] and the minimum for [$10\bar{1}0$]. Consequently b of the dislocation was [$1\bar{2}10$]. Carrying out χxl shows that the direction is [0003], parallel to the C-axis. Inspection of the topographs shows that most of the dislocation lines follow this course within a cone of about 10°. The topographs further reveal that the dislocation lines either begin in the seed or at the seed crystal interface. The interfaces were consistently characterized by dark lines parallel to the X axis [$10\bar{1}0$]. The dislocation density at the interfaces was high and individual resolution was not possible. The nature of the dislocation structure emerged, however, when the reflections approached the [$10\bar{1}0$] direction which minimized the contrast of most of the edge dislocations. Figures 14 and 15 exhibit strong contrasts, indicating a dense dislocation structure which remains in high contrast at all reflections without definite Burgers vector. Presumably misfit dislocations bridge any lattice misorientation or mismatch between the seed and crystal. Consequently, the direction of Burgers vector varies depending on local conditions. It is likely this is a highly strained region, potentially decorated by impurity atoms.

The dislocation cell structure discussed by Lang and Miuscov (1967) was also observed in this study. The observed diffraction contrast effects associated with the wall structure suggested that the cell walls are depository sites for impurities.

Another observation of this study, indicating a strong possibility of the impurity segregation at the cell walls, concerns the presence of fault surfaces. Figure 37 is a Z cut with y = [$10\bar{1}2$].



FIGURE 37 [$10\bar{1}2$]

These diffraction effects reveal large local strain accumulation. The combination of internal surfaces and local strain intensification are factors conducive to impurity segregation.

Discussion

Completed research indicates that substantial improvements are possible in reducing dislocations in cultured quartz. Although hydrogen content, growth rate, growth conditions, presence of point defects, etc. must contribute to the formation and propagation of edge dislocations resulting in etch pits and channels during etching, present evidence suggests that these are not the most critical issues. The two most important factors in producing commercial quantities of low dislocation material are the availability of low dislocation seeds and an improved understanding leading to eventual control of the seed crystal interface, especially the effect of inclusions in that area. From work here, it is not possible to make a statement regarding whether inclusions are a cause of or an effect of a highly strained, misfit dislocation boundary. Based on Croxall et al. (1982), it seems possible to grow low dislocation quartz on a natural seed without the presence of a misfit dislocation boundary. This points to labeling the inclusions as the cause and dislocations the effect.

In the event this is correct, it becomes necessary to bring the formation of these inclusions under improved control. Past efforts have focused on overall process purity, including growth in noble metal lined vessels. The complex chemistry of the system, coupled with the wide range of temperature and pressures through which the system passes, indicates that a large variety of compounds could nucleate and remain stable.

The composition of etch channels remains unidentified. Following the course of Buerger (1954) and Conen (1980), the direction of the investigation into the etch channel contents is concentrating on naturally occurring minerals composed of elements present in the culturing system having structural compatibility with a strained quartz lattice.

The cause of channel formation in some edge dislocations and not others is unknown. Available models imply that impurities are required to compensate the partially unbonded oxygen surrounding the core of the dislocation. Possible explanations of the difference in etch behavior are: the possibility some closely packed bundles exceed an undetermined critical size and differences in the impurities incorporated or the structure of the incorporation. The mechanism is not identified. However, based on the tendency of edge dislocations with similar signs to repel, the first possibility seems unlikely. Further, the large reduction in etch channel density following electrolytic sweeping could be explained by changes in the composition scheme of the dislocation core, tending to support the second possibility.

Conclusions

1. Etch channels and pits are found only in association with dislocations.

2. Most dislocations in the Z-region of cultured quartz are of the edge type directed approximately parallel to the direction of growth [0003] with Burgers vector [$1\bar{2}10$].

3. The dominant sources of dislocations in cultured quartz are the seed and seed crystal interface.

4. The seed crystal interface is composed of a variety of inclusion species and exists simultaneously with bands of misfit dislocations which by an unknown mechanism appear to increase the density of edge dislocations in the growing crystal. Additional observations concerning the interface surface are:

   A. Bubbles in the seed are a source of dislocations.

B. A combination of plates, dirt and boxes forms a debris layer very early in the growth. Combined with the bubbles this is referred to as seed veil. As the density of this layer increases additional dislocations seem to form.

C. The debris plane coexists with a band of misfit dislocations. The cause and effect relationship between misfit dislocations and debris is not determined.

5. The dominant source of solid phase inclusions in cultured quartz is through reaction of the constituents of the system.

Acknowledgements

This work was partially supported by:

National Science Foundation through Small Business Innovation Research Program, Phase I, (Grant Number ECS-8460755) and Phase II, (Grant Number ISI-8521276).

Joint funding of the Department of Defense and Lawrence Livermore National Laboratory, administered by LABCOM, contract DAALO1-86-C-0012.

The authors also acknowledge the efforts of Mr. Kelley E. Scott of Sawyer Crystal Systems in Conroe, Texas in preparing the photographs in Figures 16, 17, 19, 20, 21, 22, 24, 26, and 29, and for his many observations and discussions leading to a better understanding of the seed veil.

We acknowledge further:

Dr. Sigmund Weissmann of Rutgers University for preparation and analysis of the x-ray topographs shown in Figures 29-36.

Dr. Evelyn Willibald-Riha of Siemens, GmbH, Research Center in Munich, West Germany for preparing the x-ray topograph in Figure 6.

Honeywell, Sperry Flight Systems Division, for permission to use Figure 12.

The authors wish to thank the Program Committee of the 41st Frequency Control Symposium, especially Dr. L.E. Halliburton of Oklahoma State University, for extending an invitation to present this paper.

References

Arnold, G.W., Proceedings of the 11th Frequency Control Symposium, 1957, p. 112.

Augustine, F., Hale, D.R., and Berry, J., 5th Progress Report, U.S. Army Contract DA33-039-SC-72415, p. 15.

Armington, A.F., Larkin, J.J., O'Connor, J.J. and Horrigan, J.A., Proceedings of the 35th Frequency Control Symposium, 1981, p. 297.

Armington, A.F., Larkin, J.J., O'Connor, J.J. and Horrigan, J.A., Proceedings of the 36th Frequency Control Symposium, 1982, p. 55.

Armington, A.F., and Balascio, J.F., Proceedings of the 38th Frequency Control Symposium, 1984, p. 3.

Armington, A.F., and Balascio, J.R., Proceedings of the 39th Frequency Control Symposium, 1985, p. 230.

Balascio, J.F., and Armington, A.F., Proceedings of the 40th Frequency Control Symposium, 1986, p. 70.

Barns, R.L., Freeland, P.E., Kolb, E.D., Laudise, R.A., and Patel, J.R., Journal Crystal Growth, 43, 1978, p. 676.

Bernot, A.J., Proceedings of the 39th Frequency Control Symposium, 1985, p. 271.

Box, G.E.P., Hunter, W.G. and Hunter, J.S., Statistics for Experimenters, 1978, pp. 165-207.

Buerger, M.J., Journal American Minerals, 39, 1954, p. 600.

Cabrera, N., Levine, M.M. and Plaskett, J.S., Phys. Rev. 96, 1954, p. 1133.

Cohen, A.J., Journal Phys. Chem. Solids, 13, 1960, p. 321.

Croxall, D.F., Cristie, I.R.A., Holt, J.M., Isherwood, B.J., and Todd, A.G., Proceedings of the 36th Frequency Control Symposium, 1982, p. 62.

Gualtieri, J.G., Proceedings of the 39th Frequency Control Symposium, 1985, p. 247.

Hanyu, I., Journal Applied Phys. Soc. Japan, 19, 1964, p. 1489.

Homma, S. and Iwata, K., Journal Crystal Growth, 19, 1973, p. 125.

Iwasaki, F., Journal Crystal Growth, 39, 1977, p. 291.

Lang, A.R. and Miuscov, V.F., Journal Applied Physics, 38, 1967, p. 2477.

Martin, J.J., Bossoli, R.B., Halliburton, L.E., Subramanian, B., and West, J.D., Proceedings of the 37th Frequency Control Symposium, 1983, p. 164.

McLaren, A.C., Usborne, C.F. and Saunders, L.A., Phys. Stat. Sol. (a) 4, 1971, p. 235.

Nielsen, J.W. and Foster, F.G., Am. Mineral., 45, 1960, p. 299.

Saha, P., Annamalai, K., Baudyopadhyay, T., Mater Sc. Bull (India), 1, 1979, p. 15.

Spencer, W.J. and Haruta, J., Applied Physics, 37, 1966, p. 549.

Thicksten, H., Proceedings of the 5th Quartz Crystal Conference, 1983, p. 234.

Tsinzerling, E.V. and Mieonua, Z.A., Kristallog 8, 1963, p. 117.

Vig, J.R., LeBus, J.W. and Filler, R.L., Proceedings of the 31st Frequency Control Symposium, 1977, p. 131.

Weissmann, S., Unpublished Proprietary Report to Sawyer Research Products, 1985.

Zarka, A., Lin, L. and Buisson, M., Journal Crystal Growth, 54, 1981, p. 394.

41st Annual Frequency Control Symposium - 1987

ETCH PITS AND CHANNELS IN SWEPT AT- AND SC-CUT QUARTZ

J.R. Hunt
Piezo Technology Inc.
Orlando, Florida 32804

Summary

Crystalline quartz, both natural and cultured, can have several types of imperfections which can affect processing into frequency control devices. The etching of quartz in fluoride-based etchants can create two forms of damage within the structure of the quartz: etch pits and etch channels. Both of these can have adverse effects on device electrical and mechanical properties.

The polished surfaces necessary for the production of AT- and SC-cut VHF and UHF resonators and filters, when deeply etched, clearly show all resulting etch pits and channels on an otherwise smooth surface under varying conditions of illumination. Cultured quartz from several suppliers and natural quartz has been evaluated for etch channel and pit formation before and after sweeping. While sweeping has greatly reduced the number of etch channels, it is apparently less effective in eliminating etch pits. Substantial differences have been observed among material from different suppliers. Even greater differences have been observed in the same material by varying the etch conditions.

Introduction

The performance requirements of quartz frequency control devices continue to become more stringent. Particular emphasis is currently being placed on device performance under varying conditions of shock, vibration, temperature, radiation, and aging. As a consequence, characteristics of the crystalline quartz used in these devices have become more critical. It is imperative that anomalous behavior of the material be minimized, allowing theoretical device performance to be approached.

The controlled etching of crystalline quartz is a common process step used in the production of precision resonators and filters. Although it is used both for frequency adjustment and the removal of prior surface damage, it can actually create two forms of damage within the structure of the quartz: etch pits and etch channels. Both forms can adversely affect device performance in a number of ways.

Prior work in the field has treated the formation of etch channels and their reduction by various means, predominantly by sweeping (electrodiffusion). Very little has been said concerning etch pit formation and reduction. One reason may be that the current techniques commonly used for viewing etch channels do not readily reveal etch pits which do not have an associated channel.

We are investigating the deep etching of polished quartz crystals in connection with the development of AT- and SC-cut VHF and UHF resonators and monolithic filters. In connection with this work, we have evaluated both natural quartz and cultured quartz from several suppliers. The incidence of etch pits and channels was measured after etching both unswept and swept material. Sweeping was performed by PTI at both the bar and blank stages. Substantial differences have been observed among material from different suppliers. While sweeping appears to greatly reduce the number of etch channels, it has been less effective in eliminating etch pits.

Variations were made in etching conditions for both AT- and SC-cut crystals with significant differences in etch pit and channel counts. The differences resulted from varying the etch temperature, pH, and basic formulation. The results indicate that improvements in crystal surfaces can be obtained by tailoring the etch conditions to the particular device being fabricated.

Quartz Material Imperfections

Both natural and cultured quartz contain material imperfections. For cultured quartz, which has been the primary object of our investigations, imperfections are related to nutrient composition and growth conditions.

The principal imperfections which are found in most commercially available cultured quartz are material impurities, point defects, dislocations, and inclusions. These problems are often interrelated, as point defects and dislocations are often caused by substitutional or interstitial ions trapped in the growing quartz as well as by the agglomeration of particulate impurities such as acmite or tuhualite, which are also trapped during growth.[1] There have been indications that bundles of dislocations may be a source of etch channel formation.[2,3]

Our primary concern has been for the phenomena which occur when quartz containing one or more of these imperfections is chemically etched, especially pits and channels.

Etching of Quartz

The chemical etching of quartz blanks in fluoride-based etchants has been a common production practice for many years. Etching is performed for a variety of reasons. It can be used as a post-lapping frequency adjustment technique, as well as to remove surface damage and relieve the stresses associated with previous lapping operations. It can also be used in lieu of mechanical polishing to chemically polish lapped blanks.[4-7] Etching is used to chemically mill complex shapes from quartz plates; examples of this are the photolithographic production of tuning fork resonators and strip resonators.[8,9] The selective etching of a central diaphragm is also a technique used for the production of high frequency fundamental resonators and filters in the VHF and UHF frequency range.[10-15]

CH2427-3/87/0000-183 $1.00 C 1987 IEEE

Although chemical etching is very useful in the production of quartz devices, it can also produce adverse effects. The improper selection of etchant used on a particular cut of quartz, for instance, can roughen the surface rather than smooth it.[4 5 16] Surface scratches, edge chips, and cracks all act as stress amplifiers, whose free energy serves to accelerate local etching and thus propagate the existing flaw. This serves to further degrade the surface finish of the crystal. Chemical etching can also result in the formation of etch pits and etch channels in crystal wafers. For reasons not completely understood, accelerated preferential etching appears to occur, with small regions of the blank etching considerably faster than the bulk of the surrounding material. Figure 1 shows many small oval-shaped etch pits and etch channels (three straight dark lines) as they appear on a 6.35 mm diameter 10 MHz SC-cut polished blank etched 20 F^2 in Allied Chemical's Superwet BOE-1235 etchant at 70°C.



FIGURE 1. Etch Pits and Channels on SC-Cut Crystal (27X)

Etch Channels. Etch channels are small, tubular passages which preferentially etch through the bulk of the quartz at a much faster rate than the surrounding material. They usually form along the original quartz growth direction; thus in a pure Z-bar, they usually extend outward from the seed in the Z-direction of growth. They are thought to be related to extended dislocation defects.[2 3] In AT- and SC-cut crystal blanks, they appear as small pipes or channels which often extend from one face through the thickness of the blank to the opposite face. Their angle of penetration is related to the angle of the crystal cut.

With AT-cut crystal blanks etched in fluoride-based etchants, the etch channel usually terminates in a characteristic, fan-shaped, etch pit on both blank faces. The degree of penetration and diameter of the tubule (ranging from submicron to several microns) usually depend upon the degree of etching and on the etch conditions. Figure 2 is a photomicrograph showing a multitude of etch channels appearing as dark straight lines. Each end of a complete channel terminates in a fan-shaped etch pit. The blank is a polished 6.35 mm diameter, unswept AT-cut crystal at a frequency of 10 MHz, which has been etched 20 F^2 in a saturated solution of ammonium bifluoride at 70°C.



FIGURE 2. Etch Channels on AT-Cut Crystal (27X)

Figure 3 is an enlarged view of the central region of Figure 2. Because of the limited depth of field of the microscope the pits on the upper surface of the blank are in focus, while the pits visible on the underside of the blank are slightly blurred.

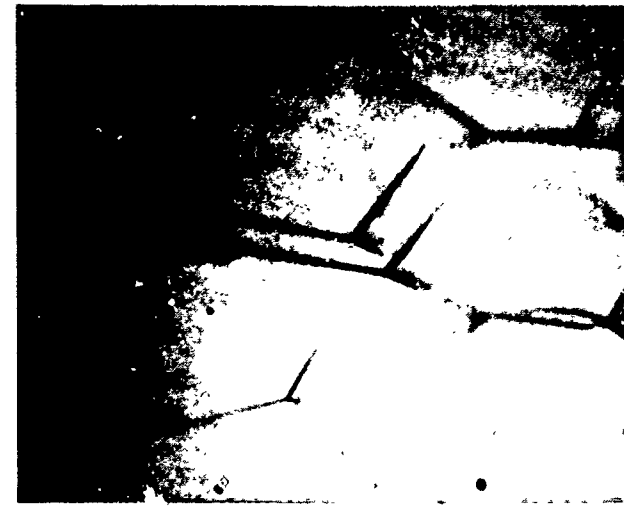


FIGURE 3. Etch Channels on AT-Cut Crystal (75X)

Because of the difference in crystallographic orientation of SC-cut crystals as opposed to AT-cut crystals, etch channels appear differently on etched SC-cut blanks. The etch channel terminates in a very characteristic etch pit which is much smaller than that formed on an etched AT-cut blank. The terminating pit usually forms on one face of the blank only; on the other face of the blank, the

channel exits the surface as a small tubular hole, identical to the tubule within the bulk of the material. Using Gualtieri's convention for SC-cut crystals (plus charge on relaxation or tension),[17] the single pit associated with the channel appears on the minus (-) blank face. No channel-associated pit appears on the plus (+) face. Figure 4 is a photomicrograph showing two etch channels as they appear on the minus face of a polished 10 MHz SC-cut crystal which has been etched 20 F^2 in Superwet HOE-1235 at 70°C. Notice the small etch pit extending obliquely from the end of each channel. Figure 5 is a view of the same two etch channels as shown in Figure 4, but viewed on the opposite side of the blank, as they exit on the plus face. There is no evidence of a pit formation on this end of the channel.



FIGURE 4. Etch Channels on SC-Cut Minus (-) Face (75X)

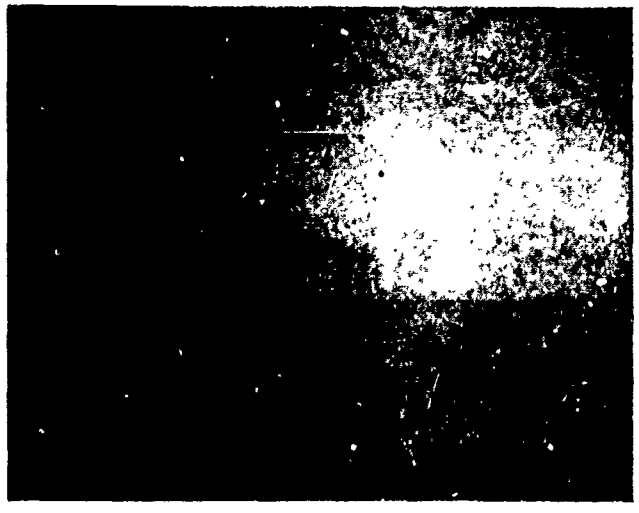


FIGURE 5. Etch Channels on SC-Cut Plus (+) Face (75X)

Etch Pits. Etch pits are small, regular depressions preferentially etched to depths of several microns into the surface of the quartz. The shape of the pit is partially dependent on the etchant and the conditions used for etching, and partially dependent on the crystal cut. Using a variety of etchants and crystal cuts, we have observed at least ten distinct shapes but have never seen more than three distinct shapes of pit on any single surface for one set of etch conditions. For AT-cut crystals the pits form on both blank faces, with and without associated channels. With SC-cut crystals however, the pits alone normally appear on the plus face, with the minus face smooth except for the characteristic pits associated with channel formation, as described above. For AT-cut crystals, the pits which appear at the termination of channels are generally similar in appearance to the pits which are not associated with any obvious channel. However, the small pits which appear on the minus face of SC-cut crystals, associated with etch channels, are distinctly different in size and shape from the larger pit which form on the plus face, not in connection with any channels. Figure 6 is a photomicrograph of the surface of a 10 MHz polished SC-cut crystal which has been etched 20 F^2 in Superwet HOE-1235 at 70°C. Two distinctively shaped pits with differing axial orientation are visible.



FIGURE 6. Etch Pits on SC-Cut Plus (+) Face (75X)

Figure 7 shows the same crystal at higher magnification. We have observed pits similar to the largest pit visible in Figure 7 which seem to exhibit a right and left handedness; that is, the conical point can project in two different directions in pits appearing on the same blank.

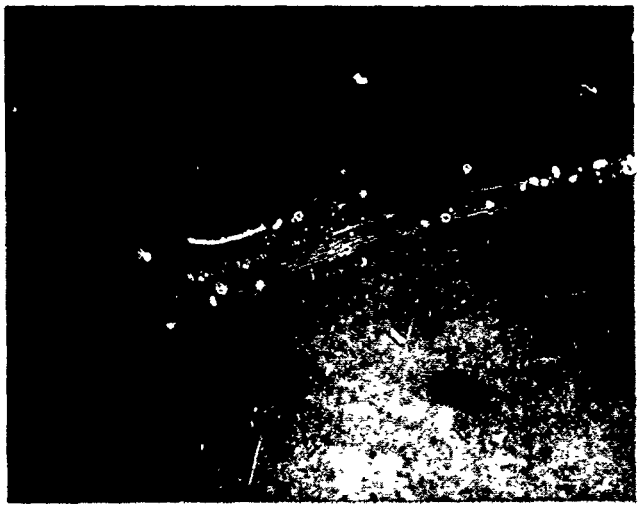


FIGURE 7. Etch Pits on SC-Cut Plus (+) Face (170X)

Consequences of Etch Defects

The formation of etch pits and etch channels can adversely affect the performance of quartz frequency control devices in a number of ways. The significance of these effects is dependent on the end use to which the devices are applied. The presence of bulk defects such as these can decrease the structural strength of quartz crystals, resulting in lower production yields and lower shock and vibration performance. The cavities formed by pits and channels can act to entrap contaminants during processing, causing a degradation in the aging performance of precision devices. The submicron and micron diameter channels can act as capillary tubes which hold liquids tenaciously after a wet processing stage. It can be very difficult to effectively rinse contaminants from these tubes.

The presence of surface and bulk voids can cause a degradation of resonator characteristics — lower device Q's and degraded mode spectra.

A current technique for the production of VHF and UHF fundamental crystal resonators is to selectively etch a central diaphragm on a crystal, leaving a thick outer ring for structural support. Prior work done at PTI resulted in crystal resonators which had central diaphragms as thin as 0.9 um.[13-15] It is obvious that the presence of pits and channels can result in catastrophic membrane rupture.

Etch Channel Determination

There has been a problem in the quartz crystal industry in past years as far as the determination of the inherent etch channel density of a given sample of quartz. Each vendor and end-user has had their own preferred technique for this determination, with little or no comparability between them. A typical procedure used for etch channel determination is shown in Figure 8; often the lapping stage is eliminated and the blanks are etched as sawn.

1) Make AT-Cut slice
2) Lap to remove saw damage
3) Etch in Ammonium Bifluoride at 75°C for 2 hrs
4) Examine at 30X with edge illumination
5) Count channels in 2.5 mm (.10") square grids
6) Compute average

FIGURE 8. Typical Lap/Etch Procedure for Etch Channel Determination

The topography which results from a lapped surface finish which has been etched is still quite rough and covered with many hillocks and depressions. Figure 9 is a photomicrograph of a quartz slice prepared as described in Figure 8, with a final lapping abrasive of 3 um used prior to etching.

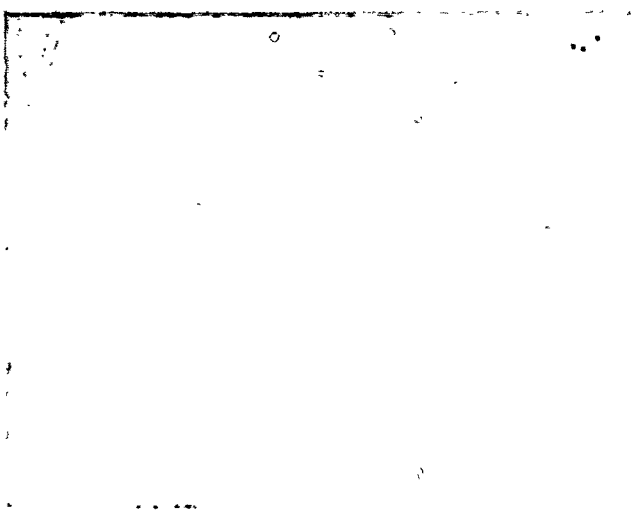


FIGURE 9. Etch Channels on a Lapped/Etched Surface

The magnification in the illustration is 27X, with edge illumination directed in from the -X direction. If this is compared with the earlier photographs showing pits and channels on a polished surface, it is obvious that small or partial etch channels may not show up on a lapped or sawn surface which has been etched. Figure 10 shows the exact same section of quartz slice as Figure 9, but with top illumination. The etch channels which appeared as bright lines in Figure 9 appear as fine dark lines in Figure 10. What is also noticeable in Figure 10 is that the surface finish is so rough that only the larger etch pits are obvious; the smaller pits are lost in the surface topography.



FIGURE 10. Etch Channels on a Lapped/Etched Surface with Top Illumination (27X)

At PTI we have been working toward the production of VHF and UHF AT- and SC-cut resonators and monolithic filters using chemical etching. In order to optimize production yields we have tested and characterized quartz from a number of vendors in a manner which was consistent with our production requirements. It has been necessary to minimize the numbers of both pits and channels in our crystal blanks. Our test for etch pits and channels therefore, must be more sensitive than the usual test for etch channels. We must perform a high quality polish operation on the particular quartz cut which we are using, and etch the blanks under conditions similar to our actual conditions. Figure 11 gives an abbreviated description of the procedure used at PTI to characterize quartz for our VHF and UHF crystal program.

1) Cut blank at appropriate angle
2) Lap to specification
3) Polish to remove surface and subsurface damage
4) Etch in appropriate etchant at desired temperature
5) Examine at 50-100X with top illumination
6) Count channels and pits in known field of view
7) Compute averages

FIGURE 11. PTI Etch Channel and Pit Measurement

As can be seen in the photographs included in Figures 1-8, the resulting finish is an extremely smooth surface which clearly shows all pits and channels without the need for edge illumination. Two groups of unswept quartz from separate vendors were tested using the procedure outlined in Figure 8 and the PTI procedure shown in Figure 11. The final abrasive used for the former test was 3 um nominal diameter. The results of the channel counts are included in Table 1.

TABLE 1

Etch Channel Measurement Comparison

Supplier	Channels/cm^2 Lap/Etch Method	PTI Method
A	310	507
B	51	121

The samples used for this test were intended to be swept, therefore the seeds had previously been split out of them. For the lap/etch method, since the entire width of the plate is used, the measurement includes the area immediately adjacent to the seed. The channel density for vendor A was quite uniformly high all the way across the sample, but vendor B had a high count within 1.25 mm of the seed area, and a considerably lower count over the balance of the sample. If the area within 1.25 mm had been excluded from the count for supplier B, the count would have dropped from 51 to 25 channels/cm^2; whereas with supplier A it would have remained essentially constant.

Material evaluations were performed on natural quartz and cultured quartz from five suppliers for the purpose of developing UHF SC-cut resonators to frequencies of 1000 MHz. All cultured samples were swept at PTI and tested in accordance with the procedure outlined in Figure 11. The blanks were SC-cut and were etched in Allied Chemical's Superwet NOE-1235 etchant at 70°C for a material removal of 20 F^2. The results are shown in Table 2.

TABLE 2

SC-Cut Etch Channel and Pit Density - Supplier Comparison

Material	Channels/cm^2	Pits/cm^2	% Blanks W/Channels
Natural	0	746	0
A	32	1196	56
B	6	418	5
C	16	601	31
D	10	830	40
E	0*	415	0*

* see text

Although supplier E had no visible channels before or after sweeping, the etch pits which are normally only associated with channels and appear on the minus face are still present. The count for these channel-related pits was 39/cm^2, higher than the channel count for any other group. Eighty-eight percent of the blanks examined from group E contained these pits. Although these channel-related pits have been observed without any associated channel before, they are normally rare. In addition, the minus face of the blanks from group E also contained an additional, small, elongated pit which has never been observed before, and is not associated with channels.

Sweeping of Quartz for Pit and Channel Reduction

Sweeping has long been used to improve the properties of cultured quartz.[18-21] Vig, et al, first observed that sweeping could be effective in reducing the incidence of etch channels in quartz.[4] The mechanism by which sweeping effects the reduction of etch channels is not fully understood. Typical sweeping conditions that FTI has used to effectively reduce the incidence of channels and pits are shown in Table 3.

TABLE 3

Typical Sweeping Conditions for Channel and Pit Reduction

- 500-540° Centigrade
- 1000-2000 Volts/cm DC applied in Z-axis direction
- Controlled atmosphere
- Sweep to determined endpoint, usually steady-state current

Material from the five suppliers of cultured quartz referred to in Table 2 was examined for pit and channel density both before and after sweeping. All samples were SC-cut blanks, prepared from matched halves of the same bars, one swept and one unswept. They were then etched in BOE-1235 at 70°C for a removal of 20 F^2. The results of etch channel counts for all groups are given in Table 4.

TABLE 4

SC-Cut Etch Channel Reduction by Sweeping

Supplier	Channels/cm^2 Unswept	Swept	Percent Reduction
A	507	32	94
B	121	6	95
C	278	16	94
D	256	10	96
E	2*	0*	100

* see text mean 95.8%

As described above, the minus face of the blanks from group E contained lone pits which are normally associated with channels, but in this case, no channels were apparent. The count for these particular pits was 121/cm^2 before sweeping, and 39/cm^2 after sweeping, giving a 68% reduction through sweeping. The significance of the occurrence of these isolated pits on the minus face is not understood.

Blanks from three suppliers were examined for etch pits in same manner as those prepared for the channel count given in Table 4. The results are given in Table 5.

TABLE 5

SC-Cut Etch Pit Reduction by Sweeping

Supplier	Pits/cm^2 Unswept	Swept	Percent Reduction
C	2,082	601	71
D	1,445	830	43
E	1,584	415	74

mean 62.7%

A comparison with the data in Table 4 shows that the incidence of etch pits is not reduced as effectively by sweeping as is the incidence of etch channels. The reasons for this are not clear.

Two groups of quartz were swept as blanks and examined for etch channels. Blanks from the same bars were processed in a similar manner except that half of the blanks in each group was swept while at a frequency of 6 MHz, a thickness of approximately 0.3 mm. All blanks were lapped and polished to 10 MHz, then etched 15 F^2; the AT-cuts in saturated ammonium bifluoride, and the SC-cuts in BOE-1235, both at 70°C. All blanks were examined for etch channel density, the results are shown in Table 6.

TABLE 6

Channel Reduction for Swept Blanks

Cut	Etchant	Channels/cm^2 Unswept	Swept	Percent Reduction
AT	Ammonium Bifluoride	436	82	81.2
SC	Superwet BOE-1235	507	18	96.5

The material used for this experiment was not from the same sources for the AT- and SC-cuts. This may account to some degree for the difference in apparent sweeping effectiveness between the two groups.

Etching Variables

The chemical etching of crystalline quartz is sensitive to a number of variables which affect the quality of the end product. Factors such as the quartz cut, etchant chemistry, temperature, and degree of agitation have all been observed to have an influence on etch quality.

A series of experiments was conducted to determine if any of the variables also affected the incidence of etch pits and etch channels. All work was performed using swept material, but different series of experiments used quartz from different suppliers. Within each experiment all quartz used was from the same bar.

The first experiment was to determine the effect of different etchants on AT-cut blanks. SC-cut blanks were not evaluated because preliminary work showed that only one etchant used, BOE-1235 was capable of maintaining a polished surface. The

other two etchants chosen were; saturated ammonium bifluoride and "Lattice Etch", a proprietary etchant manufactured by Photo Quartz, Inc. When SC-cut blanks were etched in saturated ammonium bifluoride, pits formed on the plus face and the minus face became frosted. Interestingly, when SC-cut blanks were etched in "Lattice Etch", very many small pits formed on the minus face and the plus face became frosted. In both cases they were unusable for UHF resonators.

The swept AT-cut blanks were etched 17 F^2 at 70° in three different etchants. The pits and channels were then counted, the resulting data is shown in Table 7.

TABLE 7

Etchant Effects on Channel and Pit Formation

Etchant	Etch Rate (um/min)	Channels/cm^2	Pits/cm^2
Ammonium Bifluoride	0.566	38	1,342
Superwet BOE-1235	0.023	658	1,373
Lattice Etchant	0.210	0	505

While BOE-1235 is an effective etchant for SC-cut crystals, it is not the choice for AT-cut crystals. Figure 12 is a view of one of the crystals etched in BOE-1235. The holes which are completely through the blank appear to be where etch channels were formed, each with its corresponding pair of etch pits. Figure 13 is a more magnified view of the same crystal. Most of the pits occur in pairs, with one occurring on each side of the blank, each oriented at an acute angle with its corresponding pit on the other face. Even though the blank was riddled with holes, it was quite strong and easily handled.

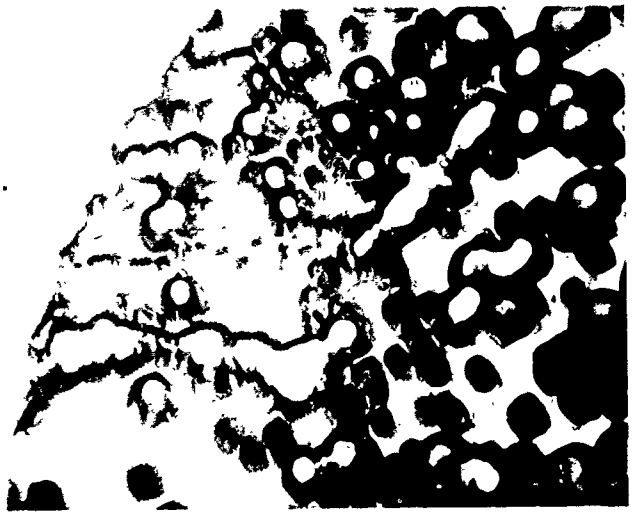


FIGURE 12. AT-Cut Crystal Deeply Etched in BOE-1235 (27X)



FIGURE 13. AT-Cut Crystal Deeply Etched in BOE-1235 (75X)

It was observed in this experiment that the etch rates varied by a factor of 25:1. Another experiment was conducted to determine the effects of small variations in etch rate with saturated ammonium bifluoride. This etchant was chosen because it is relatively easy to vary its pH, which has a marked effect on etch rate. The nominal etch rate of the ammonium bifluoride used for this experiment was 0.3-0.4 um/min. for double-sided etching, as mixed. The etch rate was lowered by the addition of small amounts of ammonium hydroxide, which raises the pH. The etch rate was raised by small additions of hydrofluoric acid, which lowers the pH. As commercial ammonium bifluoride is manufactured by the reaction of ammonia with hydrofluoric acid, our choices for addition agents introduced no foreign species.

The swept AT-cut blanks were etched 17 F^2 at 70°C in all three cases. The resulting data is shown in Table 8.

TABLE 8

Etch Rate Effects on Channel and Pit Formation

Etch Rate (um/min)	Channels/cm^2	Pits/cm^2
0.2	267	3,400
0.3	70	2,680
0.5	3	1,550

The group of blanks which were etched at 0.5 um/min. showed a very slight texturing effect on their surface. The differences in pit and channel densities are dramatic.

The last group of experiments were conducted with both AT-cut crystals and SC-cut crystals. In these experiments, only the temperature was varied, all other conditions were held constant. For both groups, the swept crystals were etched 17 F^2; satur-

ated ammonium bifluoride was used for the AT-cuts, and 60E-1235 was used for the SC-cuts. For both groups, the etch rate given is for double-sided etching.

The resulting data for the AT-cut blanks is shown in Table 9.

TABLE 9

Temperature Effects on AT-Cut Channel and Pit Formation

Temp.(°C)	Etch Rate (um/min)	Channels/cm^2	Pits/cm^2
25	.0269	7	39
70	.566	38	1,342

It is interesting to note that in this case the pit formation was suppressed considerably more than was channel formation at the lower temperature. Both channel and pit density were lower at the lower etch temperature.

The results for the same experiment performed with SC-cut blanks is given in Table 10.

TABLE 10

Temperature Effects on SC-Cut Channel and Pit Formation

Temp.(°C)	Etch Rate (um/min)	Channels/cm^2	Pits/cm^2
25	.0129	0	19,990
45	.0411	0	390
70	.130	16	601

The pits which appeared at 25°C were very small, considerably different than those which resulted at the higher temperatures.

For both groups of crystals, temperature variations alone had significant effects on the numbers of channels and pits formed by etching.

Conclusion

Two defect structures, pits and channels, which occur when crystalline quartz is chemically etched, can cause significant degradation of the performance of quartz frequency control devices. It is important for us to gain a better understanding of these phenomena, so that we may be able to make future strides in reducing their occurrence.

There is great variability between material from different suppliers of cultured quartz, and even between material from one supplier which is produced in different autoclave runs.

We have demonstrated that the occurrence of etch pits and channels in quartz is dependent on a great number of variables. Some of the factors which have been shown to have a significant effect on pit and channel density are quartz quality, angle of cut, etchant chemistry, and temperature.

The incidence of etch pits and channels, together with their particular geometries, only have meaning under a very specific set of etching conditions. For the numbers to have real meaning, the tests should be representative of the intended process used to fabricate devices. As has been shown, relatively small changes in process parameters can have significant effects on the quality of the end product.

This work was supported in part by the U.S. Army Laboratory Command under Contract DAAL01-86-C-0006.

References

[1] Halliburton, L.E., N.E. Markes, and J.J. Martin, "Point Defects in Synthetic Quartz: A Survey of Spectroscopic Results With Application to Quality Assurance," Proc. 34th AFCS, pp.1-8; 1980.

[2] Armington, A.F., "The Growth of High Quality Quartz in Commercial Autoclaves," Proc. 39th AFCS, pp.230-233; 1985.

[3] Armington, A.F., and J.F. Balascio, "The Commercial Growth of High Quality Quartz," Proc. 7th Quartz Device Conf., pp.147-151; 1985.

[4] Vig, J.R., J.W. LeBus, and R.L. Filler, "Chemically Polished Quartz," Proc. 31st AFCS, pp.131 - 143; 1977.

[5] Vig, J.R., R.J. Brandmayr, and R.L. Filler, "Etching Studies on Singly- and Doubly-Rotated Quartz Plates," Proc. 33rd AFCS, pp.351-358; 1979.

[6] Brandmayr, R.J. and J.R. Vig, "Chemical Polishing in Etching Solutions That Contain Surfactants," Proc. 39th AFCS, pp.276-281; 1985.

[7] Brandmayr, R.J. and J.R. Vig, "Further Results on the Use of Surfactants in Chemically Polishing Quartz Crystals," Proc. 40th AFCS, pp.86 - 90; 1986.

[8] Staudte, J.H., "Subminiature Quartz Tuning Fork Resonator," Proc. 27th AFCS, pp.50-54; 1973.

[9] Oguchi, K., and E. Momusaki, "+5°X Micro Quartz Resonator by Lithographic Process," Proc. 32nd AFCS, pp.277-281; 1978.

[10] Guttwein, G.K., A.D. Ballato, and T.J. Lukaszek, U.S. Patent 3,694,677; 26 Sept. 1972.

[11] Hanson, W.P., "Chemically Polished High Frequency Resonators," Proc. 37th AFCS, pp.261-264; 1983.

[12] Dowsett, J., D.P.G. Dwyer, and F. Stern, "Etch Processing of Bulk and Surface Wave Devices," Proc. 39th AFCS, pp.301-309; 1985.

[13] Hunt, J.R. and Smythe, R.C., "Chemically Milled VHF and UHF AT-Cut Resonators," Proc. 39th AFCS, pp.292-300; 1985.

[14] Smythe, R.C., M.D. Howard, and J.R. Hunt, "VHF Monolithic Crystal Filters Fabricated by Chemical Milling," Proc. 39th AFCS, pp.481-485; 1985.

[15] Smythe, R.C., J.R. Hunt, and M.D. Howard, "Chemically-Milled VHF and UHF Quartz Resonators and Monolithic Filters," Proc. IEEE Ultrasonics Symposium, pp.302-305; 1985.

[16] Bernot, A.J., "Etching Study of AT-Cut Cultured Quartz Using Etchants Containing Fluoride Salts, Hydrofluoric Acid, and Ammonium Bifluoride," Proc. 39th AFCS, pp.271-275; 1985.

[17] Gualtieri, J.G., "A Simple Method for Location of the Mounting Positions for Low Acceleration Sensitivity SC-Cut Resonators," U.S. Army ERADCOM, Res. & Dev. Tech. Rpt. DELET-TR-81-5; February, 1981.

[18] King, J.C., U.S. Patent 3,113,224; 3 Dec., 1963.

[19] Fraser, D.B., U.S. Patent 3,263,103; 26 July, 1966

[20] Fraser, D.B., U.S. Patent 3,337,439; 22 Aug., 1967

[21] King, J.C., U.S. Patent 3,932,777; 13 Jan., 1976.

FURTHER STUDIES ON ELECTRODE-DIFFUSION-SUPPRESSED-SWEPT QUARTZ

John G. Gualtieri

US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

ABSTRACT

Previously,[1] we reported that electrode-metal diffusion during sweeping could be suppressed by: 1) reduction of Z-surface damage prior to electroding, 2) application of metal-foil electrodes under pressure as an alternative to metal-evaporated or sputtered electrodes, 3) use of diffusion barrier metallizations prior to metal evaporation, and 4) by use of refractory films between metal-foil electrodes and the quartz.

Sweeping has been shown to result in significant reductions in etch-channel density.[2] Observations of etch resistance at metal-decorated dislocation sites suggest that suppression of electrode-metal diffusion results in higher etch-channel densities than when the diffusion is not suppressed. Higher etch-channel densities were found with all of the suppression techniques listed above.

We found cathode surfaces to be more resistant to etching than anode surfaces after sweeping with gold diffusion allowed, (i.e. the suppression procedures, listed above, were not used). When gold diffusion was suppressed, the etched surfaces resembled unswept samples. SC-cuts removed from gold diffusion-allowed samples showed improvement in chemical polishing when compared with gold diffusion-suppressed samples. Again, diffusion-suppressed samples resembled unswept samples.

Applying normalized-difference spectroscopy to both swept and unswept samples, we found substantial enhancement of OH^- absorption bands in samples where electrode-metal diffusion was allowed versus samples where it was suppressed. The number of singly charged ions transported during sweeping correlated with the increased H+ ion content calculated from the enhancement of OH^-absorption bands in samples representing two different grades of quartz. Neutron activation analysis of these samples revealed large decreases in the level of sodium after sweeping with gold diffusion and much smaller decreases when the diffusion was suppressed.

These findings indicate that suppression of electrode-metal diffusion usually involves suppression of H+ ion diffusion into th sample, which degrades the sweeping efficier cy. It seems very likely that differences i reported sweeping efficiency[1,3,4] are relate to the capability of the electrode to catalyz the introduction of hydrogen into the quartz.

INTRODUCTION

We have previously reported that elec trode-metal diffusion into quartz durin sweeping reduces the etch-channel density. Presumably, metal that diffuses into th quartz fills the regions (dislocations an growth tunnels) that will be the sites o future etch channels. This filling and th accompanying modification of the chemica reactivity of the region are not presentl understood. Since we have found ways to limi or suppress the electrode diffusion, we use the same techniques to determine their in fluence on etch-channel density. During th course of the etch-channel investigation, w found that, after electrode-diffusion-allowe sweeping, cathode Z-surfaces were more resis tant to etching than Z-surfaces of unswept o electrode-diffusion-suppressed material. Thi result prompted a study to determine if SC-cu surfaces cut from the differently swept bar would show similar reactivity difference whe using chemical polishing solutions. Othe studies[1,4] showed that the total transporte charge during sweeping could be related to th electrode diffusion. Kats[5] has shown ho transported charge can be related to th increase in H+ content by evaluating th difference in OH^- infrared absorption afte sweeping. Many workers have subsequentl found differences in OH^- infrared absorptio after sweeping, reinforcing the concept tha H+ replaces interstitial alkali as charg compensation for different amounts of as-grow Al in substitutional Si sites. Kats use theory developed by Crawford[6] and infrare difference measurements to determine th increase in H+ content. We decided to us infrared absorption to determine if electrode diffusion during sweeping could be related t OH^- absorption. It was apparent that th required infrared measurements and calcula tions could be accomplished automatically

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

using computer-aided Fourier transform infrared (FTIR) spectrophotometry. Our initial findings and previous experimental work of others prompted this investigation.[1,3,4] Our aim was to improve the technology of sweeping by developing methods of evaluating the electrode-diffusion suppression and its influence on etching, chemical polishing and, H^+ ion-alkali ion replacement.

EXPERIMENTAL METHOD

Cultured quartz from domestic suppliers was used in this investigation. The quartz was lumbered into Y-bars or SC-bars after seed removal.[7] Two electrode types usually resulted in diffusion of electrode material into the quartz. These were either evaporated metallizations applied directly to the Z-surfaces or Pt-foil electrodes pressed against the quartz. Gold-plated molybdenum leads were welded to the metal electrodes. To suppress electrode diffusion, a number of methods were used,[1] which included:

(1) Evaporating compounds such as TiO_2 between the gold metallization and the quartz Z-surfaces.

(2) Coating the quartz Z-surfaces with paintable boron nitride, then applying Pt-foil electrodes under pressure.

(3) Inserting sapphire plates between the electrode and the quartz Z-surfaces.

(4) Reducing Z-surface damage using mechanical and/or chemical polishing.

All sweeping was performed in air. Temperature ramped sweeping was employed, a description of the procedure was given previously.[4] At the completion of each run, a computer-generated plot of current density versus time was recorded and the charge transported per unit volume was computed. An Arrhenius plot of the product of conductivity and time versus inverse temperature was obtained for the entire run. Fourier transform infrared spectrophotometry was used to obtain room-temperature infrared absorption due to OH^- vibrations before and after sweeping. The spectrophotometer was a Perkin-Elmer model 1550.

EXPERIMENTAL OBSERVATIONS

After sweeping with electrode-metal diffusion allowed, we found that dislocations, which are the principal source of etch channels, are usually decorated with what appear to be colloidal particles of metal.[4] In Fig. 1, we show a photomicrograph of an AT-cut sample cut from a bar which was swept allowing gold diffusion. The sample was subsequently etched in ammonium bifluoride to reveal etch channels. A group of gold-decorated dislocations (colored purple), all intersecting the surface, have resisted the etchant. Close by, an undecorated dislocation has been etched. In this sample, we found 26 decorated disloca-

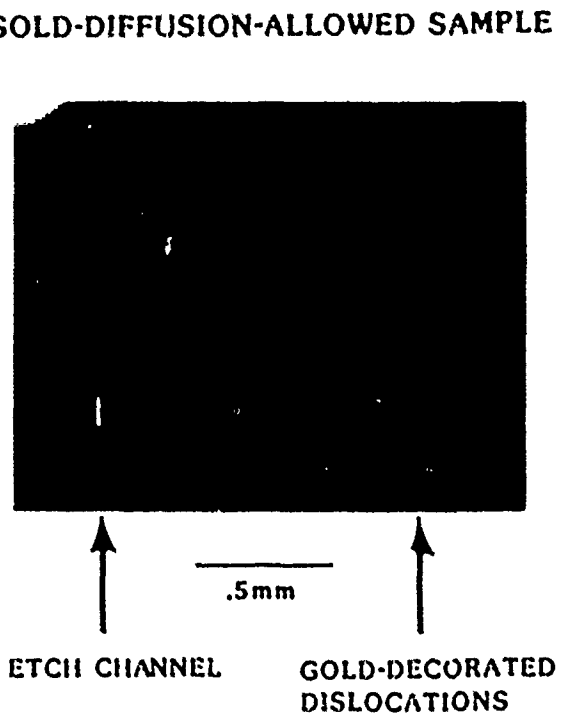


Figure 1. Photomicrograph of an AT-cut sample swept with gold diffusion allowed. Gold-decorated dislocations all resist etching. Undecorated dislocations are sites of etch channels.

Table I

SUPPRESSION OF ELECTRODE DIFFUSION vs ETCH-CHANNEL DENSITY

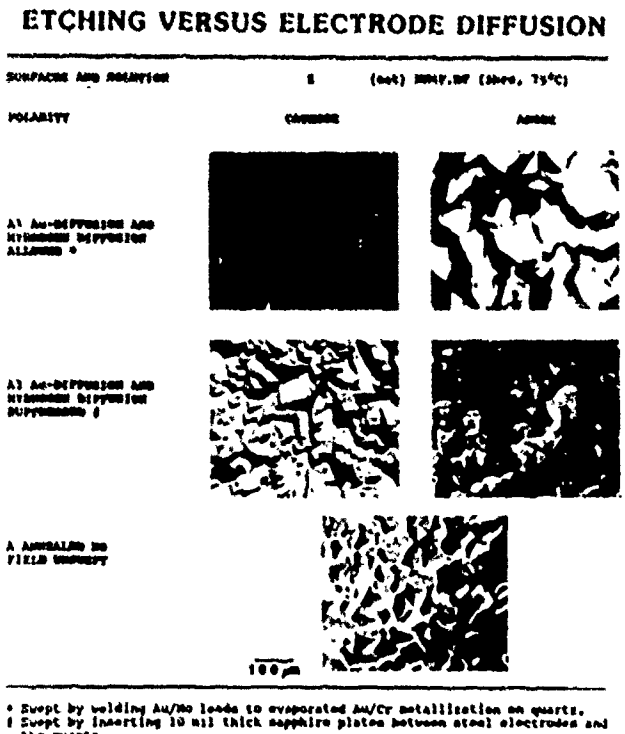
SUPPRESSION TECHNIQUE	ABSORPTION COEFFICIENT (CM^{-1})	ELECTRODE TYPE	Z-SURFACE FINISH	ETCH-CHANNEL DENSITY (CM^{-2})
DIFFUSION BARRIER TiO_2 ⟶	0.084	Au/Cr	SAWN	11
		Au/TiO_2		80
INSULATING DEPOSIT BN ⟶	0.066	Pt-foil	1 μM	0.25
		Pt-foil/BN	SAWN	35
INSULATING SAPPHIRE S ⟶	0.067	Au/Cr	3 μM	0
		Steel/S #		100
METAL FOIL vs SPUTTERING	0.040	Pt-sputt	1 μM	3
		Pt-foil		30
SURFACE FINISH	0.044	Au/Cr	SAWN	0
			1 μM	12

Swept by inserting 10 mil thick sapphire plates between steel electrodes and the quartz.

tions that all resisted etching. We also observed 13 etch channels, none of which displayed any decoration. Every suppression technique we applied, using different grades of quartz, resulted in a smaller decrease in etch-channel density. This is evident in the list compiled in Table 1.

While studying the suppression of electrode diffusion during sweeping versus etch-channel density, we noticed that, after etching, cathode surfaces were much smoother than anode surfaces. This etching difference was found to hold, even if the surfaces were ground to a depth of 100 μm, repolished, and etched again. In Fig. 2, we show photomicrographs of etched cathode and anode surfaces found after sweeping with gold diffusion, allowed and suppressed, compared to an etched unswept Z-surface. Later, we will show that suppression of electrode-metal diffusion involves suppression of hydrogen diffusion. For this reason, references to the diffusion of hydrogen are included.

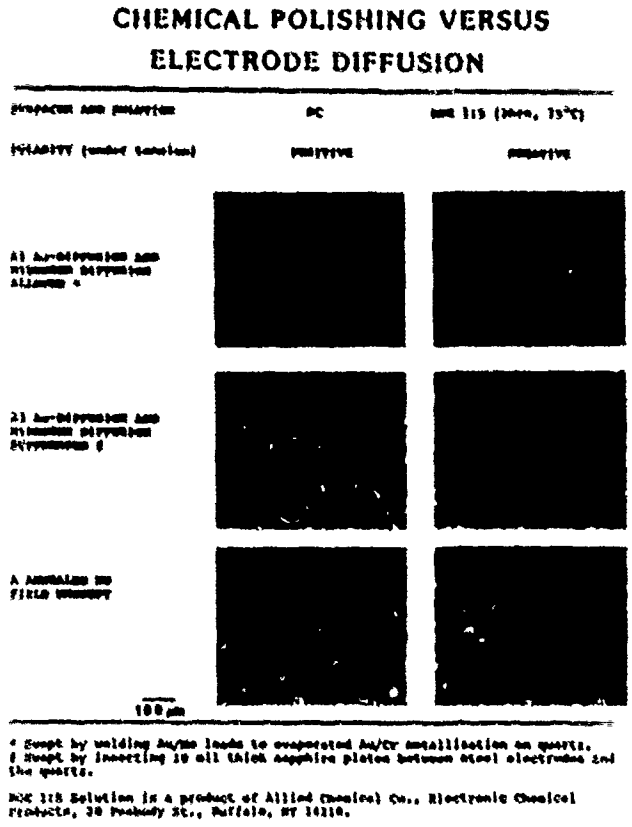
Next, we studied the chemical polishing of SC-slabs cut from the bulk of the swept quartz bars. (A discussion of polarity identification of SC-surfaces is given in the Appendix.) Using a standard chemical polish-



* Swept by welding Au/Mo leads to evaporated Au/Cr metallization on quartz.
† Swept by inserting 10 mil thick sapphire plates between steel electrodes and the quartz.

Figure 2. Photomicrograph of cathode and anode surfaces after sweeping with gold and hydrogen diffusion-allowed and -suppressed compared with unswept.

ing solution, we found that the diffusion-allowed samples were polished, but the diffusion-suppressed samples were not polished. Samples that were unswept (exposed to a similar temperature ramping without any electric field) were also not polished. (See Fig. 3.)



* Swept by welding Au/Mo leads to evaporated Au/Cr metallization on quartz.
† Swept by inserting 10 mil thick sapphire plates between steel electrodes and the quartz.

Figure 3. Photomicrograph of SC-surfaces after sweeping with gold and hydrogen diffusion allowed-and -suppressed compared with unswept.

Because the diffusion-suppressed samples resembled the unswept samples, we suspected H^+ ion-alkali exchange was being suppressed. Arrhenius plots of the temperature-ramped sweeping runs and infrared absorption before and after sweeping, confirmed the hydrogen suppression.

An Arrhenius plot of a diffusion-allowed sample is shown in Fig. 4. During the begin-

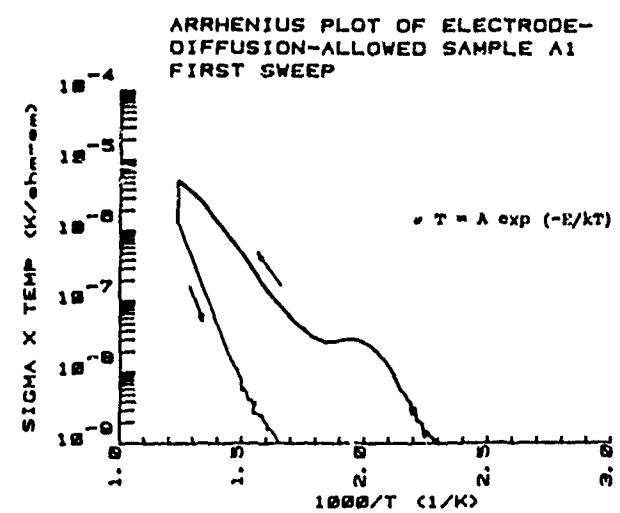


Figure 4.

ning of the heating phase, some species, possibly Na+, is responsible for the conductivity which depletes in the 200 - 270°C range. The activation energy or slope in the region before depletion is 1.2 - 1.3 eV. The next portion of the heating phase is dominated by a conducting species (possibly Li+) which usually depletes very close to our maximum temperature of sweeping (~540 deg C). The slope in the region before this depletion is 0.76 - 1 eV (usually 0.9 eV). After an isothermal phase (vertical line), the cooldown phase is very linear. Here the slope is in the range 1.50 - 1.75 eV, which is within the literature values reported for the conduction of H+ in quartz.[8] If the sample is swept a second time, the heating and cooling slopes are nearly identical (See Fig. 5). If one suppresses electrode diffusion using any of the methods given above, the Arrhenius plot in Fig. 6 resembles that of a completely swept sample.

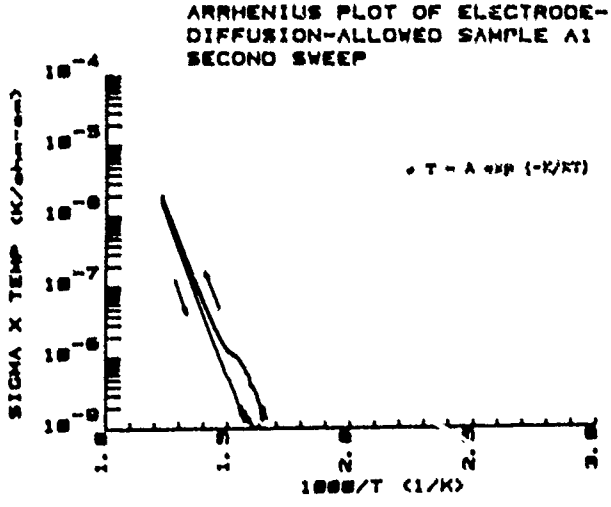


Figure 5.

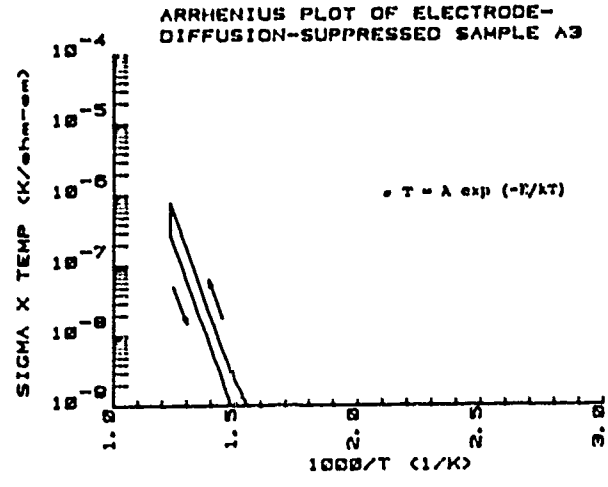


Figure 6.

Infrared absorption spectra of samples before and after electrode-diffusion-allowed sweeping are given in Fig. 7. It is obvious that the OH^- bands increase with sweeping. The difference spectrum, i.e., the diffusion-allowed spectrum minus the unswept spectrum, is shown, in Fig. 8. Infrared absorption spectra before and after diffusion-suppressed sweeping are shown in Fig. 9.

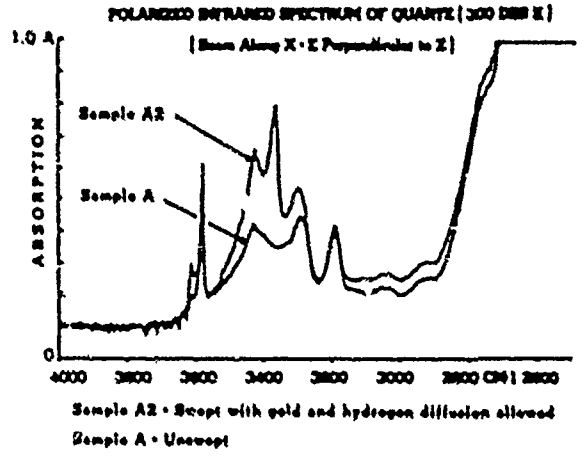


Figure 7.

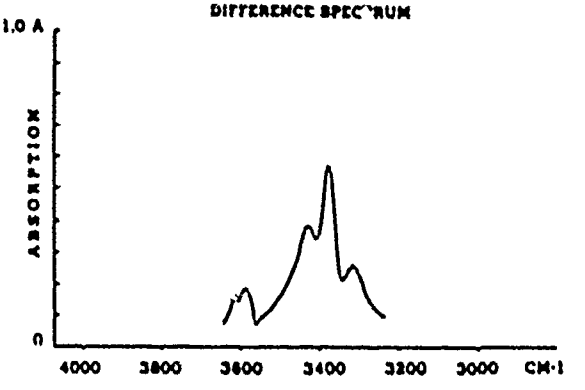


Figure 8. Difference spectrum of sample A2, swept with gold and hydrogen diffusion-allowed, minus unswept (sample A). The sample thickness was 1.95 cm. The ordinate scale should be multiplied by (2.303/1.95)= 1.181 cm-1. The area between 3645 cm-1 and 3235 cm-1 equals 57.8 cm-2. The bands showing increased absorption are located at 3614,3585,3433,3378 and 3307 cm-1

DISCUSSION

In order to estimate the amount of H+ ion-alkali ion exchange during sweeping, normalized spectral difference spectroscopy or "spectral stripping" was applied to infrared absorption spectra obtained before and after sweeping. In this technique, the difference

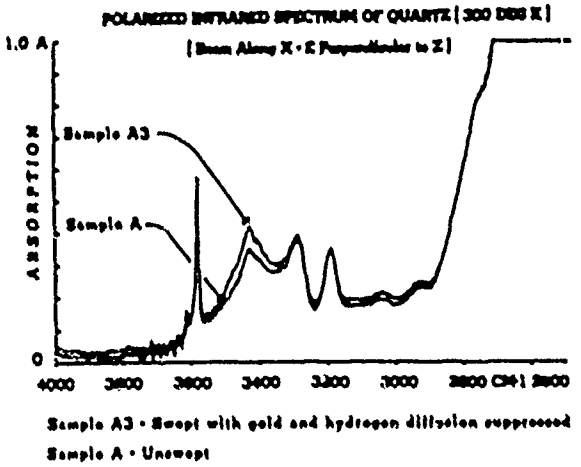


Figure 9.

between the ordinate values of both spectra are found for each data point. The difference calculation for each point is

$$H = S - (U \times F), \qquad (1)$$

where H represents the ordinate values of the difference spectrum (due to increased H+ ion content), S represents the ordinate values of the swept sample, and U represents the ordinate values of the unswept sample, F is a normalization factor applied to a lattice-vibration band of the unswept spectrum. For normalization, we selected the broad band of medium intensity centered at 3200 CM-1. Normalization compensates for any difference in thickness between the two samples.

Thus, when the normalized unswept spectrum is subtracted from the swept spectrum, only the spectrum attributable to increased H+ ion concentration should remain (See Fig. 8). Once the H+ spectrum is obtained, one can find the integral absorption

$$\bar{K} = \int \alpha(\nu)\, d\nu \ , \qquad (2)$$

where alpha is the absorption coefficient and ν is the frequency. The region of increased OH^- absorption extends from 3645 CM-1 to 3235 CM-1. The bands located at 3614, 3585, 3433, 3378, and 3307, CM-1 usually show increases. (See Fig. 8).

The difference and integral computations are automatic procedures, using the Perkin-Elmer 1550 FTIR and proprietary CDS-3 data processing software system. The transported charge during the sweeping process should be proportional to the increased OH^- absorption. We computed the transported charge per unit volume using a difference of integrals of the form

$$Q/V = 1/l_z \left(\int_{t_0}^{t_c} J\, dt - \int_{t_m}^{t_c} J_s\, dt \right), \qquad (3)$$

where J is the current density and Js is the stabilized current density before cooldown, which begins at t_c. The time when the maximum current density is reached is t_m. The onset of measurable current is designated as t_o.

The sample thickness in the z-direction is l_z. We previously reported the computation of transported charge using a different integral formula.[1] The reason for the the difference is based on plots of current density versus time for a bar which was swept a second time using temperature ramping. Initially, an infrared spectrum of an unswept bar was obtained for reference. After the first sweep, the current density was plotted as a function of time, using temperature ramped sweeping and Pt-foil electrodes. The Z-surfaces of the bar were then ground to a depth of 100 μm and repolished. The infrared spectrum was obtained and the difference spectrum (swept minus unswept) revealed that a considerable amount of hydrogen had been swept into the sample. The current density versus time plots are given in Fig. 10. The sweeping procedure

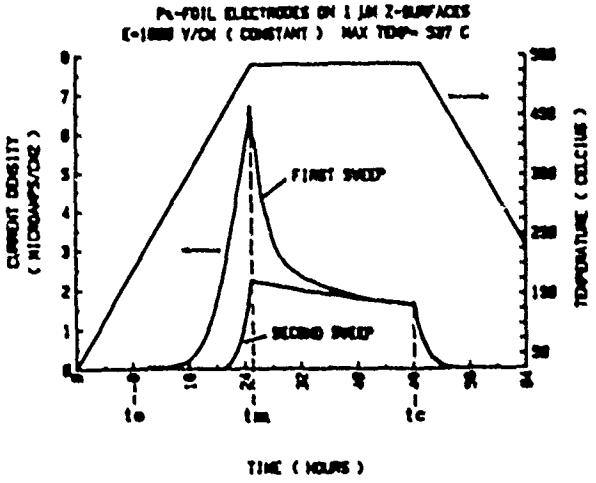


Figure 10. Temperature ramped sweeping using Pt-foil pressure electrodes. Z-surfaces were polished (average abrasive size was 1 μm). Second sweep accomplished after grinding 100 μm from each Z-surface and repolishing.

was then repeated. The difference spectrum (first sweep minus second sweep) revealed that very little additional hydrogen entered the sample during the second sweep. Referring to Eq. (3), the first integral represents all of the transported charge, including that due to hydrogen-hydrogen exchange. The second integral, which approximates the conduction due to hydrogen-hydrogen exchange is subtracted from the first integral to yield the transported charge due to hydrogen replacing alkali ions. This assures us that we are not including any current contributions due to hydrogen-hydrogen exchange. One can determine the number of H+ ions swept into the sample using the following formula developed by Kats[5]

$$N\ (H) = 3.77 \times 10^{15}\ (\bar{K}/p^2), \qquad (4)$$

where p is the relative effective charge of H+ and $\bar{K}$ is the integral absorption measured with light polarized perpendicular to the optic axis. In order to derive the effective charge for H+ in our quartz samples, we calculated the number of singly charged ions

transported during a sweeping run using

$$N(T) = (Q/ev), \qquad (5)$$

where e is the electronic charge. We then set N(H) = N(T) and found the effective charge, p, to be 0.51. This value is close to the value p = 0.52 derived by Kats, when he assumed that the difference-absorption bands had a Lorentzian profile. We used p = 0.51 for all subsequent calculations of N(H).

In Table II, we compare diffusion-allowed samples to diffusion-suppressed samples for two different grades of quartz. The increased H+ content calculated using Eq. (4), can be compared to the number of ions transported using Eq. (5). It seems that, for the suppressed samples, N(T) and N(H) are in balance, whereas, the diffusion-allowed samples show differences. Some of the differences can be accounted for by the amount of metal which diffused into the sample. (See Table III). Table III shows neutron activation analysis data for samples with and without diffusion suppression compared with an unswept sample. It is clear that, other than suppression of metal diffusion, also suppressed was the diffusion of hydrogen into the quartz via dissolution of hydrogen in the metal electrode and diffusion through the metal. This is indicated, also, by the small decrease in the level of Na in the diffusion-suppressed samples and the larger decrease in the diffusion-allowed sample. The following mechanism, suggested for the dissolution of hydrogen in palladium,[9] may apply to most metals used as electrodes for sweeping quartz.

Table II

IONS TRANSPORTED DURING SWEEPING vs INCREASED H+

SAMPLE (Alpha 3500) (CM-1)	TRANSPORTED CHARGE(Q/V) (COUL/CM2)	IONS TRANSPORTED N(T) X 10^{17} (CM-2)	SWEPT-UNSWEPT INTEGRATED ABSORPTION(δ) (CM-2)	INCREASED H+ CONTENT N(H) X 10^{17} (CM-2)
Au/Cr Diffusion and Hydrogen Diffusion Allowed *				
A1 (0.087)	0.195	12.17	78.9	11.44
A2 (0.087)	0.159	9.96	57.8	8.38
B1 (0.050)	0.038	2.40	11.7	1.70
B2 (0.050)	0.024	1.50	7.1	1.03
Au/Cr Diffusion and Hydrogen Diffusion Suppressed #				
A3 (0.087)	0.029	1.81	12.4	1.80
B3 (0.050)	0.0018	0.115	0.8	0.116

* Swept by making Au/Mo leads to evaporated Au/Cr metallization on quartz
\# Swept by inserting 10 mil thick sapphire plates between steel electrodes and the quartz

Molecular hydrogen H_2 (g) is weakly chemisorbed on the metal surface. Next, a transfer step occurs, which produces interstitial hydrogen H_i until an equilibrium is established between H_2 (g) and H_i. In the final stage of the sorption process, some hydrogen becomes strongly adsorbed on the metal surface. The interstitial hydrogen

Table III

NEUTRON ACTIVATION ANALYSIS (NAA)

PPB(Si)

SAMPLE (ALPHA 3500 CM-1)	Cr	Au	Na
A (0.087 CM-1):			
A1 Swept - Au/Cr Diffusion and Hydrogen Diffusion Allowed *	1400	80.2	463
A3 Swept - Metal Diffusion and Hydrogen Diffusion Suppressed #	<20	1.7	1460
A Unswept	<20	0.9	1600

* Swept by making Au/Mo leads to evaporated Au/Cr metallization on quartz
\# Swept by inserting 10 mil thick sapphire plates between steel electrodes and the quartz

assumes an effective charge Ze in the metal, which, at the temperatures used for sweeping, ranges from 0.24e for Fe to 1.4e for Ta.[10] For most metals, charged hydrogen will be transported through the metal to the quartz from the anode. At the quartz interface, the charged hydrogen will encounter the relatively large Z-axis channels. The field will continue to drive the charged hydrogen into the quartz for sweeping. If this process is prevented by using any effective hydrogen barrier, such as insertion of a sapphire plate and/or air gaps between the electrode and the quartz, the exiting hydrogen will reassociate into the molecular form. Since molecular hydrogen is larger and uncharged, it will not easily diffuse into the quartz. It seems that metals in contact with the Z-surface of quartz catalyze the introduction of hydrogen into the quartz for sweeping.

We know that hydrogen diffuses through Pd,V, and their alloys more easily than it diffuses through Au and Pt - which are normally used as sweeping electrodes. Moreover, these other metals, because of their ionic charge or size, may not diffuse easily into quartz. As a consequence, our future plans include an investigation of the possibility that Pd and V may prove to be superior sweeping electrodes. We also intend to investigate the radiation sensitivity of resonators made from material swept with and without electrode-diffusion suppression.

CONCLUSIONS

The neutron activation analysis shows that only a small amount of metal is diffusing into the quartz from the electrode. However, the metal agglomerates at dislocation sites. These metal-decorated dislocations resist forming etch channels, we feel this contributes to the reduction of the etch-channel density to near-zero levels.

As a result of sweeping, the exchange of H+ for metal-compensating ion occurs throughout the bulk of the sample. Thus, the etching and chemical polishing differences of

surfaces cut from the bulk of the swept material are probable evidence of suppression of the hydrogen diffusion and, less likely, the suppression of the electrode-metal diffusion.

Finally, the reduction in sweeping efficiency observed when suppression-type electrodes are used is probably the result of interference with the dissolution and diffusion of hydrogen in and through the metal-electrode into the quartz.

ACKNOWLEDGMENTS

The author thanks William Washington for the cutting and polishing of samples, Ronald Brandmayr for the etching work and Donald Boyce for the gold evaporations. Acknowledgment is made to Dr. John Vig and Dr. James Shelby, Alfred University, NY, for suggestions and stimulating conversations.

APPENDIX: POLARITY IDENTIFICATION OF SC-SURFACES

Every SC-cut plate has two crystal faces, one has the general Bravais indices $(\bar{h}k.\bar{l})$ and the other $(\bar{h}k.\bar{l})$ when considering the origin to be inside the plate. (See Fig. 11) Interpreting the 1978 IEEE standard, the $(\bar{h}k.\bar{l})$ face should develop a (+) positive charge under tension. It has been shown[11] that this face is etched smoother by using a 1:2 solution of 49% HF: 40% NH4F at 75°C for 30 minutes. It was shown[2] previously that the face which etches smoother is diffusion-controlled and etches faster. To summarize, the smooth face is the face toward which +Y'' points from an origin inside the plate, has the indices $(\bar{h}k.\bar{l})$, and develops a (+) positive charge on tension. The above applies equally well to both enantiomorphs, unless one uses a right-handed coordinate system for left-handed quartz, as the 1978 IEEE standard suggests. In this case, the shiny face would again be positive under tension, but this face would be indexed $(h\bar{k}.l)$.[12]

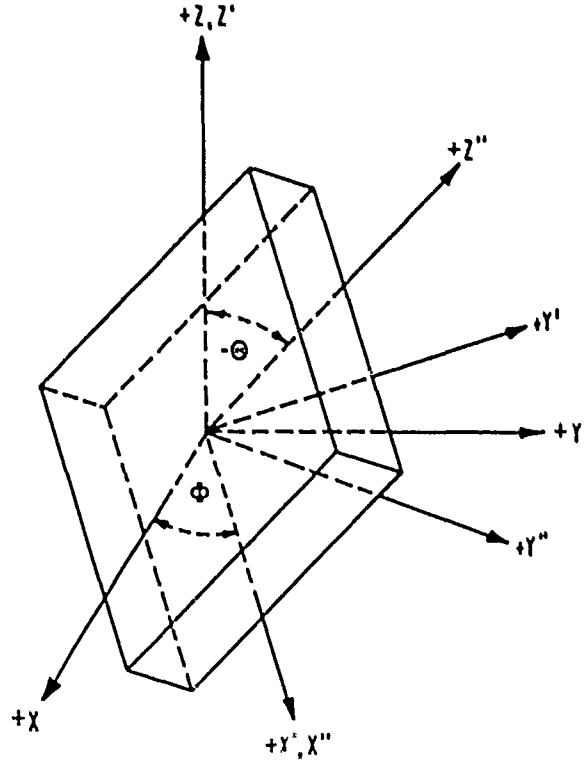


Figure 11. A doubly rotated, SC-cut, right-handed quartz crystal plate with origin inside the plate.

REFERENCES

1. J. G. Gualtieri and D. W. Eckart, "The Influence of Surface Finish and Metallization on Electrode Electromigration in Alpha-Quartz During Sweeping," Proc. 40th AFCS, pp. 115-120 (1986).

2. J. R. Vig, J. W. LeBus and R. L. Filler, "Chemically Polished Quartz," Proc. 31st AFCS, p. 131 (1977).

3. H. G. Lipson, F. Euler and A. F. Armington, "Low-Temperature Infrared Absorption of Impurities in High-Grade Quartz," Proc. 32nd AFCS, pp. 11-23 (1978).

4. 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).

5. A. Kats, "Hydrogen in Alpha Quartz," Philips Res. Repts. Vol. 17, pp. 201-279, esp. pp. 223-228 (1962).

6. B. L. Crawford and H. L. Dinsmore, J. Chem. Phys. 18, p. 983-987 (1950).

7. J. G. Gualtieri and J. Vig, "Sweeping and Irradiation Studies in Quartz," Proc. 38th AFCS, pp. 42-49 (1984).

8. J. J. Martin, R. B. Bossoli, L. E. Halliburton, Brinda Subramaniam and J. D. West, "Electrodiffusion of Charge-Compensating Ions in Alpha-Quartz," Proc. 37th AFCS, pp. 164-168 (1983).

9. R. V. Bucur, V. Mecea and E. Indrea, Jour of the Less-Common Metals V49, p. 147-158, (1976). Reprinted in, Hydrogen in Metals, Proceedings of the Meeting on "Hydrogen in Metals," Elsevier Sequoia S.A., Lausanne (1976).

10. H. Wipf, Jour of the Less-Common Metals, V49, pp. 291-307 (1976). Reprinted in, Hydrogen in Metals, Proceedings of the Meeting on "Hydrogen in Metals," Elsevier Sequoia S.A., Lausanne (1976).

11. J. R. Vig, R. J. Brandmayr and R. L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proc. 33rd AFCS, pp. 351-358 (1979).

12. J. D. H. Donnay and Yvon Le Page, "The Vicissitudes of the Low Quartz Crystal Setting or the Pitfalls of Enantiomorphism," Acta Crystallog Vol. A 34, pp. 584-594 (1978).

41st Annual Frequency Control Symposium - 1987

EFFECTS OF INITIAL QUARTZ SURFACE FINISH AND ETCH REMOVAL ON ETCH FIGURES AND QUARTZ CRYSTAL Q

K. H. Jones
Reeves-Hoffman Division of
Dynamics Corporation of America
Carlisle, Pennsylvania 17013

Abstract

The surface topography of polished, etched quartz crystal plates was compared with crystal plates whose surfaces were not polished before etching. As etching of abrasive lapped plates progressed, micro photography showed the formation of a pattern of hills and valleys characteristic of an AT cut crystal. Polished plates developed etch figures similar in size and orientation to those of the lapped plates. These etch deformations, initiated in some cases at scratch marks, were surrounded by smooth areas of quartz. When these crystals were operated on their fundamental frequency of 20 MHz, lapped plates gave better crystal performance when deeply etched. However, operation of these same crystals on the third and fifth overtones showed opposite results. In these cases, the polished crystal plates had higher Q's. It is suggested that there is a turning point between 30 and 60 MHz at which it becomes advantageous to use polished blanks for the fabrication of deeply etched quartz resonators. A perfectly polished quartz plate would be the optimum starting point for the production of thin, very high frequency crystals.

Introduction

The amplitude of oscillation of a crystal-controlled oscillator is a function of the series resistance of the quartz crystal unit. Too high an equivalent resistance can prevent oscillation from reaching the required amplitude. In order to allow the crystal resonator to have the widest range of uses, the resistance must be as low as possible.

A measure of resonator usefulness related to the resistance is the quality factor, designated Q. It provides a measure of resonator energy losses including mounting losses, losses due to dislocations or imperfections in the quartz crystalline structure, surface stresses and anything else that serves to dissipate energy in the crystal. Q is defined as

$$\frac{(2)\ (pi)\ (\text{energy stored per cycle})}{(\text{energy dissipated per cycle})}$$

and is related to the resistance by the following relationship:

$$Q = 1 / w C_1 R$$

where w = 2 pi times the frequency in hertz and C_1 is the motional capacitance of the crystal unit in farads. The capacitance is affected by the frequency of oscillation and by the electrode size. A high Q is what gives the crystal its ability to control frequency in a way that no other combination of circuit components can.[1]

Castellano[2] and others[3,4,5] have determined that surface imperfections as well as defects in the quartz itself contribute to surface stresses and energy dissipation and result in a lowering of Q. These surface imperfections also serve to raise the resistance of the crystal unit.

Disturbances in the crystalline structure of the quartz contribute to energy losses and a lowering of crystal Q. Defects such as holes in the lattice, interstitial atoms and other impurities interfere with the piezoelectric wave propagation and serve to lower the Q of the resonator produced from that piece of quartz. When quartz is deeply etched, impurities in the quartz may be etched preferentially and cause a localized erosion of the wafer surface. In some cases, these dislocations result in etch pits or in etch channels, which begin at etch pits but propagate into and sometimes all the way through the quartz plate.

Several authors[6,7,8] have conducted experiments to determine the best type of quartz to use for the least amount of etch channels. Hanson has found only natural and vacuum swept premium Q cultured quartz suitable for deep etching.[9] Hunt found that swept premium Q quartz was best for lack of etch channels, followed by natural quartz and unswept cultured quartz.[10] Vig conducted an experiment to determine the effects of vacuum sweeping of quartz on the formation of etch channels in which he had one part of a bar of cultured quartz swept while allowing the other part of the same bar to remain unswept. The results confirmed that the sweeping process, which is known to remove interstitial impurities, also reduces the formation of etch channels.[11]

Because the sweeping process is costly, quartz manufacturers are developing special materials and growth conditions for synthetic quartz which are intended to eliminate impurities and produce high Q values in the resonators which are produced from these bars of quartz. In general, these high Q bars of quartz are grown more slowly than are standard bars of the same size, and while they are less costly than swept quartz, they are still relatively expensive.

It has been shown that smooth surfaces are important for proper quartz resonator performance.[12,13] Smooth surfaces are necessary for electrical response, while a strained surface layer adversely affects resonator performance.[14]

Mechanical lapping, or grinding with a moving abrasive, results in a rough, disordered surface which may tend to lose particles.[15] These surface inhomogeneities, Castellano, Meeker and Sundahl point out, "contribute to surface stresses and energy dissipation and result in noise, rapid aging, drive sensitivity and frequency versus temperature irregularities."[16] Vig and coworkers used reflective high energy electron diffraction (RHEED) to show that when etching had changed the frequency by 1.023 F/f or removed about 1.7 microns of quartz from a crystal which had been lapped with a five micron abrasive, the damaged layer had been removed.[17] Castellano, Meeker and Sundahl also used RHEED to find that the damage had been removed when about 1.1 microns of quartz had been removed.[18] At this point, they found, the surface was entirely crystalline, no longer amorphous.

CH2427-3/87/0000-199 $1.00 C 1987 IEEE

Polishing may follow mechanical lapping in the processing of the surface of quartz plates. Polishing gives the quartz surface a visually smooth, glass-like appearance. Polishing is presently used for manufacture of the highest precision quartz devices. Properly designed devices made with polished blanks have clearly defined resonances, low energy losses (i.e. high electrical Q) and low sensitivity to the power level in the circuit.[19] Vig has shown that cerium oxide polishing, the type most commonly used for quartz polishing, is actually a chemo-mechanical process rather than a purely mechanical one. The cerium oxide must be used in a water solution to have a polishing effect. Cerium oxide and kerosene, he found, had no polishing effect.[20]

The polished quartz surface may, however, have hidden scratches which can be revealed by etching. The mechanical part of the polishing process may cause scratching or lattice disturbance of the surface. These scratches may be covered over and the surface may appear to be polished until an etchant attacks and dissolves this amorphous quartz preferentially, revealing the inhomogeneity. These surface scratches and cracks can amplify surface stresses, and cracks may propagate, degrading the performance of the crystal unit.[21] Vondeling has found that the surface strain in a polished plate is confined to a thin layer near the surface of the plate.[22] Fukuyo and co-workers have used ellipsometry to determine that polishing of fused quartz leaves a damaged layer of about 50 Angstroms, much thinner than that caused by abrasive lapping.[23] Fukuyo and Ours found that the surface layer of a polished crystal plate actually consists of three layers. The first layer is thin, and is characterized by a frequency decrease upon etching of the plate. When layer two is etched, the frequency gradually increases. The etch rate remains linear during etching of the third layer. The thickness of the first layer was found to be dependent on the pressure used during polishing, and the frequency decrease was attributed to relief of surface strain.[24]

A variety of etchants have been used for quartz. For a number of years, the main purpose of etching quartz was to reveal lattice defects in natural quartz bars so that defective sections of the bar might be avoided when fabricating quartz resonators. Val'skaya found several alternatives to hydrofluoric acid which would selectively reveal dislocation features of the quartz.[25] The purpose of the present study is the opposite: to find ways to prepare the quartz surface such that no dislocations or etch pits are revealed. Vondeling found that the addition of potassium fluoride to hydrofluoric acid reduces etch defects and gives a smooth surface.[26] Bernot also found that addition of KF to hydrofluoric acid or ammonium bifluoride at 25 degrees centigrade reduced etch pit and etch channel formation.[27] The standard etchant for AT-cut crystals is ammonium bifluoride which has been shown by Vig and co-workers to produce a chemically polished surface on the lapped surface of AT-cut plates.[28]

The reaction mechanism of etching by acidic fluoride solutions is poorly understood. The original mechanism suggested by Heising,

$$SiO_2+6HF \rightarrow SiF_4+2H_2O+(2HF) \rightarrow H_2SiF_6+2H_2O$$

is considered a simplification by Vig.[29] Judge found that Raman data indicate that no single product species is formed, although some H_2SiF_6 may be formed.[30] This product may also react with silicon dioxide, resulting in an autocatalytic reaction.[31,32] The dissolution rate was found by Judge to be dependent on the HF concentration and on the concentration of HF_2.[33] Vondeling, however, found no effects of HF_2 concentration on the reaction rate. Vondeling suggests that etching is initiated by absorption of positive ions onto the quartz surface, which changes the electronic configuration around silicon and makes possible the dissolution by HF. Vondeling[34] bases his mechanism on that proposed by Ernsberger:[35] The tetrahedral structure of quartz permits either zero, one, two or three dangling oxygen bonds to be present for each silicon exposed on the surface of the crystal. These oxygen atoms are neutralized by hydrogen ions, resulting in exposed hydroxyl groups. In actuality, only those with one and two dangling bonds exist. Those with no exposed bonds are not possible due to the twisted internal structure of quartz, and those with three dangling bonds would be too unstable. This leaves the case of one and two dangling bonds. The case of one exposed hydroxyl group is proposed to be that found on the natural faces of the quartz bar. A structure with two exposed bonds would only occur in an unnatural condition, such as that occurring when a quartz plate is ground, lapped or polished. It is proposed that these surfaces with two dangling bonds will etch preferentially, resulting in a surface whose structure has as many faces parallel to natural faces of the quartz as is possible.[35]

The surface formed after etching with ammonium bifluoride does indeed have a structure suggestive of natural face formation. Several authors have etched and photographed quartz surfaces using a scanning electron microscope showing the formation of hills and valleys or terraces.[36,37,38] The surface of a deeply etched AT-cut quartz crystal develops hills and valleys which extend about four degrees from the crystallographic x-axis.[39] Vig and Tellier have found that different crystal faces have different etching rates[40,41] and Wegner has shown that the appearance of pits and deformations in the surfaces of etched quartz varies in relation to the different faces.[42] Tellier has found that the etch pattern or surface roughness depends on the angle of cut and varies even among wafers cut close to the AT cut. Thus, the etch pattern formed on the surface, Tellier suggests, is determined by orientation rather than by a diffusion controlled process as Vig had suggested.[43,44]

A lapped, etched surface has characteristic hills and valleys. Vig has shown that the roughness of the surface after etching depends on the roughness before etching. Surfaces lapped with 12 micron and 3 micron abrasives became smoother as etching progressed, but the roughness of the 12 micron surface leveled off at a level higher than that of the 3 micron surface. Vig notes that defects etch preferentially because etching relieves the lattice strain they cause and suggests "the less deeply and the more uniformly disturbed the surface is prior to etching, the smoother will be the chemically polished surface. Accordingly, one should lap with as fine an abrasive as possible prior to etching."[45] Hunt concurs and found that the surface produced after etching a one micron surface was scratch-free and "microscopically undulating".[46] Removing the layer of disturbed matter is what improves the crystal Q as etching of lapped blanks progresses.[47]

Polished surfaces are visually smooth before etching and produce resonators with high electrical Q after only a short etching time. Polish scratches that have been covered over during the polishing process will be etched preferentially. This explains the findings of Hanson, that the "surface of a polished AT blank turns frosty when etched with ammonium bifluoride".[48] Until recently, this was also the case at Reeves-Hoffman, the author's company. Polished crystals were used after only a light etch because a deeper etch showed many scratches and degraded resonator performance. Vig also used a light

etch after polishing to give a smooth, crystalline surface.[49] When Hunt attempted deep etching of polished quartz, he found that his polishing process had to be modified to eliminate the hidden damage that etching revealed. Once scratching from the polishing process had been eliminated, Hunt found that fluoride etchants could be used to chemically mill an AT quartz crystal with no apparent degradation of the surface finish.[50] Vig also found that deeply etched polished blanks gave a surface composed only of smooth areas and scratch marks after a etch removal of 92 F_iF_f. He saw no signs of preferential etching along crystallographic axes. He also found that the surface of BT-cut plates remained featureless except at defects such as scratch marks.[51] None of the hills and valleys that appeared after the etching of lapped blanks were discovered.

The effect of the hills and valleys on the quartz crystal Q is difficult to determine. Several authors report that deep etching causes no decrease in Q. Vig chemically polished natural quartz 20 MHz resonators and found no decreasing Q with depth of etch.[52] He obtained similar results on 5 MHz third overtone crystal units and on 10 MHz third overtones.[53] Miller also found that Q did not increase with longer etching.[54] Castellano, however, found that a maximum Q was reached at 1.6 F_iF_f etch removal from a 3 micron lapped surface, and decreased with further etching.[55] This is in agreement with the findings of Jumper at Reeves-Hoffman. A minimum resistance was found after about 1.5 F_iF_f etch removal from 30 MHz fundamentals.[56]

The effect on crystal Q of etching mechanically polished wafers appears to depend on the amount and depth of scratching and other damage caused during the polishing process. When scratches are etched preferentially by the etchant, resonator performance is degraded. Hunt managed to use polished blanks by assuring that they were polished "carefully".[57] Vig has also used polished blanks for resonators which require a deep etch.[58] Deep gouges and scratches result in a higher Q than expected. Vig reports that a deep gouge in a deeply etched 5 MHz third overtone may account for its lower Q as compared to other crystal units in the lot.[59] He notes that the dimensions of etch channels and etch pits (caused by defects) are usually at least one order of magnitude larger than the dimensions of the hills and valleys formed during etching.[60]

Castellano suggests that Q degradation with continued etching of lapped plates is correlated to the depth and width of the hills and valleys formed during etching. These hills and valleys are initiated by damage during mechanical lapping and are enhanced by chemical etching. He uses a Rayleigh wave scattering model to explain the correlation between Q degradation and depth of etch and notes that the optimum etch amount varies with the crystal frequency.[61] Vig also suggests frequency dependence when he notes that he would not be surprised to find a Q degradation because at higher frequencies the wave length of the acoustic wave approaches the dimensions of the etch features.[62]

Vig etched 18-22 MHz crystal resonators with a 3 micron surface fabricated from natural quartz from two to twenty-two F_iF_f. Resistance values were from three to five ohms and Q's ranged from 140,000 to 210,000. He found no Q degradation with depth of etch. Overtone frequencies were not measured in the study because gold had been used for the electrode material and was too heavy to permit overtone operation.[63] Jumper at Reeves-Hoffman also found a difference in frequency, with 1.5 F_iF_f being the optimum etch for a three micron lapped 30 MHz third overtone crystal and 1.3 F_iF_f being optimum for a three micron lapped 60 MHz third overtone. He notes that while electrical parameters, including resistance, of the 30 MHz units were comparable to those obtained with lightly etched cerium oxide polished units, the resistance of the 60 MHz units was significantly below that obtained from cerium oxide polished blanks of the same frequency.[64]

Experimental

The following list, derived in part from Vig and co-workers[65] and Bernot[66], provides an overview of some of the parameters which affect surface topography, resistance and Q of the quartz crystal unit which is produced.

- Quartz material defects
- Crystalline orientation of surfaces
- Quartz plate diameter
- Final plate thickness and frequency
- Initial surface roughness
- Surface cleanliness
- Etchant composition
- Etch bath temperature
- Etch bath agitation
- Etching time
- Electrode characteristics
- Mounting method
- Ambient atmosphere

In order to determine the effects of polishing the surface of the quartz plate at the different etching depths, all the above factors were held as constant as possible. Initial surface roughness and etching time were systematically varied in order to determine their effects.

550 Crystals were fabricated and divided into 22 groups of 25. Half of the groups were polished, while half had a relatively rough lapped surface prior to etching. One group of each of the polished and lapped blanks was etched for each of ten increments of etch removal.

All fabrication was performed at Reeves-Hoffman Division of Dynamics Corporation of America in Carlisle, Pennsylvania.

Water Preparation. All quartz has some material defects. In order to minimize the effects of these defects, all 500 wafers were cut from the same bar of high quality cultured quartz. This quartz is certified by the supplier as having a minimum Q of 2.2 million.

The crystalline orientation of the surfaces was held constant by assuring that all the wafers were cut at the same angle from the quartz bar. All the wafers were cut at 35 degrees 16 minutes, plus or minus 30 seconds from the crystallographic x-axis. This AT angle was verified by x-ray. This angle was chosen with the understanding that these crystals would be studied on the fundamental, third, and fifth overtone modes of operation. This required a compromise in terms of frequency stability over temperature and resulted in a chosen angle that was not optimum for any of the modes of operation, yet minimized any chances of extreme sensitivity to temperature.

The crystal plate diameter was 0.250 inches, plus or minus 0.001 inches.

In the case of the AT cut, the final plate thickness is proportional to the fundamental frequency in which it operates. This means that in order to have 550 20 Megahertz crystals, all of which have had different amounts of quartz removed by etching, the frequencies (and thicknesses) prior to etching must be

systematically varied to allow the final frequency to be the same for each group. Thus there was a group of polished and a group of lapped wafers prepared at each of ten different thicknesses corresponding to the 10 different amounts of etch removal specified later in the processing.

The blanks which were to have lapped surfaces were prepared as follows. The thickness of the wafers which had been cut from the bar of quartz was gradually reduced by removing quartz by grinding (lapping) with abrasives of gradually decreasing sizes. Care was taken during each step to grind away sufficient thickness to remove the rougher damaged layer from the step before. The final surface finish was produced by lapping with an abrasive with an average particle size of three microns.

The polished wafers were prepared in the same manner as described above, with two additional steps. After lapping with the three micron abrasive, the blank surfaces were etched in a saturated aqueous solution of ammonium bifluoride to remove as much of the lapping damage as possible. They were then chemo-mechanically polished with a cerium oxide polish suspended in a water-based slurry solution. This particular brand of cerium oxide polishing compound was chosen because it resulted in the least amount of scratching when compared to several other polishing compounds, including one which had been used previously at Reeves-Hoffman.

Surface cleanliness was maintained throughout each step. Blanks were cleaned after each cutting and grinding operation to be sure that operations which use smaller abrasive sizes were not contaminated with the larger abrasive particles from previous steps which would cause scratching. After the final lapping or polishing operation, the blanks were cleaned ultrasonically in two different proprietary cleaners, rinsed and dried. Following this operation, the blanks were handled only by persons wearing finger cots to prevent finger oils and contaminants from adhering to the wafer surfaces.

The wafers were then etched in a saturated aqueous solution of ammonium bifluoride. A polypropylene etch tank was filled with enough distilled water and ammonium bifluoride flakes such that a small amount of flakes remained at the bottom of the tank during etching, assuring a saturated solution at all times. Ammonium bifluoride flakes were added during use of the bath as necessary to assure a saturated solution at all times. The etching solution was discarded after being used for one week to eliminate the effects of any impurities which may contaminate the bath. A small amount of a proprietary fluorochemical surfactant was added to the bath to assure wetting and to minimize the effects of any residue on the wafer surfaces.

The temperature of the etching solution was monitored directly by a temperature controller with a teflon probe, and kept at a constant 52 degrees centigrade. This temperature was chosen because it allows etching in a reasonable amount of time while also allowing adequate control of frequency by adjusting the etching time. Constant agitation was provided by a motor-driven rotary propeller.

After etching, the wafers are vapor-plated with a flash of chromium followed by a layer of silver sufficient to raise the frequency by approximately 300 KHz. This corresponds to a silver layer about .000475 inches thick. This electrode plating thickness was chosen based on experience that it was heavy enough for operation of a 20 MHz fundamental yet light enough to allow operation on the third and fifth overtones. It would not, therefore, provide optimum Q on any of these crystals. The electrode diameter was 0.115 inches, again chosen to allow the best operation of these crystals in all three modes of vibration.

The plated crystal wafers were then mounted in spring-type holders and pasted in with silver- filled conductive epoxy. The crystals were mounted in the holder with the positive end of the crystallographic x-axis at the top of the holder.

The crystal frequencies received their final adjustment during the next operation - electroplating. The crystals were electroplated with silver from a silver chloride plating solution to within five parts per million of 20 Megahertz on the fundamental.

The crystals were then baked under vacuum for one hour to assure complete dryness and to remove contaminates. They were then sealed under vacuum by the cold welding process.

Crystal Unit Testing. The frequency, resistance, motional capacitance and Q of each crystal unit was measured using a Saunders Automatic Testing System incorporating a Hewlett-Packard 3577A Network Analyzer at a constant 25 degrees centigrade. Each of these parameters was measured at the fundamental frequency of 20 MHz, the third overtone (approximately 60 MHz) and the fifth overtone (approximately 100 MHz). All crystal measurements were taken consecutively, on the same day, to avoid any slight differences in equipment or temperature. The frequency and resistance were also checked on a Saunders 150C Crystal Impedance Meter to verify the automatic testing system readings.

Results and Discussion

Photomicrographs. The optical microscope used in this study provided a good alternative to scanning electron microscopy when the wafer surface was coated with the reflective silver electrode. It was a convenient, low-cost method of viewing the surfaces of wafers in production. A thicker layer of silver may obscure some of the etch figures if it fills in the crevices formed during etching. The silver does have the disadvantage of obscuring any etch channels which may penetrate the quartz plate.

Figures 1 through 11 show typical surfaces of 3-micron lapped blanks after 0.02, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 F_iF_f increments of etch removal. This variety of thickness change during etching required that the initial thickness of the wafers in each group be adjusted prior to etching by the grinding procedure described previously, in order to have each group of crystals at the same thickness and frequency. As the etch time increases, the surface changes from an irregular, small-featured one to a more regular pattern of hills and valleys. These patterns are similar to those identified by Vig and others, indicating that no unusual differences in surface finish were introduced into this study.[67,68] Each of the wafers was oriented such that the positive end of the crystallographic x-axis points towards the bottom of the page. This is consistent with Vig's findings that the hills and valleys run about 4 degrees from the x-axis.[69]



Figure 1
Lapped Quartz, Etched 0.02 F_iF_f @ 1200X

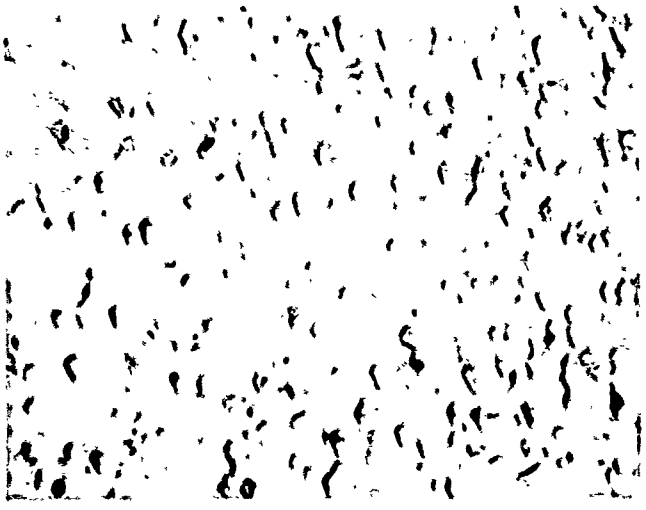


Figure 4
Lapped Quartz, Etched 3 F_iF_f @ 1200X

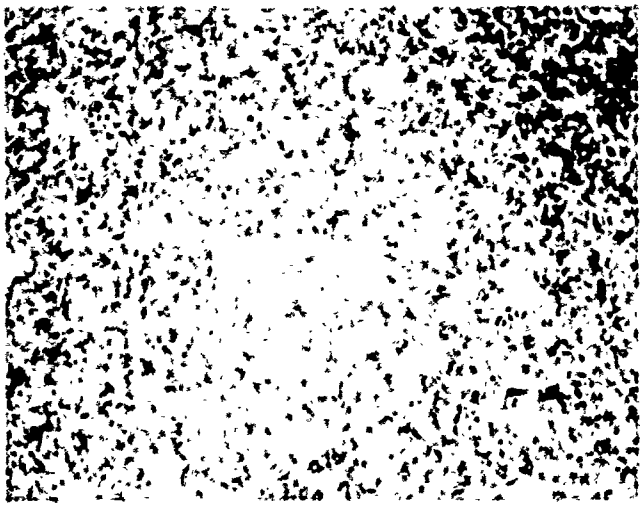


Figure 2
Lapped Quartz, Etched 1 F_iF_f @ 1200X

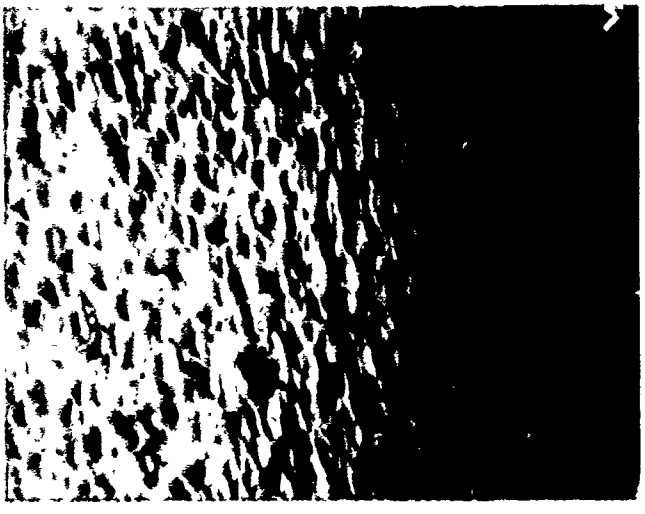


Figure 5
Lapped Quartz, Etched 4 F_iF_f @ 1200X

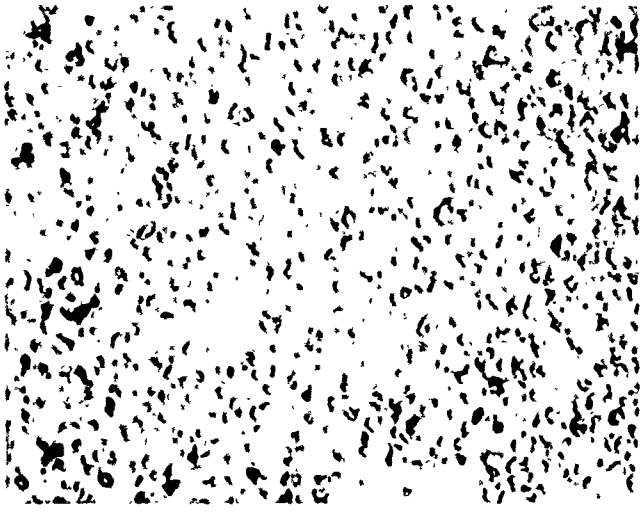


Figure 3
Lapped Quartz, Etched 2 F_iF_f @ 1200X

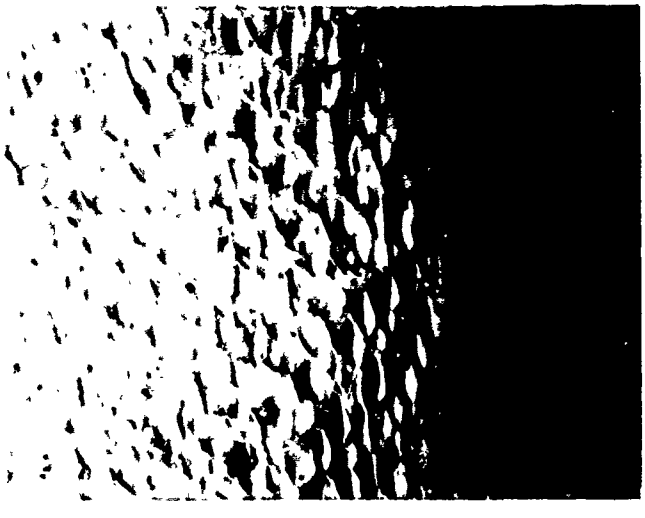


Figure 6
Lapped Quartz, Etched 5 F_iF_f @ 1200X

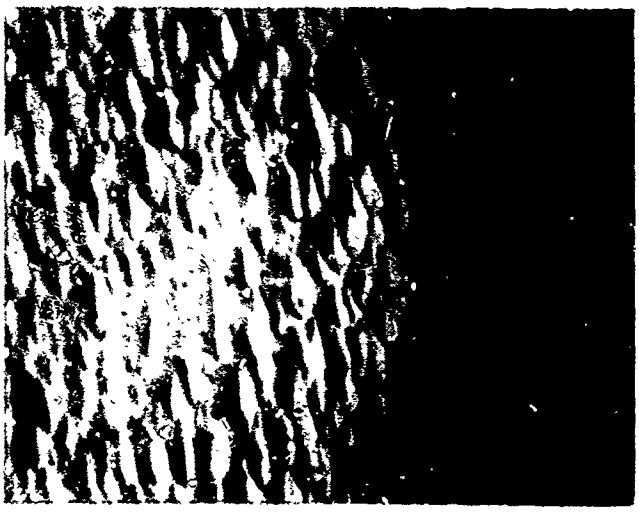


Figure 7
Lapped Quartz, Etched 6 F_iF_f @ 1200X



Figure 10
Lapped Quartz, Etched 9 F_iF_f @ 1200X

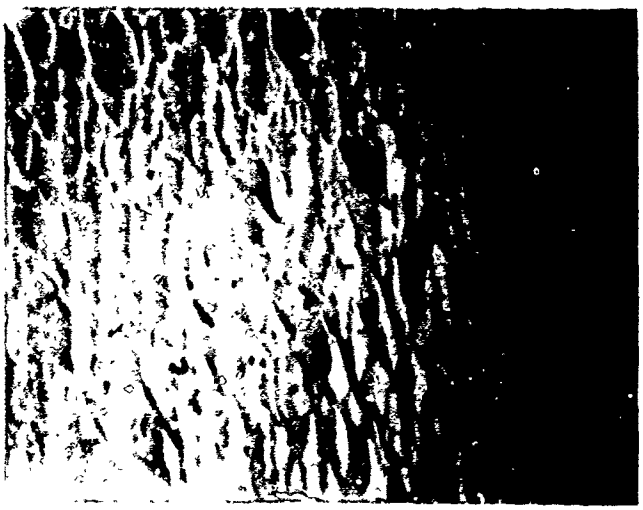


Figure 8
Lapped Quartz, Etched 7 F_iF_f @ 1200X

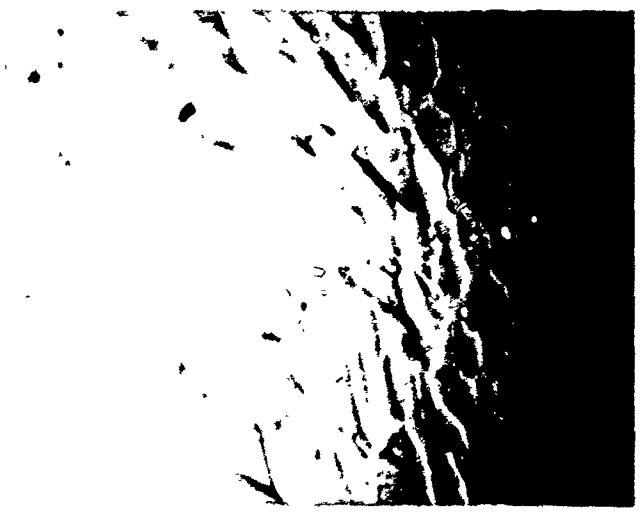


Figure 11
Lapped Quartz, Etched 10 F_iF_f @ 1200X

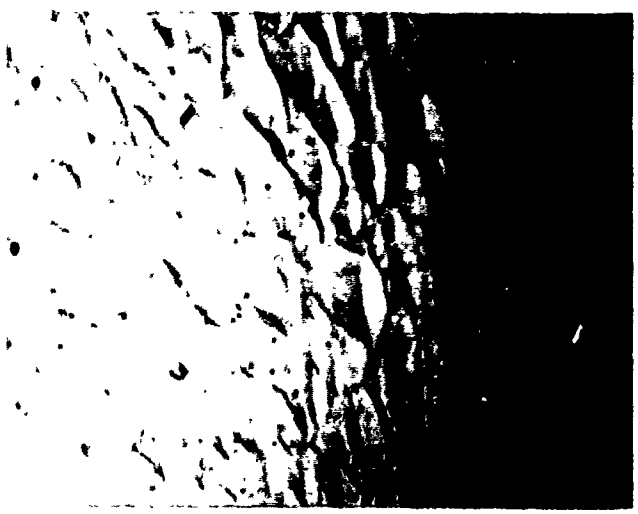


Figure 9
Lapped Quartz, Etched 8 F_iF_f @ 1200X

Figures 12 through 22 show surfaces of polished wafers after the same amounts of etch removal: 0.02, 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10 units of F_iF_f. The polished blank after 0.02 F_iF_f removal is visually smooth. This is the etch removal amount typically used for production of high quality, polished resonators, since it provides a cleaning of the surface yet does not reveal any of the polishing damage. As etching progresses, etch pits are formed and gradually progress into larger crevices which resemble the hills and valleys seen on the lapped blanks in size and orientation. Large areas of perfectly smooth surface are still present, indicating that when there is no surface damage, etched quartz remains smooth. Once damage is initiated, however, the features formed are determined by the crystallographic orientation of the quartz surfaces.



Figure 12
Polished Quartz, Etched 0.02 F_iF_f @ 1200X

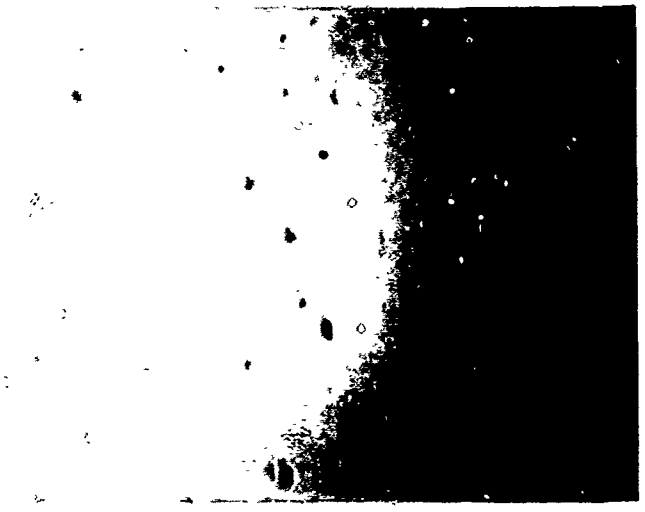


Figure 13
Polished Quartz, Etched 1 F_iF_f @ 1200X

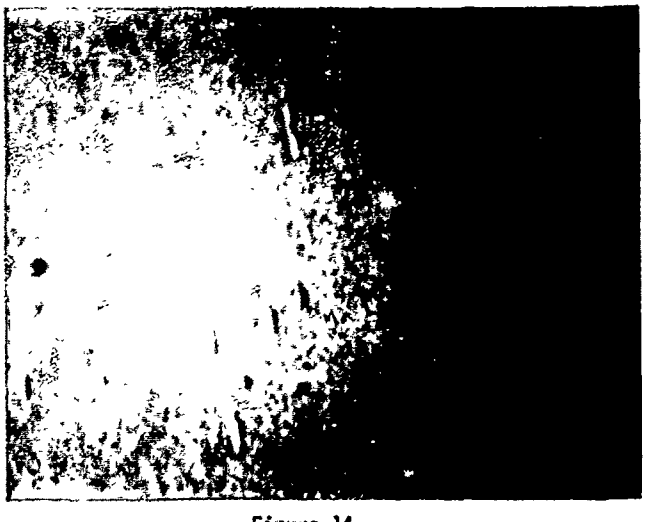


Figure 14
Polished Quartz, Etched 2 F_iF_f @ 1200X



Figure 15
Polished Quartz, Etched 3 F_iF_f @ 1200X



Figure 16
Polished Quartz, Etched 4 F_iF_f @ 1200X

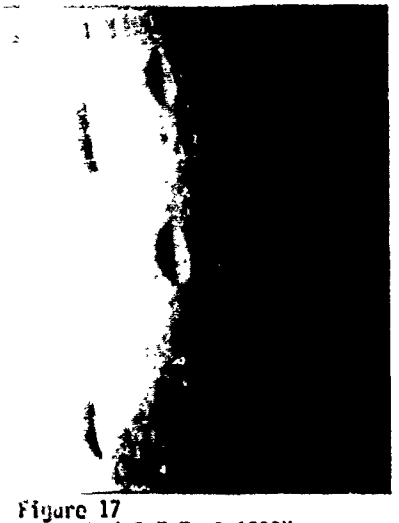


Figure 17
Polished Quartz, Etched 5 F_iF_f @ 1200X



Figure 18
Polished Quartz, Etched 6 F_iF_f @ 1200X



Figure 21
Polished Quartz, Etched 9 F_iF_f @ 1200X

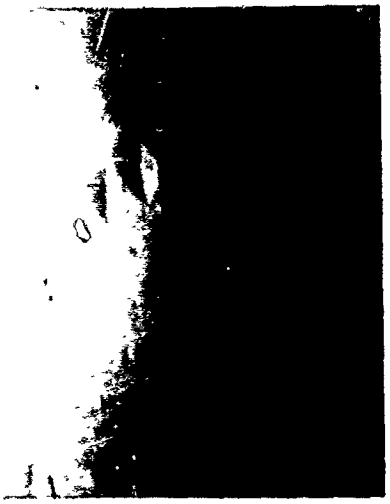


Figure 19
Polished Quartz, Etched 7 F_iF_f @ 1200X



Figure 22
Polished Quartz, Etched 10 F_iF_f @ 1200X

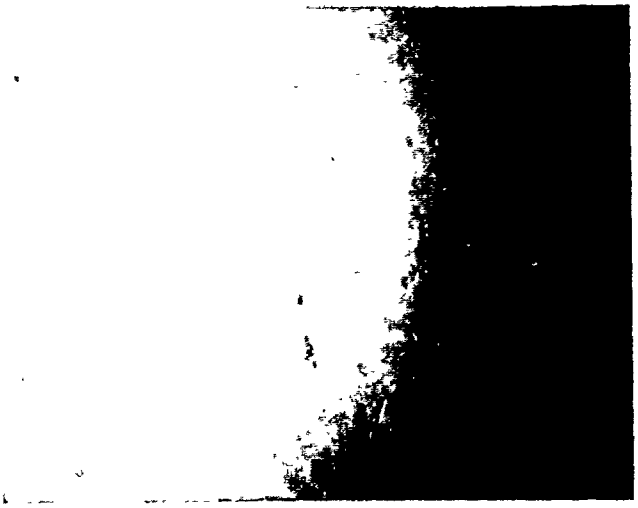


Figure 20
Polished Quartz, Etched 8 F_iF_f @ 1200X

Figure 23 shows a polished blank which has been etched for 10 F_iF_f under 200X magnification. This shows the portion of the surface area which is damaged in relation to that which is still smooth after the deep etch. Figure 24 shows the type of polish scratches which may result from edge chipping during the polishing process. Figure 25 is an example of the scratching and surface damage produced by a poor polishing process.

The fact that a better polish produced large areas of texture-free quartz surface indicates that if the polishing process were to be perfected, completely smooth surfaces could be retained after deep etching. This has important implications for the production of high frequency resonators which would have to be of thicknesses unattainable with mechanical procedures. These ultra-thin blanks require defect-free surfaces after their thickness is removed with deep etching.

Figure 26 gives an example of the different surfaces produced by different etchants. This three micron lapped blank was deep etched with a proprietary etchant reputed to have "chemical polishing" properties. Figure 26 indicates the importance of the chemistry of the etchant in the formation of etch figures.

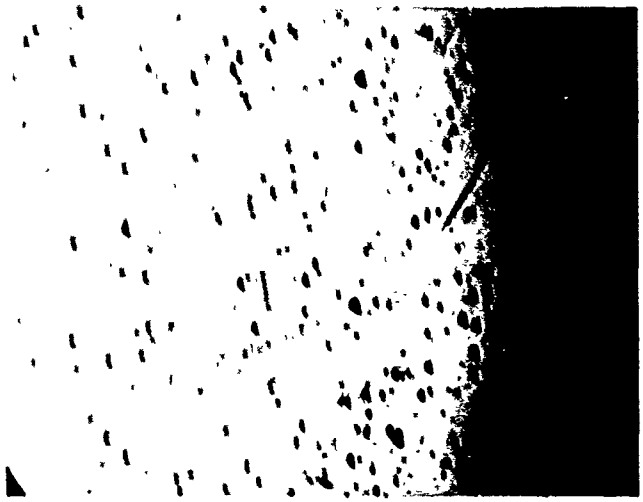


Figure 23. Polished Quartz, Etched 10 F_iF_f @ 150X

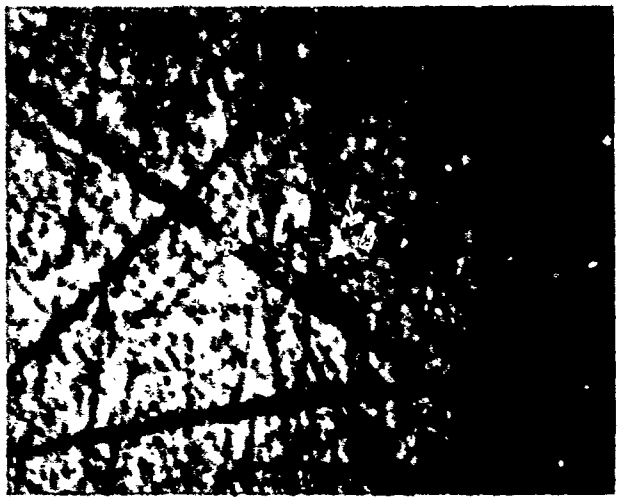


Figure 25. Poorly Polished Quartz with Major Scratches Revealed by Etching @ 1200X

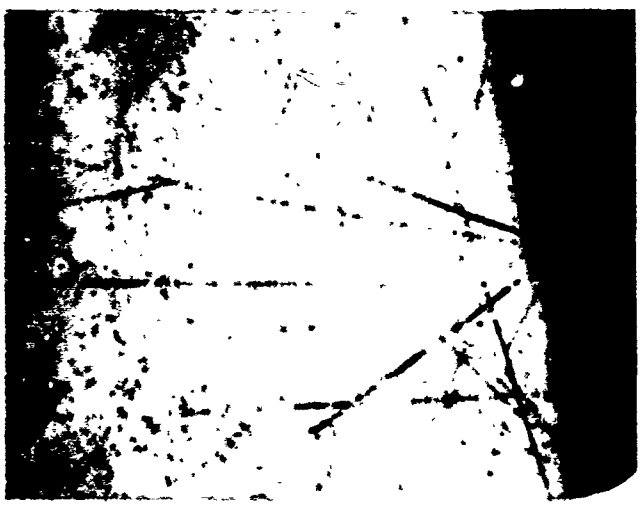


Figure 24. Polished Quartz Showing Scratches from Edge Chips @ 600X



Figure 26. Lapped Quartz, Etched with a Proprietary Etchant, Etched 9.4 F_iF_f @ 1200X

Part II: Electrical Parameters

Each data point in Figures 27 through 35 represents an average for the 18-25 crystals which were etched together.

Figure 27 shows the effect of etch removal on crystal resistance at the fundamental frequency of 20 MHz. For lapped blanks, the crystal units show an initial drop in resistance as the original lapping damage is reduced. The resistance then increases after reaching a minimum between 2 and 3 F_iF_f. This increase may be attributed to the increase in the size of the hills and valleys which is occurring at this point, as shown in the photomicrographs, or may be affected by the formation of etch channels which may be beginning to form. Polished blanks did not show a consistent change over the range of etch removals, and were advantageous over the lapped only in the first two groups. For deeper etches at this frequency, a lapped blank should be used.

Figure 28 presents the effects of etch removal on the same crystals while operating on the third overtone frequency of approximately 60 MHz. The effect here is opposite to that found on the fundamental frequency: polished plates outperformed the lapped plates over the entire spectrum of etch removals. This effect must be due in some way to the frequency of operation since no other factor can be varied: these are the same crystal units. The surface finish, electrode, design, ambient temperature, etc., must all be identical between the two groups.

Figure 29, resistance versus etch removal for the fifth overtone, 100 MHz, follows the same pattern as does the third overtone. Polished blanks again showed lower resistance values across the range of etch removals.

Figures 30 through 32 show the same effect in terms of Q, which is inversely proportional to the resistance. On the fundamental, lapped blanks have the highest value at all but the first two increments of etch removal. At 60 and 100 MHz, the polished blanks have higher Q values, and would be preferred for achieving resonator performance.

Figures 33 through 35 plot the motional capacitance against the etch removal. This value is in practice determined mainly by the electrode area. The change in the initial capacitance of the lapped units at 60 and 100 MHz is most likely due to the decrease in surface area of the quartz that occurs when the initial rough layer is removed. The remaining values

are fairly consistent and give an indication that these crystals were prepared in a uniform manner.

These data are consistent with those found by Jumper who found that deep etching of a lapped 30 MHz crystal (third overtone) was advantageous over cerium oxide polishing and lightly etching the same crystal unit. The same deep etching of a 60 MHz (third overtone) crystal produced higher resistances than did light etching of a chemomechanically polished resonator.

This is also consistent with the findings of this author that 60 MHz fundamentals have a higher Q value when prepared with polished blanks before etching to a depth of 13.5 KÅ than when prepared from three micron lapped blanks and etched to the same depth.

It appears that for chemically polished blanks, there is a point somewhere between 30 and 60 MHz where it is advantageous to switch to a polished blank for optimum Q and low electrical resistance, independent of the harmonic of operation.

As suggested by Vig, as frequencies increase, the wavelength of the acoustic piezoelectric wave is becoming smaller, and more on the order of the size of the etch features forming during etching. Thus, the etch features become more able to disturb the wave propagation.

It appears to be advantageous at lower frequencies to have an unstrained, but uniformly disturbed surface over one which has smooth areas yet irregular scratch deformations. At the higher frequencies, however, the smooth areas provide an advantage even though they are interrupted by scratch marks and pits.

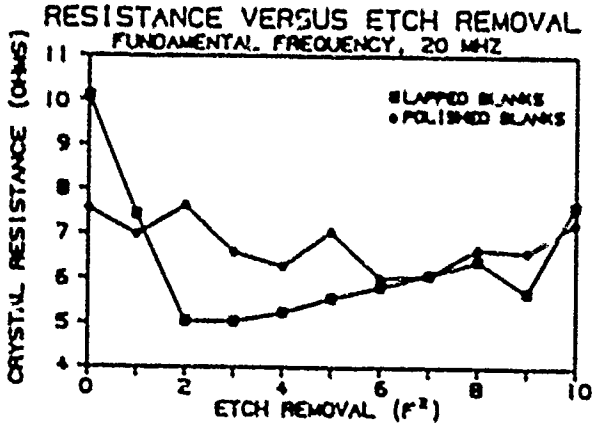


Figure 27

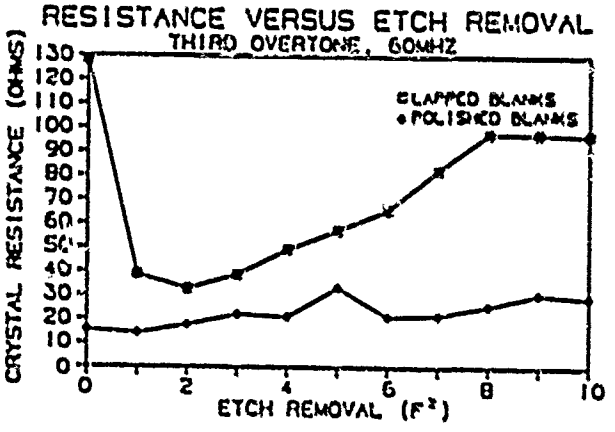


Figure 28

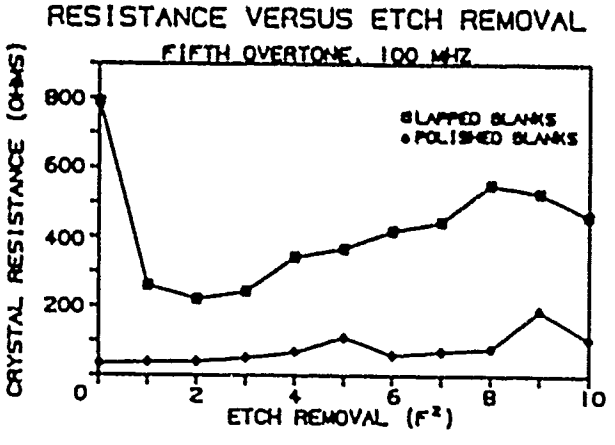


Figure 29

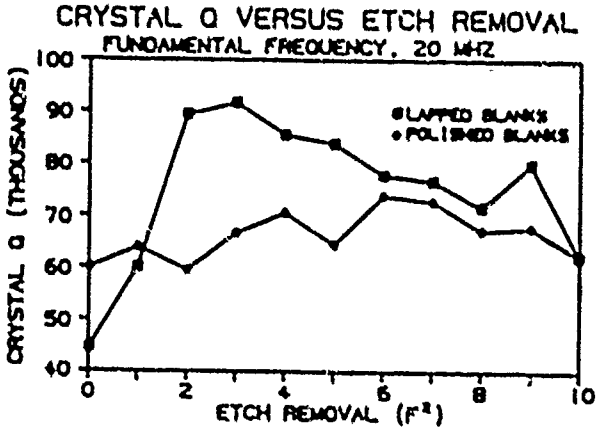


Figure 30

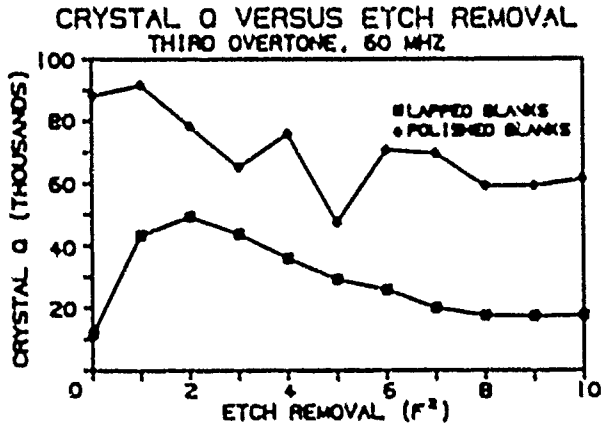


Figure 31

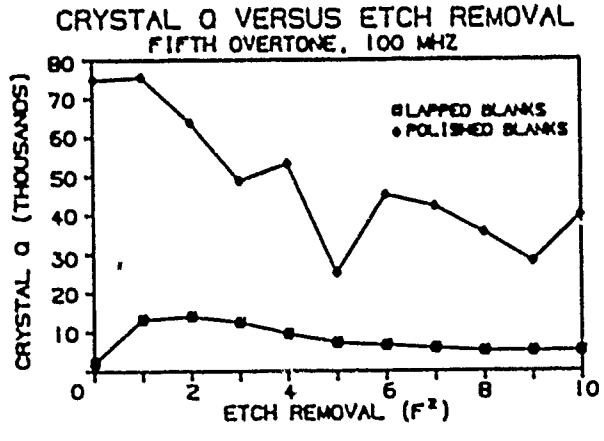


Figure 32

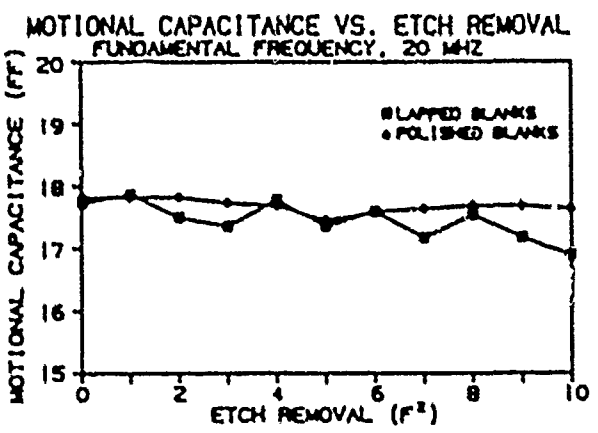


Figure 33

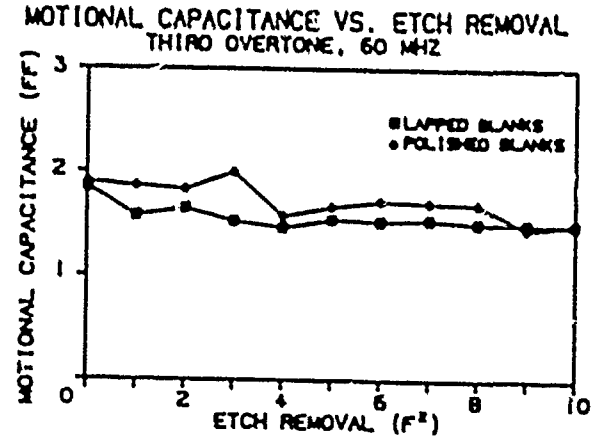


Figure 34

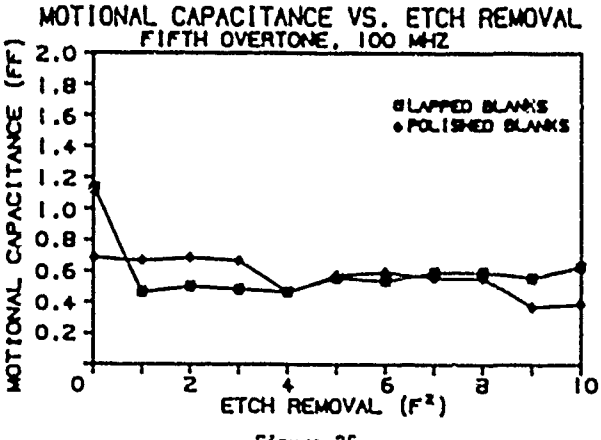


Figure 35

Summary

Photomicrographs of quartz crystal surfaces were easily made using an optical microscope when the wafer was plated with a reflective layer of silver. This technique provided an easy, less expensive alternative to scanning electron microscopy.

As quartz wafers which had been lapped with a three micron abrasive were etched with ammonium bifluoride, an initial rough, amorphous layer was removed and replaced with characteristic hills and valleys which had a specific orientation with regard to the crystallographic axes of the quartz. When polished plates were etched, pits or valleys were formed only in localized areas, starting in many cases where damage had occurred during the polishing process. Many areas remained smooth.

There is a turning point somewhere between 30 and 60 MHz at which it becomes advantageous to use polished blanks, even if there is some hidden damage, for production of deeply etched high frequency resonators. A perfectly polished wafer would be the optimum starting point for the production of thin, very high frequency AT crystals.

Notes

[1]Virgil E. Bottom, Introduction to Quartz Crystal Unit Design (New York: Van Nostrand Reinhold Company, 1982), 101-102.

[2]R. M. Castellano, T. R. Meeker, and R C. Sundahl, "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators," Proceedings of the 31st Annual Symposium on Frequency Control (1977): 126.

[3]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 292.

[4]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[5]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304-305.

[6]William P. Hanson, "Chemically Polished High Frequency Resonators," Proceedings of the 37th Annual Symposium on Frequency Control (1983): 261-264.

[7]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 294

[8]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 21-27.

[9]William P. Hanson, "Chemically Polished High Frequency Resonators," Proceedings of the 37th Annual Symposium on Frequency Control (1983): 261-264.

[10]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 294.

[11]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 21-27.

[12]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304-305.

[13]A. J. Miller, "Preparation of Quartz Crystal Plates for Monolithic Crystal Filters," Proceedings of the 24th Annual Symposium on Frequency Control (1970): 93-103.

[14]Y. Sekiguchi and H. Funakubo, "Strained Surface Layers of Quartz Plates Produced by Lapping and Polishing and Their Influence on Quartz Resonator Performance," Journal of Materials Science 15 (1980): 3066.

[15]John R. Vig, H. Wasshousen, C. Cook, M. Katz and E. Hafner, "Surface Preparation and Characterization Techniques for Quartz Resonators," Proceedings of the 27th Annual Symposium on Frequency Control (1973): 98.

[16]R. M. Castellano, T. R. Meeker, and R. C.Sundahl, "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators," Proceedings of the 31st Annual Symposium on Frequency Control (1977): 126.

[17]John R. Vig, C. F. Cook, Jr., K. Schwidtal, J. W. LeBus, and E. Hafner, "Surface Studies for Quartz Resonators," Proceedings of the 28th Annual Symposium on Frequency Control (1974): 102.

[18]R. M. Castellano, T. R. Meeker, and R.C. Sundahl, "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators," Proceedings of the 31 st Annual Symposium on Frequency Control (1977): 127.

[19]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304.

[20]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[21]John R. Vig, H. Wasshousen, C. Cook, M. Katz and E. Hafner, "Surface Preparation and Characterization Techniques for Quartz Resonators," Proceedings of the 27th Annual Symposium on Frequency Control (1973):98.

[22]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 308.

[23]Hitohiro Fukuyo, Nobunori Oura, Nobuyuki Kitajima and Hiromichi Kono, "The Refractive-Index Distribution Normal to the Polished Surface of Fused Quartz Measured by Ellipsometry," Journal of Applied Physics 50, No. 5 (May 1979): 3653.

[24]Hirohiro Fukuyo and Nobunori Oura, "Surface Layer of a Polished Crystal Plate," Proceedings of the 30th Annual Symposium on Frequency Control (1976): 254-258.

[25]O. B. Vol'skaya, "Selective Etchants for Quartz," Soviet Physics--Crystallography 13, No. 4 (Jan-Feb 1969): 620-622.

[26]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304.

[27]Anthony J. Bernot, "Etching Study of AT-Cut Cultured Quartz Using Etchants Containing Fluoride Salts, Hydrofluoric Acid and Ammonium Bifluoride," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 271-275.

[28]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 33.

[29]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[30]John S. Judge, "A Study of the Dissolution of Silicon Dioxide in Acidic Fluoride Solutions" Journal of the Electrochemical Society (November 1971): 1772-1775.

[31]John S. Judge, "A Study of the Dissolution of Silicon Dioxide in Acidic Fluoride Solutions" Journal of the Electrochemical Society (November 1971): 1772-1775.

[32]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[33]John S. Judge, "A Study of the Dissolution of Silicon Dioxide in Acidic Fluoride Solutions" Journal of the Electrochemical Society (November 1971): 1772-1775.

[34]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304.

[35]F. M. Ernsberger, "Structural Effects in the Chemical Reactivity of Silica and Silicates" Journal of the Physical Chemistry of Solids 13 (1960): 347-351.

[36]J. K. Vondeling, "Fluoride-Based Etchants for Quartz," Journal of Materials Science 18 (1983): 304.

[37]Anthony J. Bernot, "Etching Study of AT-Cut Cultured Quartz Using Etchants Containing Fluoride Salts, Hydrofluoric Acid and Ammonium Bifluoride," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 271-275.

[38]John R. Vig, C. F. Cook, Jr., K. Schwidtal, J. W. LeBus, and E. Hafner, "Surface Studies for Quartz Resonators," Proceedings of the 28th Annual Symposium on Frequency Control (1974): 96-108.

[39]John R. Vig, John W. LeBus, and Raymond L.Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 4.

[40]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[41]C. R. Tellier, "Etch Figures and Etch Rates in AT, BT, X and Y Cut Quartz Plates," Proceedings of the 38th Annual Symposium on Frequency Control (1984): 105.

[42]M. W. Wegner and J. M. Christie, "Chemical Etching of Deformation Substructures in Quartz," Physics and Chemistry of Minerals 9 (1983): 67-78.

[43]C. R. Tellier, "Surface Texture of Chemically Etched AT-Cut Quartz Plated," Surface Technology 21 (1984): 83-89.

[44]C. R. Tellier, "Effect of Crystal Orientation on the Surface Texture of Chemically Etched Quartz Plates, the Case of Cuts Close to the AT Cut," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 282-291.

[45]John R. Vig, John W. LeBus and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[46]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 294.

[47]A. J. Miller, "Preparation of Quartz Crystal Plates for Monolithic Crystal Filters," Proceedings of the 24th Annual Symposium on Frequency Control (1970): 93-103.

[48]William P. Hanson, "Chemically Polished High Frequency Resonators," Proceedings of the 37th Annual Symposium on Frequency Control (1983): 261-264.

[49]John R. Vig, H. Wasshousen, C. Cook, M. Katz and E. Hafner, "Surface Preparation and Characterization Techniques for Quartz Resonators," Proceedings of the 27th Annual Symposium on Frequency Control (1973): 98.

[50]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 294.

[51]John R. Vig, Ronald J. Brandmayr and Raymond L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proceedings of the 33rd Annual Frequency Control Symposium (1979): 351-358.

[52]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[53]John R. Vig, Ronald J. Brandmayr and Raymond L. Filler, Etching Studies on Singly and Doubly Rotated Quartz Plates," Proceedings of the 33rd Annual Frequency Control Symposium (1979): 351-358.

[54]A. J. Miller, "Preparation of Quartz Crystal Plates for Monolithic Crystal Filters," Proceedings of the 24th Annual Symposium on Frequency Control (1970): 93-103.

[55]R. N. Castellano, T. R. Meeker, and R. C. Sundahl, "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators," Proceedings of the 31st Annual Symposium on Frequency Control (1977): 126.

[56]Jim Jumper, "WR1173: Chemical Polishing," Reeves-Hoffman Technical File (1983): 1-8.

[57]J. R. Hunt and R. C. Smythe, "Chemically Milled VHF and UHF Resonators," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 294.

[58]John R. Vig and Arthur Ballato, "Method of Making Miniature High Frequency SC-Cut Quartz Crystal Resonators," U.S. Patent No. 4,554,717 (November 26, 1985).

[59]John R. Vig, Ronald J. Brandmayr and Raymond L. Filler, "Etching Studies on Singly and Doubly Rotated Quartz Plates," Proceedings of the 33rd Annual Frequency Control Symposium (1979): 351-358.

[60]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[61]R. N. Castellano, T. R. Meeker, and R. C. Sundahl, "The Relationship Between Quartz Surface Morphology and the Q of High Frequency Resonators," Proceedings of the 31st Annaul Symposium on Frequency Control (1977): 126-130.

[62]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[63]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[64]Jim Jumper, "WR1173: Chemical Polishing," Reeves-Hoffman Technical File (1983): 1-8.

[65]John R. Vig, "Etching Quartz Crystals," Proceedings of the 1981 Quartz Crystal Conference (1981).

[66]Anthony J. Bernot, "Etching Study of AT-Cut Cultured Quartz Using Etchants Containing Fluoride Salts, Hydrofluoric Acid and Ammonium Bifluoride," Proceedings of the 39th Annual Symposium on Frequency Control (1985): 271-275.

[67]John R. Vig, C. F. Cook, Jr., K. Schwidtal, J. W. LeBus, and E. Hafner, "Surface Studies for Quartz Resonators," Proceedings of the 28th Annual Symposium on Frequency Control (1974): 96-108.

[68]John R. Vig, John W. LeBus, and Raymond L. Filler, "Chemically Polished Quartz," Research and Development Technical Report ECOM-4548 (November 1977): 1-33.

[69]Jim Jumper, "WR1173: Chemical Polishing," Reeves-Hoffman Technical File (1983): 1-8.

41st Annual Frequency Control Symposium - 1987

A STUDY OF DISLOCATIONS AND INCLUSIONS IN ALPHA QUARTZ

A. F. Armington, J. A. Morrigan, M. T. Harris
RADC/ESM
Hanscom AFB, MA 01731
and
J. F. Balascio
Motorola, Inc.
Carlisle, PA 17013

Summary

The presence of dislocations and inclusions in alpha quartz and their effect upon electronic applications have received very little attention until the last few years. Some studies have been performed on dislocations, but inclusions have not been important until recently since their dimensions have not approached the thickness of the resonator. Recently however, there has been a requirement to produce thinner resonators for higher frequency applications. The dimensions of these resonators are approaching the dimensions of the inclusions in alpha quartz and may have an effect upon its electroacoustic properties.

Several techniques were employed in an effort to reduce the dislocation density in alpha quartz. Annealed and swept seeds had essentially the same dislocation density as its untreated counterpart. The utilization of seeds specifically doped with germanium, iron, aluminum and boron resulted in the growth of alpha quartz which exhibited the same dislocation density as quartz grown upon undoped seeds.

Inclusion density measurements were performed on crystals grown in lined and unlined autoclaves. A variety of nutrient sources were employed in these runs. The data obtained indicated that the inclusion density was neither a specific function of nutrient quality nor whether a liner was employed. The majority of inclusions seemed to be voids with liquid occluded in many of them.

Finally, initial results indicate that impurities other than lithium and sodium are electrically swept out of the dislocations in alpha quartz.

Introduction

Research on the purification of quartz has been extensively studied in the past few years with the result that high purity material is now available especially in terms of aluminum content. Aluminum levels of 0.5 parts per million or even lower can be routinely obtained for resonators used in radiation environments.[1] The presence of dislocations and inclusions in quartz used for piezoelectric applications has received very little attention until recently. Dislocations have been determined to be the precursors of the majority of etch channels in quartz and may have some deleterious effects on the deposition of electrodes and epitaxial devices on the quartz surface.[2] In this paper, we will present some data that dislocations may also be important in the sweeping of quartz.

Inclusions in quartz used for timing applications have not been significant since the size of the inclusions has not approached the minimum diameter of the resonator. It has been reported[3] that smaller inclusions do not affect resonator performance. Recently, however, there have been attempts to produce higher frequency oscillators resulting in smaller resonators where the size of the inclusions could effect resonator performance. Some of the initial results of our studies on quartz inclusions will also be presented in this paper.

The autoclaves used in this study have been described previously[4] as well as the procedures used. Experiments are performed in unlined and silver lined autoclaves with autoclave diameters varying from one inch to industrial scale equipment. In general, a hydroxide mineralizer was used although we employed carbonate on some of the industrial sized runs. A variety of synthetic nutrients were employed.

Results and Discussion

Dislocation Studies - It has been shown that the use of a seed cut from the +X section of synthetic quartz can result in crystals almost free of dislocations.[5] It would be desirable however, to determine a method of pretreating the more commonly used Z seed for the production of low dislocation material. We have attempted several methods so far.

Several annealing and sweeping studies were performed, both in vacuum and air. We tried to anneal a whole crystal (Fig. 1). This resulted in a highly strained crystal with most of the strain emanating from the seed-crystal interface. It also appears that the strain around the inclusions was larger after annealing. Crystals were also grown upon air swept and vacuum swept seeds. Some seeds were coated with platinum electrodes and some were uncoated. Other seeds were only annealed at 500 degrees C.

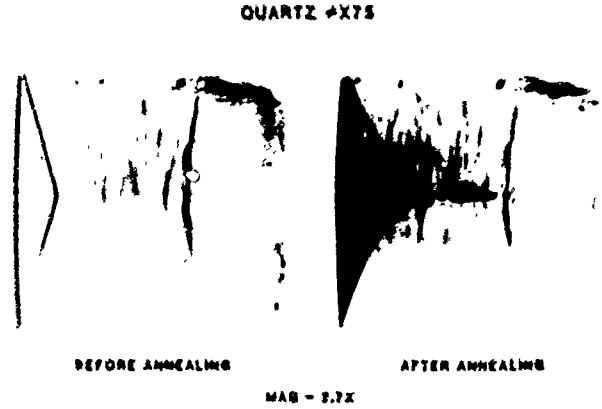


Fig. 1 Strain produced by annealing whole crystal.

One run was completed using a seed with a thin platinum coating on it. There are unofficial reports that this will lower the dislocation count. The results of some of this study are shown on Fig. 2.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

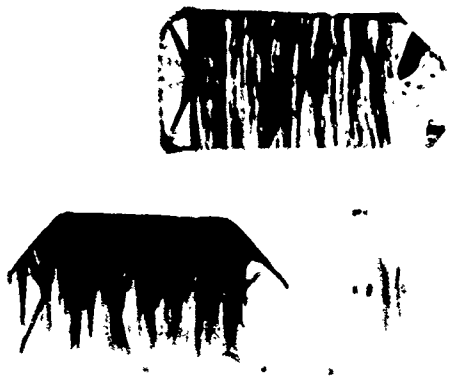


Fig. 2 Topographs of typical pretreatment results on Z seeds.

It is apparent that none of the techniques is effective in significantly lowering the dislocation count. Annealing at 600 degrees C resulted in a twinned crystal as one might expect. In another experiment, only half the seed was swept and the rest of the seed was kept outside the electrodes. X-ray topography however, showed no difference in the dislocations in the two areas.

Attempts were also made to grow from doped seeds. Germanium, iron, aluminum and boron were used to determine if a small change in the lattice spacing of the seed could decrease dislocations. All results were negative in this case with the exception of one run in which the doped germanium crystal was grown from an X seed. Some crystals grown on X-seeds were annealed both in air and vacuum at 500 and 550 degrees C (Fig. 3). All conditions produced essentially no changes in strain in low dislocation material.

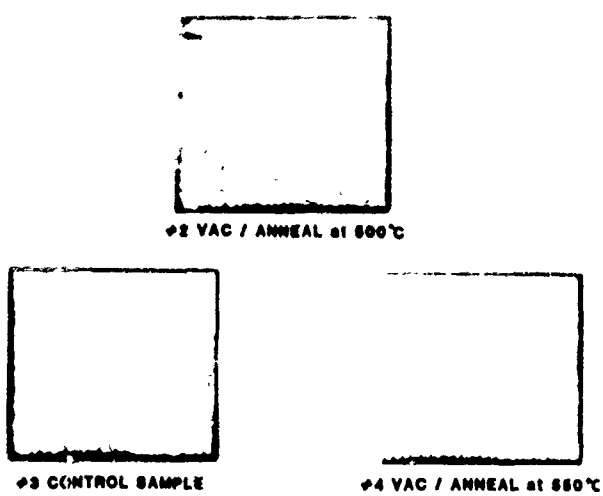


Fig. 3 Examples of crystals grown upon vacuum and air annealed X seeds.

Inclusions - A series of experiments were performed using a series of nutreints in lined and unlined runs (Fig. 4). Inclusion densities were determined in two directions utilizing a 3.5 milliwatt He-Ne laser as a light source. All of the samples were coated with an index matching fluid and 4X magnification was used during the actual counting. Only those inclusions highlighted by the laser beam were counted. Approximately 30 inclusion density measurements were made parallel to the X and Z axes. It was expected that some inclusions were from material lining the wall and therefore the unlined runs would have a higher inclusion density. Based on some of the initial data shown in the figure, this is not necessarily the only source of inclusions since the lowest inclusion density was found in a crystal from an unlined run. The nutrients tested were Z-mined synthetic, whole synthetic crystals and recrystallized glass. Again, based on initial results, the inclusion density was not a specific function of the starting nutrient. The low results for the unlined crystallized glass do indicate however, that the use of a liner may not be important. An excessive number of inclusions were found in a highly lithium-doped crystal.

QUARTZ INCLUSION RESULTS

NUTRIENT	GROWTH RATE (MILS/DAY)	LINER	INCLUSION COUNT (CM⁻³)*	
			X	Z
GLASS	18	SILVER	281.8	---
SYNTHETIC	21	SILVER	54.7	46.7
Z-GROWTH	18	SILVER	---	64.1
GLASS	18	NO	24.1	24.6
SYNTHETIC	42	NO	44.6	68.7
Z-GROWTH	18	NO	84.5	31.8

* PARALLEL TO X AND Z

Fig. 4

The nature of the inclusions is not clear. Photomicrographs in some cases indicate solid inclusions, but the majority seem to be voids with liquid present in many. The solid inclusions seem to be most prevalent near the seed. An attempt was made to anlyze some of these using scanning electron microscope with an EDS attachment. We were not successful in obtaining any meaningful results, however. Some of these inclusions are shown in Fig. 5.

QUARTZ #X51z OPTICAL MICROGRAPHS SHOWING DAMAGE AROUND ETCH TUNNELS, AFTER ANNEALING



Fig. 5

Sweeping - A number of sweeping runs were also performed to possibly determine what elements are moved or removed during sweeping. Both air and vacuum sweeping were used. In these experiments, graphfoil, a high purity material was placed between the platinum contact and the sweeping surface (Z surface). The graphfoil was then analyzed using Auger and EDS combined with SEM. Analysis was performed on all four surfaces, the two in contact with the platinum and the two in contact with the sweeping specimen.

After a 30 second argon etch, lithium was found to be present on both the positive and negative graphfoil electrode. There is some evidence of phosphorus on both electrodes. On the negative electrode a trace of iron was detected. The most surprising result was the trace of aluminum found on the positive electrode which indicates aluminum may be transported as some type of negative complex. There are small particles of quartz embedded in the graphfoil surface next to the sample (Fig. 6) which were detected by SEM. Auger analysis of the surface of these particles was generally inconclusive, but in one case iron and titanium were detected on the particles embedded in the negative electrode. Aluminum was found in one of the particles attached to the positive electrode. The tentative conclusion from these results is that most of the impurities are swept from dislocations since the migration of iron, aluminum and titanium through a perfect lattice does not seem likely. Lithium has been detected at the positive electrode before, but the possible detection of aluminum at the positive electrode indicates that aluminum in the dislocations may form a complex with lithium or hydroxide. We are continuing this investigation using materials with higher dislocation counts to increase the impurity concentrations collected during sweeping.

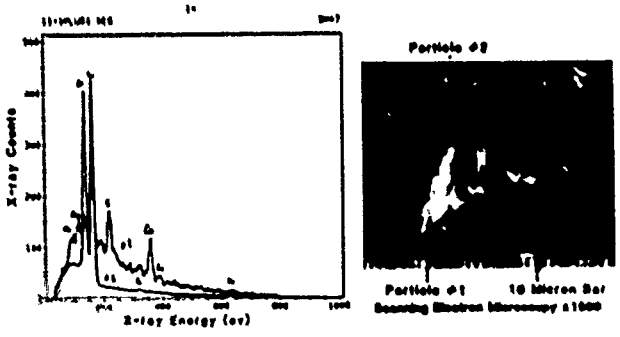


Fig. 6

References

1) J. Martin and A. F. Armington, J. Crystal Growth 62 (1983), p. 203.

2) J. F. Balascio and A. F. Armington, Proceedings of the 40th Frequency Control Symposium, IEEE 86CH2330-9 (1986), p. 70.

3) J. Brice, Reviews of Modern Physics, 57 (1985), p. 105.

4) A. F. Armington, J. J. Larkin, J. J. O'Connor and J. A. Horrigan, Proceedings of the 35th Annual Frequency Control Symposium, EIA Washington, D.C. (1981), p. 297.

5) A. F. Armington and J. J. Larkin, J. Crystal Growth, 71 (1985), p. 799.

RADIATION EFFECTS IN VACUUM-SWEPT QUARTZ

A. Kahan, F.K. Euler, and H.G. Lipson
Solid State Sciences Directorate
Rome Air Development Center
Hanscom AFB, Bedford, MA 01731

and

C.Y. Chen* and L.E. Halliburton
Department of Physics
Oklahoma State University, Stillwater, OK 74078

Summary

5-MHz, SC-cut, resonators were fabricated from cultured quartz electrodiffused (swept) in vacuum ambient and were ^{60}Co irradiated. Steady-state oscillator frequency offset $\Delta f/f$ saturates at a relatively low dose of 10 krad, and stays approximately constant, -120×10^{-9}, up to 1 Mrad. The magnitude of $\Delta f/f$ is surprisingly large compared to resonators fabricated from quartz grown in the same autoclave run and swept in air ambient, $\pm 10 \times 10^{-9}$, and exceeds even that of resonators from the unswept crystal, -65×10^{-9}. Infrared, anelastic loss, and ESR characterization of vacuum-swept material show conclusively that the aluminum-impurity compensators, alkali metal ions and protons, were removed by sweeping, and that the only observable point defect is the Al-hole center. After irradiation ESR data shows a factor-of-four increase in Al-hole concentration. We propose, that the dominant defect formed during vacuum sweeping is a negatively charged diamagnetic aluminum center. Under normal conditions this "bare aluminum" center, isoelectronic with silicon dioxide, is unstable, but vacuum sweeping "freezes" this defect into the crystal. During irradiation the defect loses an electron and is converted into the one-hole aluminum center.

Introduction

Steady-state radiation effects in resonators fabricated from high-quality quartz are attributed to radiation-induced dissociation and formation of point defects which alter the elastic, dielectric, and piezoelectric constants, and cause resonator frequency shifts. In cultured crystal, the major impurity, Al^{3+} substitutional on a Si^{4+} site, is charge compensated with a lithium or sodium ion, and forms $[AlO_4/Li^+]^o$ or $[AlO_4/Na^+]^o$ point defect centers.[1] Irradiation creates electron-hole pairs and provides an alternate compensation mechanism. After irradiation, Al^{3+} is charge compensated either by a proton released from an OH-defect, or by a hole trapped at an adjacent oxygen, forming $[AlO_4/H^+]^o$ or $[AlO_4e^+]^o$, respectively. Electrodiffusing (sweeping) quartz at high temperatures, in air ambient, dissociates $[AlO_4/Li^+]^o$ or $[AlO_4/Na^+]^o$, physically sweeps the alkali ion from the crystal, and forms $[AlO_4/H^+]^o$. In the sweeping process the proton is supplied by the air ambient. For simplicity, these centers are usually designated as Al-Li, Al-Na, Al-OH, and Al-hole. The point defect structure of quartz, its modifications by sweeping and irradiation, and radiation effects on resonators and oscillators fabricated from these materials are reviewed by Griscom,[2] Weil[3], Halliburton, Martin, and Koehler,[4] King and Koehler,[5] and Lipson and Kahan.[6]

Resonators fabricated from an air-swept crystal have smaller $\Delta f/f$ than those fabricated from an as-received crystal.[7-9] This effect was interpreted by arguing that, in as-received quartz, radiation redistributes OH-defects and alkali-metal ion compensators and causes a change in the pertinent material constants. In air-swept quartz, all aluminum impurities are compensated by protons, and radiation-induced redistribution of protons does not occur. It was conjectured that steady-state $\Delta f/f$ and transient frequency offsets can be further reduced by removing not only all alkali-metal ions but also the internally generated protons.[10] In principle, this can be accomplished by sweeping the crystal in a vacuum or inert gas ambient, the process suggested by King for reducing transient radiation sensitivity to pulsed irradiation.[11] It was further assumed that after vacuum sweeping, in the absence of Li, Na, or proton compensators, the dominant Al-defect is $[AlO_4e^+]^o$. We then define, for this paper, a "vacuum-swept" crystal as one from which all, or most, interstitial alkali-metal ions and protons have been removed. However, as shown in Ref. 10, protons can not be completely removed from crystals with high Al-impurity concentration.

In this investigation we report $\Delta f/f$ data for 5-MHz resonators fabricated from high-quality, low aluminum-impurity, vacuum-swept quartz. After vacuum sweeping the quartz was characterized by infrared, anelastic loss, and ESR measurements to insure that the crystal does not contain ionic impurities or protons. We also report ESR data on irradiated vacuum-swept crystals.

Experimental Procedures

Premium-Q quartz evaluated in this investigation was grown at Sawyer Research Products (SARP), and is designated as autoclave run D14-45. One crystal was lumbered into two pure Z-growth bars. The Z faces of one bar were clamped between platinum foil electrodes and swept at 500 °C, 10^{-5} Torr vacuum, and electric

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

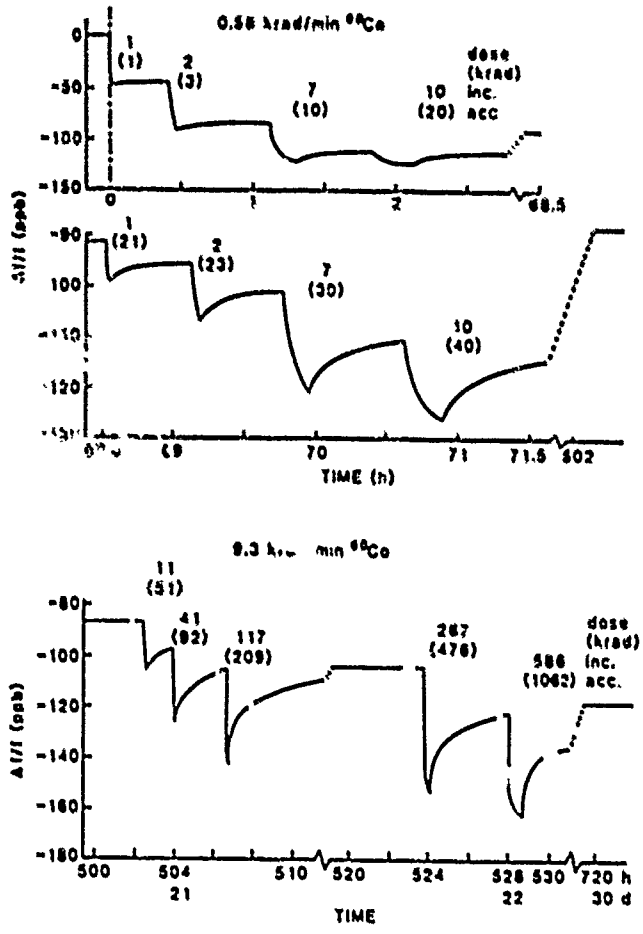


FIG. 1. Normalized radiation-induced oscillator frequency change $\Delta f/f$ as a function of time for an SC-cut resonator fabricated from vacuum-swept quartz. The time periods showing frequency decreases correspond to the radiation exposures. Incremental and accumulated doses are shown in krad. Low dose rate was used up to 40 krad accumulated dose, and high dose rate between 40 and 1062 krad.

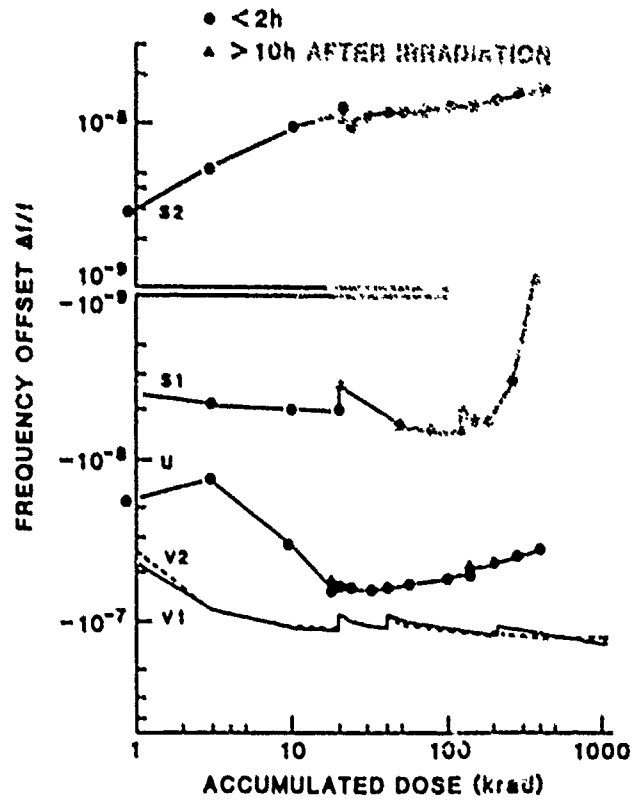


FIG. 2. Radiation-induced steady-state frequency offset $\Delta f/f$ as a function of accumulated dose of resonators fabricated from as-received (unswept), u, air-swept, s1 and s2, and vacuum-swept, v1 and v2, Premium-Q D14-45 crystals grown in the same autoclave run.

field intensity of 1000 V/cm. The electric field was maintained during the cooling period. The Al-impurity concentration of this crystal, determined from infrared measurements on a companion air-swept sample, is approximately 0.7 ppm. Previous ESR measurements on other crystals grown in this autoclave run also indicate 0.7-1.0 ppm aluminum. Compared to other commercially grown Premium-Q quartz, D14-45 has a low Al-impurity concentration and has been utilized in many other investigations.[6,9,10,12-14] Resonator units fabricated from the bar at Frequency Electronics, Inc. were 5-MHz, 5th-overtone, SC-cut, biconvex, gold electroded, and thermal compression bonded.

The experimental procedures for 85 K infrared measurements are described in Ref. 6. The 3581 cm^{-1} peak is taken as the indicator of OH-defect centers, and the 3366 cm^{-1} peak as a measure of $[AlO_4/H^+]^o$ strength. ESR experimental procedures and the relationship of $[AlO_4e^+]^o$ strength to Al-impurity concentration for as-received, irradiated, and air-swept quartz are described in Refs. 15 and 16. Experimental procedures for ^{60}Co irradiation and oscillator frequency offset measurements as a function of dose are given in Ref. 17.

Results

Figure 1 shows normalized radiation-induced oscillator frequency offset $\Delta f/f$ as a function of time for an SC-cut resonator fabricated from a vacuum-swept crystal, and operated at its lower turnover temperature, 94 °C. The time periods showing frequency decreases correspond to the radiation exposures. Incremental and accumulated doses are shown in krad. The initial ^{60}Co irradiation increments were 1, 2, 7, and 10 krad, at a dose rate of 0.58 krad/min. After each exposure the oscillator frequency was allowed to recover for a short time period and approach a steady-state condition. At the completion of the first irradiation sequence the oscillator was allowed to recover for approximately 3 days. The oscillator was again exposed to the same sequence of irradiation increments, and allowed to recover for 18 days. The oscillator was then irradiated at a dose rate of 9.3 krad/min with increments of 11, 41, 117, 267, and 586 krad to a total accumulated dose of 1.06 Mrad.

Figure 2 shows the steady-state $\Delta f/f$ as a function of accumulated dose for two resonators, v1 and

v2, fabricated from vacuum-swept quartz. Both resonators give similar results. At low irradiation doses, $\Delta f/f$ decreases rapidly, -40×10^{-9} after 1 krad and -112×10^{-9} after 20 krad accumulated dose. The vertical lines in Fig. 2 denote the effect of the long recovery periods. After 20 krad accumulated dose $\Delta f/f$ recovered to -90×10^{-9}. After 40 krad $\Delta f/f$ decreased to -110×10^{-9} and recovered to -90×10^{-9} after 18 days. After 1.06 Mrad the frequency decreased to -130×10^{-9} and recovered to -120×10^{-9} after 7 days.

Figure 2 also shows radiation sensitivity of AT-cut resonators fabricated from as-received unswept, u, and air-swept, s1 and s2, crystals grown in the same autoclave run. For the same dose range, maximum $\Delta f/f$ is -65×10^{-9} and $\pm10\times10^{-9}$, respectively. The difference between s1 and s2, negative versus positive offset, may not be intrinsic to the material but could reflect resonator fabrication differences. In addition to steady-state $\Delta f/f$ radiation sensitivity, Fig. 1 data also show rather large $\Delta f/f$ recovery after each exposure, ranging from 3 to 45×10^{-9}. It requires more than 10 days to reach a steady-state within 10^{-9}. Corresponding $\Delta f/f$ recovery for the air-swept resonators s1 and s2 ranges from 2 to 12×10^{-9}, and reaches in 2 days a steady-state within 10^{-10}. Our results then indicate that vacuum sweeping is ineffective in reducing $\Delta f/f$ sensitivity to radiation. However, the relatively low dose required for saturation implies that a 50-100 krad prophylactic irradiation is highly beneficial in reducing subsequent steady-state frequency offsets.

Figure 3 shows spectral band strengths associated with OH-defect and $[AlO_4/H^+]^0$ centers. 85-K infrared spectra were measured and scanned along the crystal x axis before and after vacuum sweeping, annealing, and irradiation. As-received, the 3581 cm^{-1} absorption was ≈ 0.05 cm^{-1} between $0 < x < 8$ mm, and increased gradually to ≈ 0.10 cm^{-1} at $x = 15$ mm. Results for 4 days of vacuum sweeping are taken from Ref. 10, where we also showed data for as-received quartz, after air sweeping, and after 3, 6, 12, and 24 h of vacuum sweeping. After 4 days of vacuum sweeping only one-half of the sample was clear of OH-defects and $[AlO_4/H^+]^0$, as measured by the 3581 and 3366 cm^{-1} peaks, respectively. After 9 days the protons were also removed from the region near the cathode but a small amount of $[AlO_4/H^+]^0$ was re-introduced at the anode. In a sample with higher Al-impurity and OH-defect concentrations, after 51 days of vacuum-sweeping the anode-half of the crystal was clear of impurities but OH-defects and $[AlO_4/H^+]^0$ increased substantially towards the cathode. Protons can be completely removed only from crystals with low Al-impurity concentration.

After the 9-day vacuum-sweeping the crystal was annealed at 500 °C. The subsequent infrared spectra show no change in OH-defect and $[AlO_4/H^+]^0$ bands, indicating that for all practical purposes the protons were removed. In the fabricated resonator, the residual $[AlO_4/H^+]^0$ in the edge region of the crystal is outside the energy-trapping portion of the disk, and has little influence on the oscillator frequency. We also performed anelastic loss measurements, $Q^{-1}(T)$, from 10 K to 100 °C, and we did not observe any bands which can be associated with the sodium compensator. Based on arguments presented in Ref. 10 regarding the formation of $[AlO_4/H^+]^0$ as an intermediate state, we are also certain that lithium ion compensators were removed. These characterization results then indicate that the processed bar is indeed vacuum-swept, and contains only a small concentration of $[AlO_4/H^+]^0$ in a geometric region which does not influence resonator results.

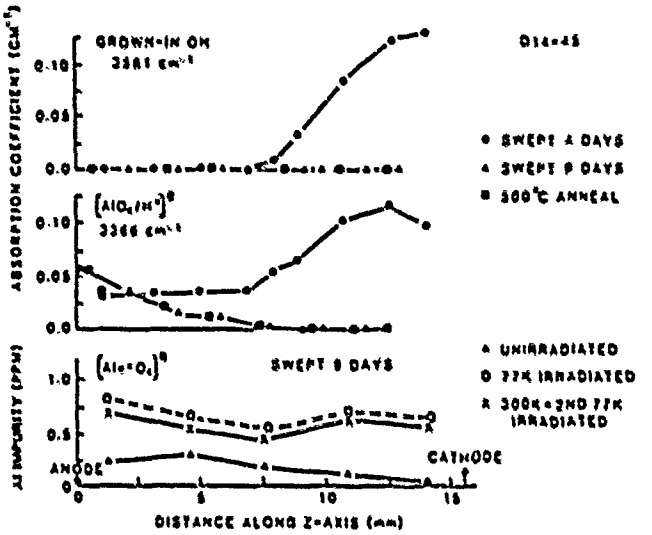


FIG. 3. 85-K infrared strength of OH-defect and $[AlO_4/H^+]^0$ associated bands, and Al-impurity concentration from ESR $[AlO_4e^+]^0$ signal strength, of vacuum-swept quartz as a function of position along the sweeping axis z.

Figure 3 also shows ESR determined Al-impurity concentration, measured by $[AlO_4e^+]^0$ strength, after the 9-day vacuum-sweeping for sample positions along the sweeping axis. $[AlO_4e^+]^0$ strength varies from 0.24 ppm at the anode to 0.04 ppm near the cathode, with an average value of 0.16 ppm. After irradiating at 77 K, the average concentration increases to 0.7 ppm. Irradiating at room temperature, followed by a second low temperature exposure, reduces the average value to 0.6 ppm.

Discussion

SCF MO results

Weil and his coworkers using Hartree-Fock self-consistent-field molecular-orbital (SCF MO) methodology, calculated atomic positions, spin densities, net charge, and total energy for the major substitutional aluminum centers in quartz.[18-20] Calculations were carried out using a 10- and a 22-atom cluster. Computer modeling for some of these point defects in

ENERGY (eV)	DEFECT CENTER		Al-O BOND LENGTHS (A)
20.0	$[Al^{+}e^{+}O_4]^{+}$	Al-2 HOLE	1.89
17.1	$[AlO_4]^{-}$	BARE Al	1.71
16.2	$[Al^{+}O_4]^{0}$	Al-HOLE	1.99
9.8	$[AlO_4/Na^{+}]^{0}$	Al-Na	1.77
9.7	$[Al^{+}O_4/Li^{+}]^{+}$	Al-Li-HOLE	1.71, 1.92
5.8	$[AlO_4/Li^{+}]^{0}$	Al-Li	1.79
4.2	$[Al^{+}O_4/H^{+}]^{+}$	Al-OH-HOLE	1.82, 1.92
0.0	$[AlO_4/H^{+}]^{0}$	Al-OH	1.91

FIG. 4. SCF MO calculated relative energy levels and bond lengths of Al-related defect centers in quartz, Refs. 19 and 20.

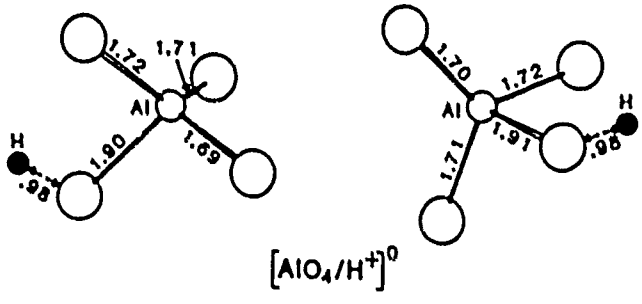


FIG. 5. Two configurations of the hydrogen-compensated diamagnetic center $[AlO_4/H^{+}]^{0}$. Energetically the model depicted on the right is slightly more favorable. Bond lengths, in Å, are based on SCF MO calculation, Ref. 20.

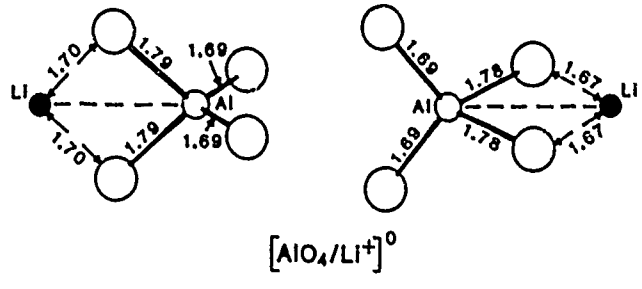


FIG. 6. Two configurations of the lithium-compensated diamagnetic center $[AlO_4/Li^{+}]^{0}$, Ref. 20. Energetically the model depicted on the right is slightly more favorable.

quartz using a 19- and a 33-atom cluster were also carried out by Wilson et al.[21,22] Limitations imposed by the more restrictive 10- and 22-atom model are described in the references. We are interested in a comparison of relative energy values of aluminum defects, and for this purpose we extract from results obtained by Weil and his coworkers. Figure 4 summarizes some of their results in the form of an "energy level" diagram for the various Al-defect centers. The corresponding lowest-energy atomic configurations are shown in Figs. 5 to 10. The compensator can be attached either to one of the two equivalent oxygens designated in Fig. 10 as O(1) and O(2), or to one of the two equivalent oxygens designated O(3) and O(4). The energy difference between the two configurations is small, 0.03 eV for proton compensation and 0.16 eV for lithium compensation, with the compensator residing on O(3) or O(4) having the lower energy state. Figures 5 and 6 show both configurations, but for simplicity we list in Fig. 4 only the lower energy state. Similarly, we omit the excited state $[AlO_4e^{+}]^{0}_{es}$, separated from the ground state $[AlO_4e^{+}]^{0}$ by 0.03 eV. $[AlO_4e^{+}]^{0}_{es}$ is very similar to $[AlO_4e^{+}]^{0}$, except that the hole is localized on O(1) or O(2).

The diamagnetic center $[AlO_4/H^{+}]^{0}$, Fig. 5, is the defect with the lowest energy, and is chosen as the energy baseline. $[AlO_4/Li^{+}]^{0}$, Fig. 6, is 5.8 eV and $[AlO_4/Na^{+}]^{0}$, Fig. 7, is 9.8 eV above $[AlO_4/H^{+}]^{0}$. Qualitatively, the increased stability as one goes from Na^{+}, via Li^{+}, to H^{+} is consistent with crystal growth and air sweeping experience. In an as-received cultured crystal the dominant defect is $[AlO_4/Na^{+}]^{0}$. If a small amount of lithium salt is added during crystal growth the dominant defect becomes $[AlO_4/Li^{+}]^{0}$. When protons become available, either from sweeping in air ambient or during irradiation from OH-defect sites, $[AlO_4/H^{+}]^{0}$ forms both in sodium and lithium compensated samples.

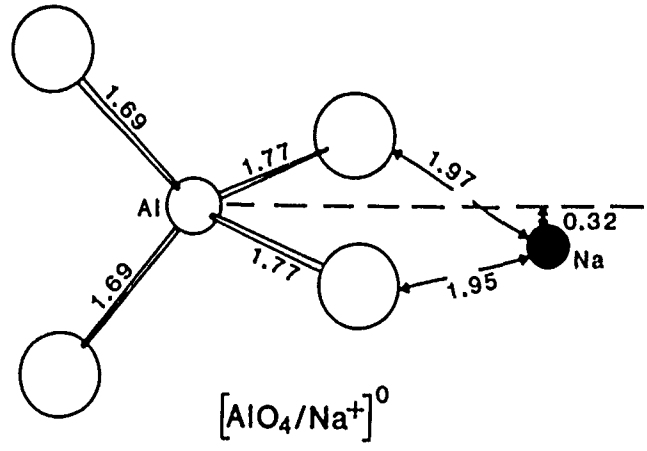


FIG. 7. Model of the sodium-compensated diamagnetic center $[AlO_4/Na^{+}]^{0}$, Ref. 20.

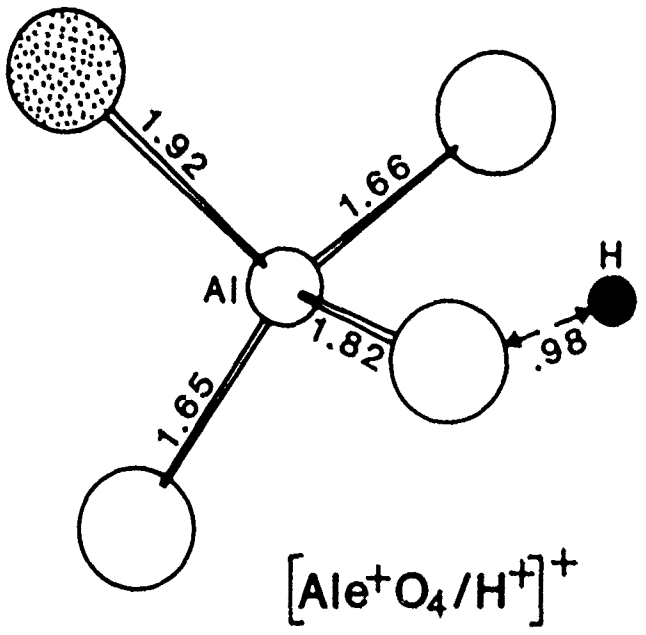


FIG. 8. Model of hydrogen-compensated paramagnetic center $[AlO_4e^+/H^+]^+$, Ref. 20. The shaded sphere indicates the hole-bearing oxygen ion.

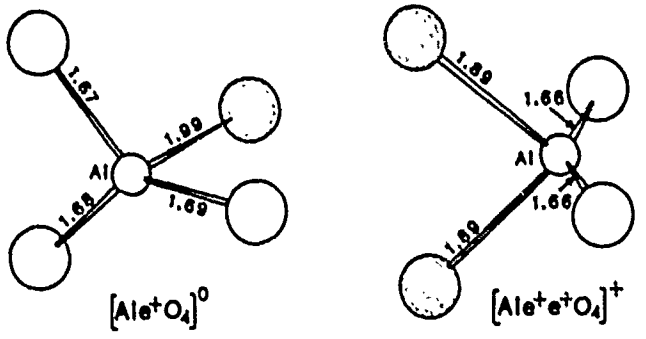


FIG. 9. Model of the aluminum one-hole $[AlO_4e^+]^0$ and aluminum two-hole $[AlO_4e^+e^+]^+$ centers, Ref. 19. The shaded spheres indicate the hole-bearing oxygen ions.

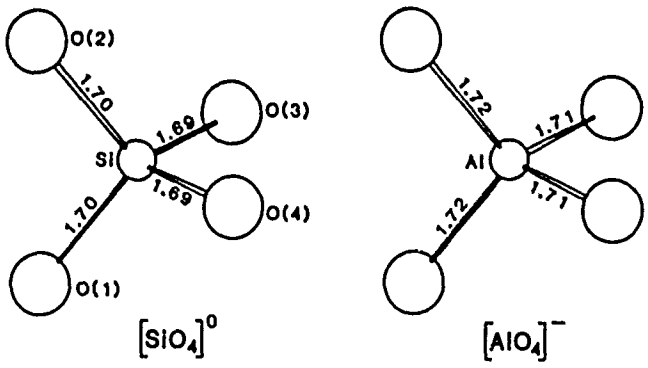


FIG. 10. Model of pure alpha-quartz unit $[SiO_4]^0$, and the diamagnetic aluminum center $[AlO_4]^-$, Ref. 19.

Figures 6 and 7 show the important result that for $[AlO_4/Li^+]^0$ the lithium ion is situated on the 2-fold C_2 symmetry axis, whereas for $[AlO_4/Na^+]^0$ the sodium ion is situated off the symmetry axis. The same configurations are also obtained by Wilson et al. using the 19-atom model.[21] These calculations confirm the model suggested by Toulouse, Green, and Nowick,[23,24] and explain the existence of $[AlO_4/Na^+]^0$ and the absence of $[AlO_4/Li^+]^0$ peaks in acoustic- and dielectric-loss spectra.

Figure 4 also lists energies for one-hole cation compensated paramagnetic centers, $[AlO_4e^+/M^+]^+$, $M^+ = H, Li$, separated from the corresponding diamagnetic defects by about 4 eV. The energy configuration for $[AlO_4e^+/Na^+]^+$ has not been calculated, but it is assumed that it exists. Figure 8 depicts $[AlO_4e^+/H^+]^+$. It has the $[AlO_4/H^+]^0$ configuration with the proton localized on O(4) and the hole localized on O(2), one of the two equivalent oxygens opposite the proton. A similar model holds for $[AlO_4e^+/Li^+]^+$, with the hole localized on one of the oxygens on the opposite side of the lithium ion. These defects are stable at low temperature but decay below room temperature. These centers may be applied in interpreting transient radiation effects for resonators irradiated at room, or higher temperatures. The diamagnetic defect $[AlO_4/H^+]^0$ acquires a temporary charge during room-temperature irradiation, and subsequently relaxes to the lower energy state.

The one-hole and two-hole aluminum centers, $[AlO_4e^+]^0$ and $[AlO_4e^+e^+]^+$, Fig. 9, lie 16.2 and 20.0 eV above $[AlO_4/H^+]^0$, respectively. $[AlO_4e^+e^+]^+$ is observed after low temperature irradiation and decays when the crystal is warmed to room temperature. Calculations also indicate the possible existence of a charged diamagnetic aluminum impurity center, $[AlO_4]^-$, Fig. 10, 17.1 eV above $[AlO_4/H^+]^0$ and 0.9 eV above $[AlO_4e^+]^0$. It is the simplest $[SiO_4]^0$ isoelectronic Al-impurity center, a "bare-Al" center. The bare-Al center has not been observed experimentally, it is considered unstable, and will relax to $[AlO_4e^+]^0$ with the acquisition of a charge, $[AlO_4]^- + e^+ \rightarrow [AlO_4e^+]^0$.

Vacuum swept quartz

There are two basic issues raised by the $\Delta f/f$ and ESR data: (1) the very large frequency offset, and (2) a consistent compensation scheme for Al-impurity in unirradiated and irradiated vacuum-swept samples. Vacuum-swept implies the absence of $[AlO_4/Li^+]^0$, $[AlO_4/Na^+]^0$, OH-defect, $[AlO_4/H^+]^0$, and radiation-induced ionic motion, yet the frequency offset is larger than either for as-received or for air-swept samples. Also, in the absence of ionic centers, the logical Al-impurity compensator for unirradiated vacuum-swept material is $[AlO_4e^+]^0$, but experimentally

we find that only 23% of the Al-impurity is in the form of $[AlO_4e^+]^0$.

In as-received high-quality cultured Premium-Q quartz the principal compensator is lithium, forming $[AlO_4/Li^+]^0$. Protons in quartz are mobile at 77 K, but lithium diffuses only above 200 K.[25] Irradiating quartz at 77 K dissociates $[AlO_4/H^+]^0$ and forms $[AlO_4e^+]^0$, but leaves $[AlO_4/Li^+]^0$ unchanged. Our infrared spectra indicate that the vacuum-swept sample is almost $[AlO_4/H^+]^0$ free, and the ESR observed increase in $[AlO_4e^+]^0$ from 0.16 ppm to 0.7 ppm after the first low-temperature irradiation can not be attributed to the dissociation of $[AlO_4/H^+]^0$ or $[AlO_4/Li^+]^0$.

In ESR investigations, the rationale for the subsequent room-temperature irradiation followed by a second low-temperature exposure is that the room-temperature irradiation dissociates $[AlO_4/Li^+]^0$, forms $[AlO_4/H^+]^0$ and $[AlO_4e^+]^0$, and the subsequent low-temperature exposure dissociates the newly formed $[AlO_4/H^+]^0$ and forms additional $[AlO_4e^+]^0$.[15,16] Consequently, the low-, room-, and low-temperature irradiation scheme for an as-received lithium compensated crystal converts all Al-impurity defects to $[AlO_4e^+]^0$. One then always observes an increase in $[AlO_4e^+]^0$ after the second low-temperature irradiation. The same scheme for an air-swept sample is used to assess sweeping completeness. If the air-sweeping process was complete, that is, if all lithium compensators were swept from the crystal, the first and second low temperature irradiations will give identical $[AlO_4e^+]^0$ strength, but if the process was incomplete, the second low-temperature irradiation will show an increase in $[AlO_4e^+]^0$. For our vacuum-swept samples, after the second low temperature irradiation, we observe a decrease in $[AlO_4e^+]^0$, from 0.7 to 0.6 ppm. This again confirms our contention that the vacuum-swept bar is lithium free, but raises the question of the identity of the Al-impurity compensator in the unirradiated and irradiated crystal.

Aluminum impurity compensator

We propose that most of our radiation effects data on vacuum-swept quartz can be interpreted in a fairly consistent manner by assuming that the dominant defect formed during sweeping is $[AlO_4]^-$. The existence of this "bare-Al" center, isoelectronic with $[SiO_4]^0$, is deduced from centers formed by nearby interstitial cations, and under normal conditions they are unstable. It is our contention that the center forms under the simultaneous high temperature and electric field conditions prevailing during vacuum sweeping, and it becomes "frozen-in" during cooling with the electric field still applied. Overall crystal charge neutrality requires an equal amount of positively charged traps at other sites. Another possibility is the formation of $[AlO_4e^+e^+]^+$, but in our vacuum-swept samples we find no experimental ESR evidence for this defect. Irradiating a crystal creates electron-hole pairs; unidentified defects trap the electrons and $[AlO_4]^-$ trap the holes and become $[AlO_4e^+]^0$.

Considering $[AlO_4]^-$ as the precursor to $[AlO_4e^+]^0$ explains the increased strength of this center, 0.16 to 0.7 ppm, between unirradiated and low-temperature irradiated vacuum-swept crystals. In our ESR experiments the samples were irradiated at low-temperatures, but this irradiation does not need to take place at 77 K. The same increase in $[AlO_4e^+]^0$ ESR signal strength was also obtained after irradiating a vacuum-swept sample directly at room-temperature. In addition, this sample was irradiated in increments, and showed $[AlO_4e^+]^0$ saturation at relatively low doses, consistent with $\Delta f/f$ results. However, the proposed $[AlO_4]^-$ model does not explain the subsequent decrease in $[AlO_4e^+]^0$ after the second low temperature irradiation.

Consistent with previous interpretations, we implied that the magnitude of resonator frequency offset is related to radiation-induced ionic redistribution, the larger the ion the greater the effect. From a more fundamental point of view, changes in material constants are due to defect structure modifications, whether caused by ionic redistribution or any other mechanism. Specifically, we associate material constant changes with radiation-induced changes in Al-O bond length and Al-M distance.

In Fig. 4, for the ionic centers, we list the Al-O bond length for the oxygen nearest to the interstitial impurity, and for the hole centers, we list the Al-O bond length for the oxygen with the missing electron. Figures 5 to 10 also show the SCF MO calculated bond lengths. The calculated Si-O bond length is 1.70 Å. For the compensated diamagnetic centers $[AlO_4/Na^+]^0$, $[AlO_4/Li^+]^0$, and $[AlO_4/H^+]^0$ the O-Na, O-Li, and O-H distances are 1.96, 1.67, and 0.98 Å, and the corresponding Al-O bond length increases from 1.70 to 1.77, 1.78, and 1.91, respectively. The almost identical Al-O(Na) and Al-O(Li) bond lengths do not necessarily imply that, for the same Al-impurity concentration, sodium and lithium compensated samples will have the same radiation sensitivity. In addition to Al-O changes, one also has to consider effects caused by large O-Na and O-Li differences. Experimentally, the frequency offset, at 1 Mrad ^{60}Co irradiation, of a resonator fabricated from an air-swept D14-45 crystal which was subsequently Na-swept, is $+760\times10^{-9}$, compared to -35×10^{-9} for the Li-doped material shown in Fig. 2.[9]

Figure 4 shows that the Al-O bond length for $[AlO_4]^-$ is 1.71 A, and for $[AlO_4e^+]^0$ the bond length, from the aluminum to the oxygen with the localized hole, relaxes to 1.99 A. Our proposed mechanism for irradiated vacuum-swept material, the transformation of $[AlO_4]^-$ into $[AlO_4e^+]^0$, causes the largest Al-O bond length change for any possible Al-defect modification. As a consequence, this causes large elastic constant

changes, and correspondingly large frequency offsets, -120×10^{-9} after 1.06 Mrad. Based on Al-O bond length consideration, overall radiation sensitivity in an air-swept sample, where all defects both before and after irradiation are in the form of $[AlO_4/H^+]^o$, should be at a minimum. This is indeed observed experimentally, Fig. 2, for resonators s1 and s2, fabricated from the same air-swept bar.

Conclusions

Oscillators utilizing resonators fabricated from low-Al-impurity quartz swept in vacuum ambient show surprisingly large radiation-induced frequency offset. Moreover, the offset saturates at relatively low radiation doses. Our results then indicate that vacuum sweeping is ineffective in reducing steady-state $\Delta f/f$ sensitivity to radiation. However, the relatively low dose required for saturation implies that a 50-100 krad prophylactic irradiation is highly beneficial in reducing subsequent steady-state frequency offsets.

ESR measurement of irradiated vacuum-swept crystals shows the puzzling result that after irradiation $[AlO_4e^+]^o$ concentration increases by a factor of four, indicating that after vacuum sweeping a large percentage of Al^{3+} was in the form of another defect, a precursor to $[AlO_4e^+]^o$.

We propose that the data can be interpreted by assuming that the dominant defect formed during vacuum sweeping is the diamagnetic $[AlO_4]^-$ center. It is our contention that under the simultaneous high temperature and electric field conditions of vacuum sweeping, $[AlO_4]^-$ forms and is "frozen" into the crystal during cooling. Irradiating a crystal creates electron-hole pairs; unidentified defects trap the electrons, and $[AlO_4]^-$ trap holes and become $[AlO_4e^+]^o$. This results in the largest possible increase in Al-O bond length, from 1.71 A to 1.99 A, which causes large elastic constant changes and, consequently, large oscillator frequency offsets.

Acknowledgments

We wish to thank our coworkers M.T. Harris for performing the sweeping experiments, J.R. Cappelli for providing radiation services, and W.L. Burke for technical assistance with oscillator measurements.

References

* Current address: Oak Ridge National Lab., Oak Ridge, TE 37830.

1. J.H. Mackey, J. Chem. Phys. 39, 74 (1963). The notation inside the square bracket identifies the substitutional impurity, the number of oxygens, the trapped carrier charge, and the presence of interstitial ions (after a virgule). The net charge of the complex is indicated by the superscript.

2. D.L. Griscom, Proc. 33th Annual Frequency Control Symposium, 98 (1979).

3. J.A. Weil, Phys. Chem. Minerals 10, 149 (1984).

4. L.E. Halliburton, J.J. Martin, and D.R. Koehler, in Precision Frequency Control, edited by E.A. Gerber and A. Ballato (Academic, New York, 1985), Chap. 1, pp. 1-45.

5. J.C. King and D.R. Koehler, Precision Frequency Control, edited by E.A. Gerber and A. Ballato (Academic, New York, 1985), Chap. 3, pp. 147-159.

6. H.G. Lipson and A. Kahan, J. Appl. Phys. 58, 963 (1985).

7. B.R. Capone, A. Kahan, R.N. Brown, and J.R. Buckmelter, IEEE Trans. Nucl. Sci. NS-17(6), 217 (1970).

8. P. Pellegrini, F. Euler, A. Kahan, T.M. Flanagan, and T.F. Wrobel, IEEE Trans. Nucl. Sci. NS-25, 1267 (1978).

9. F.K. Euler, H.G. Lipson, and P.A. Ligor, Proc. 34th Annual Frequency Control Symposium, 72 (1980).

10. H.G. Lipson and A. Kahan, IEEE Trans. Nucl. Sci. NS-31, 1223 (1984).

11. J.C. King, U.S. Patent No. 3,932,777 (Jan. 13, 1976).

12. S.P. Doherty, J.J. Martin, A.F. Armington, and R.N. Brown, J. Appl. Phys. 51, 4164 (1980).

13. A. Kahan and H.G. Lipson, Proc. 39th Annual Frequency Control Symposium, 255 (1985).

14. J.J. Martin, H.R. Hwang, and T.M. Wilson, Proc. 40th Annual Frequency Control Symposium, 32 (1986).

15. M.E. Markes and L.E. Halliburton, J. Appl. Phys. 50, 8172 (1979).

16. L.E. Halliburton, N. Koumvakalis, M.E. Markes, and J.J. Martin, J. Appl. Phys. 52, 3565 (1981).

17. F.K. Euler and A. Kahan, Phys. Rev. B 35, 4351 (1987).

18. R.H.D. Nuttall and J.A. Weil, Can. J. Phys. 59, 1696 (1981); 59, 1709 (1981); 59, 1886 (1981).

19. M.J. Mombourquette, J.A. Weil, and P.G. Mezey, Can. J. Phys. 62, 21 (1984).

20. M.J. Mombourquette and J.A. Weil, Can. J. Phys. 63, 1282 (1985).

21. T.M. Wilson, L.E. Halliburton, M.G. Jani, and J.J. Martin, Proc. 40th Annual Frequency Control Symposium 26 (1986).

22. T.M. Wilson, J.A. Weil, and P.S. Rao, Phys. Rev. B 34, 6053 (1986).

23. J. Toulouse, E.R. Green, and A.S. Nowick, Proc. 37th Annual Frequency Control Symposium, 125 (1983).

24. J. Toulouse, and A.S. Nowick, J. Phys. Chem. Solids 46, 1285 (1985).

25. K.B. Hitt and J.J. Martin, J. Appl. Phys. 54, 5030 (1983).

DIELECTRIC RELAXATION AND EPR IN QUARTZ CRYSTALS CONTAINING Fe

S. Kelleen, S. Ling and A.S. Nowick
Henry Krumb School of Mines, Columbia University
New York, NY 10027

and

L.E. Halliburton
Physics Department, Oklahoma State University
Stillwater, OK 74078

Summary

Iron-doped quartz, notably cultured amethyst, has been studied by combining the techniques of dielectric relaxation (DR) and electron spin resonance (ESR) on the same samples. In addition, samples examined were swept with Na^+, Li^+ and H^+. The DR results show four new relaxation peaks (at 20, 99, 138 and 195 K for a 1 kHz frequency) two of which (20 and 138 K) appear only for Na-swept samples. The ESR measurements show three centers S_1, S_2 and S_3. The S_1 center is believed to be due to substitutional Fe^{3+} with an adjacent interstitial Li^+ ion, while the S_2 is due to an Fe-OH center. On the other hand, the S_3 spectrum, which apparently has not been reported previously, is due to Fe-Na pairs, and is related to the 20 K and 138 K DR peaks. It appears that the S_1 and S_2 centers do not have DR equivalents. Accordingly, the 99 K and 195 K DR peaks may be related to Fe in a valence state other than 3+.

Introduction

It is well known that impurities and point defects play a large role in determining the properties of α-quartz, and thus determine its usefulness for precision frequency control in electronic devices.[1,2] There has been a long-standing interest in the effects of Fe in quartz, particularly in view of the naturally occurring forms of amethyst and citrine. In spite of a large number of studies, however, it can be said that we know relatively little about defects involving Fe as an impurity compared, for example, to those involving Al. In fact, with the aid of a wide ranges of techniques (especially infrared absorption, dielectric and anelastic relaxation and electron spin resonance: ESR) it has been found that Al^{3+} ions substitutionally occupy Si^{4+} sites at the center of a distorted tetrahedron of O^{2-} ions (see Fig. 1), and that the Al^{3+} ion is charge compensated by either an interstitial alkali (Li^+ or Na^+) or a proton (H^+). In this way centers called Al-Li, Al-Na or Al-OH are formed, the latter designation because the proton resides on one of the adjacent O^{2-} ions forming an OH ion. Finally, irradiation at room temperature can drive off an alkali and replace it with an electron hole to obtain the aluminum-hole, Al-h, center. The case of Na compensation (Al-Na center) is especially interesting, since the Na^+ ion resides off the 2-fold symmetry axis (denoted by C_2 in Fig. 1) in one of two sets of equivalent sites (denoted by α and β in Fig. 1). This defect then gives rise to two dielectric relaxation peaks, one due to the α sites, the other to the β sites, and similarly it gives rise to a pair of anelastic (internal friction) peaks.[3,4] In the case of the Al-Li defect, the Li^+ ion sits on the C_2 axis and, therefore, produces no such relaxation peaks.[3]

The Fe^{3+} ion differs from Al^{3+} in several important ways. First, it has a larger ionic radius than Al^{3+} and therefore is much larger (by ~ 50%) than the Si^{4+} ion. Second, it can change valence under irradiation conditions, while Al^{3+} probably does not. Finally, it possesses a half-filled 3d shell ($3d^5$), so that, unlike the Al^{3+} ion, it is paramagnetic and gives rise to a strong ESR spectrum. This allows one to obtain information about its crystal site and nearby defects. Thus, in contrast to Al-containing quartz, where ESR signals can only be observed following irradiation, ESR has provided the principal technique for the study of Fe-containing quartz.[5] A brief review of the literature shows that four distinct ESR spectra have received major attention. The first of these to be characterized was the S_1 center.[6,7] This defect apparently consists of an Fe^{3+} ion substituting for a Si^{4+} with an adjacent interstitial Li^+ ion providing the needed charge compensation. A similar Fe^{3+} ESR spectrum, except with a larger crystal field splitting and a weak hyperfine splitting due to an $I = 1/2$

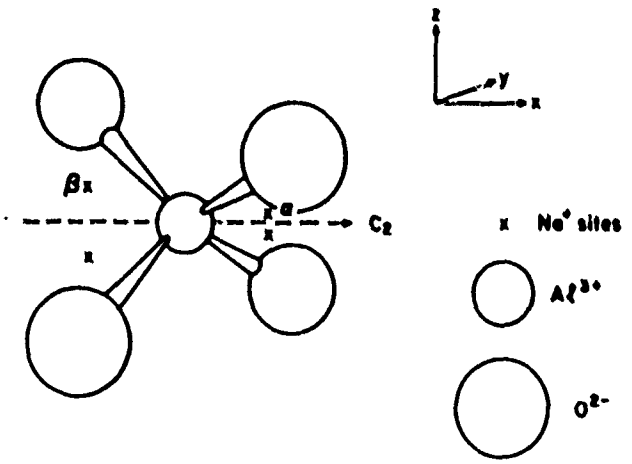


Fig. 1. Schematic diagram showing the distorted tetrahedron, the basic structural unit of α-quartz, with Al^{3+} replacing Si^{4+}. The tetrahedron contains a single twofold symmetry axis (in the x direction) designated C_2. Also shown are the two equivalent α-sites and the two equivalent β-sites. An interstitial Na^+ compensating the Al^{3+} will go into one of these four sites.

CH2427-3/87/0000-223 $1.00 C 1987 IEEE

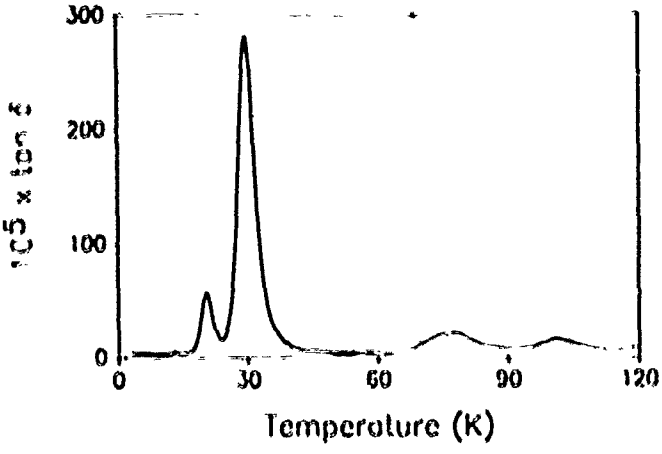


Fig. 2. Dielectric loss, tan δ, at 1 kHz as a function of temperature from 3 to 120 K for a Na-swept amethyst sample.

nucleus, has been designated the S_2 center.[8,9] This defect is a substitutional Fe^{3+} with an adjacent proton, in the form of an OH^- ion, providing the charge compensation.

A third Fe^{3+} spectrum that has been widely studied is the so-called I center.[10,12] There is significant debate over whether the I center is an interstitial or a substitutional Fe^{3+} ion. In either case, the charge compensation mechanism remains unidentified. A final ESR spectrum of importance has been reported by Cox.[13] It is assigned to Fe^{4+} ions and is directly related to the amethyst color itself.

The objective of the present work is to carry out a study of Fe-doped quartz using the combination of dielectric relaxation and ESR techniques on the same (cultured) samples. In addition, the technique of electrodiffusion (sweeping)[14,15] is used so as to introduce, as much as possible, only a single alkali or hydrogen as the principal compensator for both Al^{3+} and Fe^{3+} in a given sample. Because of these two features: the simultaneous use of two different techniques, and the control over the monovalent compensator, it is hoped that a fuller picture of Fe defects in quartz would emerge from the present work than had hitherto been obtained.

Experimental Methods

Two samples of cultured amethyst obtained from Sawyer were availble for study. They were both grown with rhombohedral seeds. The second of these was a large crystal that was oriented so that samples could be cut parallel to the c-axis. Such samples were then suitable for sweeping, which was carried out by J.J. Martin at Oklahoma State University. The samples were generally 1.0 cm^2 and 1-2 mm thick, with the thin dimension parallel to the c-axis. In addition to sweeping, samples from both crystals were subjected to irradiation and anneal treatments.

Samples for dielectric studies were irradiated at room temperature with X-rays from a tungsten target source operated at 20 mA and 40 kV. Soft x-rays were filtered out by a layer of sputtered silver electrode, and a glass filter. Other samples were irradiated at room temperature with γ-rays from a Co source at Brookhaven National Laboratory, to a total dose of 2 Mrads.

For dielectric measurements, the samples were transferred to a Super Varitemp Cryostat (Janis Corp). and cooled down to liquid helium or liquid nitrogen temperatures. An automated capacitance bridge (Andeen. Assoc.) was use to carry out the dielectric loss measurements. The measurements cover a frequency range of 10Hz to 100 kHz and a temperature range of 3-272 K.

Electron spin resonance (ESR) data were obtained from a Bruker Er200D spectrometer. The microwave frequency was 9.3 GHz and the modulation frequency was 100 kHz. All the ESR spectra described in this paper were taken at 27 K using an Oxford Instruments liquid helium flow system. Dimensions of the ESR samples were approximately 2 x 3 x 7 mm^3, and the magnetic field was always aligned parallel to the crystal's c axis (i.e., the smallest dimension). The ESR samples were also irradiated at room temperature with 2-MeV electrons from a Van de Graaff accelerator.

Results and Discussion

Dielectric Relaxation (DR)

Swept samples were either Na- or Li- swept, or swept in air to eliminate alkalis and substitute H (designated H-swept). The most interesting results were obtained for the Na-swept samples. Figure 2 shows the dielectric loss plot for a Na-swept sample in the temperature range from 3 to 110 K, while Fig. 3 shows the results for all three sweepings in the higher temperature range 80 to 220 K. First it should be noted that the Na-swept sample shows the well-known peaks at 30 K and 75 K due to the Al-Na center. (All peak temperatures are quoted for a frequency of 1 kHz.) From the height of the 30 K peak and the previous calibration of this peak,[3] it is concluded that the

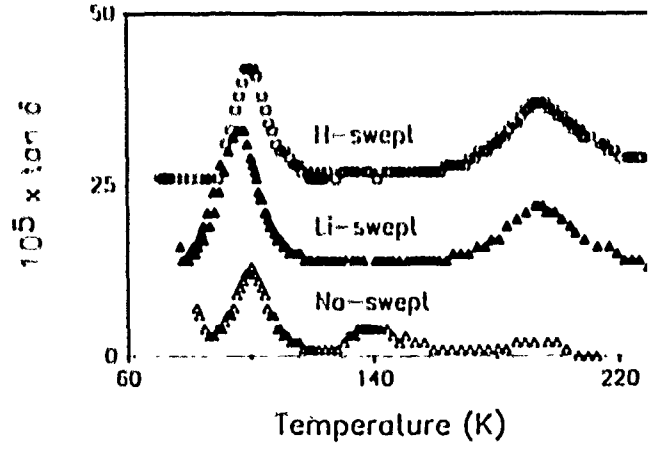


Fig. 3. Comparison of dielectric loss for Na-, Li- and H-swept amethyst in the intermediate temperature range. For clarity, the two upper curves have been displace upwards.

Table 1. Heights of Fe-related Dielectric Loss Peaks in Amethyst Samples Given Different Treatments
(Peak heights in units tan δ_{max} of 10^{-5})

Sample and Treatment	20K	99K	138K	195K
As rec'd (Ann.)	-	15	-	10
As rec'd (Y-Irr.)	-	9	-	63
H-swept	-	16	-	10
H-swept (X-Irr.)	-	13	-	25
Li-swept-A	-	20	-	8
Li-swept-B	-	17	-	3
Na-swept-A	58	11	4	-
Na-swept-B	90	10	6	-
Na-swept-B (Y-Irr.)	<1	4	1	-

concentration of Al-Na centers is 130 ppm, a rather large Al concentration. Second, it should be noted that there are two unique new peaks in the Na-swept amethyst, those at 20 K and 138 K, respectively. Third, there is a peak at 99 K that occurs for all three swept samples. This peak is ubiquitous for Fe-doped samples, and has been observed as well in citrine samples. Finally, there is a peak at 195 K only for the Li-swept and H-swept samples. A summary of all of these results is presented in Table 1 which gives the heights of all of these peaks. At the same time, Table 2 gives the activation energy, E, and pre-exponential, ν_o, for each peak, as obtained from the shift of the peak with changing frequency, ω.

All of these peaks are very close to Debye peaks, and therefore must be produced by simple point defects. Some additional peaks have been observed at still higher temperatures, but since these are very large and not always reproducible, it is concluded that they may arise from second-phase particles and interfaces.

The only additional peak produced by irradiation in all of these samples is the low temperature peak near 10 K. This is the peak that has been studied in detail and is now established to be due to the aluminum-hole (Al-h) center.[15,16] Thus, there is no new irradiation peak due to iron. The absence of an analog to the Al-h center for Fe can be explained by realizing that Fe^{3+} can change valence (to Fe^{4+}) when it captures a hole, rather than forming a dipole by having the hole on an adjacent oxygen ion, as in the case of Al^{3+}. Irradiation does, however, change the heights of the other Fe-related peaks, as shown in Table 1. Interestingly enough, the 20, 99 and 138 K peaks all are decreased by irradiation, but the 195 K peak is strongly increased.

To summarize the dielectric results, note that we have observed four relaxation peaks specifically related to the presence of Fe. Two of these: at 20 and 138 K, are also related to the presence of Na. It is tempting to regard these peaks as due to Fe-Na pairs, i.e. the analogs of the 30 and 75 K peaks due to the Al-Na center. The ubiquitous 99 K peak is difficult to identify at this stage. Finally, the 195 K peak, which is absent for Na-swept samples and increases upon irradiation, may be considered as possibly due to Fe compensated by H^+, since hydrogen is present both in unswept and H-swept samples.[17] Further, as in the case of Al-related centers in quartz,[18] room-temperature irradiation should liberate the alkali from substitutional Fe^{3+} and allow a proton to be trapped in its place.

Table 2. Activation Energies and Pre-exponentials of Fe-related Peaks

T (K) for 1 kHz	E (eV)	ν_o (sec^{-1})
20	0.030	1.5×10^{11}
99	0.17	4×10^{12}
138	0.22	5×10^{11}
195	0.37	2×10^{13}

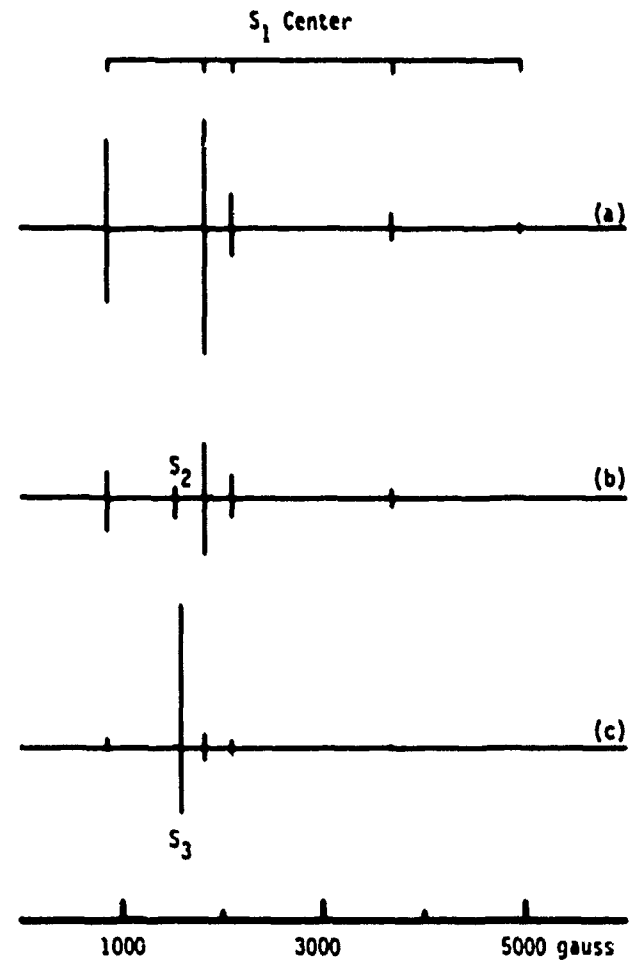


Fig. 4. ESR spectra taken at 27 K with the magnetic field parallel to the c axis. Trace (a) is from the lithium-swept sample, trace (b) is from the hydrogen-swept sample, and trace (c) is from the sodium-swept sample.

Electron Spin Resonance (ESR)

An ESR sample was cut from the center of each of the three samples that had been Li-, Na^+ and H^+-swept and first studied by DR measurements. Then, an ESR spectrum was taken at 27 K for each sample. Fig. 4 (a) shows the result obtained from the lithium-swept sample. The only spectrum present is the S_1 center. This correlates with an assignment of the S_1 center to a substitutional Fe^{3+} with an adjacent interstitial Li^+ ion. Other cultured amethyst samples which had not been swept but had been heated to the 400-500°C range exhibited a dominant S_1 ESR spectrum just like the

lithium-swept sample. This strongly suggests that the substitutional Fe^{3+} ions in the usual cultured amethyst are nearly all compensated by Li^{+} ions.

Fig. 4 (b) shows the result obtained from the hydrogen-swept sample. This spectrum contains a small S_1 center and also an S_2 center. Since the S_2 center corresponds to a proton compensating the Fe^{3+}, we would suggest that this sample was only partially swept (i.e., many, but not all, of the interstitial lithium ions adjacent to the Fe^{3+} have been replaced with protons).

Fig. 4 (c) shows the result obtained from the sodium-swept sample. This spectrum contains a very small S_1 center and a much larger spectrum which we have labeled the S_3 center. To our knowledge, the S_3 center has not been previously reported in the literature. We assign the S_3 center to a substitutional Fe^{3+} ion with an adjacent interstitial Na^{+} ion. Thus, we suggest that this sample was reasonably well-swept with the major fraction of the interstitial Li^{+} ions being replaced by Na^{+} ions. (This is supported by the appearance of the large Al-Na extra DR peak at 30 K in the Na-swept sample.) An important feature of the S_3 center is a rapid spin-lattice-relaxation time which broadens its ESR spectrum at higher temperatures. We could not observe the S_3 center above approximately 30 K because of this line broadening, whereas the S_1 and S_2 centers can be observed even at room temperature. This rapid spin-lattice-relaxation time is consistent with the small activation energy for the 20 K dielectric loss peak in Na-swept samples (see Table 2).

A striking change is observed in the ESR spectra when a sample is exposed to ionizing radiation at room temperature. The radiation (several Mrads) destroys S_1 or S_3 centers that are initially present and replaces them with S_2 centers. This is illustrated in Fig. 5 for the lithium-swept sample. This is consistent with the concept, already mentioned, that room-temperature irradiation should remove the alkali ion from an Fe^{3+}, thus allowing a proton to replace it. In addition, Al-h centers also appear after irradiation.

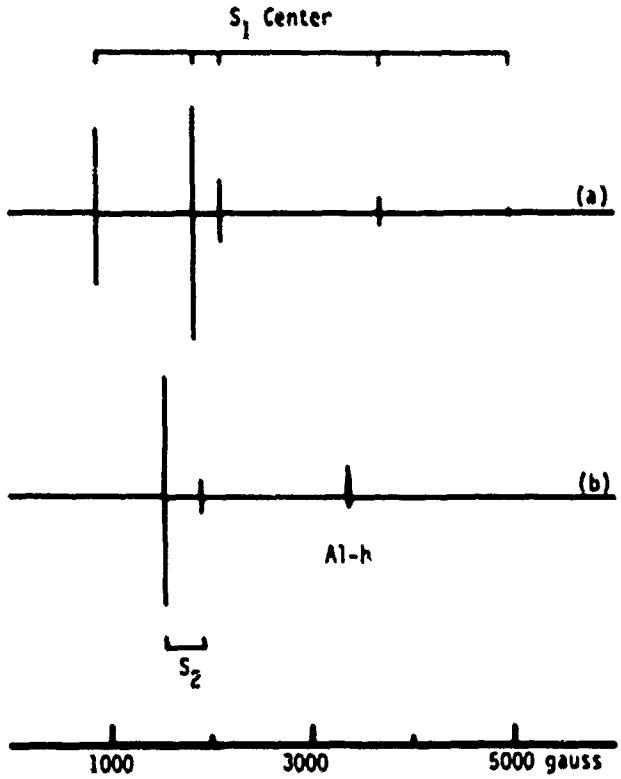


Fig. 5. Effect of room-temperature radiation on the lithium-swept sample. Trace (a) is before irradiation and trace (b) is after irradiation.

Two of the centers reported earlier, viz. the so-called I center and the center observed by Cox,[13] have not been observed in the present work. A likely explanation is that most previous work on ESR of Fe-containing quartz was carried out on natural crystals, which are probably more complex than the present (cultured) crystals.

Interrelation of DR and ESR Results

The two techniques used in this work measure very different defect properties. To obtain a dielectric relaxation (DR) response requires that the defect possess a net dipole moment and be capable of occupying more than one equivalent orientation. For electron spin resonance (ESR) the defect must possess an unpaired spin, so that the spin states can be split by a magnetic field. The ESR technique is particularly sensitive to ions with a $3d^5$ electronic structure (half-filled 3d shell), such as Fe^{3+}.

The newly observed 20 K and 138 K peaks and the newly observed S_3 center are readily linked together, as being caused by the Fe-Na center with the interstitial Na ion off-axis, similarly to the case of the Al-Na center shown in Fig. 1. This reason that this center is observed for the first time in the present work is probably because Li^{+} is the dominant alkali in both natural and cultured crystals, so that only after Na-sweeping can Na centers be readily observed. Previous workers did not examine Na-swept Fe-doped quartz.

The S_1 center does not seem to have a DR equivalent. The only possibility is the 99 K peak, but detailed analysis of the spin Hamiltonian for the S_1 spectrum shows that one principal axis is parallel to the 2-fold symmetry axis (or x-axis).[6,7] Since the S_1 center is an Fe-Li pair, this suggests that the Li interstitial resides on the 2-fold axis. But, if this is the case, there is no possibility for the low-temperature reorientation required for DR to take place. The situation for the Fe-alkali pairs then seems to be quite analogous to that for Al-alkali pairs, viz., the Fe-Li lies along the 2-fold axis while Fe-Na is off axis (as in Fig. 1), and therefore, only Fe-Na gives rise to DR. Then to what defect can the ubiquitous 99 K peak be attributed? Since it represents a substantial concentration of Fe and yet is not detected by ESR, it seems reasonable to conclude that it is due to a dipolar defect involving Fe in a valence other than 3+. Further work will be required to establish the nature of this center.

The remaining defects to be discussed are those that produce the E_1' spectrum in ESR and the 195 K peak in OR. It is tempting to regard that there are both the same, namely the Fe-OH defect. The association of the E_1' spectrum with this defect has already been established. In the case of the 195 K peak, the fact that it is present without irradiation (except for the Na-swept sample) and yet increases upon irradiation is strongly suggestive of a proton-related defect. However, the relative intensities of the E_1' spectrum for Na-, Li- and H-swept material both before and after irradiation does not correlate well with the strengths of the 195 K peak. For example, the E_1' has a strong intensity following irradiation of a Na-swept sample, while the 195 K peak is absent for this condition (see the last entry in Table 1). Accordingly, we cannot claim that the simple Fe-OH defect is responsible for the 195 K peak.

In conclusion, the simultaneous use of two major techniques for the study of Fe-related defects in quartz has better enabled us to sort out the wide complexity of defects in such crystals, but further work remains to be done to answer some of the questions raised.

Acknowledgements

The authors are grateful to Dr. B. Sawyer for growing the crystals, to Prof. J.J. Martin for carrying out the sweeping and to R. Hantehzadeh for assistance with the ESR measurements. This work was supported by the U.S. Air Force under contract F-19628-85-C-0104 with Columbia University and contract F-19628-86-C-0138 with Oklahoma State University.

References

1. J.C. King and H.H. Sander, Rad. Effects 26, 203 (1975).
2. L.E. Halliburton, Cryst. Lattice Defects and Amorphous Materials 12, 163 (1985).
3. J. Toulouse and A.S. Nowick, J. Phys. Chem. Solids 46, 1285 (1985).
4. J.J. Martin, J. Appl. Phys. 56, 2956 (1984).
5. J.A. Weil, Phys. Chem. Minerals 10, 149 (1984).
6. D.R. Hutton, Phys. Lett. 12, 310 (1964).
7. T.I. Barry, P. McNamara, and W.J. Moore, J. Chem. Phys. 42, 2599 (1965).
8. G. Lehmann and W. J. Moore, J. Chem. Phys. 44, 1741 (1966).
9. G. Lehmann, Zeits. Naturf. 22a, 2080 (1967).
10. L.M. Matarrese, J.S. Wells, and R.L. Peterson, J.Chem. Phys. 50, 2350 (1969).
11. H.D. Stock and G. Lehmann, J. Phys. Chem. Solids 38, 243 (1977).
12. M.J. Mombourquette, W.C. Tennant, and J.A. Weil, J. Chem. Phys. 85, 68 (1986).
13. R.T. Cox, J. Phys. C: Solid State Phys. 9, 3355 (1976); 10, 4631 (1977).
14. J.C. King, Bell Syst. Tech J. 38, 573 (1959).
15. W.J. De Vee and J. Volger, Physica 47, 13 (1970).
16. J. Toulouse, S. Ling and A.S. Nowick, Phys. Rev., to be published.
17. H.G. Lipson and A. Kahan, J. Appl. Phys. 58, 963 (1985).
18. L.E. Halliburton, N. Koumvakalis, M.E. Markes, and J.J. Martin, J. Appl. Phys. 52, 3565 (1981).

41st Annual Frequency Control Symposium - 1987

Transmission X-Ray Topography Of Single Crystal Quartz Using White Beam Synchrotron Radiation

William Hanson
Piezo Crystal Company
100 "K" Street
Carlisle PA
and
The Johns Hopkins University
Materials Science and Engineering

Summary

Single crystal quartz has been x-rayed in white beam synchrotron radiation yielding many topographs from each exposure. Enlarged sections of the topographs have been digitized and used for analysis. Samples included were quartz blanks, finished quartz resonators and lumbered bars. The topographs have been used to study defect density, defect strain field size, defect type, defect morphology, finished resonator mounting strain, finished resonator electrode strain, and resonator mode shape. Stereo viewing of the defects using two different but appropriately chosen topographs is possible. Defect depth and angle can be viewed.

The images have been digitized into 640 x 400 pixels. Each pixel has 16 intensity levels. Most images are displayed in grey levels. Contouring has been done by changing every third grey level to a color. The strain field size around the dislocations in the topographs has been measured by counting pixels in the digitized topographs. Dislocation densities have been measured by counting the number of dislocations in several one square millimeter areas on the same blank and doing counting statistics. Defect densities have been measured by counting pixels in the digitized topographs. Measurements taken from digitized topographs are correlated to the measured defect density of the same samples after etching. Measured defect densities vary from 3 per square cm to greater than 1200 per square cm. Defect strain field size has been measured by counting pixels in the digitized images. The measurement resolution is dependent upon the magnification prior to digitization. The resolution after digitization in this work varies from 0.6 microns to 25 microns.

Defects in crystals can be classified by group. Possible defect groups in quartz are point, line, plane, and volume (1). Each group has many possible types of defects. In quartz point defects are often Aluminum related, AL-OH, Al-hole are two examples. Possible line defects known to exist in quartz are edge and screw dislocations. Other types may exist which are combinations of these. Planar defect exist in natural quartz and may be point defects arranged in planes. Known volume defects in quartz include inclusions and voids. Anti-phase domains are also known in quartz (twinned regions). The appearance of defect type can not be explicitly determined by x-ray topography, but X-ray topography can be used with other techniques to identify the defect type. In the cultured samples the defect morphology varied considerably. The morphological differences appeared to vary as a function of the quartz grower. Possible reasons for the differences may be growth rate, position of the blanks relative to the seed and X-growth sections of the bars, nutrient solution variations, pressure, and temperature. The morphological differences reported here have been observed in blanks cut from premium Q pure Z growth bars of the same size. This should minimize the morphological differences due to variations in the distance between the seed and the blank.

Introduction

Quartz resonator parameters and defects in the crystal have been correlated with respect to Q. Warren P. Mason in (2) showed that losses in crystals are perturbated from their nominal values by defects. Perfect crystals only have losses due to phonon-phonon interactions. Crystals with defects perturbate the thermal relaxation time and phonon-phonon losses, changing the resonator Q. Recent work by J. J. Martin (3) and D. R. Koehler(4) have centered around defects associated with impurities. Sweeping (electrodiffusion) has been shown to reduce the level of some impurities. Impurities are nucleation sites for defects. Other device characteristics may also be correlated with defects. Defects are regions within the quartz crystal where the lattice parameters have been perturbated. The change in lattice parameters results in changes in the material properties. These may include all or some of the physical constants. In addition to changes in physical properties there is the possibility that material properties in an oscillating crystal are fluctuating around mean values. The fluctuations may result in increased noise in the oscillating signal.

X-ray topography has been used to study quartz resonator characteristics. R. A. Young and colleagues (5) used X-ray topography to show a correlation between surface defects induced by a manufacturing process and aging. Conventional topography techniques use X-ray tubes. These techniques produce good topographs but are time consuming techniques with limited resolution.

When manufacturing quartz resonators many fabrication processes cause variations in the finished resonator which are unrelated to material properties. These variations are well known to manufacturers of quartz devices. If variations in resonator parameters due to manufacturing processes and those due to material properties are to

be separated then the variations in defect characteristics must be measured statistically and a correlation study done with finished resonator parameters. The characterization of defects in quartz is a well established process when dealing with small samples. Once a history of a defect's morphological characteristics are known topographs may be used to identify specific defects with relative certainty. One technique which is capable of uniquely identifying a defect is transmission electron microscopy (TEM). The specific technique is to correlate a computer simulated TEM image based upon the defect model with a through focus series of an actual defect. When the through focus series (a sequence of images at different foci) matches the computer model of the defect then the TEM image can be uniquely identified. Defects appearing in x-ray topographs with a TEM image of the same region as the defect in the topograph is the basis for correlation. The compilation of defect morphologies on a large sample base of finished resonators is the first step in this process. This work presents several alternative methods to collect and analyze such a database.

EXPERIMENTAL METHODS

X-ray port 19C at The National Synchrotron Light Source (NSLS) in Brookhaven National Labs has been used to take these topographs(6,7). The measured X-ray flux at the port is 4.16 x 10 ^ 12 (photons/sec/mrad/.1% band/100 ma) . The beam size used was 40 mm wide by 7 mm high. The beam divergence is 2.5 milliradians (8.6 arc-minutes). The radiation hutch is 20 m from the tangent point of the storage ring.

Specimen and film orientation relative to the beam have been varied to evaluate the usefulness of each method in characterizing resonators. Three orientation combinations were tried. Figures one through three are diagrams of each orientation. Photos one through three are sample topographs from each orientation. Kodak type 52 film has been used in 8 inch by 11 inch sheets. Exposure times varied as a function of thickness and beam current. All film processing was done on site.

SAMPLE AND FILM PERPENDICULAR TO BEAM

The simplest geometry to use when taking topographs with the synchrotron is to have the sample and film perpendicular to the X-ray beam. This type of exposure is not only easy to set up but the defect density can be directly measured from the film. The topographs have been used for lattice plane indexing and defect density measurements. Figure one is a drawing of this configuration. In this sample orientation a beam stop is necessary to stop the direct beam from exposing the film. Photo one is a sample topograph exposed with this type of geometry. The sample was a 10 MHz 3rd overtone SC cut resonator. The quartz resonator plates (blanks) are .600 inches in diameter. Defect strain field size and defect density have been measured directly from this type of topograph.

The exposed portion of the quartz plate has been set up to be the active region of the finished resonator. The disadvantage of this sample orientation is that the whole blank is not exposed to X-ray unless the film and sample are moved together in the beam. Defect density measurements are not restricted by this disadvantage. Photo one is an example topograph from this geometry. The light central rectangle is the unexposed portion of the film under the beam stop. The small dark rectangles are the topographs. Their position is a function of the lattice plane spacing and orientation. The central darkening of the topograph is due to scattering of the central beam.

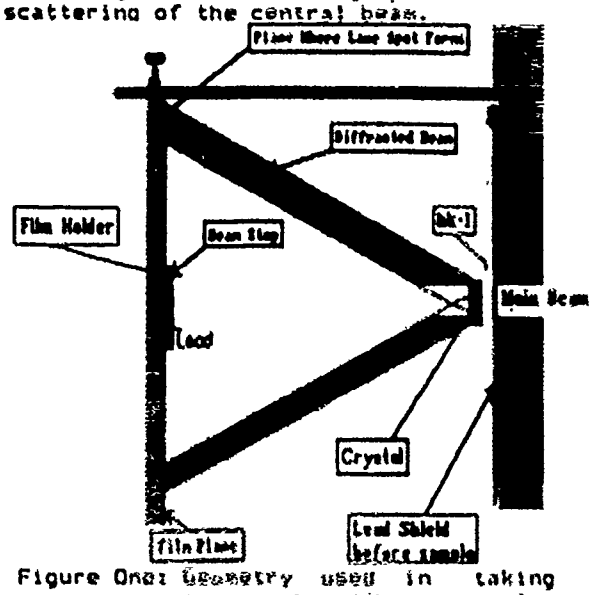


Figure One: Geometry used in taking topographs with the sample and film perpendicular to the X-ray beam

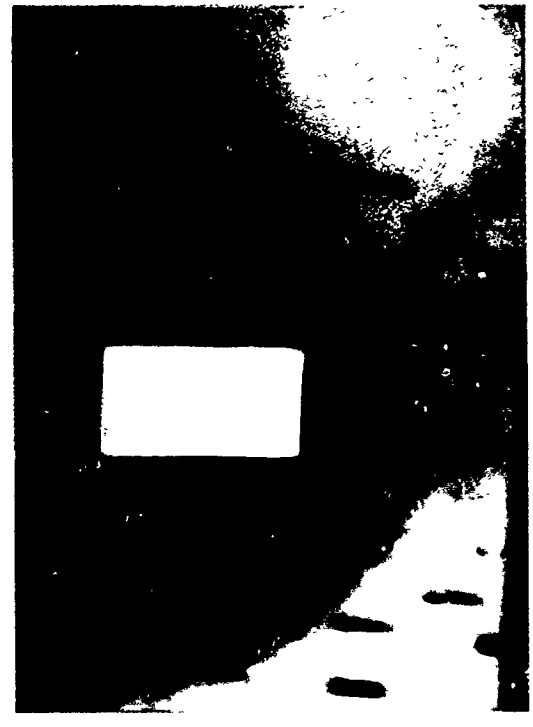


Photo One: Example topograph taken tith sample and film perpendicular to the beam

SAMPLE TILTED IN BEAM WITH FILM PERPENDICULAR TO BEAM

The tilt has been mechanically set to 82 degrees. Topograph images vary from edge images to complete images of the plate from planes nearly parallel to the plates surface. This makes it somewhat more difficult to index but results in topographs with additional information from lattice planes not recorded in the topograph with the sample perpendicular to the beam. This orientation also results in both transmitted and reflected topographs recorded on the same piece of film, both formed by Bragg diffraction. In figure two the Laue images formed on the top half of the film are transmission images through the thickness of the plate while the bottom images are formed from X-rays diffracted from the plate through the surface pointing towards the beam.

The images formed with this geometry have been useful for characterizing the strain induced by mounting of precision resonators. The strains may be related to aging problems in precision resonators. Mode shapes of oscillating resonators have also been recorded. Unsealed plated resonators have been x-rayed to characterize the mode shapes using this orientation as well. Photo two is an example topograph from this geometry. The light central circle is the portion of the topograph under the beam stop. The dark ovals with varying dimensions are the topographs. Some planar defects can be seen crossing the topograph images. The variation in intensity of the spots is mainly due to the structure factor. The position of the topograph image is dependent upon the lattice plane orientation and 'd' spacing.

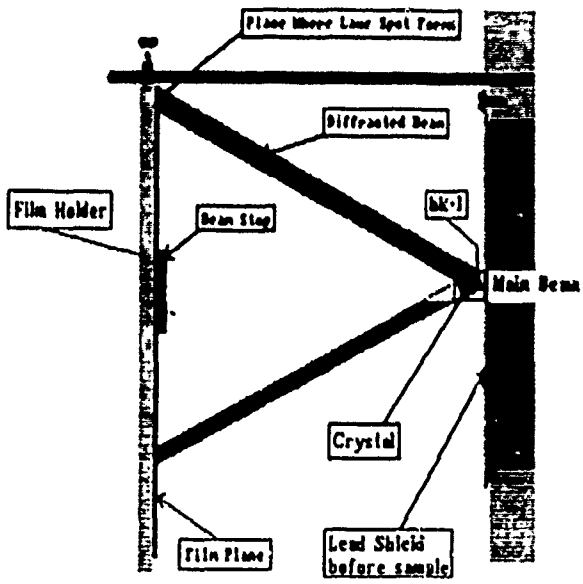


Figure Two: Geometry used in taking topographs with the sample tilted in beam and film perpendicular to the beam

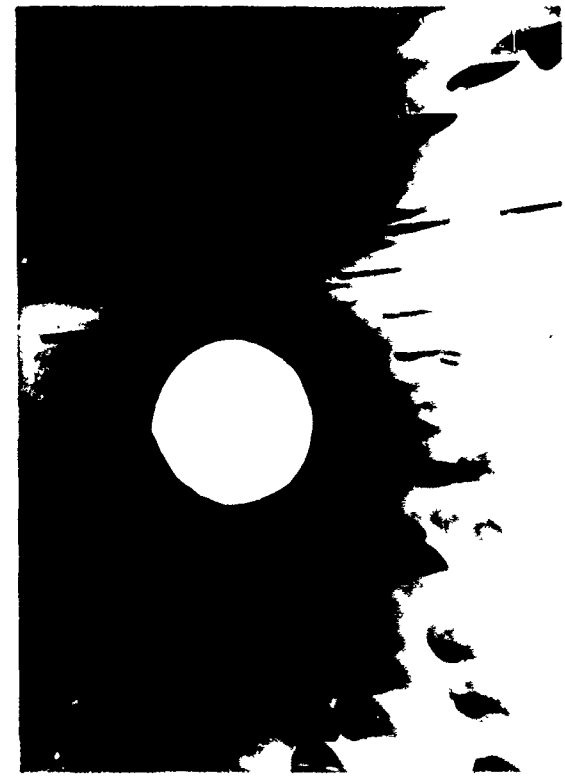


Photo Two: Example topograph taken with sample tilted in beam and film perpendicular to beam

SAMPLE TILTED IN BEAM WITH FILM PARALLEL TO BEAM

This orientation results in many complete images of the quartz plate from lattice planes nearly parallel to the surface. These images form good stereo pairs. Photo three is an example topograph from this geometry. The small central spot is an artifact due to the digitization process not from a beam stop. The dark circles are the topographs. The darkening of one side of the topograph is due to scattering of the main beam.

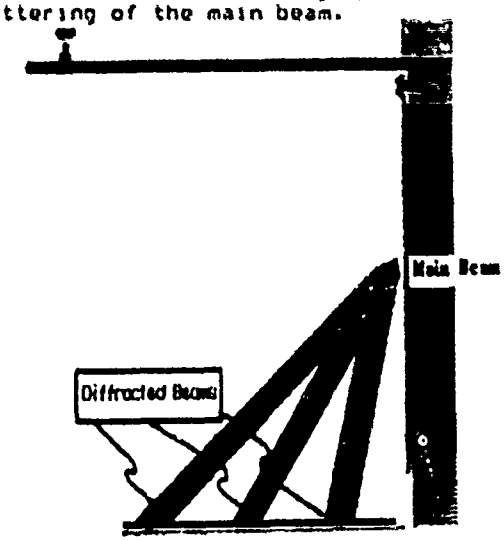


Figure Three: Geometry used in taking topographs with the sample tilted in the beam and the film parallel to the beam

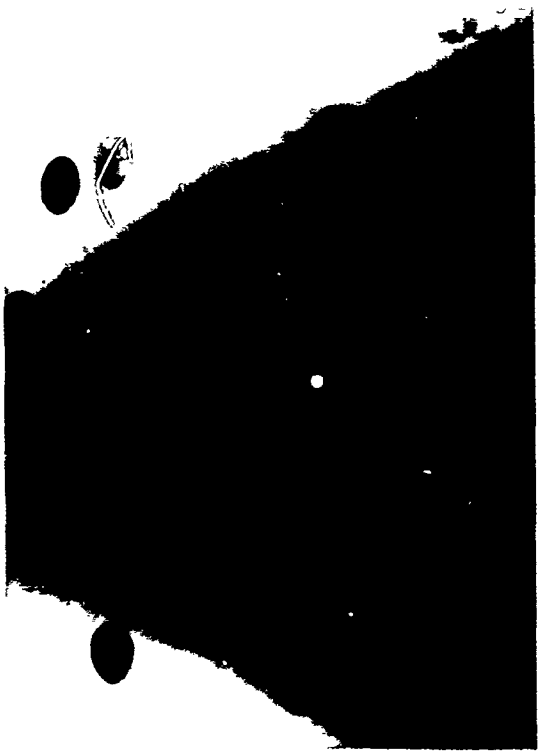


Photo Three: Example topograph with the film parallel to the beam and the sample tilted in the beam

RESULTS AND DISCUSSIONS

DEFECT DENSITY

Defect density is measured by sectioning off a square millimeter of the topograph and counting the number of defects in that area. After a minimum of ten regions have been measured the mean and standard deviation of the defect density are calculated. Photo four is a sample which has been analyzed by this method.



Photo 4: Enlarged topograph of precision quartz resonator manufactured from vendor C quartz bar.

Mean Defect Density 450 / cm^2
Standard deviation 160 / cm^2

DISLOCATION TYPE AND BURGERS VECTOR

A dislocation as defined by N. W. Ashcroft and N. D. Mermin [3] must have the following properties in the region of the dislocation.

1. Away from the region the crystal is locally only negligibly different from the perfect crystal.

2. In the neighborhood of the region the atomic positions are substantially different from the original crystalline sites.

3. There exists a non-vanishing Burgers vector.

Edge dislocations and screw dislocations are two of several defect types known to exist in quartz. Screw dislocations have a larger strain field than edge dislocations and are known to be nucleations sites for new material growth in crystals[7].

DEFECT MORPHOLOGY IN NATURAL AND CULTURED QUARTZ

All samples used in the morphological comparison were cut from bars of the same size except vendor A which has an unknown bar size. The blanks were oriented as SC cut resonators with a 21 degree 56 minute X angle and 34 degree 00 minute Z angle. The units were 10 MHz 3rd overtone units. Defect morphology appears to vary by source. Natural quartz has a greater variation in defect morphology than cultured quartz. The direction of the dislocations in the quartz plates varies but groups tend to be parallel. Some cultured samples had defect densities less then ten per square centimeter.

DEFECT STRAIN FIELD SIZE

The size of the strain field around the defect can be measured in digitized images. The strain field around the defect is a region where the lattice constant is perturbated from its nominal value. The perturbation is a change in the lattice spacing and thus the Bragg angle. The strain field is then part of the defect appearing as a change in contrast in the topograph. The measured defect density in photo four was 113 microns in diameter with a standard deviation of 43 microns.

Four sources of quartz were evaluated for defect density. The following table summarizes the results.

Source	Number of samples	Dislocation density per cm^2	Planar defects per cm^2
Natural	10	0 to 1000	0 to 30
Vendor A	1	1500	0
Vendor C	6	50 to 450	0
Vendor D	20	3 to 1000	0

DISLOCATIONS AND ETCHING

An etch pit is well known to form at the end of a dislocation line. Samples that were x-rayed were subsequently etched. Etch channels appeared in the same place as the dislocation line seen in the topographs.

NATURAL QUARTZ DEFECT MORPHOLOGY

Natural quartz had a greater variety of defects then the cultured samples. Photos 5,6,7,and 8 are all topographs of natural quartz taken with the sample and film perpendicular to the beam.

Photo 5 is a typical natural quartz specimen with planar, linear and small spherical defects. Photo 6 shows a planar defects coming to a corner. In this case they come to a corner where two 'm' faces on the natural stone came to a corner. Photo 7 shows several planar defects coming together with large variations in contrast along the planar defects. Photo 8 is less common and shows a large strained region emanating from a circular region where there is a volume defect.



Photo 5 - Natural Quartz
Planar, linear and small spherical defects



Photo 6 - Natural Quartz
Planar defects coming to the corner
of two 'm' faces

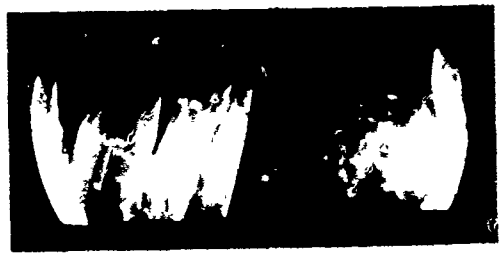


Photo 7 - Natural Quartz
Planar defects with varying contrast
with the planar defect

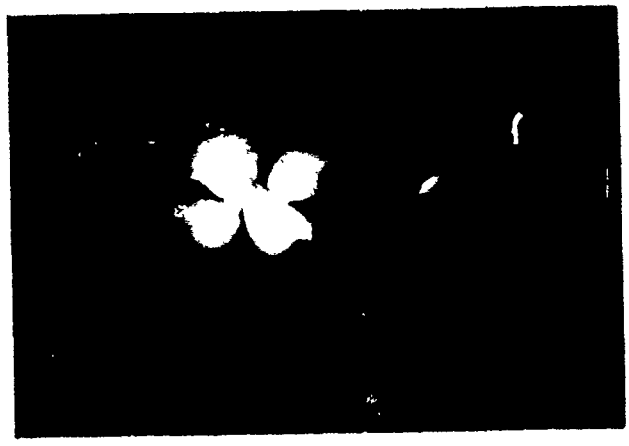


Photo 8 - Natural Quartz
Volume defect with strain field

DEFECT MORPHOLOGY OF CULTURED QUARTZ FROM VENDOR A

This is the only sample X-rayed from vendor A. The dislocation density in the topograph was measured at 1500 / cm^2. The dislocations are nearly all parallel.

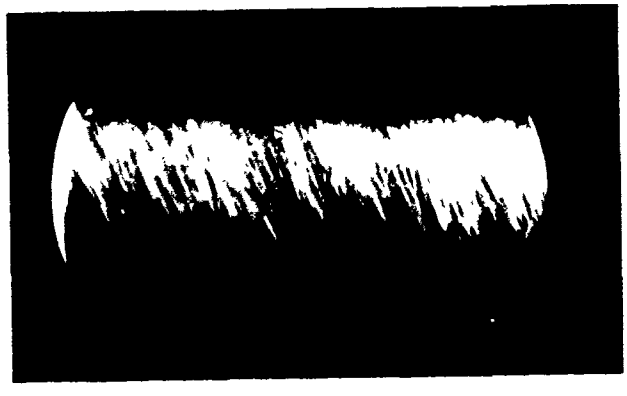


Photo 8 - Vendor A
High dislocation density quartz

DEFECT MORPHOLOGY OF CULTURED QUARTZ FROM VENDOR C

Vendor C had a small variation in defect morphology. Photos 9 and 10 are two typical samples.



Photo 9 - Vendor C
Typical defect morphology from vendor C



Photo 10 - Vendor C
Typical defect morphology from vendor C

DEFECT MORPHOLOGY OF CULTURED QUARTZ FROM VENDOR D

Vendor d quartz was one of the more interesting. The defects were arranged at the edges of cells. Photo 11 is a typical topograph from this vendor.



Photo 11 - Vendor D
Typical defect morphology from vendor D

MOUNTING STRAIN, ELECTRODE STRAIN, AND ARTIFICIALLY INDUCED STRAIN

Strain is easily observed in topographs of crystals because it perturbates the lattice spacing within the strained region. Photos 12,13 and 14 show three separate types of strain. Photo 12 is a topograph of a four point mounted 10 MHz 3rd overtone SC cut precision resonator. The mounting strains can be seen easily. Photo 13 shows the strain induced by an evaporated chromium gold electrode. The electrode had 25 angstroms of chromium under 600 angstroms of gold. Photo 14 shows strain induced in the surface of the blank from blank manufacturing processes. The strained regions in photo 14 are not visible on the blanks surface under a 30 power magnification.



Photo 12 - Mounting strain on a four point mount 10 MHz 3rd overtone SC cut



Photo 13 - Strain induced by evaporated chromium gold electrode



Photo 14 - Strain induced by manufacturing processes

LARGE BAR TOPOGRAPHS

Bars as thick as .530 inches have been x-rayed successfully. This topograph is rather jumbled with information. Thinner samples are easier to interpret.



Photo 18 - Topograph of thick bar (.530 inches thick)

STEREO VIEWING

Two images from the same topograph forming a stereo pair. These topographs were taken with the film parallel to the beam and the sample tilted in the beam. This image if from the multiple image topograph in photo 3.

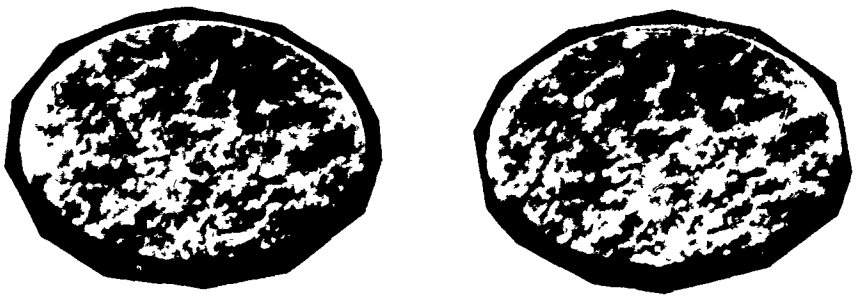


Photo 17- Stereo Pair

MODE SHAPES

This topograph is of an oscillating 10 MHz 3rd overtone SC cut resonator. The central region is lighter where the resonator is oscillating.



Photo 19- Oscillating Crystal

TOPOGRAPHS BEFORE AND AFTER SWEEPING

Several samples were X-rayed, swept, and then X-rayed again. Photos 15 and 16 are the before and after topographs from one of these. Very little change is evident.



Photo 15 - Before sweeping



Photo 16 - After Sweeping

CONCLUSIONS

1. X-ray topography is a useful technique for studying defects in precision quartz resonators

2. Etch channels can be correlated with the straight dislocations in the topographs

3. Sweeping does not appear to change the dislocation density

4. Mounting stresses, electrode stresses and artificially induced stresses can be detected in resonators

5. The defect morphology from each source was different

6. Three dimensional defect morphology can be seen in stereo images formed from different lattice planes, but with similar contrast variation.

Bibliography

1. L. V. Azaroff, "ELEMENTS OF X-RAY CRYSTALLOGRAPHY," McGraw Hill, 1968 pp 229-255

2. W. P. Mason, "PHYSICAL ACOUSTICS, Principles and Methods, Volume V," Academic Press N.Y. 1968

3. J. J. Martin, "ALUMINUM-RELATED ACOUSTIC LOSS IN AT-CUT QUARTZ CRYSTALS,"Proc. 38th ASFC*, pp 16-21 (1984)

4. D. R. Koehler, "POINT DEFECTS IN CULTURED QUARTZ: RECENT ACOUSTIC LOSS, INFARED& AND MAGNETIC RESONANCE RESULTS," Proc. 35th ASFC*, pp 317-321 1981

5. R. A. Young, R. B. Belser, A. L. Bennett, W. H. Hicklin, J. C. Meaders, and C. E. Wagner, "SPECIAL X-RAY STUDIES OF QUARTZ FREQUENCY CONTROL UNITS," Proc. 19th ASFC*, pp 23-41 (1965)

6. John C. Bilello, Haydn Chen,Anthony Hmelo, John M. Liu, Howard K. Birnbaum, Patrick J. Herley and Robert E. Green, "THE SYNCHROTRON TOPOGRAPHY PROJECT (STP) AT THE NATIONAL SYNCHROTRON LIGHT SOURCE,"Nuclear Instruments and Methods 215 (1983) 291-297, North Holland Publishing Company

7. Nicholas F. Gmur and Susan M. White-DePace, "USERS MANUAL: Guide to the VUV and X-Ray Beam Lines," The National Synchrotron Light Source, Brookhaven National Labs, May 1986

8. N. W. Ashcroft, N. D. Mermin, "SOLID STATE PHYSICS," Saunders College, 1976,

Acknowledgements

The author would like to thank R. C. Green, J. Winter, W. Samuelson and H. Hanson for helpful discussions, and Jaqueline Hanson for developing most of the topographs.

41st Annual Frequency Control Symposium - 1987

STUDIES OF QUARTZ RESONATORS BY STROBOSCOPIC TOPOGRAPHY

A. ZARKA & B. CAPELLE
Laboratoire de Minéralogie-Cristallographie, Université P. et M. CURIE, 4 place Jussieu, Paris 75252, France

and

J. DETAINT & J. SCHWARTZEL
Centre National d'Etudes des Télécommunications, PAB/RAC/HCT
Bagneux, 92220, FRANCE.

Summary

Stroboscopic X-ray topography have been carried out to observe the states of acoustic vibrations in quartz resonators. These experiments have been done using the storage ring of D.C.I. (Orsay, France) as a pulsed light source. The studied AT resonators have shown that the vibration mode is directly related to the density and the distribution of the defects.

Introduction

The first stroboscopic topography experiments have been performed on Fe(Si) crystals by means of a mechanically chopped beam with the time resolution of about one millisecond (1). Time resolved experiments on shorter time scale (between 1ms and 1ns) can be done using the intrinsic structure of synchrotron radiation originating from a Storage Ring. Stroboscopic topography, based on an exact synchronization between the observed phenomenon and the time structure of the source is very well adapted to study the resonators using synchrotron radiation because the frequency of the vibrations in the resonators and the frequency of the pulsed beam are of the same order (some MHz). So, studies of acoustic surface waves on YZ-LiNbO3 (2,3,4) and on quartz resonators (5) have been developed to study the vibration modes with a time resolution of about one nanosecond. This was done at Hasylab using the synchrotron radiation of the storage ring DORIS.

In this article we report some results obtained using the pulsed beam delivered at L.U.R.E. by the storage ring D.C.I. concerning the interactions between the BAW (Bulk Acoustic Waves) vibrations and the lattice defects observed by stroboscopic topography in quartz resonators.

EXPERIMENTAL TECHNIQUES

Synchrotron radiation from Storage Rings has a time structure because the circulating particles (positrons in the case of D.C.I.) are concentrated in bunches. Two fixed parameters of this time structure are the circulation time of an individual bunch and its length. For D.C.I. these values are 315 ns and 1ns. Another variable parameter, in some other cases, is the number of the stored bunches but in our experiments this number is always 1 because the storage ring of D.C.I. was running in single-bunch mode.

The setting of the experiment is composed of two parts :

- The topographic setting is quite simple because the topographic station at L.U.R.E. is equiped with a double axis spectrometer. This spectrometer permits, when the quartz sample is placed in the incident (white) beam, one to obtain a Laue pattern of topographs (6). This setting is simple because the resonator, mounted on a goniometer head with a support adapted for the frequency adjustment is placed perpendicular to the beam without very precise adjustment. However, the most interesting aspect of this spectrometer is the monochromatic setting. In this case, the first axis holds a (110) sample of germanium which is adjusted to select from the white beam only one wavelength. The 220 reflection was choosen and the wavelength of the beam on the second axis was 0,7 Å for a Bragg angle $\theta = 10{,}08°$. From the resonator, adjusted on this axis, monochromatized images have been obtained. By the use of very fine slit, we have also obtained "section"topographs.

- The electronic control. To obtain an exact synchronization of the X-ray pulses and of the piezoelectric vibrations of the resonators, a pulse signal obtained from the positron bunch (by a capacitive pick up) is used to generate the excitation signal of the resonators. This signal is simped in sinusoidal form by filtering at the recurrence frequency of the synchrotron (f_o = 3,169280 MHz) or at an harmonic of this frequency (n f_o, with n = 2,3...). A phase shifter permits to vary the relative phase between the synchropulse and the sinusoidal signal used to excite the resonance of the crystal (after a level adjustment).

To have an exact resonance condition (resonator voltage in phase with resonator current) the resonators were adjusted by metallization to have frequencies very close to f_o (or n.f_o). A fine adjustment (a few Hertz) is made with a large serial variable capacitance.

The electrical parameters of the resonator are measured, in situ, with a vector voltmeter (level of excitation, verification of the zero phase resonance condition). Another vector voltmeter is used to monitor the relative phase angle between the resonator current (or voltage) and the synchronization pulse. After calibration, this relative phase between the resonator current and the X-ray pulse (R.P.C.X. in the following) is known with a precision of about 3°, most of the incertitude resulting of the effects of temperature fluctuations on the resonator.

The general principle of the set-up is given in Figure 1a. An example of relative time position of the X-ray sampling pulse and of the resonator current is displayed in Figure 1b.

CH2427-3/87/0000-236 $1.00 C 1987 IEEE

EXPERIMENTS

The purpose of this study was to determine the role of the defects on the acoustical vibrations.

In the stroboscopic experiments the samples were plano-convex resonators having the basic design of high Q factor 5MHz, 5th overtone resonators (Table 1). The first resonator Q1 made with premium Q quartz was operated at third overtone at the synchrotron frequency. The second Q2 and the third Q3 crystals were made respectively with premium Q synthetic quartz and with natural quartz, they were operated as fifth overtone at two times the synchrotron frequency (Figure 2).

All resonators were electroded with Cr/Au metallization (electrode diameter 8 mm, mass loading ≠ 1%).

As established by the theory (7), the used vibration mode are much more trapped in the case of Q2 and Q3 than for Q1.

Table 1

Resonance frequency	Mode	Rc	F	2he
Q1 3.169280 MHz	Third Overtone	150mm	15mm	1,592mm
Q2 6.338560 MHz	Fifth Overtone	150mm	15mm	1,322mm
Q3 6.338560 MHz	Fifth Overtone	150mm	15mm	1,322mm

The dominating vibration mode in the AT-cut is the thickness-shear mode which corresponds to a standing transverse wave in the crystal (Figure 3).The planes parallel to the surface are displaced in the X_1 direction leading to a sinusoidal shear of planes perpendicular to the $\vec{X}_1$ axis.

The components u_2 and u_3 of the displacement are generally weak in regard to the u_1 component, when no coupling to flexure or plane shear occurs.

On the X-ray topographs, the most important constrast located under the electrodes is due to the u_1 component and depends on the level of the amplitude of the vibration applied on the crystal. This contrast depends also on the diffraction conditions. Particularly, when the product $\vec{g}.\vec{u}$, where $\vec{g}$ is the reflection vector and $\vec{u}$ the displacement due to the deformation, give zero the contrast due to the deformation vanishes.

In our cases, the most important u_1 component of the displacement is along the [100] direction. Then, on the topographs corresponding to the (0kl) reflecting planes the contrast of the deformation does not appear.

OBSERVATIONS

Different types of X-ray topographs have been obtained varing the R.P.C.X. , the level of the input signal and the topographic setting (Laue, monochromatic, section topographs...). We shall only report the most important observations we have obtained for the studied resonators.

Q1 Resonator

In Figure 4, we present three topographs of the same Laue pattern obtained for 120° R.P.C.X. and a power of 0.9 mW. It can be note, that depending on the reflection, the contrast due to the deformation changes. The central black contrast which depends on the u_1 component of the deformation is clearly visible for the $11\bar{2}$ (Figure 4.a) and $2\bar{2}3$ (Figure 4.b) reflections but vanishes for the 013 reflection (Figure 4.c). For this reflection $\vec{g}.\vec{u} = 0$. In contrary, the contrast (broad fringes) due to the u_3 component of the deformation is more visible.

In Figure 5, are presented two topographs of the crystal obtained for $11\bar{2}$ reflection for two different R.P.C.X. (with the same vibrating level). For 90° R.P.C.X. (Figure 5.a), the crystal present a contrast very similar to that obtained without vibration because the diffracting planes are not deformed. It can be observed the high density of the dislocation oriented parallel ro the $\vec{c}$ crystallographic axis. The faint circular contrast is due to stresses in the electrode metallization. The effects of the dislocations on the vibration mode are shown in Figure 5b (and also in Figure 4.a and 4.b) where it can be obviously seen that the zone of vibration has not the quasi circular symetry expected and that where a large density of dislocation exists, they unconfine the mode. This was also observed elsewhere by conventional X-ray topography (continuous exposure during the vibration).

The same features can be observed on the monochromatic topographs. In Figure 6,a monochromatic topograph (210 reflection) of the same resonator shows another interesting contrast in the central resonating zone : fringes of "moiré" type appear and their configuration change with the R.P.C.X. . These fringes are due to interferences between X-ray wavefield propagating in the crystal. They are generally located at the level of the dislocations and characterize probably very locally deformed zones.

Q2 Resonator

This crystal is characterized by a better confinement of the vibration. The four monochromatic topographs presented in Figure7 were obtained varing the level of the input signal (respectively, without vibration in Figure 7.a , 15 µW in Figure 7.b, 35 µW in Figure 7.c and 330 µW in Figure 7d). The density of the dislocation in this crystal is less important than in the previous one. However, the "moiré" type fringes are visible in the central part at the level of the dislocations. Their configuration varies with the RPCX.

The geometrical aspect of the central vibrating zone seems less deformed than the vibrating part of the Q1 crystal. For this Q2 resonator it appears, due to their lower density, that the influence of the dislocations is less important.

In other hand, circular fringes located between the black contrast and the edge of the electrodes present a contrast which varies with the RPCX. They are also due to interferences of the propagating wavefields in the crystal but reveal a probable secondary mode of vibration.

Q3 Resonator

Four monochromatic topographs are presented in Figure 8. The last three were obtained with the same level of the vibration (90 µW) but for different RPCX (respectively 45° in Figure 8.b, 0° in Figure 8.c and 135° in Figure 8.d). The topograph in Figure 8.a was obtained without vibration. This crystal is a natural one and the defects present are only growth bands.

This crystal shows different interesting features:
- The central black zone presents a less contrasted (white) part which localization depends on the RPCX.
- This is also verified for the "moiré" type fringes which configuration changes between Figure 8.b,

Figure 8.c and Figure 8.d.
- The external circular fringes (between the vibrating zone and the edge of the electrodes) appear only when the crystal is vibrating and seems to be invariable when the RFCX changes.
- The growth bands seem to have an effect on the lateral confinement of the vibration mode very less important than the dislocations.

DISCUSSION

To summarize the different observations, it can be said that the global effect of the dislocations in AT-cut resonators is to decrease the lateral confinement of the thickness shear mode when their density is important (> 10^2 to $10^3/cm^2$). The effect of the growth bands is until now not clarified.

The technique of the Laue Stroboscopic topography can reveal directly the different modes of the vibration by visualization of all the components of the (instantaneous) deformation.

It is most probable that the deformations of the Moiré fringes (produced by an X-ray interference process not already fully elucidated) correspond to local modifications of the transverse wave shape by resulting of the influence of the defects. (Surface defects appear to produce similar effects).

The dislocations seems to have the effect of increasing locally the phase velocity of the shear mode and so to contribute to untrap locally a confined mode. This hypothesis appears to be renforced by the observed fact that bunches of dislocations of similar orientations constitute a guiding structure for the shear wave at the periphery of the vibration mode. This permit , in the case of weakly trapped modes, to observe an important untrapping effect.

Although no direct experimental evidence of this other phenomenon were obtained by now, one can also wonder whether the dislocations, as other elastic discontinuities, can or not introduce some mode conversion (fast shear to slow shear or longitudinal).

The higher impurity level in growth band may also introduce local variations of the phase velocity of the shear mode, leading to increase (or decrease) locally the energy trapping parameters, but due to the complexity of the growth bands structure of the studied sample, no simple conclusion about this point can be drawn now.

Further experiments are requested, with resonators exhibiting different dislocation densities, and different energy trapping parameters to gain a full understanding of the importance of the interactions of the defects with the vibration modes.

However, the Moiré type fringes show the local deformations (associated with the defects which seem to induce perturbations in the vibration odes. In addition, "Section" topographs indicate the depth of the acoustic contrast in the crystal (the areas with maximum curvature of planes).

The stroboscopic method can also permit to obtain other informations about the piezoelectric resonators such as informations relative to the localization of acoustic dissipations, or to the influence of imperfect geometries of the surfaces.

CONCLUSION

The time structure of Synchrotron Radiation enables Stroboscopic X-ray Topography with nanosecond time resolution. With the use of the beam delivered at L.U.R.E. we have examined three quartz resonators.

By the use of different topographic setting, several types of vibration modes and the perturbations due to the defects have been observed.

Defects like growth bands and specially dislocations disturb the B.A.W. propagation in crystals and can induce losses either by untrapping a part of the energy of the mode or by mode conversion.

However, the topographic contrast due to the propagation of the X-ray wavefields in the deformation fields of acoustic waves, is not well understood and calculations of the beam trajectories in the crystal and simulation of topographs will be necessary to explain the observed contrasts.

REFERENCES

(1) J. Miltat and M. Kleman
J. Appl. Phys. 50, 7695 (1979).

(2) H. Cerva and W. Graeff
Phys. Stat. Sol.(a) 82, 35 (1984).

(3) H. Cerva and W. Graeff
Phys. Stat. Sol. (a) 87, 507 (1985).

(4) H. Cerva and W. Graeff
Phys. Stat. Sol. (a) 93, K129 (1986).

(5) C.C. Giver, W. Graeff and H. Moller
Nucl. Inst. & Meth. 208, 701 (1983).

(6) T. Fukui, K. Naukkarinen and P. Rabe
Phys. Stat. Sol. A25, 93 (1974)

(7) H.F. Tiersten and R.C. Smythe
J.A.S.A., vol. 65, 1555 (1979).

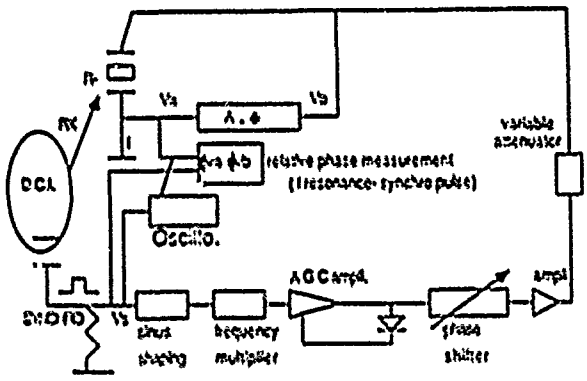


Fig. 1 a

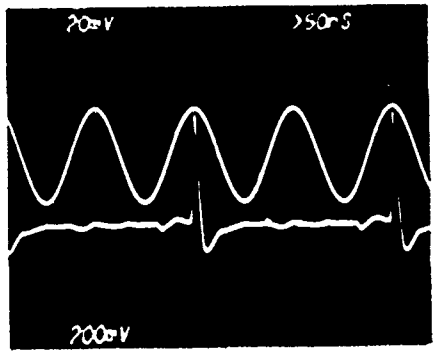


Fig. 1 b

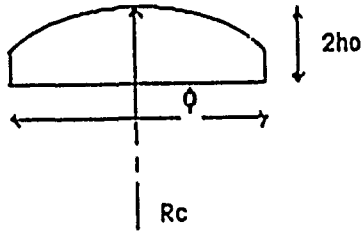


Fig. 2

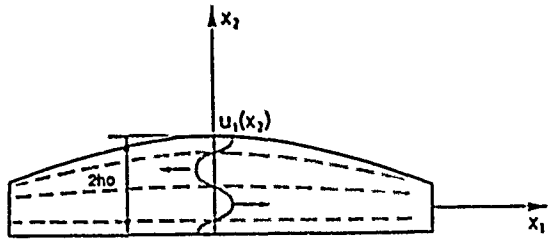


Fig. 3

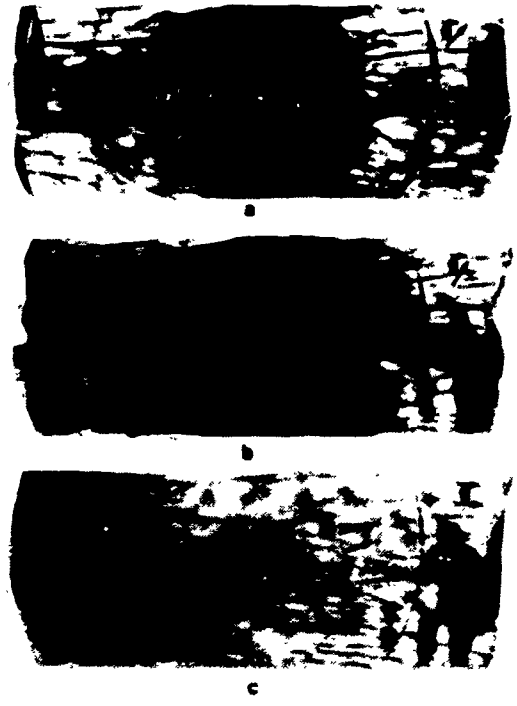


Fig. 4

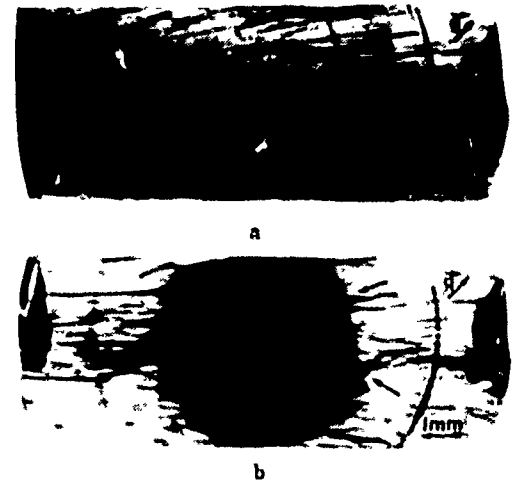


Fig. 5



Fig. 6

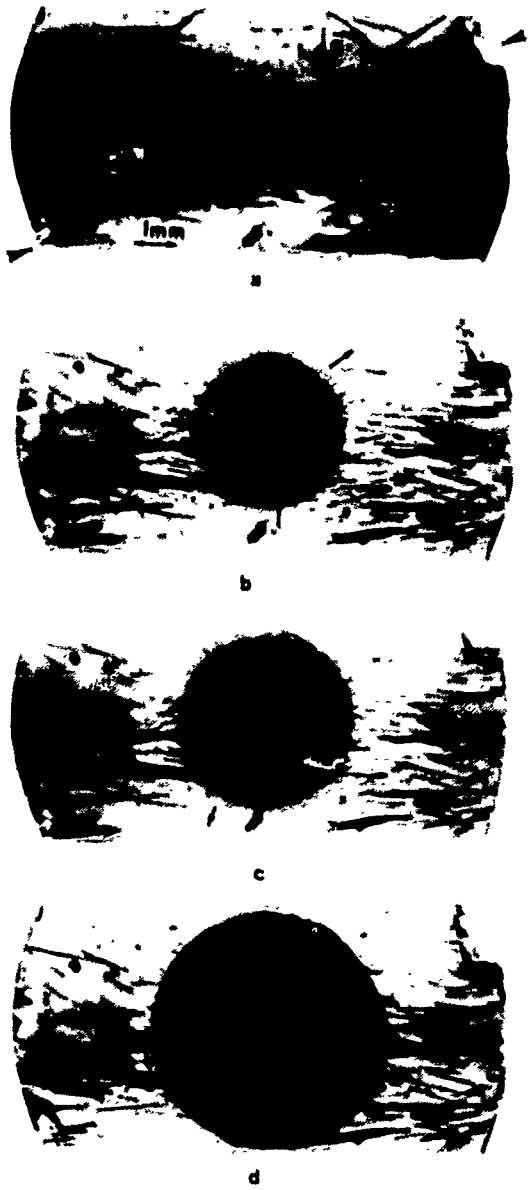


Fig. 7

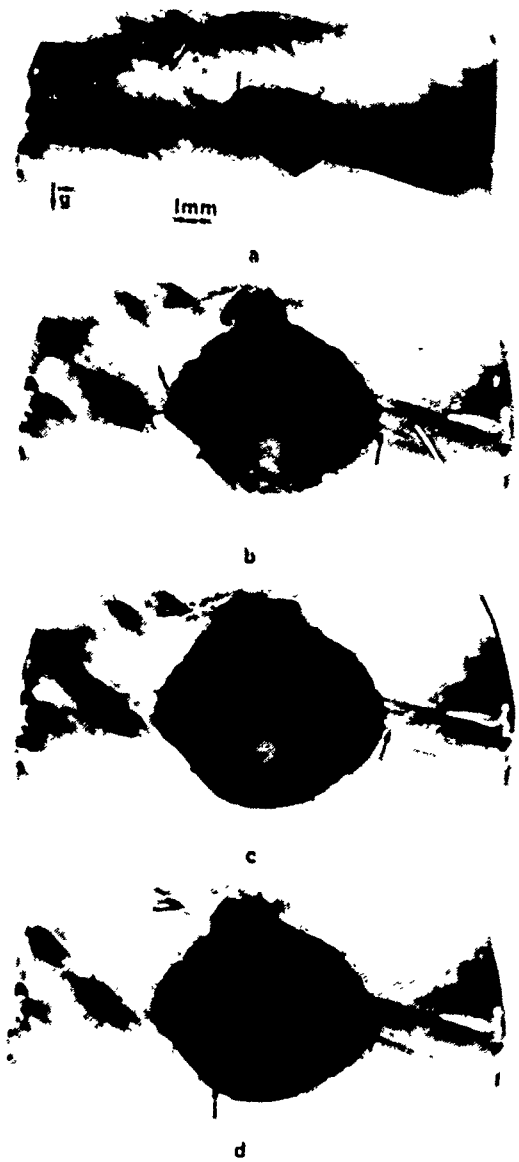


Fig. 8

41st Annual Frequency Control Symposium - 1987

CULTURED QUARTZ QUALITY STANDARDS UPDATED BY THE EIA

BALDWIN SAWYER, CONSULTANT

Sawyer Research Products, Inc.
35400 Lakeland Blvd.
Eastlake, Ohio 44094

Abstract

Revised or new standards for quality measurements on cultured quartz were adopted for: infrared absorption measurements, etch channel density, and inclusion count densities.

The Electronic Industries Association accepted in March 1987 the recommendations of its Quartz Material Study Committee to update its EIA 477-1 specification, and to propose the test procedures and standards in that specification for acceptance at the upcoming International Electrotechnical Commission meeting.

A major change in this revision is the restatement of the quality grades for cultured quartz into alpha$_{3500}$ terms instead of the 5 MHz Q terms used heretofore. Since the indicated 5 MHz Qs were based on alpha$_{3500}$ measurements, the actual operative grade limits will be, in fact, unchanged from the earlier ones.

Additional test procedures are included for obtaining standard etch channel density counts, cm^{-2}, and inclusion count densities, cm^{-3}, in specified size ranges.

The ongoing process of standardizing quartz material testing methods has progressed another step. In the USA the Electronic Industries Association (EIA) adopted, recently, three revised quartz material test specifications, designed to yield reproducible measurements in different test locations. Further, these standard test procedures are to be proposed for acceptance at the July 1987 International Electrotechnical Commission (IEC) meeting.

In normal sampling practice one or more of these tests is applied to a group of crystal bars statistically sampled from one hydrothermally grown run batch. Two of the tests need lapped and polished oriented slices from the sample bars, while the third can be used on as-grown bars. For most usual applications at moderate quality assurance levels, a low percentage sample of the batch's bars can represent the batch and meet the statistical assurance requirements.

These revised quartz material test specifications include:

1) An alpha3500 (infrared absorption coefficient) measurement which replaces Q as the main quality assurance criterion. Its measurement method is unchanged. Alpha3500 maximum tolerances replace the minimum Q limits used heretofore. Yet there is no effective change in the grading limits, because the earlier Qs were themselves indicated from alpha3500 measurements. This revised test is expected to continue serving as the quality assurance workhorse it has been for many years.

2) An etch channel density, ρ, measurement standard. This is needed when required to grade material for etch processability.

3) Inclusion density counts in standardized inclusion size ranges. Counting inclusions within specified size ranges should insure that the counts from different test laboratories will be more directly comparable with each other. These counts are expected to be critical only in certain specialized applications.

A fourth test, an analysis for aluminum content, was not specified by the committee. However the user may wish to require a maximum aluminum content, or its analysis, especially if his application involves research or requires high stability under adverse conditions that may include ionizing radiation.

The general subject of quality indications and "Qs" indicated from infrared measurements on cultured quartz has been reviewed and recalibrated in the literature during recent years[1,2]. Round robins have been conducted, first in 5MHz Q measurement, and later by the EIA committee in various infrared alpha measurements[3]. These yielded a data base for obtaining 5MHz Q indications with improved realism from alpha measurements. But to avoid possible ambiguity the EIA committee prefers to use the alpha measurements directly for quartz material test grading.

Sampling a run for the etch channel count test may have added complications because etch channels can be introduced, not only by run conditions, but also by dislocations in the seeds upon which the material is grown. Thus a good batch sample should be representive of not only the run's growth conditions, but also of the seed population that was planted in that run. A run's seed population may or may not be uniform in dislocation density from seed to seed.

Details of the specified test procedures are straight forward, and specific to insure repeatable results. Matters relating to the use of alpha measurements for repeatable quality control are discussed in a paper accepted for publication in September 1987[3]. The test procedures are to be published by the EIA and will be obtainable from them.

Measurements of basic material properties are often grouped into two broad categories: bulk properties that extend throughout the sample and have measurable values at any locality within it, and; lattice defects. These are summarized for cultured quartz in Table I. In the list of test specifications above the first and fourth are of bulk properties since they reflect impurities dissolved throughout the crystal. Specifically, the first, an infrared alpha, reflects the OH (O-bonded hydrogen) content; the fourth another impurity, aluminum, content. Other impurities are often measured but their control is not considered of critical importance to most uses of quartz material as is the

CH2427-3/87/0000-241 $1.00 © 1987 IEEE

AN OUTLINE OF CHARACTERIZATION OF CULTURED QUARTZ
for Quality Testing Use

A) Dimensions and Orientation of Bar, Locations of Seed and of Growth Region Boundaries

B) Macro Defects: Cracks, Growth Flaws and Twins

The A and B categories above, have long been effectively specified, and are unaffected by the revisions here described.

C) Bulk Properties

1) Chemical Purity

a) Infrared Transmittance: $ALPHA_{3500}$ (with $ALPHA_{3410}$ a usable alternate) INDICATIVE OF OH CONTENT.

b) Aluminum Content

c) Other Dissolved Impurities

2) Physical Properties, including their THERMAL coefficients

a) Elastic

b) Piezoelectric

c) Lattice Dimensions

d) Dielectric

e) Optical

The physical properties have been reported to converge within the accepted ranges of natural quartz properties, as its purity increases; i.e. mainly as $alpha_{3500}$ becomes less than 0.06 cm^{-1}.

D) Lattice Imperfections

1) Etch Channel Density

2) Inclusion Density, by Size Range

Table 1

A Classified Summary of Measurable Properties of Cultured Quartz. To date the main quality assurance emphasis has been on chemical purity as manifested through infrared, and sometimes aluminum content measurements.

control of the bonded hydrogen and aluminum.

The other two tests above are for: (2), those defects involving dislocations that can etch and be made visible for etch channel counting, and (3), the bits of foreign material, inclusions, to be counted according to size range. Whenever to our knowledge to date, the inclusion material has been analyzed it has been found to be a sodium iron silicate mineral, although there is no reason in principle that other materials can not also be included as indeed many are in naturally occuring quartz crystals.

When specifying cultured quartz material for specific applications most users can expect to continue to place their main reliance for quality assurance on alpha3500 through restated requirements on its maximum which are set to be equivalent to their earlier Q minimum requirements for that use. This also will continue to place effective limits on the OH content, and so insure that the physical properties, especially the frequency-temperature "angle behavior", will remain within the ranges of their normal experience. Other tests in Table 1 under the headings C1, Chemical Purity, and D, Lattice Imperfections, should be used when indicated by the application, or when a complete characterization is appropriate, as for material that will undergo experimental studies.

REFERENCES

1. J.C. Brice and A. M. Cole, "The Characterization of Synthetic Quartz by Using Infrared Absorption," Proceedings of the "32nd Annual Frequency Control Symposium" US Army Electronics Command, June 1978, 1-10.

2. B. Sawyer, "Recalibration of Q Capability Indications from Infrared Measurements on Cultured Quartz," Proceedings of the 37th Annual Frequency Control Symposium, June 1983, 151-152.

3. B. Sawyer, "Quality Indications from Infrared Absorption Measurements on Cultured Quartz," IEEE Trans on Ultrasonics, Ferroelectrics and Frequency Control, accepted for publication in September 1987.

41st Annual Frequency Control Symposium - 1987

COMPUTING CRYSTAL ORIENTATION FROM X-RAY MEASUREMENTS

John H. Sherman, Jr.
Consultant
2022 Wooderest Drive
Lynchburg, VA 24503

Abstract

The equations used to compute the orientation of doubly rotated quartz plates from Bragg X-ray data have been derived by vector methods. The solution of a two measurement data ensemble taught by Bond has been found by a convergent iteration. This paper contains an alternate derivation of the equations from considerations of solid geometry and the unit sphere, and a reduction of the solution to the solution of a quadratic equation with numerical coefficients. The derivation by vector methods of two and three measurement ensembles and their previously published methods of solution are reviewed briefly.

Introduction

X-ray measurement problems of AT-cut resonator manufacture have traditionally been solved by breaking the problem into manageable bits. Stones have been oriented one lattice plane at a time in preparation for cutting in "coring" or "lumbering" operations to generate flat surfaces containing chosen atomic planes. The tolerances of second and third angles in cut wafers thus translate into tolerances of coincidence of the surface and the chosen plane. The cutting of the θ angle about the unrotated X-axis is then made in a single oriented operation. This approach is equally applicable to that V-cut called IT and to another with a first rotation of 13.9 degrees, but not to those called FC- and ST-, in which no coincidence of a lumbered surface with a lattice plane occurs. In measuring a blank, lumbering errors, which result in non-zero second and third angles, do not get measured directly, but are assumed to be under control and of tolerable consequence by virtue of their independent control in the lumbering operation. Consequences of cutting errors have been studied [1],[2],[3] and their first order effects published in various places. These studies constitute the scientific basis of lumbering tolerances.

Another way to cut blanks long used in making AT-cuts is known as "bologna slicing", in which elaborate and ingenious procedures are used to present the stone to the saw to cut a properly oriented blank directly. When all quartz was natural quartz and all saw blades thick this method had vigorous defenders, as the quartz lost in making standardized blocks from uneven sized natural stones was expensive, as was the substantial labor demanded. Bologna sliced blanks were uneven in size and shape. They were trimmed after slicing to cut blanks of usable size out of the pieces. This was usually possible, even when the slices contained twinned regions. This method has never been a factor in cutting AT-cut blanks from cultured quartz, as the cultured stones are grown in shapes which imply efficient lumbering, and efficient gang-sawing compared to sawing with the thick blades used in bologna slicing.

Control of cutting errors traditionally acceptable for AT-cuts can hardly be tolerated in manufacturing precision SC-cut resonators, which are becoming an increasingly significant item in the repertoire of the crystal manufacturer. Cutting tolerances for SC-cuts have to be much tighter than they ever were for AT-cuts. This is not to imply that the methods of preparing blanks for SC-cut resonators are significantly different from the methods used in preparing AT-cuts. The methods are the same, even down to one maker applying a refined kind of bologna slicing [4] to the cutting of SC-cuts. The effect of cutting errors is simply more severe in SC-cut resonators, at least in high precision SC-cuts, than in AT-cuts. One needs to know the orientation in detail before fabricating the resonator so that its errors can be corrected or the blank removed from the line before the full cost of manufacture is put into a unit that could be recognized from the beginning to be a reject.

Measurement

In starting to discuss measurement, let us get one thing out of the way. What I want to be discussing is measurement by Bragg X-ray diffraction methods. There is another class of X-ray diffraction usable to determine the orientation of a crystal plate, known as Laue diffraction. Bragg diffraction is diffraction of a monochromatic beam of X-rays, while Laue diffraction is diffraction of a continuous spectrum. Laue diffraction occurs because there are distinct and unique path directions through a crystal along which only a single wavelength of X-rays can pass. The explanation of these paths is the same as the explanation of the diffraction. Each path selects a wavelength which can traverse it. Along all other directions the X-rays are either absorbed or scattered incoherently. When that crystal is illuminated by a narrow beam of continuous spectrum X-rays there will result a characteristic transmission or reflection of X-rays such that beams will leave the crystal in only certain directions, each beam being of a single wavelength. A photographic film mounted in a plane perpendicular to the incident beam will intercept transmitted or reflected beams forming a pattern of spots displaying a symmetry which depends upon the symmetry of the crystal. The spacing of the spots in the pattern depends upon the spacing of the lattice planes and the distance between the film and the crystal. When the spots are at their brightest and the symmetry of the pattern is perfect, the incident beam is perpendicular to a lattice plane. This method of X-ray examination is being used, to my best knowledge and belief, with exemplary success, by a single maker of crystals [5],[6] in the western world. This, incidentally, is the same maker who has successfully adapted bologna slicing to cutting SC-cut blanks.

CH2427-3/87/0000-243 $1.00 C 1987 IEEE

What is known as Bragg diffraction follows from the mechanism by which Bragg explained Laue diffraction. When it was clear that the pattern of Laue spots was due to monochromatic paths through the crystal it became clear that a beam originally monochromatic would be selectively diffracted in only particular directions depending upon the wavelength and the lattice plane spacing of the crystal. Machines which depend upon Bragg diffraction use characteristic X-rays of a metal, in our case of copper, as a functionally monochromatic source, though it is sometimes necessary to recall that the beam is actually a doublet of very nearly but not exactly a single wavelength. A Bragg diffractometer is a machine which measures the angle between the surface of the plate and the lattice plane for which the machine was set up. This measurement is really a protractor measurement between two positions of a holding fixture, as it holds a plate having its surface parallel to the lattice plane and as it holds the unknown plate being measured. This protractor is capable of resolving a fraction of a minute of angle.

Various means to hold and position the plate under measurement [7],[8],[9] are described in the literature. Since this paper is not concerned with making the measurement, but with the use of the measurements to calculate the orientation of a blank oblique to several planes and simply related to none, there will be no detailed discussion of the techniques of measurement themselves.

Of Dimensions and Coördinates

The measurements made by a Bragg spectrometer are of the angles between the plane surface of a crystalline plate and a lattice plane of the crystalline substance. In the measurement of a totally unknown plate the angles between the plate and three lattice planes are found. This information is subject to several different descriptions, all of which must be ultimately equivalent. Discussing several descriptions may lead to easier visualization and understanding.

A plane surface of a crystal plate is clearly a single surface, immediately apprehended. To speak of a lattice plane does not, however, imply that we are trying to discuss a particular plane within the crystal in the same sense as we discuss the surface, but to characterize by their parallelism all planes parallel to the plane named.

We really consider a geometric plane, not a true lattice plane at all, but a plane parallel to real lattice planes, a plane tangent to a sphere of unit radius about an arbitrary origin. This geometric plane is perpendicular to the radius of the unit sphere at the point of tangency and the coördinates of the point of tangency are the cosines of the three angles between that radius and the three Cartesian axes. All planes perpendicular to this radius are parallel, and are conventionally considered the same plane. For Bragg diffraction to occur, a multiplicity of many thousands of real lattice planes are involved.

Three dimensional space is described by various coördinate systems. In this paper we use a blend of two, rectangular Cartesian coördinates and spherical coördinates. These are paired in the standard way of designating crystal orientation in the designation of X-, Y-, and Z- axes of quartz and angular rotations of ϕ and θ degrees around Z- and X- in a sphere of unit radius.

In three dimensional space a plane can be designated by its intercepts on the three Cartesian axes. These three sets of data values contain all the information to define the plane uniquely and completely. The plane is equally defined by the coördinates of any one point in the plane and the three angles formed between the Cartesian axes and the radius from the origin perpendicular to the plane. Usually the angles themselves are not stated, but their cosines. The nicest point to know is the foot of the perpendicular on the plane. For this point the coördinates are proportional to the cosines of the three angles. The sum of the squares of the three cosines is identically unity.

The regular structure of a crystal implies directions which are obvious candidates axes in a coördinate system to describe the crystal. The natural increments along these axes are commonly not of the same size, though Cartesian coördinates are characterized by uniform steps along all three axes. If unit steps along each axis in a crystal be taken as the spacing of the planes in that direction and the unit distances be enclosed by the smooth surface most resembling a sphere, that surface would have to be some kind of ellipsoid. Only in a very few cases of the very simplest crystals would the surface be a sphere. In order to discuss a crystal in Cartesian coördinates it is usually necessary to account for this. Thus in quartz, if the 4.9029 Å lattice spacing along an X-axis, called "a_0" (there are three axes 120° apart, each equally validly considered an X-axis) is defined to be 1 unit, then the lattice spacing along a Y-axis, perpendicular to the X-axis, is $\sqrt{3}/2$ units. The spacing of the planes along the Z-axis, called "c_0", perpendicular to both X- and Y- has to be measured, and is found to be 1.10009 units.

Intercepts of lattice planes with these axes are all integer or rational fraction values of these natural units of spacing. The 120° separation between the three X-axes of quartz assures that any plane which intercepts a Y-axis will also intercept a + end and a - end of the two X-axes which lie only 30° away from that Y-axis. A Y-axis is not even a necessary thing in quartz, given the three X-axes.

The 120° between the X-axes has a number of interesting consequences which are simply properties of a hexagon. A lattice plane not perpendicular to Z must intersect one or another X-axis at an integral value of the natural unit. It will then necessarily intersect the other two X-axes at points having values of integers, rational fractions or infinity. Furthermore, the sum of the reciprocals of the three intercepts on the X-axes will be identically zero. The reciprocals of the intercepts in lattice units of that particular plane of a family of parallel planes which lies closest to, but not containing, the origin constitute an assemblage of four digits called "Miller-Bravais" indices for the plane. The cluster has the form $hki\bar{l}$. The first three indices are the reciprocals of the intercepts on the three X-axes. Since the sum of these three (two positive and one negative or vice versa) is necessarily zero it is conventional to omit the value of the third one, substituting for it a point. The minus sign of a

negative index, if the index is written at all, is written over its number in the cluster rather than before it, making for a neat and compact array.

These indices are used to code the intercepts of that particular representative plane. From them and the actual values of the natural crystal dimensions are computed the location of the plane and properties of its family in Cartesian space.

In the crystallography of quartz the natural unit along the X-axis, the X- dimension of the unit cell, is called "a_0" while the Z- dimension is called "c_0". They are subject to measurement. The best values of these [11] at 18° C are:

$$a_0 = 4.90290 \pm 0.00003 \text{ Å}$$

$$c_0 = 5.39365 \pm 0.00003 \text{ Å}$$

and their ratio:

$$c_0/a_0 = 1.10009 \pm 0.00001$$

In order to make the conversion of lattice dimensions to Cartesian coördinates it is convenient to define a scale factor, S, based on the geometry of the conversion, as follows:

$$S^2 = 4(h^2 + hk + k^2) + \ell^2/(c/a)^2$$

Then the spacing of the lattice planes hk.ℓ is given by:

$$d = a_0/S$$

and the direction cosines of the normal to this plane, given by p1, p2, p3 respectively, are:

$$p1 = h/S \qquad (1a)$$

$$p2 = (h + 2k)/\sqrt{3}S \qquad (1b)$$

$$p3 = c_0\ell/a_0S \qquad (1c)$$

The direction cosines of the normal to a plate P having rotation angles Φ and Θ are:

$$p1 = -\sin\Phi\cos\Theta \qquad (2a)$$

$$p2 = \cos\Phi\cos\Theta \qquad (2b)$$

$$p3 = \sin\Theta \qquad (2c)$$

The Nature of the Data

The X-ray goniometer measures a scale reading for the location of the surface of the unknown crystal plate. The calibration of a conventional machine is made using the reading of a plate having the lattice plane used for the reflection lying in the surface of the plate. If the plane actually lies in the surface then the plate can be rotated in its own plane into any position whatever and the X-rays will be diffracted into the detector with no further adjustment of the machine required. If the plane and the surface are not separated from each other by an excessive angle, the location of the plane is found at the average of the two azimuth settings of the plate holder of maximum separation which result in diffraction of the beam into the detector. These occur as the plate is rotated in its own plane at points 180° apart. Since the difference between two azimuths can be found independent of the accuracy or constancy of a direct scale reading, the best choice of plane is one which allows a good strong diffracted beam to be directed into the detector from two positions of the plate 180° apart.

Bond [7] discussed measurement of SC-cut plates as an aspect of cutting them. He had occasion to consider diffraction from two different planes with the same Bragg angle. In order to mount the plate properly in the X-ray machine he generated in the lumbering operation an oblique edge of the plate of appropriate orientation to allow the placement of the plate in the machine, thus to use a single detector to measure the angles between the plate and the two different planes with the same Bragg angle. In the absence of a reference edge, as occurs in measuring a round blank, a rotating chuck [8][10] is indicated. This can, of course, be used also to hold a blank with straight edges. There will be four positions which result in detection of the beam.

A single datum, then, is an angle reading which either equals the angle between the surface of the plate and the diffracting lattice plane or equals just twice that angle. The designer of the equipment makes a conscious decision which kind of data is to be obtained from that machine. Thus there were simultaneously being used in the early years of the American crystal industry, crystal goniometers made by General Electric which measured AT-cut plates by detecting a single diffraction and converted the scale reading thus obtained to an angle by a calibration, and goniometers made by Philips which detected two diffraction peaks approximately six degrees apart from a plate placed sequentially in two positions and determined the plate angle as half the position difference below the known angle of the lattice plane.

In general, an unknown may be oblique to all lattice planes in the crystal. In the case of AT-cut plates the X-axis nominally lies in the surface of the plate. Only one angle is measured. In the case of SC-cut plates no axis lies in the surface of the plate so the angle between the surface and more than one plane must be measured to locate the surface in space. To measure a known doubly rotated cut, measurements against at least two planes must be made. To measure a complete unknown, at least three are required.

Interpreting the Data Geometrically

All that is known from a datum is the angle formed by a chosen lattice plane with the surface of the plate and the orientation of the lattice plane. The plane and the surface intersect in a straight line the orientation in space of which is enough, if known, to allow determination of the orientation of the surface. Unfortunately, the X-ray goniometer does not supply any information about the orientation of that line. All we know is that the surface is tangent somewhere to a cone which makes the measured angle with the lattice plane and that the line of tangency is perpendicular to the intersection, as shown in Figure 1.

The most picturesque and most fruitful way to view the content of this datum is to consider the normal to the lattice plane as a line projecting from the origin out into space. The direction cosines of this normal are computed from the

formulas (1) above. Also projecting from the origin, forming the measured angle from the first line, is found the normal to the unknown surface. Since its actual direction is as yet unknown, all possible directions must be considered to be implied by the datum. Thus the normal to the surface lies on a circular cone about the normal to the lattice plane, with its apex at the origin and with the measured angle as the generating apex angle, also as shown in Figure 1. This normal and the lines of tangency and intersection comprise a natural coördinate system for the surface.

A measurement made against another plane will define another cone having a different axis and a different apex angle. Since the normal to the surface of the plate undoubtedly is in the same relation to the surface at all times, even when the angle is being measured, then the normal is on both cones, and the two cones undoubtedly intersect each other. In general there are two lines of intersection of two cones having a common apex, just as there are two points of intersection of two circles in a plane. There is, therefore a certain amount of uncertainty as to which of the two intersections is actually the normal, an uncertainty which must be handled by other considerations. The simplest consideration is that a particular cut was desired to be made, so the intended normal is probably the appropriate intersection.

We surround the pair of cones by a sphere of unit radius centered on the apices, as illustrated in Figure 2. The intersection of this sphere with the two cones is made the focus of our attention to compute the intersections of the cones, thus to find the plate normal and the orientation of the plate.

For convenience we have called one lattice plane plane A and the other plane B. The unknown surface is called plane X. The portions of the normals contained within the unit sphere are called the cone axes A and B, and we mention radii A, B, and X. The direction cosines of the normal to plane A are a1, a2, a3. The cone is cone A and the apex angle is angle A. Coördinate values are called x, y, z. No confusion will be found to exist from any of this.

Each cone intersects the sphere in a circle smaller than a great circle. The center of each circle is one of the plate normals. The centers of the circles are joined by the great circle AB. The circles intersect at two points, X1 and X2 which are on great circle X1X2 which is perpendicular to great circle AB. One of these intersections is the true normal to the surface X.

Figure 3 displays the significant arc of great circle AB. Along this great circle are the overlapping circular arcs which are the surface diameters of the circles A and B, of length 2A and 2B. The ends of these arcs are ends of chords which are plane diameters of the same circles, of length 2sinA and 2sinB. The midpoints of the chords are cosA and cosB distant from the origin. The chords intersect at the midpoint of straight line X1X2. The figures of rotation of these two arcs and their chords about their own axes are the two sectors of the spherical surface subtended by the cones, the planes of the circles of the intersection with the sphere, and the circles.

A plane tangent to the unit sphere is perpendicular to the radius to that point. The tangent plane at the end of axis A is, therefore, parallel to the lattice plane A. The plane formed by the rotation of the chord is perpendicular to axis A and is therefore also parallel to lattice plane A. Both are equivalent, if desired, to lattice plane A

Planes and Lines in Space

A point in a plane is located by a set of two numbers, in space by a set of three. An expression of the form Ax + By + C = 0 is the equation for a straight line in a plane. An expression of the form Ax + By + Cz + D = 0 is not the equation for a straight line in space, however, but the equation for a plane. An equation for a straight line in space cannot be written in a single closed expression with one equals sign. A line can be defined by two number triples locating two points on the line. The radii which coincide with the lattice plane normals we have been discussing are defined thus by the two number triples, the origin, 0,0,0, and their intersections with the unit sphere, the points a1,a2,a3 and b1,b2,b3.

Two planes intersect in a line. This line is identified by the two equations in three variables taken together. By elimination between the two it is possible to convert the two equations in three variables into an equivalent three equations in two variables. Any given line may be considered the intersection of any two of an infinite number of non-parallel planes.

The plane tangent to the unit sphere at the end of axis A is given by the equation:

$$a1x + a2y + a3z = 1 \qquad (3a)$$

and the tangent at the end of axis B is:

$$b1x + b2y + b3z = 1 \qquad (3b)$$

The plane containing the circle A has the equation:

$$a1x + a2y + a3z = \cos A \qquad (4a)$$

and the plane containing the circle B:

$$b1x + b2y + b3z = \cos B \qquad (4b)$$

Every point on the tangent plane is distant from the origin by an amount equal to or greater than 1 unit of length. On the plane containing a circle of intersection of a cone with the sphere, every point inside the circle is nearer to the origin than 1 unit, every point outside the circle is farther from the origin than 1 unit, and every point on the circle is just 1 unit from the origin.

The line of intersection of the two planes containing circles is written by writing the two equations (4) together. This line contains all the common points of the two planes, including the two points which are the intersections of the circles (which circles, we remember, are circles both of their planes and of the sphere). These points are the sphere ends of the two radii candidates to be the normal to the unknown plate.

If we divide equation (4a) by a2 and equation (4b) by b2 we eliminate y and find an equation for x in terms of z. Similarly eliminating z yields an equation for x in terms of y. Since things equal to the same thing are equal to each other, we have a third relating y and z. These three equations in two variables were referred to earlier. Written out in detail they are:

$$\frac{x}{(a2b3-a3b2)} = \frac{y-(b3\cos A-a3\cos B)}{(a3b1-a1b3)} \qquad (5)$$

$$= \frac{z+(b2\cos A-a2\cos B)}{(a1b2-a2b1)}$$

The expressions within parentheses have the form of determinants of two rows and two columns. They are, of course, simply numbers. It will be convenient to use single characters for these 5 determinants:

$$D1 = \begin{vmatrix} a1 & a2 \\ b1 & b2 \end{vmatrix}, \quad D2 = \begin{vmatrix} a2 & a3 \\ b2 & b3 \end{vmatrix}, \quad D3 = \begin{vmatrix} a3 & a1 \\ b3 & b1 \end{vmatrix}$$

$$D4 = \begin{vmatrix} \cos A & a2 \\ \cos B & b2 \end{vmatrix}, \quad D5 = \begin{vmatrix} \cos A & a3 \\ \cos B & b3 \end{vmatrix} \qquad (6)$$

Solving for the Orientation

Having equations for the line containing the intersections of the two circles, we proceed to find the coordinates of these intersections. The unknown radii meet the unit sphere at x,y,z. Direction cosines to that point are x, y and z. The sum of the squares of the direction cosines and also of the coördinates is 1.

Equations (5) are equivalent to:

$$y = \frac{D3}{D2}x + D5 \qquad (7a)$$

and

$$z = \frac{D1}{D2}x - D4 \qquad (7b)$$

which are substituted into:

$$x^2 + y^2 + z^2 = 1 \qquad (8)$$

to yield:

$$(D1^2+D2^2+D3^2)x^2-2D2(D1D4-D3D5)x+D2^2(D4^2+D5^2-1)=0 \qquad (9)$$

This is a quadratic equation in x having two solutions readily found by the quadratic formula. The solutions are substituted into equations (7) to find the corresponding coördinate values of y and z. The set appropriate to the intended orientation is converted, in turn, into the rotation angles of the plate surface using the formulas of equations (2).

Vector Methods

A completely different line of reasoning for using the goniometer data comes out of simple vector theory. The radii are treated as radius vectors. The dot product of the unknown with each cone axis vector is taken to generate the equations to be solved, resulting in:

$$A'X = a1x1 + a2x2 + a3x3 = \cos A \qquad (10a)$$

and

$$B'X = b1x1 + b2x2 + b3x3 = \cos B \qquad (10b)$$

These equations relate direction cosines. They look very similar to equations (4) which relate coordinate values. When combined with (8) the same problem of solution exists. The vector derivation is described by Bond and Kusters [11] who offer an alternate approach to solving the set of equations. Their method proceeds by assuming a solution, which is surely the set of direction cosines of the normal to the intended orientation. This set of values is substituted into the set of equations, resulting in a calculation of errors. A second and improved approximation is found from these errors. After several cycles of computation the unknown cosines can be computed to any desired number of significant figures. This kind of repetitive convergent computation is performed by the commercial microcomputer program TKISolver™ [12] and similar general mathematical programs.

If the plate being measured is completely unknown there is no basis in measurement by two diffracting planes to determine which of the two solutions is correct. This ambiguity is readily removed by making a third measurement against a third lattice plane, plane C. When this is done a third equation results:

$$C'X = c1x1 + c2x2 + c3x3 = \cos C \qquad (10c)$$

Strictly formally the equations for the two and three measurement complexes can be written as follows:

$$\begin{Bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \end{Bmatrix} \times \begin{Bmatrix} x1 \\ x2 \\ x3 \end{Bmatrix} = \begin{Bmatrix} \cos A \\ \cos B \end{Bmatrix} \quad \begin{Bmatrix} a1 & a2 & a3 \\ b1 & b2 & b3 \\ c1 & c2 & c3 \end{Bmatrix} \times \begin{Bmatrix} x1 \\ x2 \\ x3 \end{Bmatrix} = \begin{Bmatrix} \cos A \\ \cos B \\ \cos C \end{Bmatrix}$$

which appear deceptively symmetrical. These are the vector equations but the equations from the planes in space are similarly written, simply by writing x,y,z where we have written x1,x2,x3.

The set of three linear equations yields an unequivocal single solution for the direction of the normal. Geometrically the introduction of the third plane introduced a third cone C having the third normal C as axis, and a third circle. Three planes can intersect in three parallel lines, but if the lines of intersection are not parallel then they must meet in a single point. The physical circumstance clearly requires that the planes of the three circles must so meet. The intersection of the planes of the three circles is the end of the radius vector X, the normal to the surface of the unknown.

The quality of the measurements can be evaluated by calculating the sum of the squares of the values obtained for x1, x2 and x3. The sum should be identically unity. If the sum is not sufficiently close to unity the readings may be recalculated in pairs by either method to determine if one of the angle measurements is clearly associated with discrepant results.

Conclusions

Determination of the orientation of doubly rotated crystal blanks entails measurement against more than one lattice plane and a computation from a moderately complex configuration in three dimensional space. Two methods exist to perform the computation from measurements against two planes, one a direct solution of a quadratic equation, the other a process of convergent iteration. These methods both yield a spurious solution as well as the one sought. Measurement against three lattice planes yields a set of three linear equations the solution of which is an unambiguous result.

References

[1] C. Barclay and L. T. Sogn, "Reference Data for Orienting Quartz Plates by X-ray Diffraction", N. B. S. Circular 543, 1953

[2] R. Bennett (ed), "Quartz Resonator Handbook", PB-171839, 1960, Figs. 70, 71

[3] A. Ballato and G. J. Iafrate, "The Angular Dependence of Piezoelectric Plate Frequencies and their Temperature Coefficients", Proc 30th AFCS, 1976, pp141-156, esp. Fig. 2

[4] A. W. Warner, J. Tsaclas and J. Korman, "Cut and Grind", Proc 39th AFCS, 1985,pp 342-344

[5] J. L. Chambers, "An Instrument for Automated Measurement of the Angles of Cut of Doubly Rotated Quartz Crystals", Proc 37th AFCS, 1983, pp 275-283

[6] A. W. Warner personal communication, 1986

[7] W. L. Bond and J. A. Kusters, "Making Doubly Rotated Crystals", Proc 31st AFCS, 1979, pp 153-158

[8] J. F. Darces and H. Merigoux, "Final X-Ray Control of the Orientation of Round or Rectangular Quartz Slides for Industrial Purposes", Proc 32nd AFCS, 1978, pp 304-309

[9] J. Clastre, C. Pegeot and P. Y. Leroy, "Conometric Measurements of the Angles of Cut of Doubly Rotated Quartz Plates", Proc 32nd AFCS, 1978, pp 310-316

[10] J. H. Sherman, Jr., "An X-ray Machine for Generally Rotated Blanks", Proc 8th Quartz Device Conf. 1986, pp 208-220

[11] C. Frondel, "Dana's System of Mineralogy", Volume III, John Wiley & Sons, New York 1962

[12] TKISOLVER is a registered trademark of Software Arts, Inc.

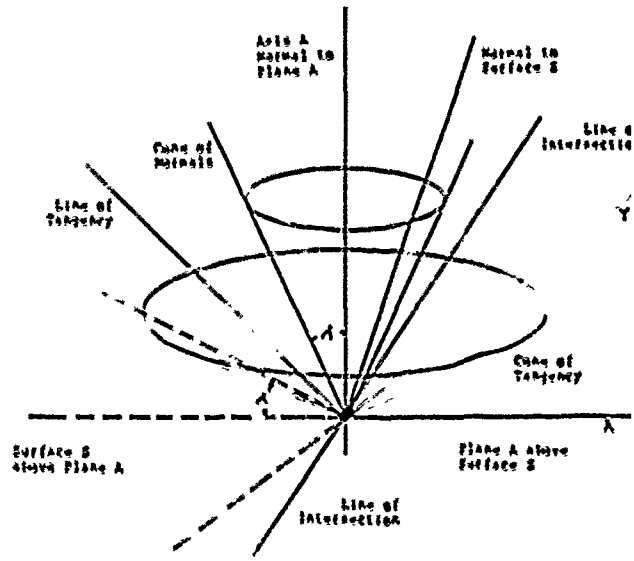


Figure 1. Intersection of surface S with plane A at the angle A and the two cones thus implied.

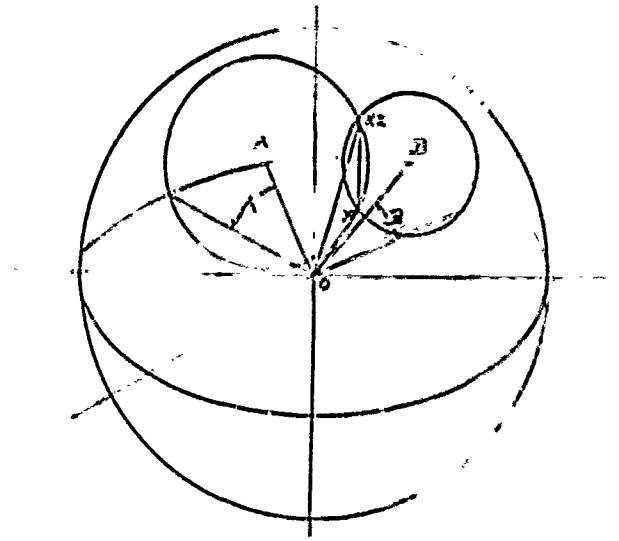


Figure 2. Cones A and B intersecting within and on the surface of the unit sphere.

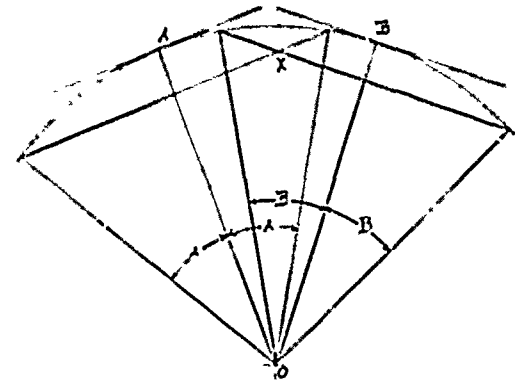


Figure 3. Intersection of great circle AB with cones A and B and their associated planes.

X-RAY TECHNOLOGY - A REVIEW

Charles A. Adams,* D. Canon Bradley,* and John A. Kusters*

*Hewlett Packard Company, Santa Clara, California 95051
*Colorado Crystal Company, Loveland, Colorado 80537
*Ball Corporation, Efratom Division, Irvine, California 92718

ABSTRACT

This paper is a tutorial overview of general x-ray technology as applied in the manufacturing of quartz crystals. Topics covered include general x-ray principles including the Bragg condition and refraction, single axis x-ray systems including choices of collimating crystals, and multiple axis systems where two or more x-ray beams are used either to simultaneously measure a quartz blank's current crystallographic orientation, or to angle correct a surface of a crystal blank to a specific orientation. Applications and illustrations are drawn from the authors' experience in all three companies involved.

INTRODUCTION

In their employment in several different manufacturing facilities of quartz resonators over the past two decades, the authors have experienced a wide range of different approaches, ideas, and techniques in the precision measurement of the crystallographic orientation of a quartz crystal blank.

We have also seen many applications of x-ray technology which were based on erroneous assumptions and which led to, in our opinion, unnecessary complications in manufacturing processes.

As a result of the above, much has been developed and is still under development in our respective companies to improve the state-of-the-art in x-ray orientation. Virtually none of this has been presented in past symposia.

I. GENERAL PRINCIPLES

The Bragg Condition

Under proper conditions, a beam of x-rays incident on the surface of a crystal may be scattered coherently. In this case we usually call the coherent scattering x-ray reflection even though, strictly speaking, the coherent scattering occurs only at discrete, specific angles to the surface. The specific angles are a function of the x-ray wavelength and lattice spacing and are defined by the Bragg relation [1], which is:

$$n\lambda = 2d \sin\theta \qquad (1)$$

where n = 1, 2, 3, ... and is the order of reflection
λ = x-ray wavelength
d = lattice separation in same units as λ
and θ = the Bragg angle

Figure 1 shows an orthographic projection of the primitive region in quartz. Also shown are dots that represent normals to x-ray diffraction planes. The planes shown are those which are allowable using copper Kα radiation with a wavelength of 1.54Å.

In practice, the allowable diffraction planes are governed by the choice of x-ray target material. Table 1 gives a relationship between material chosen, the minimum lattice spacing that can be resolved, and the equivalent order of reflections possible in quartz. As can be seen, copper is a reasonable choice in that a wide variety of planes is available for use with order of reflections up to about 5 (00·6 for the Z-plane). The longer wavelength that copper provides is also a safety factor in that it is not as biologically damaging as tungsten x-ray radiation.

X-RAY TARGET MATERIAL

Material	K-alpha wavelength	d-min	Equiv n
Aluminum	8.33934	4.1697	1
Copper	1.540562	0.7697	5
Silver	0.55941	0.2797	15
Tungsten	0.20901000	0.1045	40

Wavelength and spacing in angstrom units

Table 1. Minimum lattice spacing and reflection order observable as a function of target material

X-Ray Line Shapes

A series of calculations which takes into account all of the known extinction, absorption, and refraction effects of quartz as a function of the atomic diffraction plane predicts that the actual peak of reflection energy occurs not at the Bragg angle as predicted in Eq. 1, but at an angle slightly less than the Bragg angle. The primary cause for this is refraction. The change in the index of refraction for Kα radiation in α-quartz compared to air is of the order of 1 part in 10e5. The Bragg angle calculation applies inside the quartz while angles are measured outside of the quartz blank. Figures 2 through 6 show computed lineshape curves for several of the more common planes in α-quartz. [2]

In this series of calculations, the relative amplitude, which is simply the computed area under the lineshape relative to the 01·1 plane (Fig. 2) whose intensity is defined as 1.00, is also given. Of interest is that the relative amplitude of the 01·$\bar{1}$ plane (Fig. 3) is computed to be 0.65, vs. 0.75 as given in Heising [1], and the 02·2 (Fig. 4) is computed to be 0.26 vs. 0.26 in Heising. This is remarkably good agreement given the difficulties of measuring relative x-ray intensity.

CH2427-3/87/0000-249 $1.00 C 1987 IEEE

The lineshape data is interesting theoretically but is of practical interest primarily in determining the intensity of possible x-ray reflection intensities.

The offset from the computed Bragg angle (Eq. 1) as shown on these figures is not of much practical interest, as it is usually masked by variations in the atomic lattice constants which can vary several to many parts per million from growth batch to growth batch, or even from piece to piece, especially in natural material. It however becomes significant if one is attempting to obtain extremely precise x-ray measurements or alignment.

Kα Doublet

In most practical applications of x-ray reflection from the quartz diffraction planes the choice of the Kα peak poses a problem in that this peak is a doublet as shown in Figure 7. (3)

When this doublet is used to reflect energy off of an atomic plane as shown in Figure 8, the detected x-ray energy shows a double peaked response corresponding to the different Bragg angles for the two x-ray lines. The angular separation for the two peaks is given by: (4)

$$d\theta = \tan \theta \, d\lambda/\lambda \qquad (2)$$

Thus, given a wavelength difference of 0.0038Å, the angular separation after one reflection is from 2 to 10 minutes of arc given a practical range of from 10° to 50° for the Bragg angle.

This points to two methods of eliminating the difficulties associated with the Kα doublet. Either make the angular separation so small that it cannot cause any confusion, or, make it so large that it is easy to distinguish one peak from another. Both alternatives will be discussed further in later sections.

II. SINGLE BEAM X-RAY MACHINES

Single Crystal X-Ray Diffraction System

This is the simplest form of x-ray system, consisting of a power supply, x-ray tube, collimating slits, a nickel filter, a holder for the crystal being measured, a goniometer head for measuring the reflection angles, and a suitable x-ray detector. Such a system is shown in Figure 9. (5)

Accuracy of this type of system is governed by the allowable slit widths, the x-ray intensity available, the path length from the source to the crystal blank, and the resolution of the goniometer. With care, measurements to about 5 minutes of arc are possible, but rarely achieved in practice. The Kα doublet is usually not resolvable in such a system.

Dual Crystal X-Ray Diffraction Systems

In this system, (6) an additional crystal, usually called the reference or collimating crystal, is mounted such that it provides a highly directional, monochromatic x-ray beam. Such a system is shown in Figure 10. The radiation from the x-ray tube is directed at the surface of the collimating crystal. Only that energy which satisfies the Bragg condition will be reflected from the crystal and directed towards the blank to be measured. The collimating crystal is usually mounted to reflect the copper Kα, x-ray line. This configuration has become the accepted practice in the quartz industry.

In this manner, no collimating slits or filters are needed. Although energy is lost at the collimating crystal, substantial gain is made in resolution. With a proper goniometer head for measuring angles, accuracies approaching 5 seconds of arc are possible. While further resolution and accuracy is available in measuring heads, the natural linewidth as shown above, and uncertainty in lattice constants place a limit on the actual orientation accuracy that can be achieved.

Choice of Collimating Crystals - With accuracies approaching 5 arc-seconds and resolution approaching 1 arc-second, the Kα doublet is easily resolved. Therefore, the choice of a collimating crystal becomes important.

The effect of the doublet can be eliminated by making the angular separation so small between the two peaks of the doublet that the x-ray system cannot resolve the peaks. If the collimating crystal plane is exactly the same as in the crystal being measured, then the situation as shown in Figure 11 occurs. When the corresponding atomic planes in the collimating and measured crystal are exactly parallel, reflections at different wavelengths occur simultaneously and no separation of the doublet is apparent. (4)

Another possible method is to make the separation large enough so that the resolution of the system can easily identify the difference between the two peaks. As shown previously, choosing an atomic plane for the collimating crystal that has a Bragg angle that is large when compared with the measured crystal plane Bragg angle will result in angular separation approaching 10 arc-minutes. As the two peaks also differ significantly in amplitude, identification of the Kα, peak is usually quite simple.

A third method is to simply ignore the doublet and train the x-ray operators in methods that eliminate any such problem. Such an approach is used in at least three crystal facilities.

The third method is also of some practical importance as the new doubly-rotated cuts, such as the SC-cut, pose a special problem. A single measurement using one atomic plane is no longer sufficient. At least two measurements using two atomic planes are required. Routine measurements which require matching atomic planes as in the first method above, require either two x-ray systems, or a change of the collimating crystal. Given further the difficulties of mounting the crystal to be measured accurately and repeatedly, and the additional alignment problems with a change of collimating crystal, methods which are developed to measure a crystal blank orientation without worrying about the Kα doublet lead to significant productivity improvements.

Rocking Curves - With advances in automation and computer systems, it becomes possible to further enhance x-ray resolution and accuracy by providing a controlled angular rotation, or "rocking" motion of the crystal being measured. Diagrams of possible systems for the AT-cut and SC-cut are shown in Figures 12 and 13 respectively. For the system shown in Figure 13, actual rocking curves for the two orthogonal positions are shown in Figures 14 and 15.

The virtue of these systems is that the x-ray intensity data and position data of the crystal being measured can be taken simultaneously and the peak of the intensity curve calculated through advanced algorithmic techniques to provide a more accurate measurement of the reflection angles than that which is possible for a human operator.

Psi Error - A more serious area for consideration is the possible error that arises because of blank misalignment during measurement.

In every case of practical interest in the making of quartz resonators, the atomic planes used in x-ray reflection do not correspond with the surface being measured. Thus, as the blank being measured is rotated about its surface normal, the actual goniometer angle measured varies as shown in Figure 16. The angle δ shown is the angle between the surface normal and the normal to the atomic plane used. Thus the goniometer traces a curve about $\theta = \delta$.

The angle between the atomic plane used and the surface of several standard crystal cuts is shown in Table 2. As can be seen, the angle varies from about 3° to about 16°. In practice, angles up to about 25° can be used.

CUT	*PLANE*	*ANGLE (Deg)*
AT	01.1	2.95
	03.2	7.57
	12.2	16.62
BT	10.1	6.78
	20.3	4.73
	21.3	14.22
SC	12.2	4.12
	12.3	7.97
	11.2	10.33
	11.1	11.95

Table 2. Angle between the atomic plane indicated and the surface of common quartz crystal cuts

The optimum points for measuring a crystal are at the extreme points of Figure 16, i.e., at its minimum or its maximum. Practical considerations, especially when using a collimating crystal plane which is not the same as the plane being measured, indicates that the minimum of the curve is preferred.

In virtually every single beam system, the blank being measured is oriented through the use of one or more reference edges on the blank. Figure 17 indicates that to obtain an orientation accuracy of 5 arc-seconds or better implies that the allowable edge orientation error is one degree or less. This is usually not a problem in AT or BT blanks cut from well oriented cultured material, but can be serious in doubly-rotated blanks or blanks poorly cut from natural material.

III. MULTIPLE BEAM X-RAY MACHINES

Although far more complicated than a single beam x-ray system, and hence also not very versatile, multiple beam x-ray systems have been in production use for nearly 20 years.

These systems have their start in the realization that:

> "A complete determination of the errors in a quartz plate includes the measurement of angle errors about three mutually perpendicular axes such as, for example, the plate edges. Therefore, in correcting a plate we must use three different x-ray 'shots'; usually two on the major surface with the plate rotated 90° between the two 'shots' and a third on a surface normal to the major surface, commonly called an edge."
>
> W.L. Bond
> E.J. Armstrong [7]

In practice, obtaining two 'shots' 90° apart on the major surface is quite difficult without going to extremes in correcting for all of the errors which can arise when not at the minimum (or maximum) of Figure 16.

For any of the most commonly used crystal cuts, AT, BT, or SC, of all the useful planes sufficiently parallel to the surface of the cut to minimize most of the known sources of error, none form the required 90° relationship with the idealized cut when viewed on a stereographic projection [1]. Many are close. For example, the 03·2 plane and the 12·2 plane for the AT cut.

Even more difficult is satisfying the final requirement, a 'shot' at the blank edge. This condition can be satisfied by finding an atomic plane which is not nearly parallel to the desired surface, but is almost perpendicular to the desired surface. Planes exist lying within 5° of being perpendicular to all of the major cuts. All investigated have useful x-ray reflection intensity.

The perpendicular plane is used primarily to establish the correct 'psi' rotation of the blank. This greatly minimizes the psi errors on the other two atomic planes as discussed above. Also, as discussed above, not extreme accuracy is required in the alignment or determination of the actual x-ray peak position. Accuracies on the order of 0.5 degrees of rotation are usually sufficient.

By the proper determination of a set of planes which meet the above requirements, it is possible to construct a system which can be used to either measure or to angle-correct a quartz blank. Such a system, designed for angle-correcting SC-cut blanks, is shown in Figure 18. It is unique in its use of only two x-ray sources to generate the necessary 3 x-ray beams. This system has been functional for nearly a decade and repeatedly angle-corrects blanks to a total 2-sigma scatter of less than 5 arc-seconds. [8]

Drawbacks to multiple beam systems are their complexity, cost, and lack of versatility. The SC-cut system shown in Figure 18 is limited by its internal construction to operation within 30 arc-minutes of the original TTC-cut. Changing to another crystal cut requires an entirely new machine. Indeed, the one shown got its start almost 2 decades ago as an AT-cut

machine, then was completely redesigned and rebuilt in its present form.

IV. APPLICATIONS

Sorting

This is the most common usage of x-ray reflection systems in crystal production today. Although much precision can be obtained in the sawing operations that yield quartz blanks, in the final analysis, most crystal operations depend upon sorting crystal blanks to some form of specification. This usually takes the form of x-ray goniometer readings. With a knowledge of operating temperature vs. orientation angle, this is usually sufficient to obtain crystal blanks with the proper performance. This is especially true for the AT-cut.

Recent advances include automation, sorters, and autohandlers, all under computer control. [9, 10, 11, 12]

Incorporation of rocking curve calculations as discussed above also improve the overall accuracy of the x-ray system.

Angle Correction

In several crystal production facilities, a very limited number of different resonator blanks are in production. For example, all units produced might be required to have turnovers between 101° and 105° C. If the yield due to sorting is not sufficiently high, a more difficult production process may be required. Angle correction is a process whereby the actual surface of the crystal blank can be changed slightly in orientation angle to provide the correct crystallographic orientation in the final blank. Blanks can be produced in this manner to an accuracy of ±3 arc-seconds with a yield greater than 99%.

In one method, the crystal is mounted on a lapping fixture with adjustable diamond feet (see Figure 19) which rest against a reference surface on the x-ray system shown in Figure 20. Instead of actually measuring the surface orientation, the diamond feet are adjusted until the crystal is at the desired orientation. The lapping fixture with the crystal blank attached is placed on an abrasive lapping machine. Quartz removal continues until the diamond tipped feet prevent any further lapping. The new, lapped, crystal surface is now at the desired orientation.

In another method, the crystal is carefully measured using a rocking curve system. Correction factors are computed based on the measured x-ray angles. The crystal blank is then inserted into a special vacuum chuck which is adjustable in rotation about 3 perpendicular axes. The correction factors are dialed into the chuck settings, and the surface of the crystal ground with a diamond wheel to the correct crystallographic orientation.

Other methods which depend upon asymmetric surface damage, etch rates, etc., have been tried but have not been incorporated into full-scale production operations.

ACKNOWLEDGMENTS

The authors wish to acknowledge the work of Dr. Hugo von Fellner for the x-ray lineshapes presented, and for the efforts, encouragement, and direction of Dr. Virgil Bottom, without whom, many of us would not be in this business.

REFERENCES

1. Heising, R.A., Quartz Crystals for Electrical Circuits, Electronic Industries Association, Washington D.C., 1970, pg. 99.

2. Private communication, Dr. Hugo von Fellner.

3. Heising, pg. 97.

4. Bottom, V.E., Introduction to Quartz Crystal Unit Design, Van Nostrand, Reinhold, 1982, pp. 186-187.

5. Heising, pg. 108, ff. See also: Bond, W.L., Crystal Technology, John Wiley & Sons, New York, 1976, pp. 86-91.

6. Bond, W.L., "A Double-Crystal X-Ray Goniometer for Accurate Orientation Determination," Proc. IRE, Aug. 1950, pp. 886-889.

7. Heising, pg. 128.

8. As measured in the system described in: Chambers, J.L., "An Instrument for Automated Measurement of the Angles of Cut of Doubly Rotated Quartz Crystals," Proc. of the 37th Annual Symposium on Frequency Control, IEEE 83CH1957-0, 1983, pp. 275, ff.

9. Darces, J.F. and Merigoux, H., "Final X-Ray Control of the Orientation of Round or Rectangular Quartz Slides for Industrial Purposes," Proc. of the 32nd Annual Symposium on Frequency Control, EIA, 1978, pp. 304-309.

10. Asanuma, N. and Asahara, J., "Highly Precise Measurement of Orientation Angle for Crystal Blanks," Proc. of the 34th Annual Symposium on Frequency Control, EIA, 1980, pp. 120-130.

11. Knolmayer, E., "X-Ray Goniometry of the Modified (simplified) Doubly Rotated Cuts," Proc. of the 35th Annual Symposium on Frequency Control, EIA, 1981, pp. 56, ff.

12. Merigoux, H. and Darces, J.F., "Ψ Angle in a Triply Rotated Cut, Determination and Control," Proc. of the 40th Annual Symposium on Frequency Control, IEEE 86CH2330-9, 1986, pp. 140-144.

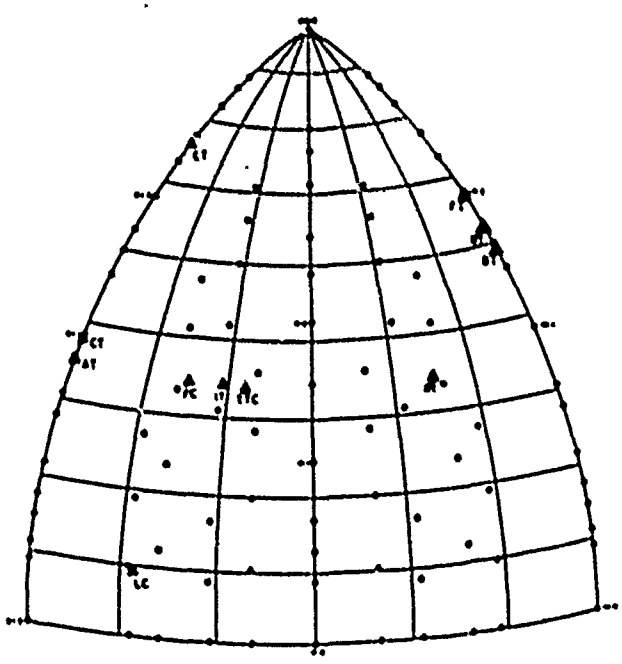


Figure 1. Orthographic projection - Primitive region in α-quartz.

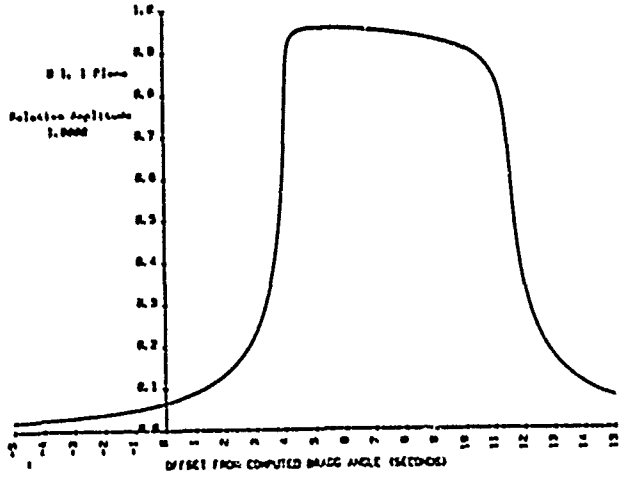


Figure 2. Calculated lineshape - 01·1 plane

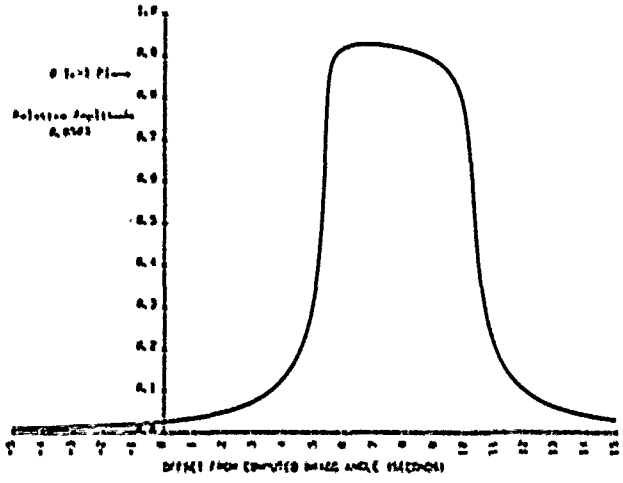


Figure 3. Calculated lineshape - 01·$\bar{1}$ plane

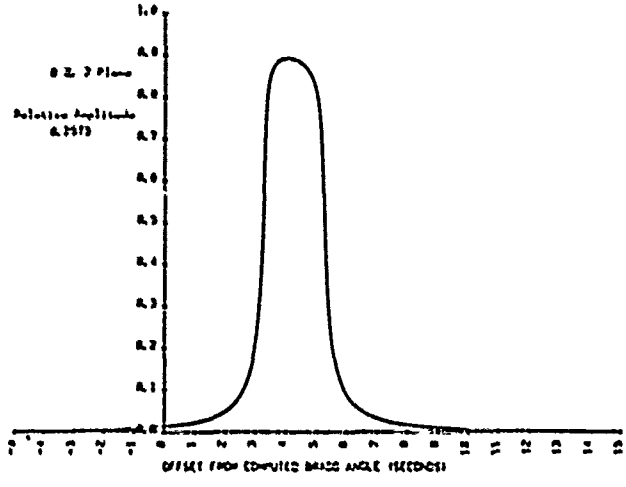


Figure 4. Calculated lineshape - 02·2 plane

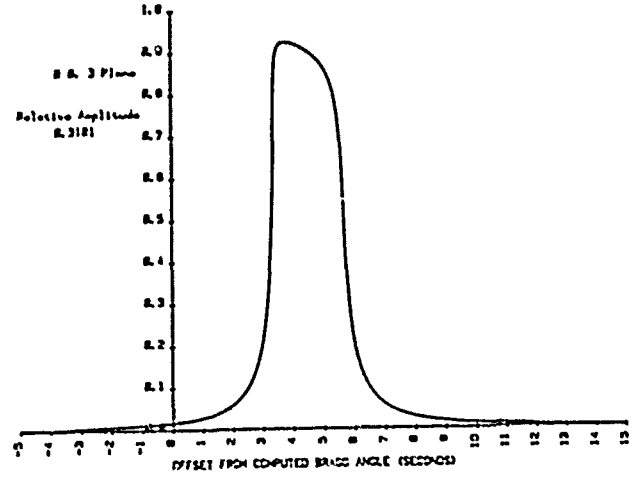


Figure 5. Calculated lineshape - 00·3 plane

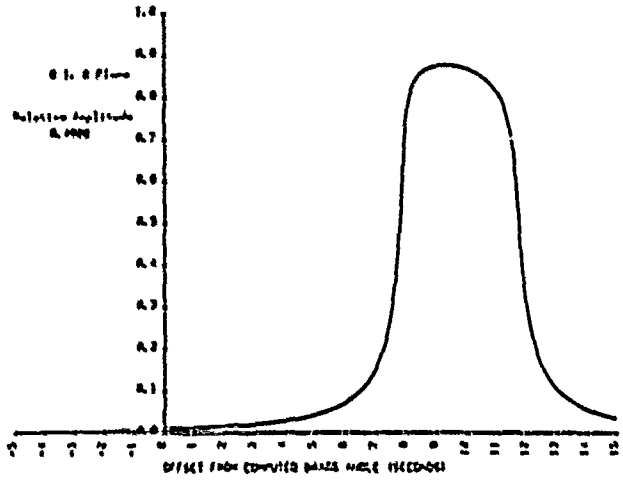


Figure 6. Calculated lineshape - 01·0 plane

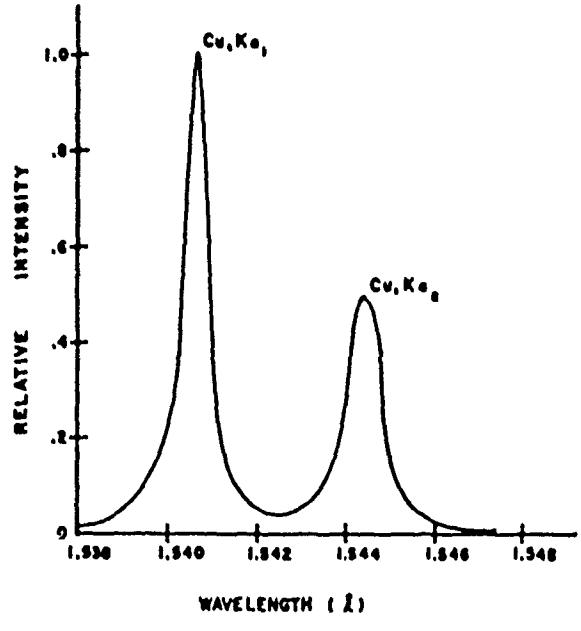


Figure 7. Energy plot - Kα doublet

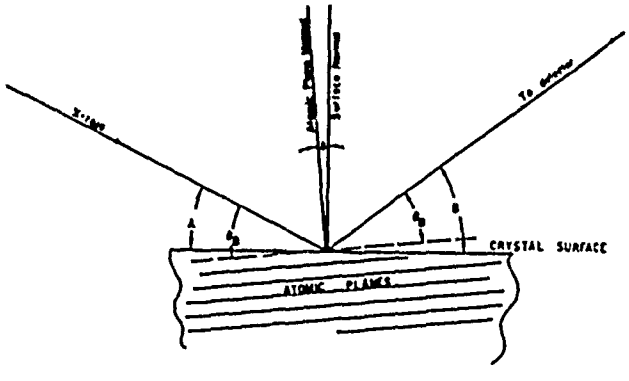


Figure 8. X-ray reflection

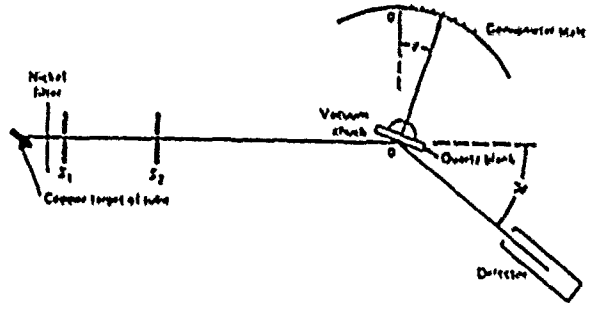


Figure 9. Single crystal x-ray diffraction system

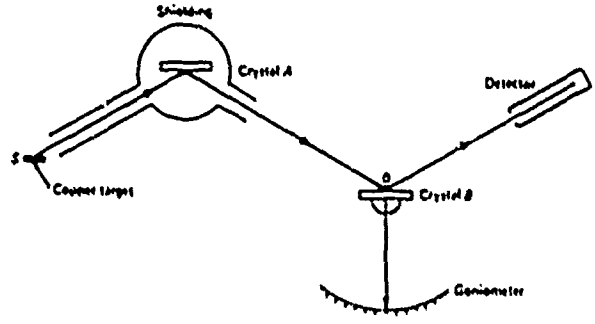


Figure 10. Dual crystal x-ray diffraction system

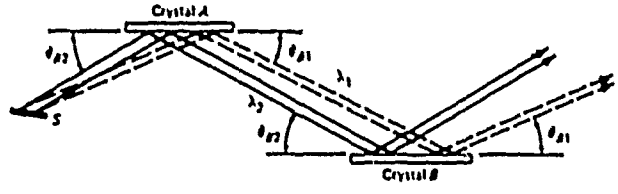


Figure 11. Elimination of Kα doublet effect by use of a collimating crystal with the same Bragg angle.

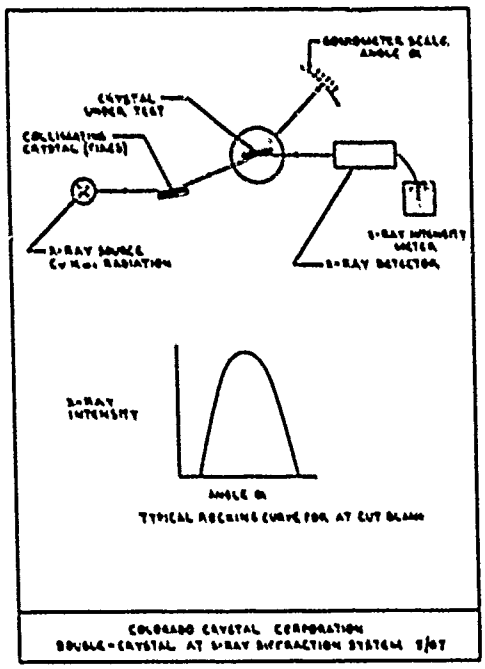


Figure 12. Rocking curve system - AT cut

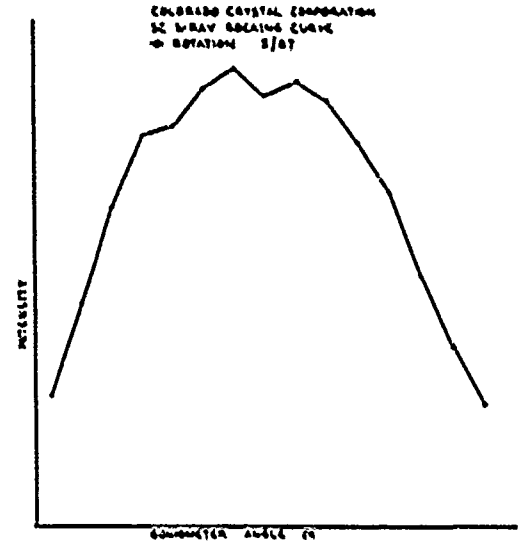


Figure 14. Rocking curve response - SC cut, first position

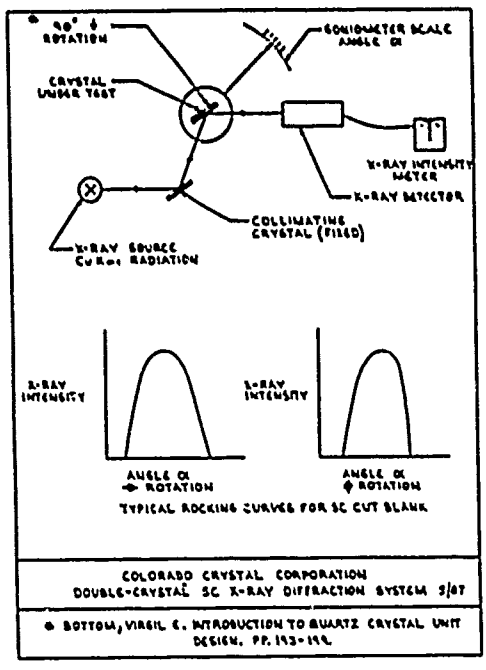


Figure 13. Rocking curve system - SC cut

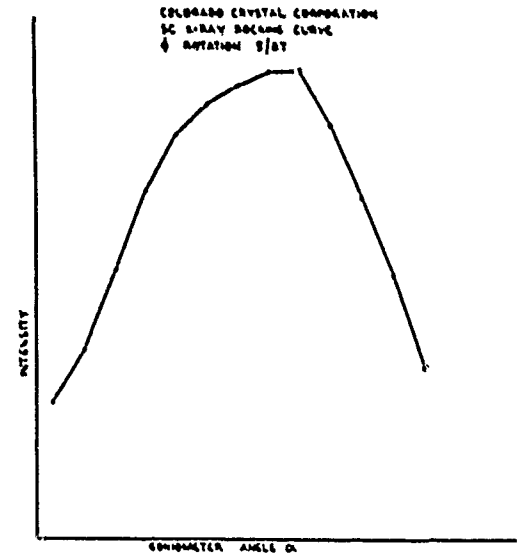


Figure 15. Rocking curve response - SC cut, second position

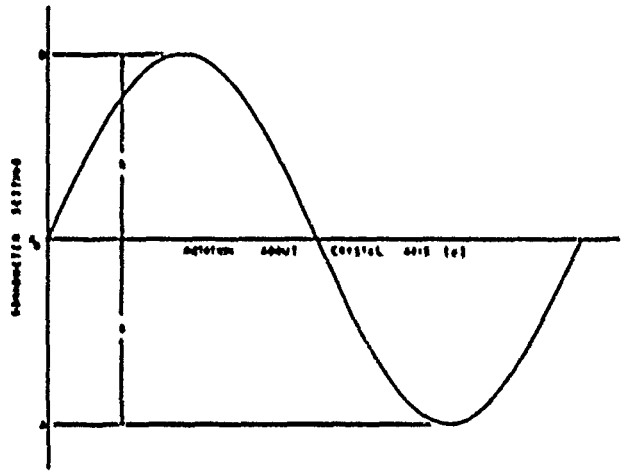


Figure 16. X-ray goniometer reflection angle vs. rotation of quartz blank

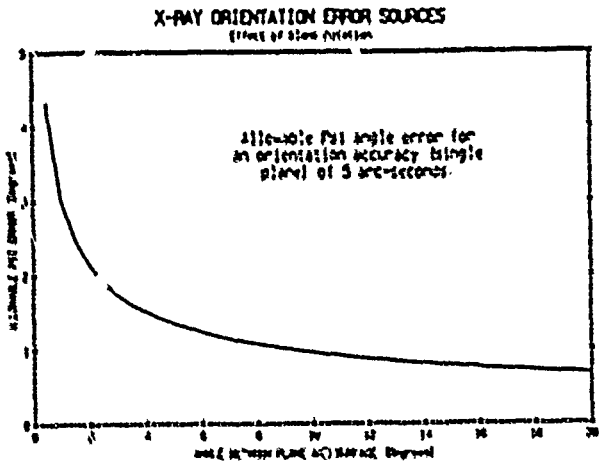


Figure 17. Allowable edge orientation error as a function of the Bragg angle to obtain blank orientation within 5' of arc



Figure 18. Multiple beam x-ray system - "the 3-headed monster"

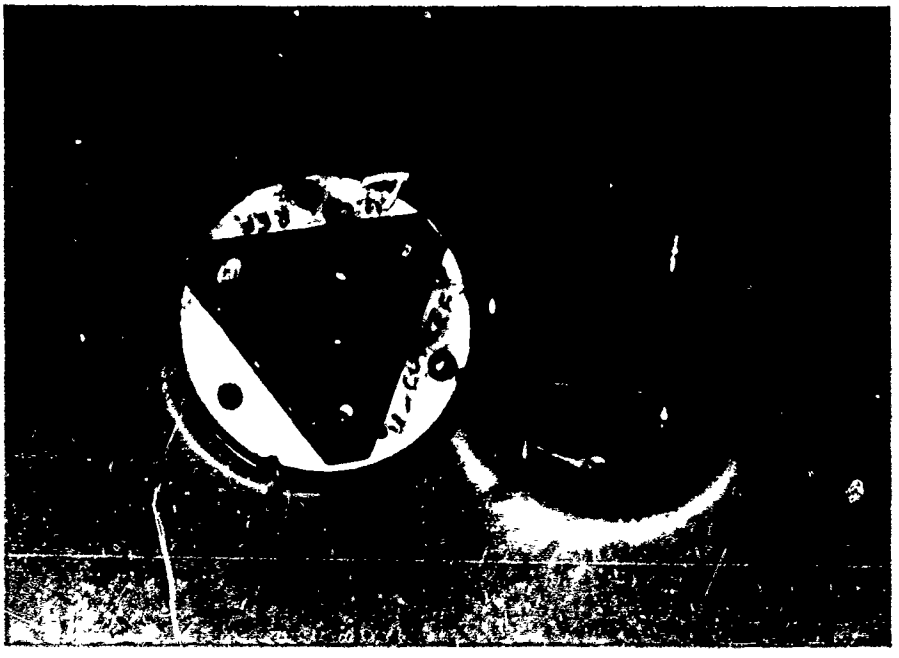


Figure 19. Adjustable lapping fixture

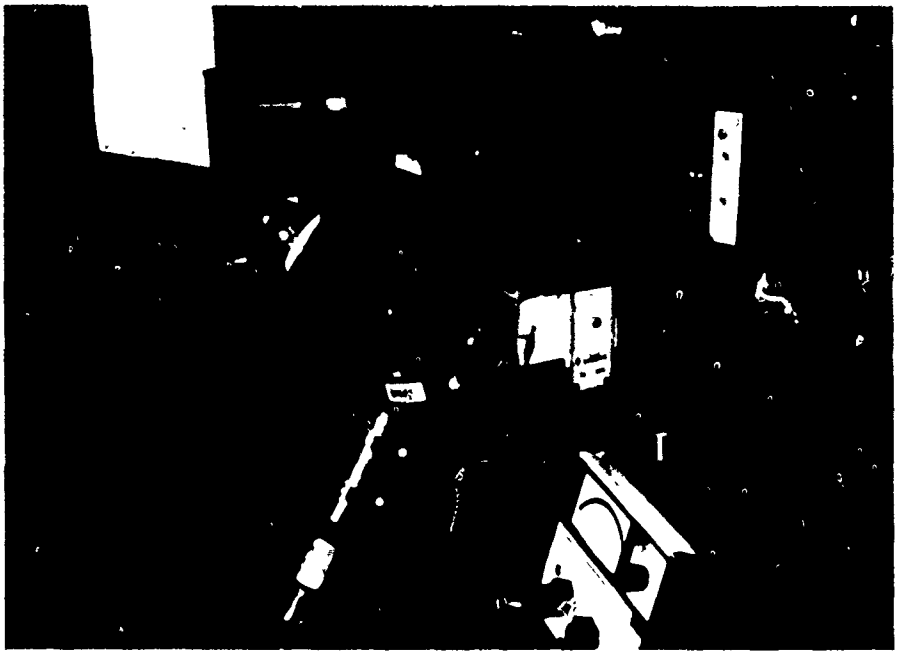


Figure 20. Single axis x-ray system used for angle correction.

AN UPDATE OF SURFACE MOUNT PACKAGES FOR QUARTZ CRYSTAL PRODUCTS

Clifford Mercer
Reeves-Hoffman Division
Carlisle, PA 17013

Abstract

In the 8th (1986) Quartz Devices Conference & Exhibition Proceedings, the author reported on the past, present and future of surface mount components for quartz frequency control products. Industry standards for electronic components were reviewed for possible application to the quartz industry. The impact of surface mount board processing on component package design was discussed. Available suppliers of surface mountable quartz crystal products and descriptions of those products were listed. Specifications for environmental testing of surface mount quartz products were proposed.

This paper concentrates on the availability of surface mount quartz products.

Introduction

Approximately fifty (50) manufacturers of quartz frequency control products were contacted to determine whether or not they offered for sale surface mount crystals, oscillators or filters. Those companies that did were requested to send sample packages and product data sheets. Photographs of those devices are presented. Details on package materials, construction, dimensions and applications are presented if the information was made available by the manufacturers.

Although every effort was made to accurately present this information, final package details and product data should be obtained from the respective companies. Their telephone numbers and addresses are listed.

Product Descriptions

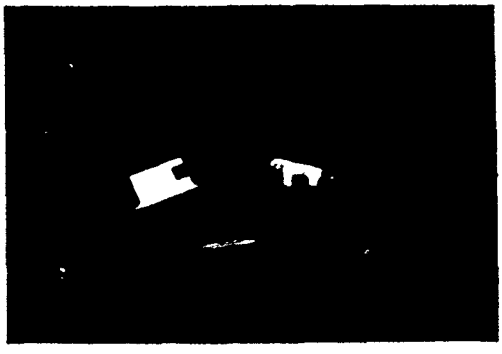


Figure 1. Ceramic Packages from Statek Corporation

Figure 1 is a picture of two (2) miniature ceramic packages with glass covers hermetically soft solder sealed. The larger package measures .335"x.160"x.080". The smaller measures .275"x.108"x.075". In the larger package AT strip resonators from 10 to 30 MHz can be supplied. Tuning fork type resonators from 10 KHz to 2 MHz are available in the smaller package. The larger package can be supplied on 16 mm tape while the smaller package is supplied on 12 mm tape.Both are available on 7" or 13" reels as well as in trays or in bulk. The available data sheet specifies that these parts will withstand 270°C for up to 30 seconds. Statek recommends gold plated tabs for vapor and infrared reflow systems and nickel-tin plating for wave solder. For parts requiring good solderability after a long shelf life, a nickel-tin plating plus a solder dip is recommended. All are available when requested. For more details contact:

Statek Corporation
512 North Main St.
Orange, CA 92668
(714) 639-7810

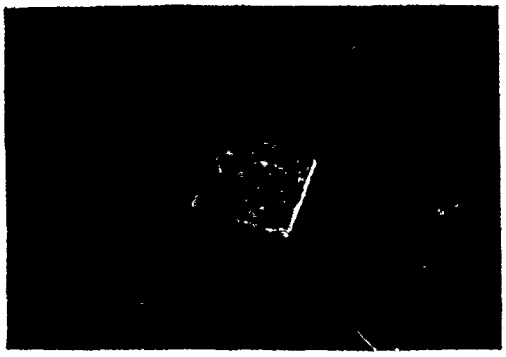


Figure 2. LCCC from Statek Corporation

Figure 2 shows a .400"x.400" LCCC. With a ceramic lid the package measures .090" in height. An .080" high option is available with a metal lid. This is a 24-pin device with .050" lead centers. TTL and CMOS oscillators from 10 KHz to 2 MHz are available in this package. Statek advertises that this package is suitable for wave solder and vapor phase reflow systems. (Reference Figure 1 description for company address.)

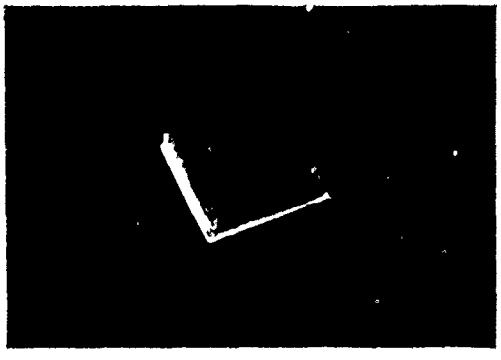


Figure 3. LCCC from Oscillatek Corporation

Shown in Figure 3 is a .480"x.480" 40-pin LCCC. The pins are on .040" centers. The package measures .120" high. Oscillatek advertises that TTL and CMOS oscillators are available in this package.

Oscillatek Corporation
620 North Lindenwood Drive
Olathe, Kansas 66062
(913) 829-1777

CH2427-3/87/0000-258 $1.00 © 1987 IEEE

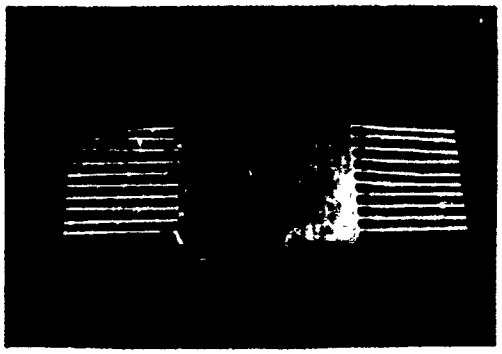


Figure 4. Flatpack from Oscillatek Corp.

Figure 4 shows a .625"x.625"x.150" metal flatpack. The 20 pins are on .050" centers. TTL and CMOS oscillators can be supplied in this package. (Reference Figure 3 description for company address.)

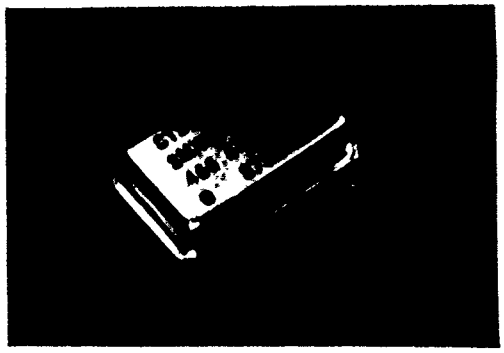


Figure 5. 4-Pin from CTS Knights, Inc.

Figure 5 is a picture of a .500"x.800", 4-pin package. The body height is .200". The leads add .050" to the overall height. The leads are flattened and formed into a gullwing configuration. The package is made of cold rolled steel and is hermetically sealed by resistance welding. The leads are nickel plated Kovar for good solderability and lead compliance. Parts can be supplied on tape and reel. TTL, CMOS, ECL clocks and TTL type VCXO oscillators are available. Consult the factory for details. CTS claims that these parts will meet a 16-hour solderability steam test. CTS also advertises that the package is suitable for wave solder and vapor phase reflow systems.

CTS Knights, Inc.
400 Reimann Avenue
Sandwich, IL 60548
(815) 786-8411



Figure 6. .500" Square 4-Pin from CTS Knights, Inc.

CTS Knights, Inc. also supplies a .500" square version (Figure 6) of the package shown in Figure 5. TTL oscillators from 307 KHz to 70 MHz and HCMOS oscillators from 4 MHz to 25 MHz are available in this package. (Reference Figure 5 description for company address.)

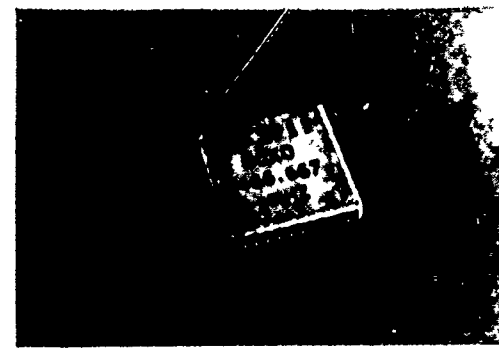


Figure 7. LCCC from CTS Knights, Inc.

CTS offers a .480" square LCCC (Figure 7) with 40 leads on .040" centers. The package height is .075". Not pictured but also available are .650" square and .450" square LCCC packages. Consult the factory for oscillator specifications on all three packages. (Reference Figure 5 description for company address.)

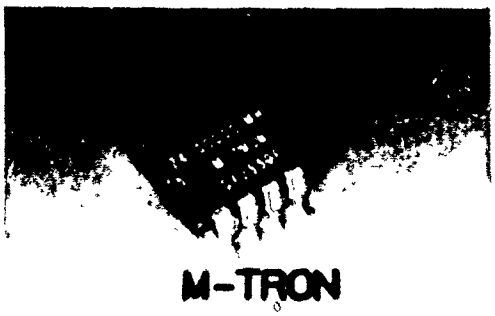


Figure 8. 8-Pin Gullwing Epoxy Package from M-Tron Industries, Inc.

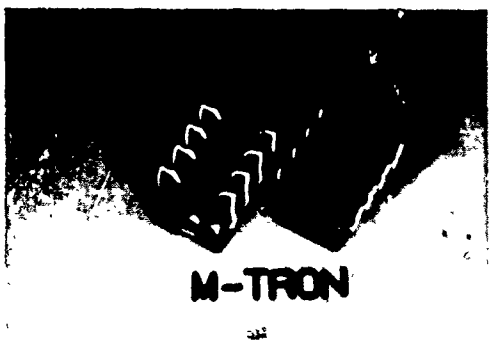


Figure 9. 8-Pin Butt and J Lead Epoxy Package from M-Tron Industries, Inc.

M-Tron Industries, Inc. offers a .500"x.300" epoxy package with gullwing leads (Figure 8) and butt or J leads (Figure 9). All three (3) package bodies are .160" in height. Total package height with leads is approximately .210". Hermeticity is achieved by encapsulating a metal hermetically sealed package in epoxy. AT strip resonators are used. Frequencies between 4 to 24 MHz can be supplied. M-Tron advertises that this package is suitable for all solder reflow systems. Consult the M-Tron product data sheet for solvent resistivity specifications.

M-Tron Industries, Inc.
100 Douglas Avenue
Yankton, S.D. 57078
(605) 665-9321



Figure 10. TO Header from Piezo Crystal Company

Piezo Crystal Company has just developed an HC-37/U size header for the high reliability crystal market (Figure 10). The package measures .510" in diameter and .200" in height. The package is gold plated. Piezo claims helium leak rates better than 1 x 10-10 atm cc/sec. Parts are coldweld sealed. Unique features include wraparound contact pads for easier solder joint inspection and a notched glass base for auto placement orientation.

Piezo Crystal Company
P. O. Box 619
Carlisle, PA 17013
(717) 249-2151

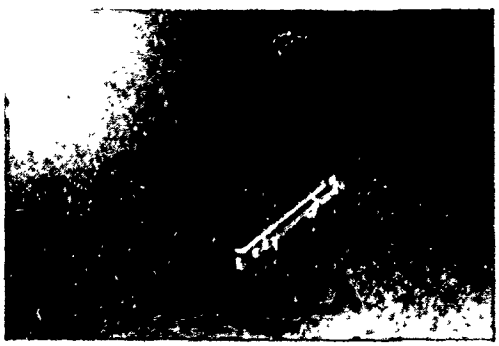


Figure 11. Ceramic Package from Motorola, Inc.

Motorola, Inc. offers a 4-leaded ceramic hermetically sealed package which measures .560"x.370". The sample pictured in Figure 11 and Figure 12 incorporates the optional J lead with an overall height of .155". The standard product (not shown) comes with a C lead and overall height of .125". Leads are copper pretinned with solder. Parts can be supplied on 24 mm tape in either 7" or 13" reels or in tubes. CMOS and TTL clocks between 1.25 to 35 MHz using AT strip resonators can be supplied in this package. The very complete product data sheet states thermal shock, vibration, mechanical shock and temperature cycle environmental test conditions this product will pass. Pad geometrics, land patterns and solder paste screening specifications are also listed. Motorola states this part is suitable for vapor phase and wave solder applications.



Figure 12. J-Lead Details of Motorola Package

Motorola, Inc. Components Div.
2553 N. Edgington St.
Franklin Park, IL 60131
(314) 451-1000

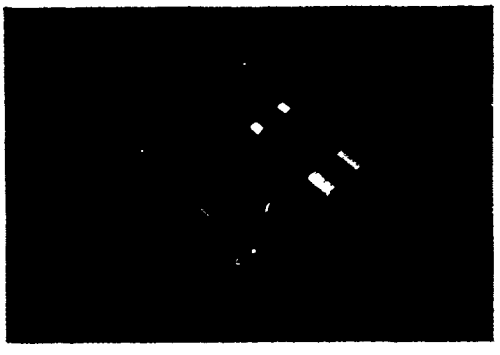


Fig. 13. Plastic Package from Std. Communication Corp.

Figure 13 shows a top and bottom view of a plastic package measuring .510"x.200". Four (4) leads are bent in a C configuration. A hermetically sealed AT strip resonator in a 3 x 8 mm metal package is encapsulated in the plastic as shown in Figure 14. The frequency range available is 4 to 20 MHz.



Fig.14.Cross Section of the Std. Comm.Plastics Package

Standard Communications Corp.
P. O. Box 92151
Los Angeles, CA 90009
(213) 532-5300



Figure 15. HC-49/U Style Holders from Valpey-Fisher Corporation



Figure 16. Short, Leadless HC-49/U Style Holder from Valpey-Fisher Corporation

Valpey-Fisher Corporation offers a standard HC-49/U package with formed leads for surface mount. A third lead is attached to the top of the crystal cover as shown in Figure 15. A short version (.200" maximum height) is shown in Figure 16. This has been mounted on a printed circuit board for leadless surface mount applications.

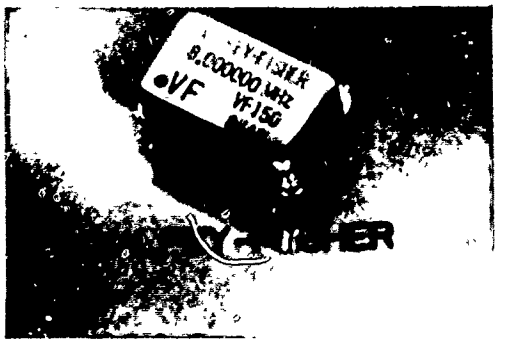


Figure 17. 4-Pin Header from Valpey-Fisher Corp.

The 4-pin header in Figure 17 measures .800"x.500" and has the leads formed into a gullwing configuration. Unlike the CTS part (Figure 5), the leads of the Valpey-Fisher package are not flattened.

Valpey-Fisher Corporation
75 South Street
Hopkinton, MA 01748
(617) 435-6831

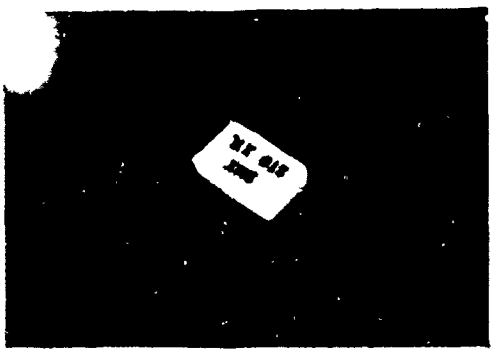


Figure 18. Ceramic Package from Standard Crystal Corp.

The package in Figure 18 consists of an all ceramic base and cover. It measures .400"x.250"x.070". AT resonators from 4 MHz to 200 MHz can be supplied in this package. Not shown are two gold fired contact pads on the underside of the holder.

Standard Crystal Corp.
9940 Baldwin Place
El Monte, CA 91731
(818) 443-2121

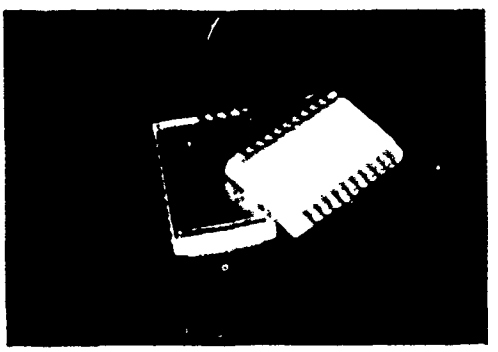


Fig. 19. Ceramic Flatpack from McCoy Electronics Co.

The ceramic flatpack in Figure 19 measures .600"x .400"x.100" high. The twenty (20) leads are on .050" centers. In this package McCoy offers clock oscillators from 14 to 60 MHz, crystals from 10 to 300 MHz and filters from 10 to 30 MHz.

McCoy Electronics Company
100 Watts Street
P. O. Box B
Mount Holly Springs, PA 17065
(717) 486-3411

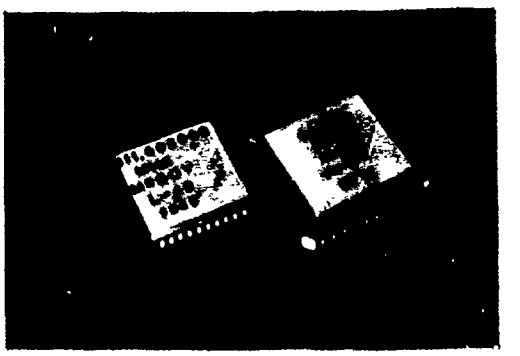


Figure 20. LCCC's from McCoy Electronics Company

The LCCC on the left, Figure 20, measures .480" square by .085" and has 40 pins on .040" centers. Clock oscillators from 10 to 100 MHz, crystals from 10 to 300 MHz and filters from 10 to 30 MHz are available. The package on the right in Figure 20 measures .650" square by .085" high. It is a 40-pin device on .050 centers. McCoy advertises TTL clocks from 1 to 100 MHz in this package. (Reference Figure 19 description for company address.)

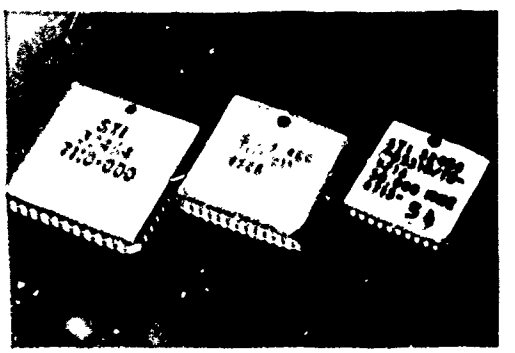


Figure 21. LCCC's from Spectrum Technology, Inc.

The package on the left in Figure 21 measures .650" square by .085" high. It has 44 pins on .050" centers. Spectrum Technology advertises this package for MIL-O-55310/20 oscillators.

The package in the center of Figure 21 measures .560" square by .105" high. It has 48 pins on .040" centers.

The package on the right in Figure 21 measures .480" square by .085" high. It has 40 pins on .040" centers. Spectrum Technology uses this package for MIL-O-55310/19 oscillators.

Spectrum Technology, Inc.
P. O. Box 948
Goleta, CA 93116
(805) 964-7791

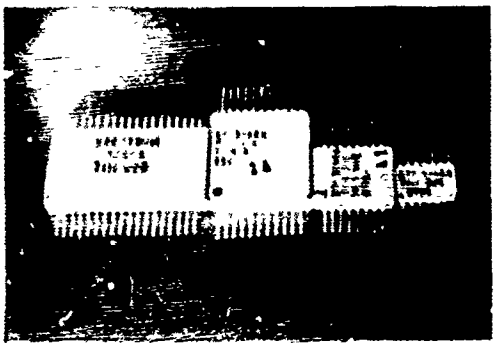
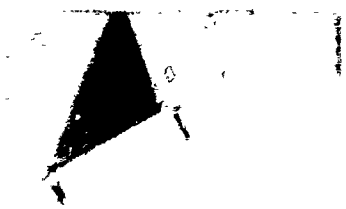


Figure 22. Metal Flatpacks from Spectrum Technology

The four flatpacks in Figure 22 range from 1.00"x .500"x.200" (left), .625" square x .120" (2nd from left), .375x.500"x.120" (2nd from right) to .250"x.375" x.060" (right). All leads are on .050 centers. TTL and CMOS oscillators are available in all packages. The .625" square package is used primarily for MIL-O-55310/21 oscillators. Consult the factory for application details. (Reference Figure 21 description for company address.)



MURATA ERIE

Figure 23. 4-Pin Header from Murata-Erie North America

The 4-pin from Murata Erie measures .500"x.600". The body package height is .185", not counting the leads. The leads are flattened and formed into a gull-wing configuration. TTL and CMOS oscillators are available in this package.

Murata Erie North America
Carlisle Division
453 Lincoln St.
P. O. Box B
Carlisle, PA 17013
(717) 249-2232

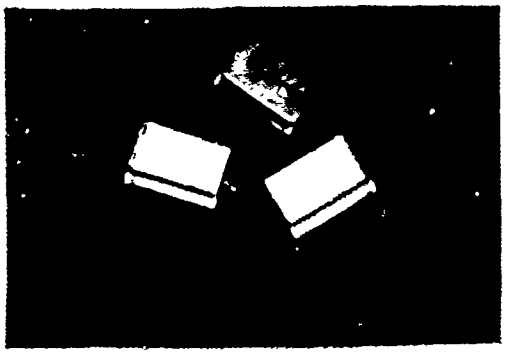


Figure 24. HC-49/U Type Headers from Murata Erie North America

The HC-49/U type headers in Figure 24 are supplied by Murata Erie with formed leads for surface mount applications. A third lead is optional if a chip component is not used for the third support. (Reference Figure 23 description for company address.)

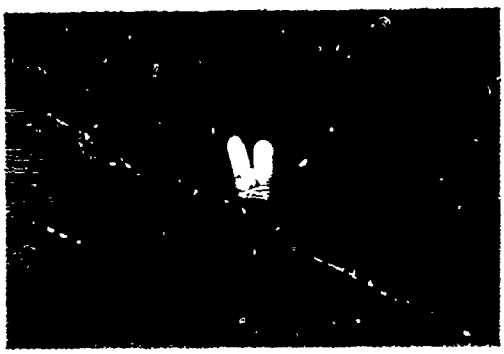


Figure 25. Tubular Ceramic Package from Murata-Erie North America



Fig. 26. Plastic Packages from Murata Erie N. America

The ceramic tube shown in Figure 25 is presently used by Murata Erie to package ceramic resonators. It measures .275" long by .110" in diameter. The plastic packages shown in Figure 25 measure .315" square by .140" high and is also presently used to house ceramic resonators. (Reference Figure 23 description for company address.)

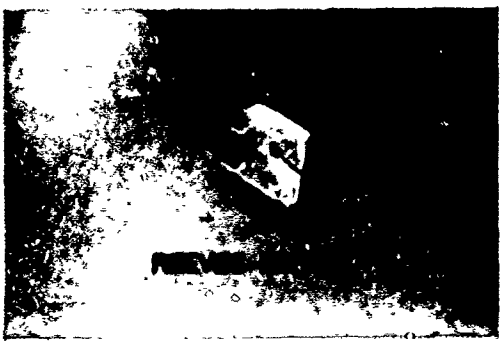


Figure 27. 4-Pin Package from Reeves-Hoffman Division

The package supplied by Reeves-Hoffman Div. and shown in Figure 27 measures .500" square. Total height including leads is .265". The leads are flattened and formed into a gullwing configuration. Base material is cold rolled steel. The leads are kovar. Parts are hermetically sealed with a resistance weld process. Nickel plating is used to improve corrosion resistance. Crystals from 4 to 200 MHz are supplied in this package.

Reeves-Hoffman Division
400 W. North St.
Carlisle, PA 17013
(717) 243-5929

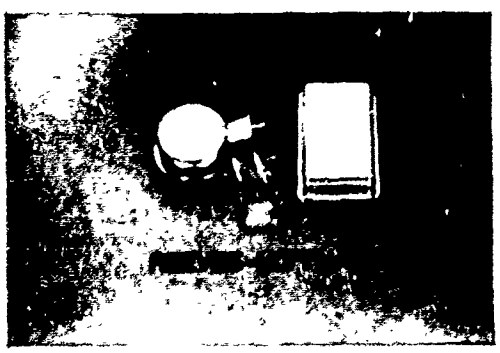


Figure 28. Butt Lead Packages from Reeves-Hoffman Division

Butt lead packages are available from Reeves-Hoffman Div. in both the HC-37/U style and 14-pin DIP style packages as shown in Figure 28. The leads on both packages measure .017" in diameter with a .040" head. The standard lead length is .050". Leads are made of alloy 52 to improve compliance. Base material is cold rolled steel. Excellent coplanarity results with this unique design. The round header is .510" in diameter and .230" high including leads. Crystals from 4 to 200 MHz can be supplied in this package. The rectangular package (Figure 28) measures .800"x .500" and can be supplied with up to 14 leads. CMOS, TTL, ECL (10K) and ECL (100K) clocks and TTL VCXO's are supplied in the rectangular package. Consult the factory for details. (Reference Figure 27 description for company address.)

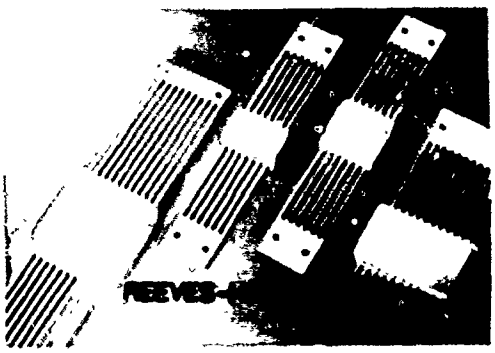


Figure 29. Metal Flatpacks from Reeves-Hoffman Div.

Figure 29 illustrates a sample of more than twenty (20) different metal flatpacks available from Reeves-Hoffman Div. Consult the factory for package dimensions. (Reference Figure 27 description for company address.)

Additional Sources of Surface Mount Components

The following companies also offer surface mount quartz frequency control products. In each case, the manufacturer should be contacted for details.

Company	Packages
Midland Ross Corporation 357 Belort St. Burlington, WI 53105 (414) 763-3591	Modified HC-18/U and HC-33/U Packages
Monitor Products Co., Inc. 502 Via del Monte Oceanside, CA 92054 (619) 433-4510	Modified HC-49/U .500" x .800" Gullwing 4-Pin
Bliley Electric Co. 2545 W. Grandview Blvd. Erie, PA 16506 (814) 838-3571	.520 Square Gullwing 4-Pin, Modified HC-18/U
Connor-Winfield Corp. 114 West Washington P. O. Box L West Chicago, IL 60185 (312) 231-5270	.500" x .800" Gullwing 4-Pin
Dale Electronics 1155 W. 23rd St. Tempe, AZ 85282 (602) 967-7874	.500" x .800" Gullwing 4-Pin, .480" Square LCCC
Hughes Aircraft Co. 500 Superior Avenue Newport Beach, CA 92663 (714) 759-2430	Metal Flatpacks, Custom Ceramic
Q-Tech Corporation 2201 Carmelina Ave. Los Angeles, CA 90064 (213) 820-4921	.625" Square and .500" x .375 Flatpacks, .650" Square and .480" Square LCCC's
Saronix 4010 Transport St. Palo Alto, CA 94303 (415) 856-6900	Leadless .200"x.500" Plastic Package, .160"x .335" Leadless Ceramic Package

Company	Packages
Sokol Crystal Products 121 Water St. P. O. Box 249 Mineral Point, WI 53565 (608) 987-3363	Various LCCC's
Daiwa Crystal Corp. 3838 Carlson St. Suite 301 Torrance, CA 90503	Short HC-49/U, .200" x .400" Leadless Plastic, .181" x .417" Glass
Vectron Laboratories, Inc. 166 Glover Ave. Norwalk, CT 06850 (203) 853-4433	Various Flatpacks, .500" x .800" Gullwing 4-Pin

Surface Mount Package Registration

Companies interested in registering their surface mount packages for the purposes of eventual standardization should request information from:

P-4 Committee on Mechanical Outlines
c/o Electronics Industry Association
Engineering Departme-t
2001 Eye Street, N.W.
Washington, D.C. 20006

The following chain of events is required before a component can be registered:

a) Formal Request to EIA
b) EIA Approval to Ballot
c) P-4 Committee Ballots
d) P-4 Committee Makes Recommendation
e) EIA Parts Panel Chairman Acts on Recommendation
f) Final Drawings Are Requested
g) EIA Publishes Registered Component

Acknowledgement

The author would like to thank all of those companies that allowed photographs of their products to be shared with the industry. Without their cooperation, this report would not have been possible.

References

1. C. D. Mercer, Proc. 8th Quartz Devices Conference, Electronic Industries Association, Washington, D.C., p. 78, (1986).

2. 1986-87 Quartz Devices Directory, Great Southern Marketing, Ft. Lauderdale, FL.

3. 1986-87 Electronic Engineers Master Catalog, Hearst Business Communications, Inc., Garden City, N.Y.

NONLINEAR CONSTANTS AND THEIR SIGNIFICANCE

J.J. Gagnepain

Laboratoire de Physique et Métrologie des Oscillateurs du C.N.R.S.
associé à l'Université de Franche-Comté - Besançon
32 avenue de l'Observatoire - 25000 Besançon - France

Abstract

Nonlinearities are at the origin of most of the instabilities of resonators and oscillators, and a large effort during the two last decades was achieved to reduce their effects. At the opposite they can be used in some applications like sensors or signal processing. The paper introduces the different nonlinear fundamental constants which can be encountered in a crystal in relation with its elastic, piezoelectric, dielectric, optic, properties. The knowledge of these different constants and the confidence one can have is commented. The unknown constants and the need of measuring some of them is also indicated. The influence of these constants on the behavior of resonators and other devices is presented.

Introduction

The properties of acoustic waves and therefore the characteristics of resonators and other piezoelectric devices can be influenced by internal or external parameters on account of the nonlinear properties of the crystal. These nonlinearities are at the origin of harmonic generation, amplitude-frequency/velocity effect, intermodulation, in case of interactions between waves, and of the sensitivities to various physical quantities, like temperature, forces, accelerations, pressures, electric fields, ... when the acoustic wave is propagating in a medium with a bias. It must be remembered, in addition, that nonlinearities find their origin in the anharmonic nature of the interatomic forces of the crystal, and that well known phenomena like acoustic attenuation, thermal expansion, heat diffusion result from phonon interactions due to the nonlinear coupling between the lattice modes.

Since all these different kinds of nonlinear couplings are in fact a cause of instabilities in the piezoelectric devices, and mainly in resonators, large efforts have been made during the two last decades in trying to reduce them. However the description of the nonlinear behavior of the device requires the knowledge, of not only the regular second order fundamental constants of the material, but also of the higher order constants of the third order and in some cases of the fourth order.

At the opposite, nonlinearities can be used to make, for example, a resonator selectively sensitive to a given physical quantity. Thus it can be used as a sensor. Temperature, pressure, acceleration, etc, sensors have been developed based on this principle. Nonlinearities also enable to achieve convolution products or correlation function calculation. In those applications the knowledge of nonlinear constants also is necessary for improving the sensitivity or the efficiency.

This paper will be related with the phenomena resulting mainly of elastic and piezoelectric nonlinearities.

Nonlinear fundamental equations and nonlinear fundamental constants

It is not the purpose of this paper to give an exhaustive presentation of the nonlinear theory of elasticity and piezoelectricity. The details of this theory can be found in different papers of Murnaghan, Toupin, Thurston and Tiersten.[1-4] However it seems important to describe the different nonlinear contributions, and to recall that the nonlinear fundamental constants are not the only cause of nonlinearities, and that the fundamental equations have an inherent nonlinear nature, which cannot be neglected.

The behavior of a solid submitted to a finite deformation (and not to an infinitesimal one) is described by equilibrium and constitutive equations, which are nonlinear. These nonlinearities are due to the deformations themselves, which induce local changes of the specific mass, of the elementary surface areas, which modify the direction of the surface normal vectors, etc. As a consequence, the internal stresses (which are forces per unit surfaces) cannot be proportional to the strains.

A point of the solid with coordinates a_j, when at rest (initial state) comes to a new position of coordinates x_j (final state) after a deformation.

The mechanical equilibrium is represented in the final state by the equation

$$\rho \frac{dv_i}{dt} = \frac{\partial T_{ij}}{\partial x_i} \qquad (1)$$

where v_j is the instantaneous velocity of the elementary particle ($v_j = dx_j/dt$), ρ is the specific mass after deformation, and T_{ij} represent the stresses in the final state. Therefore it is obvious that this equation is nonlinear.

The corresponding boundary condition, in case of a stress free surface, is

$$n_i T_{ij} = 0 \quad \text{on the surface} \qquad (2)$$

where the cosine director n_i of the normal to the surface, in the final state, depends on the spatial coordinates.

The stresses T_{ij} are related to the internal energy χ (per unit mass) by the constitutive equation

$$T_{ij} = \rho \frac{\partial x_i}{\partial a_k} \frac{\partial x_j}{\partial a_\ell} \frac{\partial \chi}{\partial \eta_{k\ell}} \qquad (3)$$

where $\eta_{k\ell}$ corresponds to the strain tensor. This relation also is nonlinear, whatever the expression of the internal energy χ can be.

CH2427-3/87/0000-266 $1.00 C 1987 IEEE

The strain tensor η_{kl} is given as a function of the mechanical displacement $u_j = x_j - a_j$ by the relation

$$\eta_{kl} = \frac{1}{2}\left(\frac{\partial u_k}{\partial a_l} + \frac{\partial u_l}{\partial a_k} + \frac{\partial u_a}{\partial a_k}\frac{\partial u_a}{\partial a_l}\right) \qquad (4)$$

The third term of this relation is nonlinear. It is also to be noticed that this tensor is written with respect to the initial state coordinate system.

This brief presentation was limited for simplicity to the pure mechanical equations. For a piezoelectric solid, electric equations are to be added, which also will give additional nonlinearities.[4]

These different nonlinearities were introduced with almost no assumption on the nature of the solid (at the exception of the specific mass).

The peculiar properties of the solid are defined by the expression of the internal energy. If one considers a medium submitted to both mechanical and electrical fields, respectively noted η_{ij} and E_m in the initial state, the development in a power series gives for the internal energy per unit mass in the initial state

$$\begin{aligned}\rho_o X = {} & \frac{1}{2} C_{ijkl}\,\eta_{ij}\,\eta_{kl} + \frac{1}{6} C_{ijklmn}\,\eta_{ij}\,\eta_{kl}\,\eta_{mn} \\ & + \frac{1}{24} C_{ijklmnpq}\,\eta_{ij}\,\eta_{kl}\,\eta_{mn}\,\eta_{pq} + \dots \\ & - \frac{1}{2} e_{mn}\,E_m\,E_n - \frac{1}{6} e_{mnp}\,E_m\,E_n\,E_p \\ & - \frac{1}{24} e_{mnpq}\,E_m\,E_n\,E_p\,E_q + \dots \\ & - e_{mij}\,E_m\,\eta_{ij} - \frac{1}{2} e_{m,ijkl}\,E_m\,\eta_{ij}\,\eta_{kl} \\ & - \frac{1}{2} e_{mn,ij}\,E_m\,E_n\,\eta_{ij} + \text{higher order terms}\end{aligned} \qquad (5)$$

C_{ijkl}, C_{ijklmn}, $C_{ijklmnpq}$ represent respectively the 2nd, 3rd and 4th order elastic constants (the order corresponds to their rank in the energy).

e_{mn}, e_{mnp} and e_{mnpq} are the 2nd, 3rd and 4th order dielectric constants.

e_{mij} are the regular 2nd order piezoelectric constants.

$e_{m,ijkl}$ and $e_{mn,ij}$ correspondant to nonlinear 3rd order piezoconstants. They are respectively known as electroelastic and electrostrictive constants (these last ones can exist even in a nonpiezoelectric medium).

The previous equations indicate that the behavior of the solid will involve two nonlinear contributions. On the one hand will appear the 3rd order constants combined with the linear part of the equations, and on the other hand the nonlinear part of the equations combined with the linear 2nd order constants. As it will be shown further these two contributions are of the same order of magnitude and therefore none can be neglected.

In the microscopic model of the vibration of chains of atoms, the Grüneisen parameter is used for describing anharmonic effects like acoustic attenuation or thermal expansion.[5-9] This parameter can be related to the 2nd and 3rd order elastic constants.[10]

Nonlinear elastic constants

We will consider the simplified pure elastic case. The different problems, which generally are encountered, can be classified into two main families : propagation of a finite amplitude wave in a nonlinear medium, and propagation of an infinitesimal wave in a prestrained nonlinear medium.

In the first case, the previous equations (1 to 5 without any electrical term), lead to the wave propagation equation

$$\begin{aligned}\rho_o \frac{\partial^2 u_j}{\partial t^2} = {} & C_{ijkl}\frac{\partial^2 u_k}{\partial a_i \partial a_l} + \gamma_{ijklmn}\frac{\partial}{\partial a_i}\left(\frac{\partial u_k}{\partial a_l}\frac{\partial u_m}{\partial a_n}\right) \\ & + \delta_{ijklmnpq}\frac{\partial}{\partial a_i}\left(\frac{\partial u_k}{\partial a_l}\frac{\partial u_m}{\partial a_n}\frac{\partial u_p}{\partial a_q}\right) + \dots\end{aligned} \qquad (6)$$

where γ_{ijklmn} and $\delta_{ijklmnpq}$ are effective nonlinear elastic constants of the 3rd and 4th orders.[11,14]

γ_{ijklmn} is a function of the 2nd and 3rd order constants (δ_{km} is the Kronecker symbol)

$$\gamma_{ijklmn} = \frac{1}{2} C_{ijnl}\,\delta_{km} + C_{inkl}\,\delta_{jm} + \frac{1}{2} C_{ijklmn} \qquad (7)$$

and $\delta_{ijklmnpq}$ is in addition function also of the 4th order constants

$$\begin{aligned}\delta_{ijklmnpq} = {} & \frac{1}{2} C_{iqnl}\,\delta_{km}\,\delta_{jp} + \frac{3}{4} C_{ijklqn}\,\delta_{pm} \\ & + \frac{1}{4} C_{ijqlmn}\,\delta_{kp} + \frac{1}{2} C_{iqklmn}\,\delta_{pj} + \frac{1}{6} C_{ijklmnpq}\end{aligned} \qquad (8)$$

Therefore the nonlinear behavior of the wave is entirely defined by these effective elastic constants.

In the present problem these nonlinearities are at the origin of harmonic generation,[11-14] with its consequences : amplitude-frequency effect[14-23] and intermodulation.[24-26]

Amplitude-frequency effect results from the interaction between 2nd and 3rd harmonics, which gives a correcting term at the fundamental frequency. It induces a change of the wave velocity, and thus for resonators a change of their resonance frequency. As shown in Fig. 1, the resonance frequency of a quartz crystal resonator depends on its driving level.

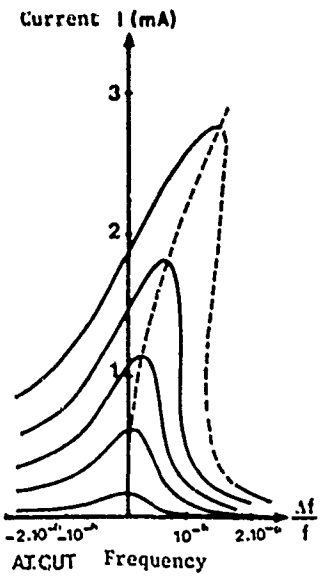


Fig. 1
Amplitude-frequency effect of a quartz crystal resonator due to 3rd and 4th order nonlinearities

For a one-dimensional resonator with thickness 2h and infinite lateral sizes, without energy trapping, the amplitude-frequency effect is represented by the relation

$$\frac{\Delta\omega}{\omega_o} = \frac{\pi^2 U_o^2}{128h^2}\left(-\frac{21\Gamma^2}{2\lambda^2} + \frac{9\Delta}{\lambda}\right) \qquad (9)$$

where U_o is the mechanical vibration amplitude. λ, Γ and Δ are the 2nd, 3rd, and 4th order effective elastic constants for the vibration mode and the crystal cut which are considered[14]

$$\lambda = C_{ijkl}\, N_i N_l \ell_j \ell_k$$
$$\Gamma = Y_{ijklmn}\, N_i N_l N_n \ell_j \ell_k \ell_m \qquad (10)$$
$$\Delta = \delta_{ijklmnpq}\, N_i N_l N_n N_q \ell_j \ell_k \ell_m \ell_p$$

where the N_i's represent the cosine directors of the propagation direction, defined by the crystal cut, and the ℓ_j's are the cosine directors of the mechanical polarization for the mode which is excited.

This shows that the amplitude-frequency effect depends not only on the 2nd and 3rd order fundamental elastic constants, but also on the 4th order ones. This is a serious difficulty for evaluating the amplitude-frequency effect, from such models, because the values of the 4th order fundamental elastic constants are almost completely unknowns, and even those of quartz crystal. The energy trapping and the resonator contour influence also the A-F behavior. Therefore measurements performed on resonators give access to effective nonlinear constants,[27] which can be evaluated by using appropriate models.[28] However the dispersion of the data, which is observed, indicated the presence of additional phenomena.[27]

For surface acoustic wave resonators the problem is different, because the contribution of the 4th order constants becomes negligible for a propagation length larger than the wave length. Therefore the intensity of the A-F effect can be calculated, as a function of the crystal anisotropy.[14,29-31]

Comparison has been made between BAW and SAW resonators as shown in Table I. The A-F effect is given in term of the k coefficient defined by the relation $\Delta f/f = kI^2$ where I is the intensity of the current through the resonator.

	cut	Frequency	Q-factor	A-F coef. k
SAW resonator	ST	110 MHz	26 000	$1.1\ 10^{-3}$ /A
BW resonator	AT	100 MHz	63 000	$2.5\ 10^{-1}$ /A
BW resonator	AT	5 MHz	2.10^6	$2\ 10^{-1}$ /A

Tableau I
Comparison of the amplitude-frequency effect of BAW and SAW quartz resonators (from Ref. 14)

Intermodulation corresponds to a mixing of frequencies when two (or more) signals at different but close frequencies are applied on a resonator. If the two frequencies, ω_1 and ω_2, are within the resonator bandwidth the cubic nonlinearities will generate two frequencies $2\omega_1-\omega_2$ and $2\omega_2-\omega_1$ (among the other ones like $2\omega_1+\omega_2$, $3\omega_1$, $3\omega_2$, $2\omega_2+\omega_1$, etc), which will be located also within the bandwidth, and therefore will not be filtered. The generation of these intermodulation frequencies involves as previously 2nd, 3rd, and 4th order fundamental elastic constants.[24-26] For the same reason, intermodulation cannot be evaluated from the models, but these last ones can be used to approximate values of some 4th order constants from the intermodulation measurements. This has be done with thickness shear quartz resonators[25] involving the C_{6666} constant, and also for X-cut lithium tantalate.[26]

Comparison between quartz and lithium tantalate indicates lower nonlinearities for this last one. Intermodulation measurements were also achieved on quartz SAW resonators and delay lines, which confirmed the A-F results.[29,30]

The second family of nonlinear elastic problems corresponds to the propagation of a small amplitude wave in a prestrained medium. The theoretical analysis is simplified by considering that the wave has no influence on the static deformation. Only the modifications of the wave characteristics by the predeformation are considered. This realistic hypothesis enables to linearize the equilibrium and constitutive equations, by using a power series development, generally limited at the 3rd order of nonlinearities.[3,32]

This linearization leads to a new wave equation which can be written

$$\rho_o \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial}{\partial a_s}\left[\overline{A_{iskr}}\frac{\partial u_k}{\partial a_r}\right] \qquad (11)$$

where $\overline{A_{iskr}}$ are effective elastic constants, which depend on the static deformation, through 2nd and 3rd order elastic constants.

$$\overline{A_{iskr}} = C_{iskr} + \overline{H_{iskr}} \qquad (12)$$

$\overline{C_{iskr}}$ are the regular 2nd order constants and $\overline{H_{iskr}}$ appear as perturbating terms

$$\overline{H_{iskr}} = \delta_{ik}\,\overline{T_{sr}} + C_{\ell skr}\frac{\overline{\partial U_i}}{\partial a_\ell} + C_{iskr}\frac{\overline{\partial U_k}}{\partial a_\ell} + C_{iskruv}\,\overline{S_{uv}} \qquad (13)$$

$\overline{T_{sr}}$, $\overline{S_{uv}}$ and $\overline{\partial U_i}/\partial a_\ell$ are the static stresses, strains and displacement gradients. If these quantities are homogeneous, and do not depend on the space variables, the equilibrium equation can be directly solved after replacing the regular 2nd order constants by the effective ones. If the $\overline{A_{iskr}}$'s are functionally dependant on the space variables, a perturbation method is used, which gives directly the velocity or frequency shifts, i.e. the sensitivity to the perturbation.

Since these problems involve only 2nd and 3rd order fundamental elastic constants, which are known for various materials, the models can be usefully applied to the determination of the sensitivities of resonators to different physical quantities like forces, pressures and accelerations. As it will be shown further the same procedure applies to problems involving electrical phenomena in piezoelectric materials.

As it can be seen from equation (13) the sensitivity to external perturbations depends on the crystal anisotropy. Therefore particular configurations can be expected to minimize the sensitivities when some compensation between the different terms of $\overline{H_{iskr}}$ occurs.

The problem of force-sensitivity was extensively studied.[33-52] In Fig. 2 is presented an example of the calculated sensitivities of bulk and surface wave resonators and delay lines to external diametrically applied forces, which shows the dependence of the sensitivity to the direction of the force application.

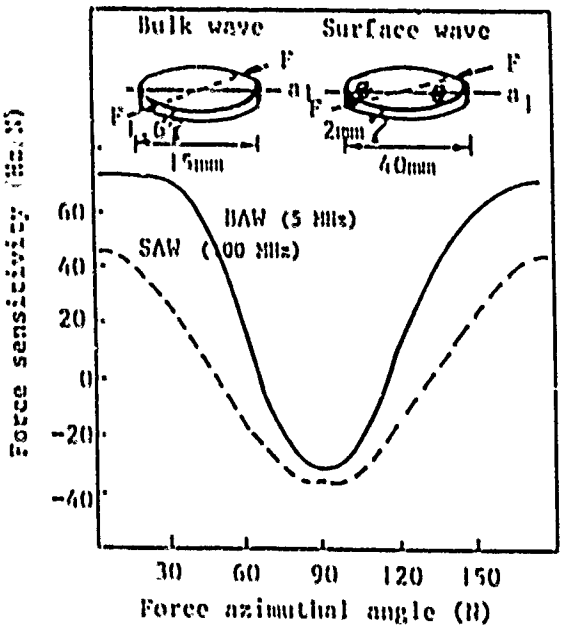


Fig. 2
Force-sensitivity of BAW and SAW resonators and delay lines

Sensitivity to hydrostatic pressure is an indirect problem for BAW resonators. The crystal resonator is in principle under vacuum in its enclosure. Therefore there is no direct effect of the surrounding gas on the vibrating crystal. However the gas pressure acts on the enclosure and deforms it. Forces are transmitted to the crystal by the mounting supports and a frequency shifts occurs due to the resulting stresses and strains. For regular quartz resonators sensitivities of the order of 10^{-9}/bar are observed. With appropriate mounting this sensitivity can be reduced by almost a factor 10.[53]

Configurations with larger values of the Π_{1skr} coefficient are also studied for pressure sensor applications.[54-62] An example of a SAW pressure sensor is shown in Fig. 3. This sensor is made of a thin diaphragm supporting one or two SAW delay lines or resonators located at the maxima of sensitivity on the substrate.

Acceleration sensitivity remains one of the most important problems. Under an acceleration field the quartz crystal is submitted to a system of body forces due to the applied field and reaction forces exerted by the supports.[63-66] Large efforts wer undertaken in order to reduce this g-sensitivity, because of its implication for spatial and tactical quartz oscillators. Reduction of the sensitivity from a few 10^{-9}/g to a few 10^{-10}/g was achieved on BAW resonators.[67-73] With SAW devices the situation was worse (10^{-7}/g)[74,75] but improvements also were obtained, and sensitivities of the order of a few 10^{-9}/g are obtainable today.[76,77]

Quartz accelerometers are also in progress.[78-80] But they do not use the g-sensitivity of the crystal resonator, which would be too low for a sensor. The acceleration field is transformed into a force, by means of an inertial mass, which then is applied to deform the crystal. In Fig. 4 is shown an example of a SAW cantilever beam accelerometer with its main and spurious sensitivities.[78]

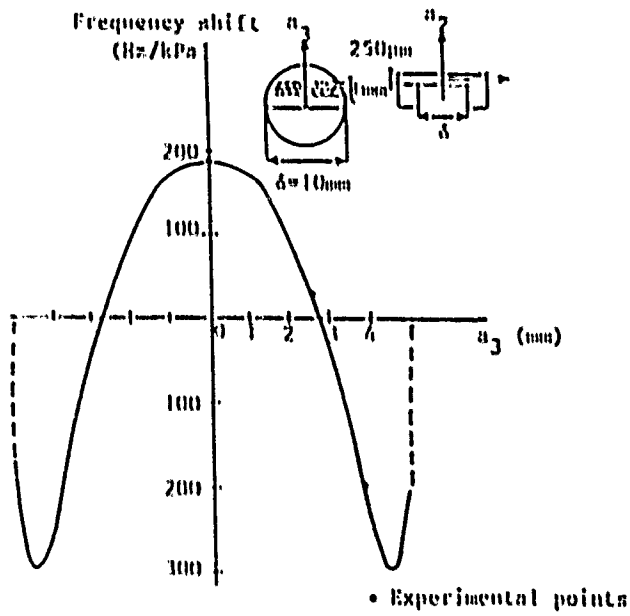
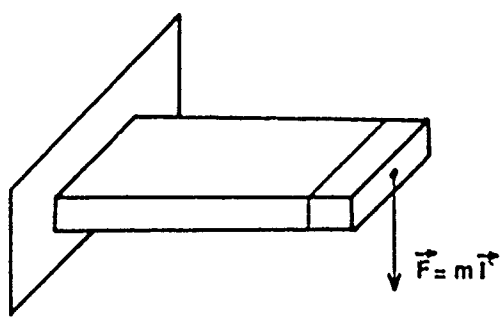


Fig. 3
SAW pressure sensor



SC-cut (ϑ = 0) at 100 MHz	Theoretical values (Hz/g)	Experimental values (Hz/g)
Main g-sensitivity	1387	1318
Transverse g-sensitivity	3.3×10^{-3}	no measurabl
Compression g-sensitivity	8	10

Fig. 4
SAW cantilever beam accelerometer

All these sensitivities to forces, pressures, or accelerations are related to effective 3rd order constants, which involve 2nd and 3rd order fundamental elastic constants.

The values of the 2nd order elastic constants are well known, and are available in the form of the compliance or of the stiffness constants, defined at constant electric field or constant electric displacement.

3rd order elastic constants of quartz has been measured by Thurston et al in 1965.[81] These values correspond to the only complet set which is available. There are 31 non zero 3rd order constants, and among them 14 are independent. The values of the 14 independent constants are given in table II.

Coeff.	Value	Standard error	Coeff.	Value	Standard error
C_{111}	-2.10	0.07	C_{134}	+0.02	0.04
C_{112}	-3.45	0.06	C_{144}	-1.34	0.07
C_{113}	+0.12	0.06	C_{155}	-2.00	0.08
C_{114}	-1.63	0.05	C_{222}	-3.32	0.08
C_{123}	-2.94	0.05	C_{333}	-0.15	0.18
C_{124}	-0.15	0.04	C_{344}	-1.10	0.07
C_{133}	-3.12	0.07	C_{444}	-2.76	0.17

Table II
Values of the 14th independent 3rd order fundamental elastic constants of quartz (in 10^{11} N/m^2)
from Ref. 81

The confidence one can have in these values is high according to the good agreement which was generally observed when comparing the calculated force, pressure, or acceleration sensitivities of the devices to the experimental values.

3rd order constants have also been measured for many materials to cubic, trigonal, tetragonal crystallographic systems. The values can be found in Tables of Ref. 82.

At the opposite 4th order elastic constants are almost completely unknown. For quartz crystal only $C_{6666} \approx 8\times10^{12}$ N/m^2,[25] $C_{1111} = 1.6\times10^{13}$ N/m^2, $C_{3333} = 1.8\times10^{13}$ N/m^2 [83] and some compliance constants[84,85] are known.

Temperature coefficients

The behavior of resonators submitted to temperature variations is generally described by means of thermal expansion coefficients and temperature coefficients of the elastic constants. These last ones have been measured by Bechmann and coauthors[86] from the temperature dependence of BAW quartz crystal resonator frequency. Such coefficients therefore are phenomenological coefficients, which can be used with confidence only for calculating the behavior of resonators of the same type and in the same conditions. They cannot be considered as fundamental constants at all. In fact they implicitly contain the contribution of nonlinear elastic effects due to the crystal deformation related with the thermal expansion.

The temperature derivative of the fundamental elastic constants were determined by Sinha[87] by substracting the contribution of the 3rd order elastic nonlinearities from Bechmann's coefficients. The values are given in Table III. More recently the same calculation was made by Lee and extended to the second order temperature derivatives.[88] These last ones are effective coefficients which are a combination of the fundamental 2nd temperature derivatives, of the temperature derivatives of the 3rd order elastic constants (unknown), and of the 4th order elastic constants (unknown).

pq	c^E_{pq} 10^9 N/M^2	$(1/c^E_{pq})dc^E_{pq}/dT$ $10^{-6}/°C$	α $10^{-6}/°C$
11	86.74	10.16	-
33	107.2	- 66.60	14.5
12	6.98	- 1 222	-
13	11.91	- 178.6	34.2
44	57.94	- 89.72	-
66	39.88	126.7	-
14	- 17.91	- 49.21	-

Table III
First order temperature derivatives of the fundamental elastic constants of quartz (from ref. 87)

The nonlinear elastic constants combined with the fundamental temperature derivatives coefficients of the resonator are necessary for describing nonuniformly heated resonators. When temperature is not uniform, temperature gradients induce thermal stresses and strains which, in addition of thermal expansion and temperature dependence of the fundamental elastic constants, contribute to the frequency shifts by nonlinear elastic effects. The so-called "dynamic temperature behavior" has been studied in details experimentally and theoretically[89-97] in BAW resonators and SAW devices.

The complete frequency-temperature dependence of resonators is summarized by the relation

$$\frac{f-f_o}{f_o} = a_o\,(T-T_o) + b_o\,(T-T_o)^2 + c_o\,(T-T_o)^3 + \tilde{a}\,\frac{dT}{dt} \qquad (14)$$

where a_o, b_o, c_o represent the 1st, 2nd and 3rd order elastic temperature coefficients of frequency, and $\tilde{a}$ the dynamic temperature coefficient. a_o, b_o and c_o generally are evaluated from the phenomenological Bechmann's constants. $\tilde{a}$ must be calculated by using the temperature derivative of the fundamental elastic constants and nonlinear elastic couplings involving the 3rd order elastic constants.

Since temperature gradients induce in-plane thermal stresses,[98] their compensation consists in determining a crystal orientation such that the two in-plane axial stress components cancell out. This was the definition given by Cernisse of SC-cut.[99]

Electroelastic and electrostrictive constants

The electroelastic constants are the 3rd order constants noted $e_{m,ijkl}$, which in the expression of internal energy couple one component of the electric field with two components of the strain tensor, corresponding to the term $1/2\ e_{m,ijkl}\,E_m\,\eta_{ij}\,\eta_{kl}$. Since the strain derivative of this quantity gives the stress tensor, it can be written

$$T_{ij} = e_{m,ijkl}\,E_m\,\eta_{kl} \qquad (15)$$

Therefore electroelastic effect appears as a nonlinear coupling between electric fields and mechanical strains. If, for instance, $\mathcal{E}_m$ is a static quantity noted $\bar{\mathcal{E}}_m$, and T_{ij} and η_{kl} dynamic quantities (noted $\tilde{T}_{ij}$ and $\tilde{\eta}_{kl}$) relation (15) becomes

$$\tilde{T}_{ij} = (e_{m,ijkl}\,\bar{\mathcal{E}}_m)\,\tilde{\eta}_{kl} \qquad (16)$$

which represents a linearized form of stress-strain relation ; $(e_{m,ijkl}\,\bar{\mathcal{E}}_m)$ in that case is equivalent to a correcting term of some 2nd order elastic constants. In this example electroelastic effect induces a direct change of the elastic constants under the application of a DC electric field.

The electrostrictive constants correspond in the energy to the coupling of two electric field components with one strain component in the term (1/2) $e_{mn,ij}\,\mathcal{E}_m\,\mathcal{E}_n\,\eta_{ij}$.

After strain derivation it follows

$$T_{ij} = e_{mn,ij}\,\mathcal{E}_m\,\mathcal{E}_n \qquad (17)$$

which is a stress component proportional to a quadratic form of the electric field. This property, called electrostriction, can be observed in most of the materials, and in particular in materials which are not piezoelectric.

Therefore the application of a DC field on a piezoelectric resonator introduces a rather complicate situation :

- if the DC field is applied in a direction for which the crystal is piezoelectric a deformation follows which can modify the plate thickness and the specific mass, and therefore the resonance frequency. This is a direct effect, but which occurs only for particular directions.

- then is the nonlinear coupling between the static strains and the high frequency wave through the third order nonlinearities as for the previous mechanical bias.

- electroelasticity also changes the resonance frequency by the direct modification of the elastic constants.

- in addition electrostriction introduces correcting terms in the stress-field relation, but electrostriction can be easily distinguished from electroelasticity because of its quadratic nature.

The electric-field dependence of quartz resonators was presented for the first time by Kusters.[100] As shown in Fig. 5, a fast frequency shift due to the previously described effects follows the application of the DC field. The relaxation phenomenon which appears after the initial frequency shift is due to ionic impurity migration in the crystal. The electroelastic effect was studied by Hruska,[101-103] who evaluated the values of linear combinations of electroelastic constants. Complete sets of the 8 independent constants (there are 23 nonzero constants) were determined by different authors. Kusters,[100] Brendel,[104] Hruska, Brendel[105] used frequency measurements of resonators under DC fields. Reider,[106] Kittinger[107,108] determined the same constants with a pulse echo method. The different values obtained by these authors are given in Table IV.

It must be noticed that not all of these values correspond to material constants, and that some are in fact phenomenological.

The agreement between the different values is not so satisfactory, and the reason of the discrepancies is not sufficiently understood yet.

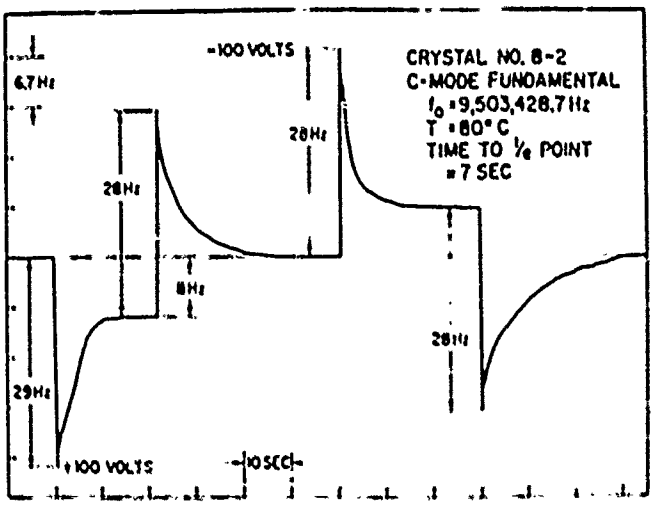


Fig. 5
Influence of a DC electric field on a quartz crystal resonator (from Ref. 24)

	Kusters Ref. 100	Brendel Ref. 104	Hruska Ref. 101	Hruska Brendel Ref. 105	Kittinger Tichy Friedel Ref. 107	Reider Kittinger Tichy Ref. 106
$e_{1,11}$	- 2.97	- 2.73	- 2.96	- 2.86	- 2.18	2.61
$e_{1,12}$	-49.05	-12			0.5	0.49
$e_{1,14}$	-28.01	33.4			- 0.28	0.39
$e_{1,22}$	-114.76	3.05			1.1	- 1.47
$e_{1,24}$	-85.49	34.6			- 0.78	1.13
$e_{3,14}$	2.12	50.5			- 1.63	1.35
$e_{3,44}$	57.77	- 0.6			- 0.05	- 0.15
$e_{3,15}$	70.72	24.9			0.9	- 0.64
$\mathcal{E}_{1,11}$			- 0.68	- 0.42		
$\mathcal{E}_{1,14}$			- 0.90	- 1.99		
$\mathcal{E}_{1,24}$			- 0.88	- 2.41		
$\mathcal{E}_{1,44}$			0.90	1.02		
$\mathcal{E}_{3,15}$			1.01	0.72		
	phenom.	material	phenom.	material	material	phenom.

$\mathcal{E}_{111} = e_{111} - 0.42\,e_{122} + 0.25\,e_{134}$
$\mathcal{E}_{114} = e_{114} - 0.26\,e_{122} - 0.68\,e_{134}$
$\mathcal{E}_{124} = e_{124} - 0.75\,e_{122} - 0.68\,e_{134}$
$\mathcal{E}_{144} = e_{144} + 0.5\,e_{122}$
$\mathcal{E}_{315} = e_{315} + 0.6\,e_{122} - 0.51\,e_{134}$

Tableau IV
Electroelastic constants of quartz crystal (in C/m^2)

The method used by Kittinger and coauthors also gave the electrostrictive constants, which are presented in Table V.

$\chi_{11.1}$	$\chi_{11.2}$	$\chi_{11.3}$	$\chi_{11.4}$	$\chi_{33.1}$	$\chi_{33.3}$	$\chi_{13.5}$	$\chi_{13.6}$
-4.8	1.1	10.2	-2.2	0.8	-3.9	-4.1	1.3

Tableau V
Electrostrictive constants of quartz crystal (in F/m) from Ref. 107

The electroelastic effect found also application in signal processing and in particular in the acoustic convolver. The principle of this device is shown in Fig. 6. Two signals $E_1(t)$ and $E_2(t)$ are applied on the two input transducers, which generate surface acoustic waves in opposite directions. These two waves are mixed by the substrate nonlinearities and the wave component with frequency $\omega_1+\omega_2$ and wave number k_1-k_2 is detected by the interdigital transducer at the middle of the plate. In particular if $\omega_1 = \omega_2$, then $k_1-k_2 = 0$; the nonlinear signal is uniformly distributed, and can be detected by replacing the output IDT by a simple electrode. The output signal is proportional to the convolution product of the input signals (this device is called degenerated convolver).

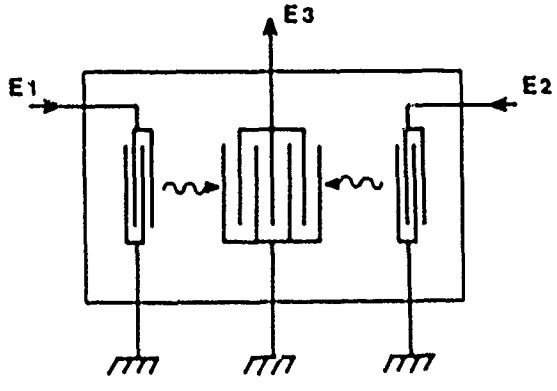


Fig. 6
Acoustic convolver

The theoretical analysis of the acoustic convolver was made by Ganguly.[109] The efficiency of the convolver can be improved by increasing the waves energy densities by means of acoustic wave guides. The guided wave convolver was studied by Planat and coauthors.[110,111]

The mixing of the two waves results from the elastic (3rd order constants), piezoelectric (electroelastic, electrostrictive), and dielectric (3rd order constants) nonlinearities. The detection of the convolution product is due to the 2nd order piezoelectric constants. Therefore for improving the efficiency, material with high electromechanical coupling factor, like lithium niobate, is used.

It was also proposed to use the electroelastic effect to create gratings in $LiNbO_3$ and $LiTaO_3$ with the interference electric field of two laser beams.[112]

The third order elastic constants of $LiNbO_3$ have been measured by Nakagawa.[113] The electroelastic constants are the more important ones for the convolver ; the 13th independent constants were measured by different authors, but it must be noticed that the different values do not always correspond to fundamental constants and rather to effective ones.[114-118] The electrostrictive constants were also evaluated and the 8 independent constants can be found in Ref. 109.

Dielectric constants

The nonlinear dielectric constants do not take a prominent part in the acousto-electric models of piezoelectric devices (at least in the radiofrequency range), since they essentially appear as correcting terms of the elastic and piezoelectric effects. However 3rd and 4th order dielectric constants of quartz have been measured.[85] Values of 3rd order constants of lithium niobate are also available. They were obtained from electro-optic coefficients.[109]

Electro-optic and elasto-optic constants

The electro-optic and elasto-optic interactions involve fundamental constants which already appeared in the previous sections of this paper.

Electro-optic effect corresponds to the interaction of a light beam, i.e. an electromagnetic wave at very high frequency with a low frequency electric field. Even if this last one is in the microwave range it can be considered as a DC field in comparison to the light frequency. Since the mathematical representation of the interaction links the modified light wave to the initial light wave multiplied by the electric field, the electro-optic effect must involve in the expression of internal energy (5) three components of the electric field and thus corresponds to the quantity (1/6) $e_{mnp} E_m E_n E_p$. Therefore the electro-optic constants are similar to the 3rd order dielectric constants. If the 4th order constants are considered, they will correspond to a quadratic electro-optic effect (Kerr effect)

When the crystal is piezoelectric, mechanical deformations occur which can be coupled with the light beam. The corresponding term in the energy involves two electrical components with one mechanical one. This is (1/2) $e_{mn.ij} E_m E_n \eta_{ij}$. The elasto-optic constants correspond at low frequency to the electrostrictive constants.

As previously pointed out some of the 3rd order dielectric and electrostrictive constants of $LiNbO_3$ were deduced this way.

A rigourous presentation of these different coefficients would have to take into account the differences between electro-optic constants defined at constant strains or at constant stresses.[119] A similar remark applies to the elasto-optic constants defined at constant electric field or electric displacement. These distinctions have not been made in this short presentation.

Conclusions

The nonlinear behavior of piezoelectric devices is described not only by the nonlinear fundamental constants, but rather by effective constants which involve 2nd order, 3rd order, and sometimes 4th order fundamental constants. This means that a crystal with zero

nonlinear fundamental constants -irrealistic hypothesis- would still exhibit nonlinear properties. On account of the important anisotropy of piezoelectric crystals, and the development today of doubly rotated cuts almost all the different independent constants are involved in the relations describing this or that effect. This requires that the values of all the constants must be known.

For quartz crystal the 3rd order elastic constants can be used with confidence, but :

the 4th order elastic constants are not known. This is a lack, which does not enable to determine orientations with lower amplitude-frequency effect or intermodulation, for instance.

the temperature derivatives of the 3rd order elastic constants are not known. These derivatives would be interesting for determining the influence of temperature on the sensitivity of force, pressure, or acceleration sensors, or the values of the 2nd order temperature derivatives of the fundamental elastic constants.

the electroelastic constants have been measured, but the disagreement between the different authors is too large.

Even in quartz, which is supposed to be best known crystal, there is still efforts to do in order to improve the knowledge of its nonlinear properties. For crystals like lithium niobate, lithium tantalate, there are more unknows, and "a fortiori" the situation is worse for other crystals.

Acknowledgements

The author is very gratefull to Drs R. Brendel and C. Bigler from LPLO-CNRS, who provided most of the tables of the nonlinear constants.

References

1. F.D. Murnaghan, "Finite deformation of an elastic solid", Wiley and Sons Inc., (1951).
2. R.A. Toupin, "The elastic dielectric", J. Ration. Mech. Anal., 5, 849 (1956).
3. R.N. Thurston, "Wave propagation in fluids and normal solids", Physical Acoustics, W.P. Mason Ed., vol. 1, part A, Academic Press (1964).
4. H.F. Tiersten, "On the nonlinear equations of thermo-electroelasticity, Int. J. Eng. Sci., 9, 587 (1971).
5. P.G. Klemens, "The thermal conductivity of dielectric solids at low temperatures", Proc. Roy. Soc. A 208, 108 (1951).
6. A.A. Maradudin, "Thermal expansion and phonon frequency shifts", Phys. Stat. Sol., 2, 1493 (1962).
7. A.J. Haris, "Interaction of sound waves with thermal phonons in dielectric crystals", Physical Acoustics, vol. VIII, 275, Academic Press, N.Y. (1971).
8. L. Landau, G. Rumer, Phys. Z. Sowjet Union, 11, 18 (1937).
9. A. Akhieser, "On the absorption of sound in solids", J. Phys. (USSR), 1 (1939).
10. R. Brugger, "Generalized Grüneisen parameters in the anisotropic Debye model", Phys. Rev., 137, n° 6A (1965).
11. H.F. Tiersten, J.C. Baumhauer, "Second harmonic generation of surface waves in isotropic elastic solids", Ultrasonics Symposium, IEEE cat. 73 CHO 8078SU (1973).
12. R.B. Thompson, H.F. Tiersten, "Harmonic generation of longitudinal elastic waves", J. Acoust. Soc. Am., 62, n° 1, 33 (1977).
13. P.J. Vella, G.I. Stegeman, U.M. Ristic, "Analysis of parametric mixing and harmonic generation of surface acoustic waves", J. Appl. Phys., 50, 81 (1979).
14. M. Planat, G. Théobald, J.J. Gagnepain, "Propagation non linéaire d'ondes élastiques dans un solide anisotrope"
    I Ondes de volume Onde Electrique, 60, n° 8-9, 33 (1980).
    II Ondes de surface Onde Electrique, 60, n° 11, 61 (1980).
15. H.F. Tiersten, "Analysis of nonlinear resonance in rotated Y-cut quartz thickness-shear resonators", Proc. of the 28th AFCS (1974).
16. H.F. Tiersten, "Analysis of nonlinear resonance in thickness-shear and trapped-energy resonators", J. Acoust. Soc. Am., 59, n° 4, 866 (1976).
17. V.E. Bottom et al. "Mounting techniques for improved heat dissipation in quartz crystal units", Contract DA 36-039 sc-5485, Final Report (1952).
18. A.W. Warner, "Design and performance of ultraprecise 2-5 mc quartz crystal units", Bell System Tech. Journal, 39, 1193 (1960).
19. D.L. Hammond, C. Adams, L. Cutler, "Precision crystal units", Proc. of the 17th AFCS (1963).
20. A. Seed, "Nonlinear effects in quartz resonators", 4th Int. Congress on Acoustics, Copenhagen (1962).
21. A. Smolarski, "Free oscillation-frequency method for investigating quartz resonators", Bull. Acad. Pol. Sci., XII, n° 6, 11 (1964).
22. J.J. Gagnepain, R. Besson, "Nonlinear effects in piezoelectric quartz crystals", Physical Acoustics, W.P. Mason Ed., vol. XI, 245, Academic Press (1975).
23. J.J. Gagnepain, J.C. Poncot, C. Pégeot, "Amplitude-frequency behavior of doubly rotated quartz resonators", Proc. 31st AFCS, 17 (1977).
24. H.F. Tiersten, "An analysis of intermodulation in rotated Y-cut quartz thickness-shear resonators", Proc. 28th AFCS (1974).
25. R.C. Smythe, "Intermodulation in thickness-shear resonators", Proc. 28th AFCS (1974).
26. M. Planat, G. Théobald, J.J. Gagnepain, P. Siffert "Intermodulation in X-cut lithium tantalate resonators", El. Lett., 16, n° 5, 174 (1980).
27. R.C. Smythe, P.E. Morley, "Experimental evaluation of the effective nonlinear elastic constant for trapped energy and contoured resonators", Proc. of the 39th AFCS (1985).
28. H.F. Tiersten, D.S. Stevens, "The evaluation of the coefficient of nonlinear resonance for SC-cut quartz resonators, Proc. of the 39th AFCS (1985).
29. M. Planat et al., "Nonlinear propagation of surface acoustic waves on quartz"; Proc. of the 34th AFCS, 255 (1980).
30. M. Planat et al., "Nonlinear characteristics of SAW grooved resonators", Ultrasonics Symposium, IEEE 80 CH 1602.2 (1980).

31. M. Planat, D. Hauden, "Nonlinear properties of bulk and surface acoustic waves in piezoelectric crystals", in Ferroelectricity and related phenomena, vol. 4, edited by G.W. Taylor, J.J. Gagnepain, T.R. Hecker, T. Nakamura and L.A. Shuvalov ; Gordon and Breach Science Publishers, N.Y., 1985, p. 277.

32. J.C. Baumhauer, H.F. Tiersten, "Nonlinear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., 54, N° 4, 1017 (1973).

33. A.D. Ballato, "Effects of initial stress on quartz plates vibrating in thickness modes, Proc. of the 14th AFCS, 89 (1960).

34. E.A. Gerber, "Reduction of frequency-temperature shift of piezoelectric crystals by application of temperature-dependent pressure", Proc. IRE, 48, 244 (1960).

35. E.A. Gerber, H.N. Miles, "Reduction of the frequency-temperature shift of piezoelectric resonators by mechanical stress", Proc. IRE, 49, n° 11, 1650 (1961).

36. R.W. Keyes, F.W. Blair, "Stress dependence of the frequency of quartz plates", Proc. IEEE, 55, n° 4, 565 (1967).

37. J.M. Ratajski, "The force sensitivity of AT-cut quartz crystals, Proc. of the 20th AFCS, 22 (1966)

38. J.M. Ratajski, "Force-frequency coefficient of singly-rotated vibrating quartz crystals", IBM J. Res. Dev., 12, n° 1, 92 (1968).

39. C.R. Dauwalter, "The temperature dependence of the force sensitivity of AT-cut quartz crystals, Proc. of the 26th AFCS, 108 (1972).

40. P.C.Y. Lee, Y.S. Wang, X. Markenscoff, "Elastic waves and vibrations in deformed crystal plates", Proc. of the 27th AFCS, 1 (1973).

41. P.C.Y. Lee, Y.S. Wang, X. Markenscoff, "Effects of initial bending on the resonance frequencies of crystal plates", Proc. of the 28th AFCS, 14 (1974)

42. P.C.Y. Lee, Y.S. Wang, X. Markenscoff, "High frequency vibrations of crystal plates under initial stresses", J. Acoust. Soc. Am., 57, n° 1, 95 (1975).

43. P.C.Y. Lee, Y.S. Wang, X. Markenscoff, 'Nonlinear effects of initial bending forces on the vibrations of crystal plates", J. Acoust. Soc. Am., 59, 90 (1976).

44. A. Ballato, E.P. Eernisse, T. Lukaszek, "The force-frequency effect in doubly-rotated-quartz resonators", Proc. of the 31st AFCS, 8 (1977).

45. E.P. Eernisse, T. Lukaszek, A. Ballato, "Variational calculation of force-frequency constants of doubly rotated quartz resonators", IEEE Trans. Sonics and Ultrasonics, SU-25, 132 (1978).

46. D. Janiaud, L. Nissim, J.J. Gagnepain, "Analytical calculation of initial stress effects on anisotropic crystals : application to quartz resonators", Proc. of the 32nd AFCS, 169 (1978).

47. E.P. Eernisse, "Rotated X-cut quartz resonators for high temperature applications", Proc. of the 32nd AFCS, 255 (1978).

48. A. Ballato, "Force-frequency compensation applied to four-point mounting of AT-cut resonators", IEEE Trans. Sonics and Ultrasonics, SU-25, 233 (1978).

49. E.P. Eernisse, "Temperature dependence of the force-frequency effect for the rotated X-cut", Proc. of the 33rd AFCS, 300 (1979).

50. E.D. Fletcher, A.J. Douglas, "A comparison of the effects of bending moments on the vibrations of AT and SC (or TTC) cuts of quartz", Proc. of the 33rd AFCS, 346 (1979).

51. P.C.Y. Lee, H.S.H. Tang, "Initial stress field and resonance frequencies of incremental vibrations in crystal resonators by finite element method, Proc. of the 40th AFCS (1986).

52. H. Hizata, A. Ballato, "The stress coefficient of frequency of quartz plate resonator", Proc. of the 37th AFCS (1983).

53. R. Delaite, "Résonateurs à quartz pour environnement sévère", Rev. Phys. Appl., 741 (1985).

54. D.L. Hammond, A. Benjaminson, "The crystal resonator, a digital transducer", IEEE Spectrum, 6, n° 4, 53 (1969).

55. E. Karrer, R. Ward, "A low-range quartz resonator pressure transducer", ISA Trans., 16, n° 2, 90 (1977).

56. J.F. Dias, E. Kat.er, "Stress effects in acoustic surface wave circuits and applications to pressure and force transducers", IEEE Int. Solid-state Conf., 166 (1974).

57. D. Hauden, S. Rousseau, J.J. Gagnepain, "Sensitivities of SAW oscillators to temperature, forces and pressure", Proc. of the 34th AFCS, 312 (1980).

58. T.M. Reeder, D.E. Cullen, H. Gilden, "SAW oscillator pressure sensors", 1975 Ultrasonics Symp., IEEE 75 CHO 994-4SU.

59. S. Rousseau, "Sensibilité à la pression de lignes à ondes de surface : étude d'un capteur à quartz", Thesis n° 104, Besançon, LPHO (1980).

60. D. Hauden, M. Planat, J.J. Gagnepain, "Nonlinear properties of SAW : applications to oscillators and sensors", IEEE Trans. Sonics and Ultrasonics, vol. SU-28, n° 5 (1981).

61. B.K. Sinha, Propagation characteristics of surface waves in quartz and their influence on device performance, 1985 Ultrasonics Symp., IEEE 85 CH 2209-5.

62. R.M. White, "Surface acoustic wave sensors", 1985 Ultrasonics Symp., IEEE 85 CH 2209-5.

63. M. Valdois, J. Besson, J.J. Gagnepain, "Influence of environment on a quartz resonator", Proc. of the 28th AFCS, 19 (1974).

64. P.C.Y. Lee, Kuang-Ming Wu, "Effects of acceleration on the resonance frequencies of crystal plates", Proc. of the 30th AFCS, 1 (1976).

65. P.C.Y. Lee, Kuang-Ming Wu, "The influence of support-configuration on the acceleration sensitivity of quartz resonator plates, Proc. of the 31st AFCS, 29 (1977).

66. D. Janiaud, "Modélisation de l'influence d'une accélération sur la fréquence des résonateurs à quartz", Thesis report. ONERA, Chatillon/Bagneux, March (1978).

67. R. Besson et al., "Design of a bulk wave quartz resonator insensitive to acceleration", Proc. of the 33rd AFCS, 337 (1979).

68. J.J. Gagnepain, F.L. Walls, "Quartz crystal oscillators with low acceleration sensitivity", Techn. Report. NBSIR 77-855.

69. A. Ballato, "Resonators compensated for acceleration fields", Proc. of the 33rd AFCS, 322 (1979).

70. J.M. Przyjemski, "Improvement in system performance using a crystal oscillator compensated for acceleration sensitivity", Proc. of the 32nd AFCS, 426 (1978).

71. D.A. Emmons, "Acceleration sensitivity compensation in high performance crystal oscillators, Proc. 10th PTTI, Nasa Techn. Memo. 80250, 55 (1978).

72. A. Debaisieux, J.P. Aubry, J. Groslambert, "Design of SC-cut 10 MHz U.O. crystal with G-sensitivity better than 2.10^{-10}/G", Proc. 15th Ann. PTTI Meeting (1983).

73. R.D. Weglein, "The vibration sensitivity of VHF quartz crystals for missile applications", Proc. of the 38th AFCS (1984).

74. M. Valdois, P. Levesque, P. Hartemann, "Acceleration dependence of the surface acoustic wave oscillator frequency", 1977 Ultrasonics Symp., Proc. IEEE 77 CH 1264-1SU, 936.

75. P. Levesque et al., "Theoretical and experimental analysis of SAW quartz oscillator acceleration sensitivity", 1979 Ultrasonics Symp., Proc. IEEE 79 CH 1482-9, 896.

76. G.K. Montress, T.C. Parker, J. Callerame, "A miniature hybrid circuit SAW oscillator using an all quartz package resonator", 1985 Ultrasonics Symp., IEEE cat. n° 85 CH 2209-5.

77. D.V. Shick, H.F. Tiersten, "An analysis of the acceleration sensitivity of ST-cut quartz surface wave resonators supported along the edges", Proc. of the 40th AFCS, 1986.

78. D. Hauden, F. Bindler, R. Coquerel, "SAW cantilever beam accelerometer sensitivities", 1985 Ultrasonics Symp., IEEE cat. n° 85 CH 2209-5.

79. P. Hartemann, P.L. Meunier, "Tensioned or flexured SAW accelerometers", 1983 Ultrasonics Symp., IEEE cat. n° 83 CH 1947-1.

80. M.J. Kass, G.S. Snow, "Double-ended tuning fork quartz accelerometer", Proc. of the 40th AFCS (1986).

81. R.N. Thurston, J.J. McSkimin, P. Andreatch, "Third order elastic coefficients of quartz", J. Appl. Phys., 37 (1966).

82. Landolt-Börnstein, "Elastic, piezoelectric and related constants of crystals, III, vol. 11, Springer-Verlag (1979).

83. R. Fowles, Geophys. Res., 72, 5729 (1967).

84. R. Besson, "Measurement of nonlinear elastic, piezoelectric, dielectric coefficients of quartz crystal, Proc. of the 28th AFCS (1974).

85. J.J. Gagnepain, R. Besson, "Nonlinear effects in piezoelectric quartz crystals, Physical Acoustics, vol. XI, Academic Press (1975).

86. R. Bechman, A.D. Ballato, T.J. Lukaszek, "Higher-order temperature coefficients of the elastic stiffness and compliances of alpha-quartz", Proc. IRE, 50, n° 8, 1812 (1962).

87. B.K. Sinha, H.F. Tiersten, "Temperature derivatives of the fundamental elastic constants of quartz", Proc. of the 32nd AFCS, 150 (1978).

88. P.C.Y. Lee, Y.K. Yong, "Temperature derivatives of elastic stiffnesses derived from the frequency-temperature behavior of quartz plates, Proc. of the 33rd AFCS (1983).

89. R. Holland, "Non uniformly heated anisotropic plates :
I Mechanical distoration and relaxation", IEEE Trans. Sonics and Ultrasonics SU-21, n° 3 (1974).

90. R. Holland, "Non uniformly heated anisotropic plates :
II Frequency transients in AT and BT quartz plates", 1974 Ultrasonics Symp., IEEE cat 74 CHO 896-1SU.

91. J. Kusters, "Transient thermal compensation for quartz resonators", IEEE Trans. Sonics and Ultrasonics, SU-23, 273 (1976).

92. J.A. Kusters, J.G. Leach, "Further experimental data on stress and thermal gradient compensated crystals", Proc. IEEE, 202 (1977).

93. A. Ballato, J.R. Vig, "Static and dynamic frequency-temperature behavior of singly and doubly rotated oven-controlled quartz resonators", Proc. of the 32nd AFCS, 180 (1978).

94. A. Ballato, "Static and dynamic behavior of quartz resonators", IEEE Trans. Sonics and Ultrasonics, SU-26, n° 4, 299 (1979).

95. G. Théobald et al., "Dynamic thermal behavior of quartz resonators", Proc. of the 33rd AFCS (1979).

96. B.K. Sinha, H.F. Tiersten, "Transient thermally induced frequency excursions in doubly-rotated quartz thickness mode resonators", Proc. of the 34th AFCS, 391 (1980).

97. D. Hauden, G. Théobald, "Dynamic thermal sensitivity of SAW quartz oscillators" 1980 Ultrasonics Symp., 264, IEEE 80 CH 1602.2.

98. J.P. Valentin, G. Théobald, J.J. Gagnepain, "Temperature induced frequency shifts in quartz resonators, J. of Appl. Phys., 58(3) (1985).

99. E.P. Eernisse, "Calculations on the stress compensated (SC-cut) quartz resonator, Proc. of the 39th AFCS (1985).

100. J. Kusters, "The effect of static electric fields on the elastic constants of a quartz", Proc. of the 24th AFCS (1970).

101. C.K. Hruska, "The electroelastic tensor and other second order phenomena in quasilinear interpretation of the polarizing effect with thickness vibrations of α-quartz plates", Proc. of the 31st AFCS, 159 (1977).

102. C.K. Hruska, "Second order phenomena in α-quartz and the polarizing effect with plate of orientation (xzl)Ψ", IEEE Trans. on Sonics and Ultrasonics, vol. SU-25, n° 4, 198-203 (1978).

103. C.K. Hruska, "Polarizing effect with doubly rotated α-quartz plates vibrating in thickness", IEEE Trans. on Sonics and Ultrasonics, vol. SU-25, n° 6, 390-392 (1978).

104. R. Brendel, "Material nonlinear piezoelectric coefficients for quartz", J. of Appl. Phys., vol. 54, n° 9, 5339-5346 (1983)
vol. 55, n° 2, 608 (1984).

105. C.K. Hruska, R. Brendel, "The material electroelastic constants of quartz determined by the resonator method", 1er Forum Européen Temps-Fréquence, Besançon, 1987.

106. G.A. Reider, E. Kittinger, J. Tichy, "Electroelastic effect in α-quartz", J. of Appl. Phys., vol. 53, n° 12, 8716-8721 (1982).

107. E. Kittinger, J. Tichy, W. Friedel, "Nonlinear piezoelectricity and electrostriction of α-quartz", J. of Appl. Phys., vol. 60, n° 4, 1465-1471 (1986).

108. E. Kittinger, J. Tichy, "Mapping of the electroelastic effect in quartz based on fundamental material constants", 1er Forum Européen Temps-Fréquence, Besançon (1987).

109. A.K. Ganguly, K.L. Davis, "Nonlinear interaction in degenerate surface acoustic wave elastic convolvers", J. of Appl. Phys., 51 (1980).

110. M. Planat, G. Vanderbeck, H. Gauthier, C. Haerfeld, "Linear and nonlinear analysis of a piezoelectric waveguide convolver using a finite element method, 1982 Ultrasonics Symp., IEEE cat. 82 CH 1823-4.

111. M. Planat, G. Vanderbeck, H. Gauthier, C. Haerfeld, "Calculation of the figure of merit of a piezoelectric waveguide", 1983 Ultrasonics Symp., IEEE cat. 83 CH 1947-1.

112. P.G. Gottschalk, D.E. Oates, P.V. Wright, "Measurement of electroacoustic coefficients in $LiNbO_3$ and $LiTaO_3$ and application to signal-processing devices", 1983 Ultrasonics Symp., IEEE cat. 83 CH 1947-1.

113. Y. Nakagawa, K. Yamanouchi, K. Shibayama, "Third order elastic constants of lithium niobate", J. of Appl. Phys., 44 (1973).

114. R.B. Thomson, C.F. Quate, "Nonlinear interaction of microwave electric fields and sound in $LiNbO_3$" J. of Appl. Phys., 42(3) (1971).

115. A.I. Korobov, V.C. Lyamov, "Nonlinear coefficients of $LiNbO_3$", Sov. Phys. Sol. State, 17(5) (1975).

116. B.A. Agishev, V.V. Lemanov, N.K. Yushni, "Electroacoustic coefficients of lithium tantalate and niobate crystals", Sov. Phys. Sol. State, 20(9) (1979).

117. S.I. Chizhikov, N.G. Sorokni, U.S. Petrokov, "The elastoelectric effect in the noncentrosymmetric crystals", Ferroelectrics, 41 (1982).

118. A. Palma, L. Palmieri, G. Socino, E. Verona, "Nonlinear electroacoustic effect in Lamb wave propagation in $LiNbO_3$ plates", 1984 Ultrasonics Symp., IEEE cat 84 CH 2112-1.

119. W.P. Mason, "Crystal physics of interaction processes", Academic Press (1966).

ACCELERATION EFFECT ON THE THICKNESS VIBRATIONS OF DOUBLY ROTATED CRYSTAL RESONATORS

P. C. Y. Lee and N. S. H. Tang
Department of Civil Engineering
Princeton University
Princeton, NJ 08544

Summary

Changes of resonance frequencies of thickness vibrations in doubly rotated circular disks of quartz due to steady accelerations as body forces are studied. The plate is of uniform thickness and supported at the edge by flexible metal ribbons with rigid contact.

The initial fields of displacement, strain, and stress, caused by the reactions of the ribbon supports to the body force of arbitrary direction, are obtained by a finite element method based on Mindlin's two-dimensional, first-order equations of equilibrium for crystal plates.

The frequency equation for incremental thickness vibrations superposed on the initial deformations is obtained from the previously derived two-dimensional equations of motion for incremental vibrations. In the frequency equation, the initial strain and initial deformation tensors appear as given functions of x_1 and x_3 of the plate. Then, by a perturbation method, changes of thickness frequencies as functions of the direction of acceleration are computed for SC-cut circular quartz resonators. The effects of plate thickness, support configurations, and support structure on the acceleration sensitivity are studied.

1. Introduction

Effects on frequency changes in crystal resonators due to applied forces or accelerations have been studied since the early 1970's as small vibrations superposed on finite deformations induced by forces or accelerations at an initial state.[1]

Three-dimensional equations of finite elasticity in Lagrangian formulation, including nonlinear anisotropic stress-strain relations, are employed as the governing equations for fields at both the initial and final states. By taking the differences of the field equations corresponding to the final and initial states, three-dimensional linear equations for small vibrations superposed on initial deformations are obtained. By expanding initial and incremental displacements in a power series of the thickness coordinate and by Mindlin's[2] general procedure, two-dimensional nonlinear governing equations for initial fields and two-dimensional linear equations for incremental vibrations are obtained for crystal plates.[1]

These equations are employed to study the force-sensitivity of circular crystal resonators subject to a pair of in-plane, diametral forces. Predicted results for rotated Y-cuts[1] and doubly rotated cuts[3,11] are in good agreement with experimental results.[4,5,6]

Force sensitivity of circular cantilever plates subjected to transverse loading was studied for rotated Y-cuts[7] and then for doubly-rotated cuts.[8] Predicted results are reasonably close to experimental data for AT-[5] and SC-cuts[9], but less so for BT-cuts.[6]

Acceleration sensitivity of circular disks, with three- and four-point mount, subject to in-plane accelerations was studied for rotated Y-cuts[11] and then extended to doubly rotated cuts.[11] Predicted result for disks with three-point mount had good agreement with experimental result, but no consistent agreement between the predicted and observed results for disks with four-point mount.[12,13]

Effect on frequency shift due to transverse accelerations in circular quartz disks has not yet been studied analytically so far as we know.

These above-mentioned analytical studies have contributed valuable information and fundamental understanding of the force and acceleration effect on the vibrations of crystal resonators. However, these chosen problems had to be limited to simple and idealized cases, such as plate being of uniform thickness and the connection of the plate to the metal ribbon support assumed being simply-supported. In obtaining analytical solutions, further mathematical approximations were introduced, such as replacing the anisotropic initial stress field in a crystal disk by the stress field obtained from an isotropic disk.

To remove some of these limitations and approximations, the finite element method has been employed to develop a computer code for solutions of Mindlin's six two-dimensional first-order plate equations for initial stresses.[14] Force sensitivity coefficients for circular disks subject to diametral forces are recalculated based on the initial fields obtained by the FEM code. Predicted results match the experimental data very closely for all the measured doubly rotated cuts.[15] The improvement is mainly due to the removal of the isotropic approximation of initial stress field.

To establish the accuracy and flexibility of the numerical-analytical approach further, changes in frequencies in the circular cantilever plates subject to transverse loading are also recalculated. Predicted results are closer to the experimental data[8,9] than the previously calculated results for both AT- and SC-cuts.[16]

In the present paper, acceleration sensitivity of SC-cut quartz resonators is studied systematically for various plate and support parameters as described in the subsequent sections.

2. Four-Point Mount SC-Cut Resonators

We consider SC-cut ($\theta = 34^\circ$, $\varphi = 21.93^\circ$) circular quartz resonators with diameter $d = 14.0$ mm and thickness $2b = 1.1469$ mm (fundamental). The disk is supported by four metal ribbons, 90° apart, with two located on the x_1 axis and the other two on the x_3 axis. The ribbon is made of molybdenum ($E = 27.579 \times 10^9$ dynes/mm^2, $\nu = 0.32$) with length $\ell = 1.143$ mm and rectangular cross section ($t = 0.0177$ mm, $w = 1.524$ mm). The contact condition between the plate and the ribbon is assumed to be "rigid", i.e., it can transmit both moments and forces.

We denote the resonance frequency of a resonator without any stress bias by f_0, and the resonance frequency under 1 g acceleration in the x_i direction ($i = 1,2,3$) by f_i. The we define

CH2427-3/87/0000-277 $1.00 C 1987 IEEE

$$\Gamma_i \equiv \frac{f_i - f_0}{f_0} = \frac{\Delta f_i}{f_0}$$

$\Gamma_{\parallel} = \sqrt{\Gamma_1^2 + \Gamma_3^2}$ ... lateral acceleration sensitivity (per g)

$\Gamma_{\perp} = |\Gamma_2|$ ... thickness acceleration sensitivity (per g)

$$\Gamma = \sqrt{\Gamma_1^2 + \Gamma_2^2 + \Gamma_3^2}$$

$$= \sqrt{\Gamma_{\parallel}^2 + \Gamma_{\perp}^2} \quad \text{... (total) acceleration sensitivity (per g).} \qquad (1)$$

In the finite element method, a circular disk modelled by a mesh of 228 elements is shown in Fig. 1.

We denote the direction of the body force $\underset{\sim}{\Pi}$ due to acceleration in the x_1x_3 plane by the azimuth angle ψ_2, a right-hand rotation about the x_2 axis starting from the x_1 axis. Similarly, the direction of a body force in the x_2x_3 plane is denoted by ψ_1, a right-hand rotation about the x_1 axis starting from the x_2 axis.

Changes of frequencies ($\Delta f/f_0 \times 10^{-10}$/g) as functions of ψ_2 and ψ_1, for vibrational modes a (thickness stretch), b (thickness-shear in x_3) and c(thickness-shear in x_1) are computed and shown in Figs. 2 and 3, respectively. In Fig. 2, the measured values for mode c by Filler, Kosinski, and Vig are also shown for comparison.[17]

From Figs. 2 and 3, we obtain, according to the definitions given in (1), the acceleration sensitivities for mode c: $\Gamma_{\parallel} = 2.85 \times 10^{-10}$/g, $\Gamma_{\perp} = 4.85 \times 10^{-12}$/g, $\Gamma = 2.85 \times 10^{-10}$/g. These values are comparable to the measured values: $\Gamma_{\parallel} = 3.62 \times 10^{-10}$/g, $\Gamma_{\perp} = -1.56 \times 10^{-11}$/g, $\Gamma = 3.62 \times 10^{-10}$/g.[17]

3. Six-Point and Three-Point Mounts

To examine the effect of different support configurations on the acceleration sensitivity, two additional mounting arrangements are considered: (1) Six-point mount (60° apart), and (2) Three-point mount (a four point mount with the support at the $+x_1$ axis removed).

The $\Delta f/f_0$ vs. ψ_2 curves for the six-point mount and three-point mount are shown in Figs. 4 and 5, respectively. For comparison, values of acceleration sensitivities of mode c for these three support configurations are listed below. From this point on in this paper, acceleration sensitivity shall mean acceleration sensitivity in mode c, unless it is specifically indicated otherwise.

Support Configuration	$\Gamma_{\parallel}$	$\Gamma_{\perp}$	Γ (in 10^{-10}/g)
Six-point	0.55	0.09	0.55
Four-point	2.55	0.05	2.55
Three-point	7.50	1.27	7.60.

4. Acceleration Sensitivity versus Mount Orientation

Let the orientation of a four-point and 90° apart mount configuration be denoted by angle α, a right-hand rotation angle about the x_2 axis starting from the x_1 axis.

The $\Delta f/f_0$ vs. ψ_2 curves are computed for a series values of α and are given in Fig. 6. And total acceleration sensitivity Γ as a function of α is shown in Fig. 7. It may be seen in Fig. 7 that Γ has the lowest value at $\alpha = 15°$ but varies little between $\alpha = 0°$ to $\alpha = 45°$, and it has the highest value at $\alpha = 75°$. The predicted values of Γ for $\alpha = 0°$ to $45°$ appears to represent the lower bound of the measured values of Filler, Kosinski and Vig in Fig. 8 of Ref. 18.

5. Acceleration Sensitivity versus Plate Thickness

Acceleration sensitivities $\Gamma_{\parallel}$, $\Gamma_{\perp}$, and Γ for resonators with four-point mount and $\alpha = 0$ are computed for a series of values of plate thickness 2b. The results are shown in Fig. 8, in which the measured values, given in Fig. 4 of Ref. 18, are also included for comparison.

It may be seen in Fig. 8 that for symmetrically placed four-point mount $\Gamma_{\parallel}$ is approximately linearly proportional to the thickness of the plate and $\Gamma_{\perp}$ is approximately inversely proportional to the plate thickness. However, if one of the supports is not aligned along one of the orthogonal axes, for instance, it is off the $+x_1$ axis by an angle $\Delta\alpha_1 = 7.18°$, the values of $\Gamma_{\perp}$ are no longer predominant for very small thickness of the plate. Hence the total acceleration sensitivity Γ is dominated by $\Gamma_{\parallel}$ and therefore is practically proportional to the plate thickness as shown in Fig. 9.

6. Acceleration Sensitivity Affected by Support Structure

A resonator with four-point mount and $\alpha = 0$ is shown schematically in Fig. 10(a). If the lower end of a ribbon support is displaced by a distance $\Delta = 10^{-3}$ mm in the x_1 direction from its stress-free position, stresses and strains will be induced throughout the plate and supports. The convention for the forces and moments induced at the other end of the support is depicted in Fig. 10(b). We denote the changes of frequency, caused by support displacement $u_1^{(0)} = \Delta = 10^{-3}$ mm, by $\Sigma_1 \equiv \Delta f/f_0$ which shall be called displacement sensitivity. Letter symbols representing the dimensions of a general support structure are shown in Fig. 10(c). By varying the values of these dimensions, five types of support structure are defined as listed below.

Support Type	ℓ	ℓ_1	ℓ_2	w	w_1 (in cm)
A	1.143	0	0	1.524	0
B	1.143	0	0.2	1.524	0
C	1.143	0.2	0.2	1.524	0.505
D	1.143	0.4	0	1.524	0.505
E	1.143	0.4	0	0.761	0.505

The acceleration sensitivity Γ, induced force F_i and moments M_i ($i = 1,2,3$), and displacement sensitivity Σ_1, Σ_2, Σ_3 are computed for the above listed five types of support structure and are shown inTable I.

We note that support type A is the one being employed in Sections 2 - 5 of the present paper. It may be seen from Table I that support type E has the lowest combined acceleration sensitivity and displacement sensitivities.

Acknowledgement

We wish to express our deep appreciation to Prof. J. H. Prevost of Princeton University for valuable advice and assistance in carrying out finite element method for computations.

This work was supported by the U.S. Army Research Office, Contract No. DAAG29-84-0133.

References

1. P. C. Y. Lee, Y. S. Wang and X. Markenscoff, Proc. 27th Ann. Freq. Cont. Symp., pp. 1-6, 1973, also J. Acoust. Soc. Am., Vol. 57, pp. 95-105, 1975.
2. R. D. Mindlin, "An Introduction to the Mathematical Theory of Vibrations of Elastic Plates," U.S. Army Signal Corps Engineering Laboratories, Fort Monmouth, NJ, 1955.
3. P. C. Y. Lee and K. H. Wu, Proc. 34th Ann. Freq. Control Symp., pp. 403-411, 1980.
4. A. D. Ballato, Proc. 14th Ann. Freq. Cont. Symp., pp. 89-114, 1960; also A. D. Ballato and R. Bechmann, IRE, 48, pp. 261-262, 1960.
5. C. R. Mingins, L. C. Marcus, and R. W. Perry, Reports 1-20, Lowell Tech. Inst. Found., 1961-1966.
6. A. Ballato, E. P. Eernisse, and T. Lukaszek, Proc. 31st Ann. Freq. Cont. Symp., pp. 8-16, 1977.
7. P. C. Y. Lee, Y. S. Wang, and X. Markenscoff, Proc. 28th Ann. Freq. Cont. Symp., pp. 14-18, 1974; also J. Acoust. Soc. Am., 59, pp. 90-96, 1976.
8. P. C. Y. Lee and C. S. Lam, Proc. 36th Ann. Freq. Cont. Symp., pp. 29-36, 1982.
9. E. D. Fletcher and A. J. Douglas, Proc. 33rd Ann. Freq. Cont. Symp., pp. 346-350, 1979.
10. P. C. Y. Lee and K. H. Wu, Proc. 30th Ann. Freq. Cont. Symp., pp. 1-7, 1976, also P. C. Y. Lee, K. H. Wu, and Y. S. Wang, J. Acoust. Soc. Am., 63(4), pp. 1039-1047, 1978.
11. P. C. Y. Lee and K. H. Wu, J. Acoust. Soc. Am., 75(4), pp. 1105-1117, 1984.
12. A. W. Warner, Interim Reps. 10-11, Bell Telephone Lab, 1959. And W. L. Smith, Interim Reps. 12-13 and Final Rep., Bell Telephone Lab., 1960.
13. B. Goldfrank and A. Warner, Research and Development Tech. Rep. DELET-TR-79-0272-3, U.S. Army Electronics Research and Development Command, Fort Monmouth, NJ, 1981.
14. P. C. Y. Lee and H. S. H. Tang, Proc. 30th Ann. Freq. Cont. Symp., pp. 152-160, 1966.
15. P. C. Y. Lee and H. S. H. Tang, Research Report SM-5-86, Dept. of Civil Engineering, Princeton University, 1986. Also accepted for publication by IEEE Trans. on Ultrasonics, Ferroelectrics and Frequency Control, (Nov. 1987).
16. P. C. Y. Lee and H. S. H. Tang, presented at the 41st Ann. Freq. Cont. Symp., 1987.
17. R. L. Filler, J. A. Kosinski and J. R. Vig, Two-g Tip-over Test Curves for Bendix-3, CR-692, private communication.
18. R. L. Filler, J. A. Kosinski and J. R. Vig, Proc. 37th Ann. Freq. Cont. Symp., pp. 265-271, 1983.

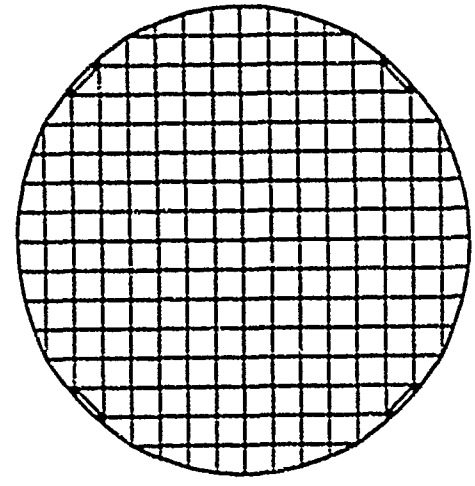


Fig. 1 A circular plate modelled by a mesh with 228 elements.

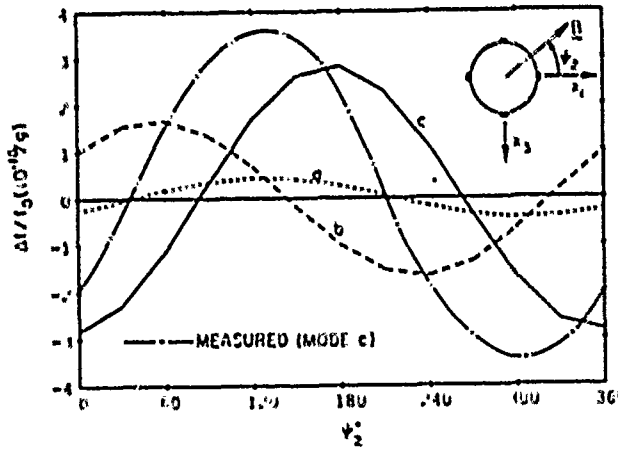


Fig. 2 $\Delta f/f_0$ vs. ψ_2 for four-point mounted SC-cut resonator.

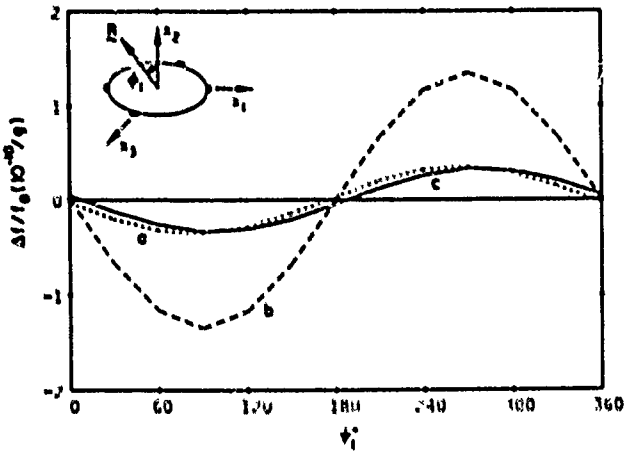


Fig. 3 $\Delta f/f_0$ vs. ψ_1 for four-point mounted SC-cut resonator.

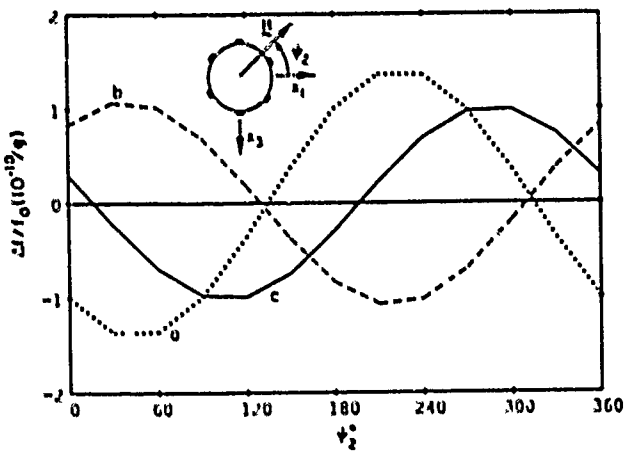


Fig. 4 $\Delta f/f_0$ vs. ψ_2 for six-point mounted SC-cut resonator.

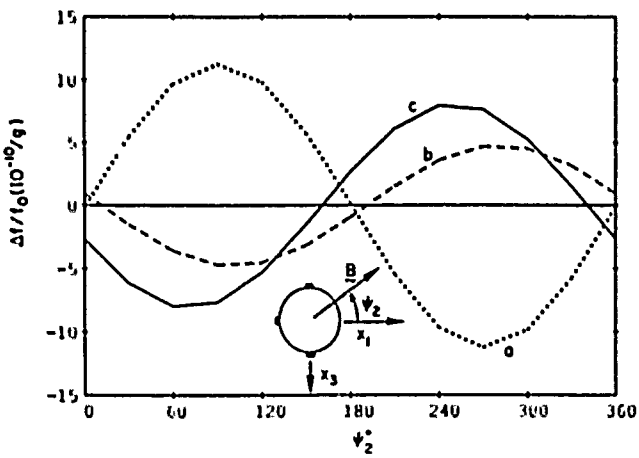


Fig. 5 $\Delta f/f_0$ vs. ψ_2 for three-point mounted SC-cut resonator.

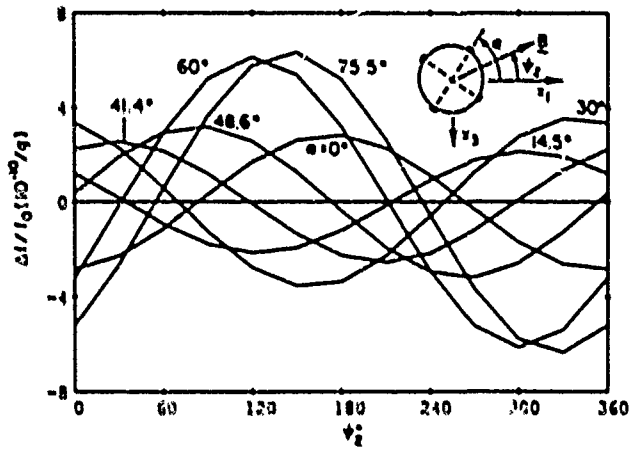


Fig. 6 $\Delta f/f_0$ vs. ψ_2 for various mount orientation angle α of four-point mounted SC-cut resonators.

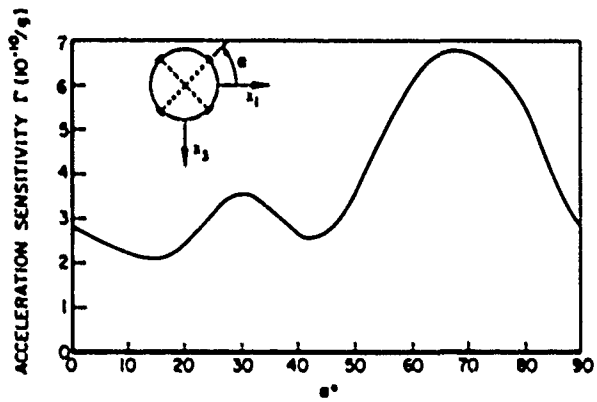


Fig. 7 Acceleration sensitivity Γ vs. α for four-point mounted SC-cut resonators.

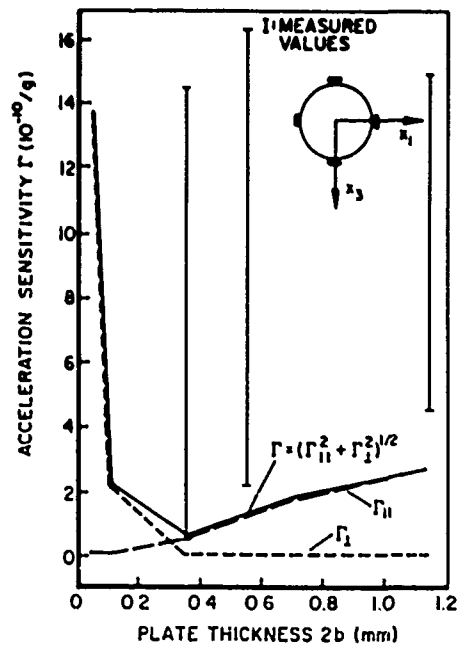


Fig. 8 Γ vs. plate thickness 2b for SC-cut resonators with the 90° apart, four-point mount.

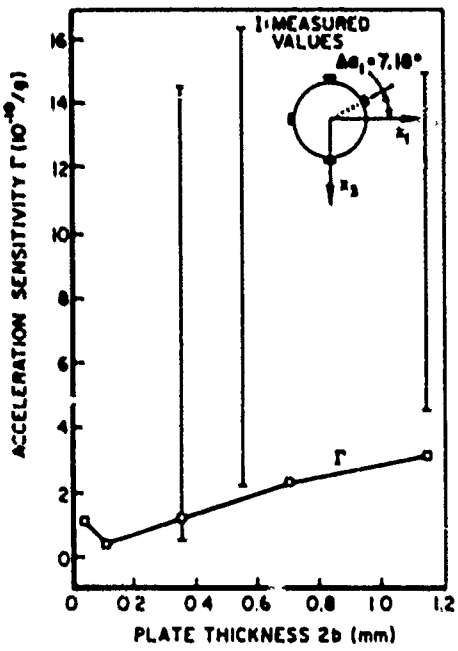


Fig. 9 Γ vs. 2b for SC-cut resonators with slightly deviated 90° apart, four-point mount.

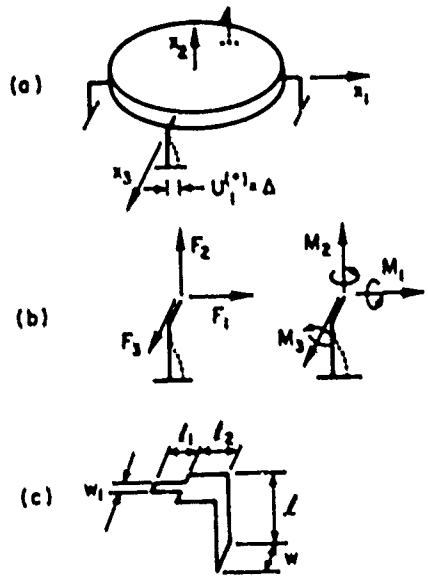


Fig. 10 (a) Four-point mounted resonator. (b) Induced forces and moments due to initial displacement. (c) Letter symbols for a general support structure.

Table I Acceleration and Displacement Sensitivities for Various Types of Support Structure

SUPPORT TYPE	ACCELERATION (g)	SUPPORT DISPL. (10^{-3} mm)	INDUCED FORCES (Dynes) F_1	F_2	F_3	INDUCED MOMENTS (Dynes-mm) M_1	M_2	M_3	ACCELERATION SENSITIVITY $\Gamma(10^{-10}/g)$	DISPLACEMENT SENSITIVITY $\Sigma_i(10^{-10})$
A	1								2.85	
		$U_1^{(0)} = 1$	-1.78×10^{5}	-4.27×10^{3}	-1.00×10^{-1}	8.85×10^{-2}	-1.31×10^{-11}	-8.78×10^{4}		$\Sigma_1 = 6221$
		$U_2^{(0)} = 1$	-4.27×10^{3}	-2.32×10^{4}	5.21	-5.09	-3.94×10^{-13}	-3.45×10^{3}		$\Sigma_2 = 156$
		$U_3^{(0)} = 1$	-1.00×1^{-1}	5.20	-1.58×10^{2}	9.04×10^{1}	-2.86×10^{-17}	-4.01×10^{-1}		$\Sigma_3 = -2.84$
B	1								1.91	
		$U_1^{(0)} = 1$	-1.21×10^{5}	-2.10×10^{2}	1.11×10^{1}	3.36×10^{1}	-2.35×10^{4}	-4.23×10^{4}		$\Sigma_1 = 4217$
		$U_2^{(0)} = 1$	-2.10×10^{2}	-3.04×10^{3}	1.65×10^{2}	4.81×10^{2}	-4.08×10^{1}	-1.80×10^{2}		$\Sigma_2 = -35.6$
		$U_3^{(0)} = 1$	1.11×10^{1}	1.65×10^{2}	-1.18×10^{2}	2.70×10^{1}	2.15	9.36		$\Sigma_3 = 0.44$
C	1								0.22	
		$U_1^{(0)} = 1$	-5.21×10^{4}	1.61×10^{-1}	-7.25×10^{-2}	-2.04×10^{-2}	-1.09×10^{4}	-1.99×10^{3}		$\Sigma_1 = 1812$
		$U_2^{(0)} = 1$	1.61×10^{-1}	-2.96×10^{2}	7.26×10^{1}	6.26×10^{1}	3.38×10^{-2}	-1.16		$\Sigma_2 = -24.6$
		$U_3^{(0)} = 1$	-7.25×10^{-2}	7.26×10^{1}	-8.86×10^{1}	8.35	-1.52×10^{-2}	2.62×10^{-1}		$\Sigma_3 = 5.46$
D	1								-0.15	
		$U_1^{(0)} = 1$	-4.82×10^{4}	2.09×10^{-1}	-7.07×10^{-2}	-4.05×10^{-2}	-1.17×10^{4}	-9.91×10^{2}		$\Sigma_1 = 1678$
		$U_2^{(0)} = 1$	2.09×10^{-1}	-2.35×10^{2}	5.07×10^{1}	5.51×10^{1}	5.09×10^{-2}	-5.48×10^{-1}		$\Sigma_2 = -24.2$
		$U_3^{(0)} = 1$	-7.07×10^{-2}	5.07×10^{1}	-7.33×10^{1}	5.45	-1.72×10^{-2}	1.08×10^{-1}		$\Sigma_3 = 4.9$
E	1								0.59	
		$U_1^{(0)} = 1$	-9.71×10^{3}	-3.90×10^{-1}	6.50×10^{-2}	1.06×10^{-1}	-3.28×10^{3}	-1.30×10^{3}		$\Sigma_1 = 334$
		$U_2^{(0)} = 1$	-3.90×10^{-1}	-1.87×10^{2}	3.39×10^{1}	4.90×10^{1}	-1.32×10^{-1}	-4.06×10^{-1}		$\Sigma_2 = -23.3$
		$U_3^{(0)} = 1$	-6.48×10^{-2}	3.39×10^{1}	-4.50×10^{1}	5.72	2.19×10^{-2}	6.73×10^{-2}		$\Sigma_3 = 3.80$

AN ANALYSIS OF THE NORMAL ACCELERATION SENSITIVITY OF ST-CUT QUARTZ SURFACE WAVE RESONATORS RIGIDLY SUPPORTED ALONG THE EDGES

H.F. Tiersten and D.V. Shick
Department of Mechanical Engineering,
Aeronautical Engineering & Mechanics
Troy, New York 12180-3590

Abstract

An analysis of the normal acceleration sensitivity of grooved ST-cut quartz surface wave resonators rigidly supported along the edges is performed. The variational principle with all natural conditions for anisotropic flexure, which is required for the new approximation procedure used in the calculation of the flexural biasing state, is presented. The flexural biasing state is employed in the existing perturbation equation along with the proper continuous representation of the acoustic surface wave mode shape for arrays of reflecting grooves to calculate the normal acceleration sensitivity. It is shown that by appropriate selection of the planar aspect ratio of an ST-cut quartz substrate the normal acceleration sensitivity can be made to vanish.

1. Introduction

In earlier work[1] on the normal acceleration sensitivity of ST-cut quartz surface wave resonators simply-supported along the edges, which produces flexure in the quartz substrate, it was shown that the acceleration sensitivity can be made to vanish by proper selection of the planar aspect ratios. The sensitivity vanishes even though the influence of flexure on the frequency of surface wave resonators is large because the changes in frequency caused by the flexure in the two spanning directions are of opposite sign and cancel. The unrealistic assumption of simple supports was made because this enabled an exact analysis for the flexural biasing state, which was necessary because the flexural variational principle used in the approximation procedure employed in this work was not available at the time. In addition, an assumed surface wave mode shape along the transmission path was employed because the proper continuous representation[2,3] of the mode shape along the transmission path for arrays of reflecting grooves did not yet exist. The surface wave mode shape that was employed[1] has been shown[2,3] to be inaccurate and results in an overestimate of the normal acceleration sensitivity.

In this work an analysis of the normal acceleration sensitivity of grooved ST-cut quartz surface wave resonators rigidly supported along the edges is performed. The variational principle for anisotropic flexure, in which both constraint-type and natural-type conditions[4,5] appear as natural conditions[6,7], is presented. The new approximation procedure[1] is used along with the variational principle to calculate the flexural biasing state for rigid edge supports. The biasing state is employed in a perturbation equation[8] along with the proper continuous representation of the surface wave mode shape for arrays of reflecting grooves[2,3] to calculate the normal acceleration sensitivity.

2. Perturbation Equations

For purely elastic nonlinearities the equation for the first perturbation of the eigenvalue obtained from the perturbation analysis[8] 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

$$H_\mu = -\int_V \hat{K}^n_{LY} g^\mu_{Y,L}\, dV\,, \tag{2.2}$$

where V is the undeformed volume of the piezoelectric plate. In (2.2) g^μ_Y denotes the normalized mechanical displacement vector, and $\hat{K}^n_{LY}$ denotes the portion of the Piola-Kirchhoff stress tensor resulting from the biasing state in the presence of the g^μ_Y, and is given by

$$\hat{K}^n_{LY} = \hat{c}_{LYM\alpha} g^\mu_{\alpha,M}\,, \tag{2.3}$$

where

$$\hat{c}_{LYM\alpha} = T^1_{LM}\delta_{Y\alpha} + c_{3LYM\alpha KN}E^1_{KN} + c_{2LYKM}w_{\alpha,K} + c_{2LKM\alpha}w_{Y,K}\,, \tag{2.4}$$

and

$$T^1_{LM} = c_{2LMKN}E^1_{KN}\,, \quad E^1_{KN} = \frac{1}{2}(w_{K,N} + w_{N,K})\,. \tag{2.5}$$

The quantities T^1_{LM}, E^1_{KN} and w_K denote the static biasing stress, strain and displacement field, respectively. Thus, in this description the present position $\underset{\sim}{y}$ is related to the reference position $\underset{\sim}{X}$ by

$$\underset{\sim}{y}(X_L, t) = \underset{\sim}{X} + \underset{\sim}{w}(X_L) + \underset{\sim}{u}(X_L, t)\,. \tag{2.6}$$

The coefficients c_{2LMKN} and $c_{3LYM\alpha KN}$ denote the second and third order elastic constants, respectively.

The normalized eigensolution g^μ_Y and $\hat{f}^\mu$ is defined by

$$g^\mu_Y = \frac{u^\mu_Y}{N_\mu}\,, \quad \hat{f}^\mu = \frac{\tilde{\varphi}^\mu}{N_\mu}\,, \quad N^2_\mu = \int \rho u^\mu_Y u^\mu_Y\, dV\,, \tag{2.7}$$

where u^μ_Y and $\tilde{\varphi}^\mu$ are the mechanical displacement and electric potential, respectively, which satisfy the equations of linear piezoelectricity

$$\tilde{K}^\ell_{LY} = c_{2LYM\alpha}u^\ell_{\alpha,M} + e_{MLY}\tilde{\varphi}^\ell_{,M}\,,$$
$$\tilde{D}^\ell_L = e_{LMY}u^\ell_{Y,M} - \epsilon_{LM}\tilde{\varphi}^\ell_{,M}\,, \tag{2.8}$$
$$\tilde{K}^\ell_{LY,L} = \rho\ddot{u}^\ell_Y\,, \quad \tilde{D}^\ell_{L,L} = 0\,, \tag{2.9}$$

subject to the appropriate boundary conditions, and ρ is the mass density. Equations (2.8) are the linear piezoelectric constitutive relations and (2.9) 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.8, as is the remainder of the notation in this section.

CH2427-3/87/0000-282 $1.00 C 1987 IEEE

The substitution of (2.3) in (2.2) yields

$$H_\mu = -\int_V \hat{c}_{LYM\alpha} E^\mu_{\alpha,H} K^\mu_{Y,L}\, dV . \qquad (2.10)$$

Since E^μ_α denotes the normalized surface wave mode shape and from (2.4) $\hat{c}_{LYM\alpha}$ depends on the biasing state, H_μ can be evaluated when the surface wave mode shape and biasing state are known.

3. Equations for Anisotropic Static Flexure

With the aid of Mindlin's plate equations[9-11] 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[12]

$$K^{(1)}_{AB,AB} + F^{(1)}_{B,B} + F^{(0)}_2 = 0 , \qquad (3.1)$$

where we have introduced the convention that A, B, C, D take the values 1 and 3 and skip 2 and

$$K^{(1)}_{AB} = \int_{-h}^{h} X_2 K_{AB}\, dX_2 , \quad F^{(1)}_B = h[K_{2B}(h) + K_{2B}(-h)] , \qquad (3.2)$$

and in this instance

$$F^{(1)}_B = 0, \quad F^{(0)}_2 = K_{22}(h) - K_{22}(-h) - 2h\rho a_2 = -2h\rho a_2 , \qquad (3.3)$$

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.12 the constitutive equations for the stress-resultants take the form

$$K^{(1)}_{AB} = \frac{2}{3} h^3 Y_{ABCD} E^{(1)}_{CD} , \qquad (3.4)$$

where Voigt's anisotropic plate elastic constants are given by

$$Y_{RS} = c_{RS} - c_{RW} c^{-1}_{WV} c_{VS} , \quad R,S = 1,3,5;\ W,V = 2,4,6, \qquad (3.5)$$

in the compressed notation, and where we have introduced the scheme shown. The plate strains $E^{(1)}_{CD}$ in (3.4) are given by

$$E^{(1)}_{CD} = \frac{1}{2}\left(w^{(1)}_{C,D} + w^{(1)}_{D,C}\right) , \qquad (3.6)$$

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

$$E^{(1)}_W = -c^{-1}_{WV} c_{VS} E^{(1)}_S , \quad w^{(1)}_A = -w^{(0)}_{2,A}, \quad E^{(1)}_{CD} = -w^{(0)}_{2,CD} \qquad (3.7)$$

which, respectively, are given in Eqs.(4.24), (3.34) and (3.37) of Ref.12. Since from Eqs.(3.31) of Ref.13 we have $E^{(0)}_{KL} = 0$, when the plate deflection $w^{(0)}_2$ has been found, we know the three-dimensional strain field from the relation

$$E_{KL} = \frac{1}{2}(w_{L,K} + w_{K,L}) = X_2 E^{(1)}_{KL} . \qquad (3.8)$$

as in Eqs.(3.32) of Ref.13.

Although we now have the plate strains from (3.8), we cannot yet determine the $\hat{c}_{LYM\alpha}$ from (2.4) because, as noted in Ref.13, we need the displacement gradients $w_{L,K}$ or equivalently the rotations Ω_{KL}, which are defined by

$$\Omega_{KL} = \frac{1}{2}(w_{L,K} - w_{K,L}) . \qquad (3.9)$$

As in Ref.13, we require that the associated plate rotations that accompany the plate strains that arise from the relaxation of the plate stress resultants $K^{(1)}_{2L}$ satisfy the appropriate three-dimensional rotation gradient-strain gradient relations, i.e.,

$$\Omega_{KL,M} = E_{ML,K} - E_{MK,L} . \qquad (3.10)$$

From (3.22) and (3.23) of Ref.13 we have the expressions for the plate rotations

$$\Omega^{(n)}_{AB} = \frac{1}{2}\left(w^{(n)}_{B,A} - w^{(n)}_{A,B}\right),$$
$$\Omega^{(n)}_{2A} = \frac{1}{2}\left[(n+1)w^{(n+1)}_A - w^{(n)}_{2,A}\right] . \qquad (3.11)$$

From (3.33) of Ref.13, we have

$$\Omega^{(0)}_{13} = \text{arbitrary constant} = 0, \quad \Omega^{(1)}_{13} = 0 , \qquad (3.12)$$

and from (3.35) and (3.34), respectively, of Ref.13, we have

$$\Omega^{(0)}_{2A} = -w^{(0)}_{2,A}, \quad \Omega^{(1)}_{2A} = E^{(1)}_{2A} , \qquad (3.13)$$

which with (3.24) of Ref.13 enables us to write

$$\Omega_{2A} = -w^{(0)}_{2,A} + X_2 E^{(1)}_{2A}, \quad \Omega_{13} = 0 , \qquad (3.14)$$

which is the same as (3.36) of Ref.13. Now, from (3.8), (3.14) and

$$w_{K,L} = E_{KL} + \Omega_{LK} , \qquad (3.15)$$

we have the desired three-dimensional displacement gradients when the plate deflection $w^{(0)}_2$ has been found.

4. Unconstrained Variational Principle for Static Flexure of Plates

In this section we present the unconstrained variational principle for static flexure of thin plates, in which all conditions, including constraint-type edge conditions, appear as natural conditions. This principle is required for the termination of the flexural biasing state for rigid edge supports using the new approximation procedure we employ. Before proceeding with the variational principle we consider it advisable for clarity to discuss the plan view of a plate shown in Fig.1. The edge of the plate consists of two smooth curves with outward unit normal N_A, which intersect at C_1 and C_2, and we denote the entire closed circuit by c. Some portions of the edge are supported and other portions are subject to prescribed loadings, which may vanish in any part. The portions of the edge which are supported have $\bar{w}^{(0)}_2$ and $\partial\bar{w}^{(0)}_2/\partial n$ prescribed where $n = N_A X_A$, and are denoted c^C. The portions of the edge which are subject to prescribed loadings, which consist of bending moments $\bar{m}$, twisting moments $\bar{t}$ and vertical shearing forces $\bar{v}$, are denoted c^N. The corner C_1 is subject to a prescribed vertical force $\bar{T}$ and the corner C_2 is subject to a prescribed corner

displacement $w_2^{(0)C_2}$. As usual, in order that all variations may be regarded as independent when constraints exist, we introduce each constraint as a zero times a Lagrange multiplier[6,7,14,15] in the variational principle. Accordingly, the variational principle for static flexure of thin plates may be written in the form[13,16]

$$\delta \int_S \left[-U + F_2^{(0)} w_2^{(0)} - F_B^{(1)} w_{2,B}^{(0)}\right] dS + \delta \oint_{c^N} \left[\left(\bar{v} + \frac{\partial \bar{t}}{\partial s}\right) w_2^{(0)} - \bar{m} \frac{\partial w_2^{(0)}}{\partial n}\right] ds + \delta \oint_{c^C} \left[\lambda(\bar{w}_2^{(0)} - w_2^{(0)}) + \mu\left(\frac{\partial \bar{w}_2^{(0)}}{\partial n} - \frac{\partial w_2^{(0)}}{\partial n}\right)\right] ds - \delta\left[-\bar{T} w_2^{(0)C_1} + \lambda^{C_2}(\bar{w}_2^{(0)C_2} - w_2^{(0)C_2})\right] = 0 . \quad (4.1)$$

where S denotes the area of the plate,

$$U = \frac{1}{2} K_{AB}^{(1)} w_{2,AB}^{(0)}, \quad K_{AB}^{(1)} = -\frac{2h^3}{3} Y_{ABCD} w_{2,CD}^{(0)}, \quad (4.2)$$

and λ, μ and λ^{C_2} are Lagrange undetermined multipliers.

Taking the variations in (4.1) and integrating by parts and using the surface divergence theorem[17] twice, we obtain

$$\int_S \left[K_{AB,AB}^{(1)} + F_2^{(0)} + F_{B,B}^{(1)}\right] \delta w_2^{(0)} \, dS + \oint_{c^N} \left[\left(-M_{Ns,s} - N_B K_{BA,A}^{(1)} - N_B F_B^{(1)} + \bar{v} + \frac{\partial \bar{t}}{\partial s}\right)\delta w_2^{(0)} + (N_A K_{AB}^{(1)} N_B - \bar{m}) \delta \frac{\partial w_2^{(0)}}{\partial n}\right] ds + \oint_{c^C} \left[\left(-M_{Ns,s} - N_B K_{BA,A}^{(1)} - N_B F_B^{(1)} - \lambda\right) \delta w_2^{(0)} + (N_A K_{AB}^{(1)} N_B - \mu) \delta w_{2,n}^{(0)} + \delta\lambda(\bar{w}_2^{(0)} - w_2^{(0)}) + \delta\mu(\bar{w}_{2,n}^{(0)} - w_{2,n}^{(0)})\right] ds + (M_{Ns}^+ - M_{Ns}^- - \lambda^{C_2}) \delta w_2^{(0)C_2} + (M_{Ns}^+ - M_{Ns}^- + \bar{T}) \delta w_2^{(0)C_1} + \delta\lambda^{C_2}(\bar{w}_2^{(0)C_2} - w_2^{(0)C_2}) = 0 , \quad (4.3)$$

where

$$M_{Ns} = N_D K_{DC}^{(1)} \nu_C , \quad (4.4)$$

and ν_C denotes a unit vector tangent to c in the counterclockwise direction. Since on account of the use of Lagrange multipliers with the constraint conditions all variations in (4.3) are arbitrary, we obtain (3.1), the natural edge and corner conditions

$$M_{Ns,s} + N_B K_{BA,A}^{(1)} + N_B F_B^{(1)} = \bar{v} + \frac{\partial \bar{t}}{\partial s} \text{ on } c^N , \quad (4.5)$$

$$N_A K_{AB}^{(1)} N_B = \bar{m} \text{ on } c^N , \quad (4.6)$$

$$M_{Ns}^- - M_{Ns}^+ = \bar{T} \text{ at } C_1 , \quad (4.7)$$

the natural form of the constraint-type edge and corner conditions

$$w_2^{(0)} = \bar{w}_2^{(0)}, \quad \partial w_2^{(0)}/\partial n = \partial \bar{w}_2^{(0)}/\partial n \text{ on } c^C , \quad (4.8)$$

$$w_2^{(0)} = \bar{w}_2^{(0)} \text{ at } C_2 , \quad (4.9)$$

and the Lagrange multipliers

$$\lambda = -M_{Ns,s} - N_B K_{BA,A}^{(1)} - N_B F_B^{(1)} , \quad (4.10)$$

$$\mu = N_A K_{AB}^{(1)} N_B , \quad \lambda^{C_2} = M_{Ns}^+ - M_{Ns}^- . \quad (4.11)$$

Thus it is clear that the variational principle (4.1) with unconstrained variations yields the differential equation (3.1), the natural edge and corner conditions (4.5) - (4.7) and the constraint-type edge and corner conditions (4.8) and (4.9). In addition, the Lagrange multipliers λ, μ and λ^{C_2} have been expressed in terms of derivatives of $w_2^{(0)}$ in (4.10) and (4.11). Consequently, the variations $\delta\lambda$, $\delta\mu$ and $\delta\lambda^{C_2}$ may be obtained from (4.10) and (4.11) and substituted[15] in (4.3), which may then be used for obtaining an approximate solution for the flexural biasing state without any a priori conditions[4,5] on the approximating functions.

5. Flexure of Rectangular ST-Cut Quartz Plate Rigidly Supported Along the Edges

A plan view and cross-section of the rigidly-supported rectangular plate is shown in Fig.2 along with the coordinate system. 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

$$\frac{2}{3} h^3 Y_{ABCD} w_{2,CDAB}^{(0)} + 2h\rho a_2 = 0 . \quad (5.1)$$

Since the plate is rigidly supported along the edges, the boundary conditions take the form

$$w_2^{(0)} = 0, \; w_{2,1}^{(0)} = 0 \text{ at } X_1 = \pm a , \quad -b < X_3 < b ,$$
$$w_2^{(0)} = 0, \; w_{2,3}^{(0)} = 0 \text{ at } X_3 = \pm b , \quad -a < X_1 < a . \quad (5.2)$$

Since as already indicated the problem defined in (5.1) and (5.2) cannot be solved exactly, an approximation procedure is employed. To this end we first transform the inhomogeneity from the differential equation (5.1) into the boundary conditions (5.2) by writing

$$w_2^{(0)} = \hat{w}_2^{(0)} + A(X_1^2 - a^2)(X_3^2 - b^2) , \quad (5.3)$$

which when substituted into (5.1) yields

$$Y_{11} \hat{w}_{2,1111}^{(0)} + 2(Y_{13} + 2Y_{55}) \hat{w}_{2,1133}^{(0)} + Y_{33} \hat{w}_{2,3333}^{(0)} = 0 , \quad (5.4)$$

since for ST-cut quartz $Y_{15} = Y_{35} = 0$ and A has been selected as

$$A = -3\rho a_2 / 8h^2 (Y_{13} + 2Y_{55}) . \quad (5.5)$$

The further substitution of (5.3) into (5.2) yields the edge conditions

$$\hat{w}_2^{(0)} = 0, \; \hat{w}_{2,1}^{(0)} = -2AX_1(X_3^2 - b^2) \text{ at } X_1 = \pm a , \quad -b < X_3 < b , \quad (5.6)$$

$$\hat{w}_2^{(0)} = 0, \; \hat{w}_{2,3}^{(0)} = -2AX_3(X_1^2 - a^2) \text{ at } X_3 = \pm b , \quad -a < X_1 < a . \quad (5.7)$$

As a solution of (5.4) consider the finite sum

$$\hat{w}_2^{(0)} = \sum_{n=1}^{N} D_n \cosh \eta_n X_1 \cos \nu_n X_3 , \qquad (5.8)$$

where $\nu_n = n\pi/2pb$ and p is chosen to be an irrational number in order that neither $\cos \nu_n b$ or $\sin \nu_n b$ vanish. The substitution of (5.8) into (5.4) yields

$$\eta_n^2 = \nu_n^2\left[(\gamma_{13} + 2\gamma_{55}) \pm \sqrt{(\gamma_{13} + 2\gamma_{55})^2 - \gamma_{11}\gamma_{33}}\right]/\gamma_{11} , \qquad (5.9)$$

for each n. Equation (5.9) has two positive independent roots. The two negative roots yield the same respective solutions as the two positive roots because of the form of (5.8). Hence, as a solution of the boundary value problem we may write

$$\hat{w}_2^{(0)} = \sum_{n=1}^{N} (D_n^{(1)} \cosh \eta_n^{(1)} X_1 + D_n^{(2)} \cosh \eta_n^{(2)} X_1) \cos \nu_n X_3 , \qquad (5.10)$$

where $D_n^{(1)}$ and $D_n^{(2)}$ are amplitude coefficients still to be determined and

$$\eta_n^{(1)} = g_1 \nu_n , \quad \eta_n^{(2)} = g_2 \nu_n , \qquad (5.11)$$

provided

$$g_{\frac{1}{2}} = \left[\frac{(\gamma_{13} + 2\gamma_{55}) \pm \sqrt{(\gamma_{13} + 2\gamma_{55})^2 - \gamma_{11}\gamma_{33}}}{\gamma_{11}}\right]^{1/2} . \qquad (5.12)$$

By requiring $(5.6)_1$ to be satisfied for each n, we obtain

$$D_n^{(2)} = S_n D_n^{(1)} , \quad S_n = -\cosh(g_1 \nu_n a)/\cosh(g_2 \nu_n a) , \qquad (5.13)$$

the substitution of which in (5.10) enables us to write

$$\hat{w}_2^{(0)} = \sum_{n=1}^{N} D_n (\cosh g_1 \nu_n X_1 + S_n \cosh g_2 \nu_n X_1) \cos \nu_n X_3 , \qquad (5.14)$$

as the approximate solution function, where we have taken the liberty of eliminating the superscript (1) on the D_n.

Since the solution (5.3) with (5.14) satisfies the differential equation (5.1) exactly and all edge conditions are of constraint type but in natural form all that remains of the variational principle given in (4.3) is

$$\oint_C \left[\delta\lambda(\bar{w}_2^{(0)} - w_2^{(0)}) + \delta\mu\left(\frac{\overline{\partial w_2^{(0)}}}{\partial n} - \frac{\partial w_2^{(0)}}{\partial n}\right)\right] ds + \delta\lambda^{C_2}(\bar{w}_2^{(0)C_2} - w_2^{(0)C_2}) = 0 , \qquad (5.15)$$

in which

$$\bar{w}_2^{(0)} = \frac{\overline{\partial w_2^{(0)}}}{\partial n} = \bar{w}_2^{(0)C_2} = 0 \quad \text{on } c^C , \qquad (5.16)$$

and since $(5.2)_1$ is satisfied, we have $w_2^{(0)C_2} = 0$ and the term in (5.15) which is not under the integral sign vanishes. The required variations $\delta\lambda$ and $\delta\mu$ are readily obtained from (4.10) and $(4.11)_1$. Substituting from (5.3), (5.16) and $(5.2)_1$, which has been satisfied exactly, into (5.15) and employing $(3.3)_1$, (4.10) and $(4.11)_1$, we obtain[15]

$$\int_{-a}^{a}\left[-(2\delta\hat{K}_{13,1}^{(1)} + \delta\hat{K}_{33,3}^{(1)})\hat{w}_2^{(0)} + \delta\hat{K}_{33}^{(1)}\left(\hat{w}_{2,3}^{(0)} + 2\lambda(X_1^2 - a^2)X_3\right)\right]_{X_3=b} dX_1$$

$$-\int_{-a}^{a}\left[-(2\delta\hat{K}_{13,1}^{(1)} + \delta\hat{K}_{33,3}^{(1)})\hat{w}_2^{(0)} + \delta\hat{K}_{33}^{(1)}\left(\hat{w}_{2,3}^{(0)} + 2\lambda(X_1^2 - a^2)X_3\right)\right]_{X_3=-b} dX_1$$

$$+\int_{-b}^{b}\delta\hat{K}_{11}^{(1)}\left[\hat{w}_{2,1}^{(0)} + 2\lambda X_1(X_3^2 - b^2)\right]_{X_1=a} dX_3$$

$$-\int_{-b}^{b}\delta\hat{K}_{11}^{(1)}\left[\hat{w}_{2,1}^{(0)} + 2\lambda X_1(X_3^2 - b^2)\right]_{X_1=-a} dX_3 = 0 , \qquad (5.17)$$

where the $\delta\hat{K}_{\lambda\beta}^{(1)}$ are obtained from $(4.2)_2$ with $\hat{w}_2^{(0)}$ instead of $w_2^{(0)}$ since λ is fixed and the variation of prescribed quantities vanishes. Substituting from (5.14) into (5.17) and performing the integrations, we obtain

$$\sum_{n=1}^{N}\sum_{m=1}^{N} a_{mn} D_m \delta D_n + \sum_{n=1}^{N} b_n \delta D_n = 0 , \qquad (5.18)$$

where the expressions for the a_{mn} and b_n are too lengthy to present here[15]. Since the variations δD_n are arbitrary, we obtain

$$\sum_{m=1}^{N} a_{mn} D_m = -b_n , \quad n = 1, 2, \ldots N , \qquad (5.19)$$

which constitute N inhomogeneous linear algebraic equations in the N 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.

The center displacement and the edge conditions which are satisfied approximately in the variational equation (5.17) versus N are shown in Table I. It can be seen from the table that $w_2^{(0)}(0,0)$ has converged extremely well before N = 20. The quantities in the other three columns, i.e., $w_2^{(0)}(0,b)$, $w_{2,1}^{(0)}(a,0)$ and $w_{2,3}^{(0)}(0,b)$, are supposed to vanish. The table shows that for N = 20 the quantities have been reduced by three and two orders of magnitude, respectively, from the values calculated from the lowest N considered. However, while $w_2^{(0)}(0,b)$ is down to the order of 10^{-15} m, which is four orders of magnitude below $w_2^{(0)}(0,0)$, $w_{2,1}^{(0)}(a,0)$ and $w_{2,3}^{(0)}(0,b)$ are down only to the orders of 10^{-12} and 10^{-10}, respectively. This is as expected since differentiation always reduces the rate of convergence[18]. The oscillation in sign shown in the last three columns of Table I is a result of

the fact that each of the quantities is evaluated at one point from a sum of oscillating functions. The calculated centerline deflection of the plate, i.e., at $X_3 = 0$, is plotted as a function of X_1 in Fig.3.

6. Resonant Surface Wave Mode Shape

In this section we present the proper continuous representation of the acoustic surface wave mode shape in resonators with grooved reflectors, which was obtained in recent work[2,3]. The straight-crested surface wave displacement field may be written in the known form[19,20],

$$u_j = \alpha_j(X_2)e^{i\xi(X_1-Vt)}, \quad \alpha_j = \sum_{m=1}^{4} C^{(m)}A_j^{(m)}e^{i\beta_m X_2}. \quad (6.1)$$

A plan view of the resonator showing the reflecting arrays of grooves, the coordinate system and the associated planar geometry is shown in Fig.2, and we note that the plane $X_2 = 0$ denotes the ungrooved surface of the substrate and the axis points down. It has been shown[2,3] that the variable-crested resonant surface wave mode shape with variable amplitude along the transmission path is very accurately approximated by

$$u_j = \cos\frac{\pi X_3}{2W}\,\mathrm{Re}\Big[\alpha_j(X_2)\hat{C}^R(X_1)e^{i\xi(X_1-s)} + \alpha_j^*(X_2)\hat{C}^L(X_1)e^{-i\xi(X_1-s)}\Big]e^{-i\omega t}, \quad (6.2)$$

where the variations along the transmission path are given by

$$\hat{C}^R(X_1) = \Big[r_1 e^{-\alpha_1 N} e^{-\beta_2 X_1} - r_2 e^{-\alpha_2 N} e^{-\beta_1 X_1}\Big]\bar{C}/d,$$

$$\hat{C}^L(X_1) = r_1 r_2\Big[e^{-\alpha_1 N} e^{-\beta_2 X_1} - e^{-\alpha_2 N} e^{-\beta_1 X_1}\Big]\bar{C}/d. \quad (6.3)$$

In (6.3) $\bar{C}$ denotes the amplitude of the input wave, N denotes the number of grooves,

$$d = r_1 e^{-\alpha_1 N} - r_2 e^{-\alpha_2 N}, \quad \beta_{1,2} = \alpha_{1,2}N/(\ell - s), \quad (6.4)$$

and $\alpha_{1,2}$ and $r_{1,2}$ are given in Eqs.(26) and (28) of Ref.2. The amplitude of the standing surface wave mode, i.e., the part multiplying cos ωt, along the transmission path is plotted in Fig.4.

7. Acceleration Sensitivity

From Sec.6 we now know g_γ^μ and from Secs.3 - 5 we know $\hat{c}_{LYM\alpha}$ for normal acceleration, i.e., for flexure with rigidly-supported edges. Hence, we can now evaluate H_μ in Eq.(2.10). Such calculations have been performed using the known values of the second order[21] and third order[22] elastic constants of quartz for the case of normal acceleration, i.e., flexure with rigid edge supports, and the acceleration sensitivity is plotted as the solid curve in Fig.5 as a function of the planar aspect ratio a/b. It can be seen from the figure that the acceleration sensitivity goes through zero for a value of a/b of about 4.9. In addition, the normal acceleration sensitivity for the unrealistic case of simple edge supports obtained from Fig.5 of Ref.1 is plotted as the dotted curve in Fig.5 of this work. It can be seen from the figure that, except for a narrow vicinity of the zero crossing, the normal acceleration sensitivity calculated for the case of simple-supports is an unrealistic significant overestimate, as noted in Ref.1.

Acknowledgements

This work was supported in part by the Army Research Office under Contract No. DAAG-29-85-K-0075.

References

1. D.V. Shick and H.F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along the Edges," Proceedings of the 40th 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. 86CH2330-9, 262 (1986).
2. H.F. Tiersten, J.T. Song and D.V. Shick, "A Variational Analysis of the Reflection of Surface Waves by Arrays of Reflecting Grooves," 1986 Ultrasonics Symposium Proceedings, IEEE Cat. No. 86CH2375-4, Institute of Electrical and Electronics Engineers, New York, 29 (1986).
3. H.F. Tiersten, J.T. Song and D.V. Shick, "On a Continuous Representation of the Acoustic Surface Wave Mode Shape in Arrays of Reflecting Grooves," J. Appl. Phys., to be published (1986).
4. L.V. Kantorovich and V.I. Krylov, Approximate Methods of Higher Analysis (Interscience Publ. Inc., New York, and P. Noordhoff Ltd., Groningen, The Netherlands, transl. by C.D. Benster from 3rd Russian ed., 1964), pp.258-260, 272-273 and 279-281.
5. L. Collatz, The Numerical Treatment of Differential Equations, transl. by P.C. Williams (Springer-Verlag, Berlin, 1960), 2nd ed., pp.202-207 and 213-216.
6. H.F. Tiersten, "Natural Boundary and Initial Conditions from a Modification of Hamilton's Principle," J. Math. Phys., 9, 1445 (1968).
7. H.F. Tiersten, Linear Piezoelectric Plate Vibration (Plenum, New York, 1969), Chap.6, Sec.4.
8. H.F. Tiersten, "Perturbation Theory for Linear Electroelastic Equations for Small Fields Superposed on a Bias," J. Acoust. Soc. Am., 64, 832 (1978).
9. 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.
10. R.D. Mindlin, "High Frequency Vibrations of Crystal Plates," Quart. Appl. Math., 19, 51 (1961).
11. Ref.7, Chap.13.
12. 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).
13. 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).

14. We choose to write the functional for the applied corner force $-\bar{\tau}w_2^{(0)}C_1$ and the integrated form for the applied twisting moments $+(\partial\bar{\tau}/\partial s)w_2^{(0)}$ rather than the unintegrated functional for the applied twisting moments $-\bar{\tau}w_{2,s}^{(0)}$ and then integrate by parts. Both yield the same result and the one presented seems more direct.

15. For more detail see H.F. Tiersten and D.V. Shick, "On the Normal Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported Along Rectangular Edges," to be issued as a technical report, Rensselaer Polytechnic Institute, Troy, New York.

16. The generalization to more complicated cases is clear and we do not bother to consider internal surfaces of discontinuity here because they are not needed in this work.

17. L. Brand, Vector and Tensor Analysis (Wiley, New York, 1947), p.222.

18. S.L. Sobolev, Partial Differential Equations of Mathematical Physics (Pergamon Press, London and New York, 1964; Addison-Wesley Publishing Co., Reading, Mass. and London), transl. by E.R. Dawson from 3rd Russian ed., Lect.6 and Lect.22, Sec.2.

19. J.J. Campbell and W.R. Jones, "A Method for Estimating Optimal Crystal Cuts and Propagation Directions for Excitation of Piezoelectric Surface Waves," IEEE Trans. Sonics Ultrason., SU-15, 209 (1968).

20. B.K. Sinha and H.F. Tiersten, "Elastic and Piezoelectric Surface Waves Guided by Thin Films," J. Appl. Phys., 44, 4830 (1973). Sec.II.

21. R. Bechmann, "Elastic and Piezoelectric Constants of Alpha-Quartz," Phys. Rev., 110, 1060 (1958).

22. R.N. Thurston, H.J. McSkimin and P. Andreatch, Jr., "Third Order Elastic Constants of Quartz," J. Appl. Phys., 37, 267 (1966).

TABLE I - CENTER DISPLACEMENT AND VARIATIONALLY SATISFIED EDGE CONDITIONS VS. NUMBER OF TERMS IN SERIES

($a = b = 5$ mm; $a_2 = 1$ g)				
N	$w_2(0,0)$	$w_2(0,b)$	$w_{2,1}(a,0)$	$w_{2,3}(0,b)$
2	-7.35×10^{-11} m	3.75×10^{-12} m	3.27×10^{-9}	3.53×10^{-8}
4	-7.76	-1.23×10^{-11}	-3.22×10^{-9}	2.63×10^{-8}
6	-7.54	-9.05×10^{-12}	1.65×10^{-10}	2.60×10^{-8}
8	-5.12	1.75×10^{-12}	2.16×10^{-9}	6.71×10^{-9}
10	-4.66	3.00×10^{-13}	2.89×10^{-10}	1.03×10^{-9}
12	-4.64	4.13×10^{-15}	-1.01×10^{-10}	3.22×10^{-10}
14	-4.617	-8.26×10^{-13}	-1.45×10^{-11}	-2.14×10^{-9}
15	-4.6411	-3.65×10^{-15}	-1.54×10^{-11}	2.77×10^{-10}
16	-4.6413	2.92×10^{-14}	-6.19×10^{-12}	3.43×10^{-10}
17	-4.6410	-1.78×10^{-14}	-1.18×10^{-12}	2.54×10^{-10}
18	-4.6404	-4.11×10^{-15}	-1.41×10^{-12}	2.12×10^{-10}
19	-4.6406	4.11×10^{-15}	-2.55×10^{-12}	2.50×10^{-10}
20	-4.6406	-3.40×10^{-15}	2.82×10^{-12}	2.16×10^{-10}

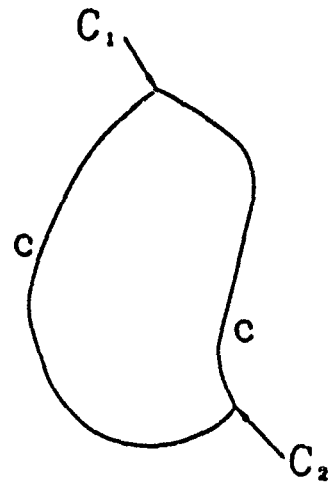
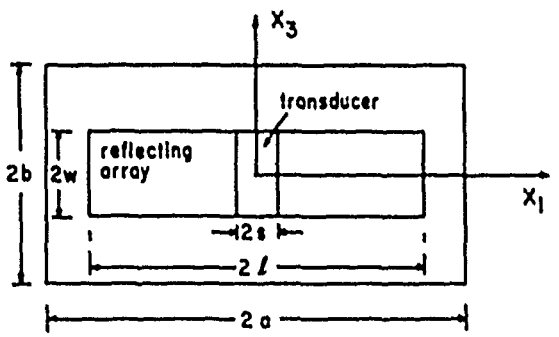


Figure 1 Plan View of a Typical Plate with an Arbitrary Curved Edge and Corners



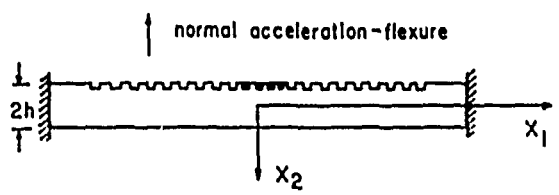


Figure 2 Plan View and Cross Section of Rectangular ST-Cut Quartz Plate Rigidly Supported at the Edges

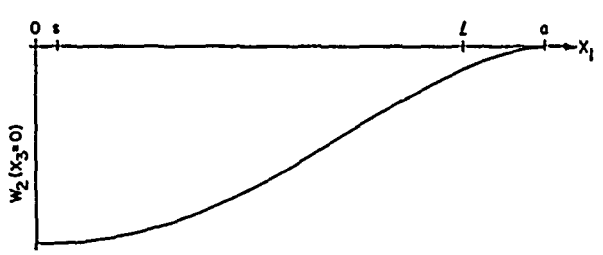


Figure 3 Centerline Deflection of Plate ($X_3 = 0$) Versus X_1

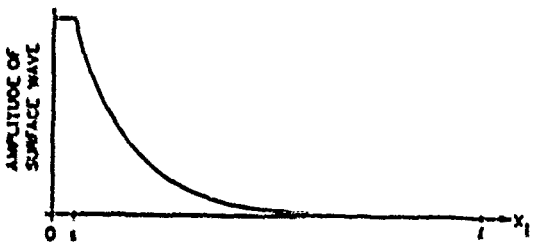


Figure 4 Amplitude of Resonant Surface Wave Mode Along Transmission Path

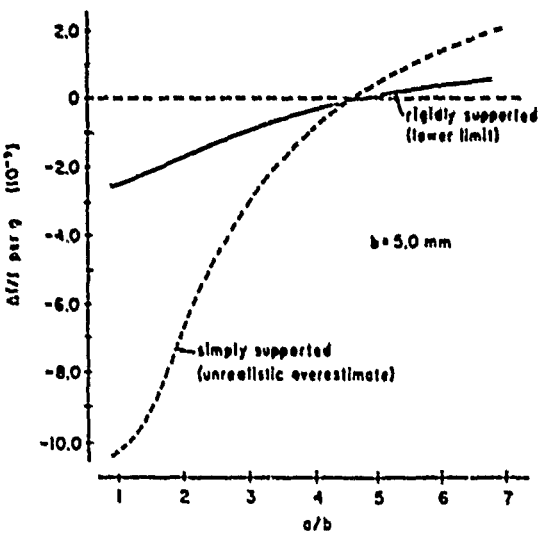


Figure 5 Calculated Normal Acceleration Sensitivity Versus Planar Aspect Ratio a/b for Rigid Edge Supports (Solid Line) and Simple Edge Supports (Dotted Line)

41st Annual Frequency Control Symposium - 1987

FORCE SENSITIVITY OF TRAPPED ENERGY VIBRATIONS IN A CONTOURED RESONATOR

R. BOURQUIN and B. DULMET
Ecole Nationale Supérieure de Mécanique et des Microtechniques
La Bouloie, Route de Gray, 25030 BESANCON CEDEX, FRANCE

ABSTRACT

The force sensitivity of thickness shear mode in quartz resonators is an essentially non linear effect arising from the non linearities of the stress-strain relations for the material, as well as from the superposition of the static and dynamic strains. That sensitivity can accurately be predicted by a first order perturbation integral wich depends on both of the static and dynamic fields in the resonator.
Here, we study the case of a circular contoured resonator submitted to a diametral compression. The study is extended to the case of AT-Cuts (C mode) and SC-Cuts (B and C modes), for overtone modes as well as for anharmonic modes, the maximum amplitude of which does not occur at the center of the disk. The perturbation integral is performed over the whole bulk of the resonator, in order to take into account the non-homogeneity of both of the vibration and the static strain.
Theoretical results giving the Ratajski's coefficient K_f versus the azimuth of the acting force are presented and compared to experiments.

INTRODUCTION

Some previous studies have predicted the force sensitivity of the quasi thickness shear vibrations [1,2] and also of other vibrations [3,4] . Such a sensitivity arises from the acoustoelastic effect, which is defined as the dependence of the speed of an acoustic wave versus a static bias on which the vibration is superimposed [5]. Then, the motion of any infinitesimal element of the body results from the stresses increments off their static values, according to the main law of dynamics. For a convenient solution of the corresponding balance equations, the dynamic mechanical displacement is selected as main field of unknowns. In this purpose, the incremental dynamic stresses should be expressed in terms of the partial derivatives of the dynamic displacement. The corresponding expansion is governed by a set of effective elastic coefficients, which depend on the biasing state. The material elastic coefficients which characterize the propagation medium and are unaffected by the bias, may be defined as the partial derivatives of the strain energy density (scalar) versus the total Lagrangian strains (second rank tensor) in a Taylor series expansion. Such a definition ensures correct tensorial properties for the material coefficients. When the vibration tends to be infinitely small, the above-mentionned effective coefficients may be explicited in terms of the material elastic coefficients of the second and third orders, and in terms of the partial derivatives of the biasing (static) displacement [6].

The preceeding topics are involved on elaborating the wave propagation equations as second order partial differential equations of the dynamic displacement. The pertinent conditions at the boundary surface of the studied finite resonator should be added to the former equations in order to determine the resonant frequencies.

The boundary surface shape is affected by the biasing state. This fact would strongly increase the difficulty of properly describing the boundary conditions, unless we refer the stress tensor at any point on the boundary surface, and also inside the bulk of the body, to the corresponding material surface element, evaluated in the natural, unstrained, state of the body [7] . Such a choice precisely defines the kind of elastic coefficients needed for this study [6].

The subject of the present paper consists in studying the frequency shift of a thickness shear vibration in a quartz resonator submitted to a diametral compression : the static bias is not homogeneously distributed in such a case, thus leading to local expressions of the effective elastic coefficients inside the resonator. Nevertheless, the relative difference between effective and material coefficients exhibits the same order of magnitude as the static strain, i.e. less than 10^{-4} (a corresponding value to the breaking stress). Therefore, it appears very logical to adopt a perturbation method to determine the small frequency shift induced by the biasing strain [8] . The force sensitivity is then given by a bulk integral, the parameters of which are the second and third order material coefficients, the mass density and the partial derivatives of the static and dynamic displacements with respect to the material coordinates taken in the natural state. Because only trapped vibrations, induced either by contouring the plate or by the mass loading effect [9] , are of interest for high Q vibrations, the vibrating region of the most commonly employed modes is confined in the central region of the plate. As a consequence, instead of performing a complete integration over the whole bulk of the resonator, previous modellings commonly calculate the force sensitivity by means of the local equations of the acoustoelastic effect, merely evaluated at the precise center of the plate.

In the present paper, the whole integration is retained in the modelling. Our field of interest is restricted to the case of contoured resonators, but from any cristallographic orientation. We stick to enlightening the differences between the results obtained by means of the here presented method and the most generally employed one, rather than attempting to improve the solution of either the dynamic or the static separate parts of the problem [10]. Especially, some investigations have been made about the behavior of the anharmonic modes, the extent of which far exceeds the restricted central region of the plate. Actually, these modes are not expected to exhibit the same behavior as the well known "metrologic" modes, mostly employed in pratical applications. This point is confirmed by first experimental results, included in the present study.

I Basic Equations

Disregarding the piezoelectric interaction, the equations of the more general electroelasticity for small fields superposed on a bias [6] reduce into the following equations, valid for acoustoelasticity problems :

CH2427-3/87/0000-289 $1.00 C 1987 IEEE

$$(1)\quad \frac{\partial \tilde{P}_{ij}}{\partial a_i} = \rho_o \ddot{u}_j$$

$$(2)\quad \tilde{P}_{ij} = (C_{ijkl} + \hat{C}_{ijkl}) \frac{\partial \hat{u}_l}{\partial a_k}$$

$$(3)\quad \hat{C}_{ijkl} = (\delta_{jl} C_{ikmn} + C_{ijklmn}) E_{mn} + C_{ijsk} \frac{\partial w_l}{\partial a_s} + C_{ipkl} \frac{\partial w_j}{\partial a_p}$$

where $\tilde{P}$ denotes the dynamic stress Piola tensor [6,7] ρ_o is the mass density in the natural state, C_{ijkl} and C_{ijklmn} are components of the second and third order material elastic tensors (tabulated) and $\hat{C}_{ijkl}$ is the increment of the effective coefficient from the material coefficient, while E_{mn} indicates a component of the static strain tensor a_i denotes a current coordinate of a material point of the body in the natural state, any stress free. The biasing state transforms it into $a_i + w_i$, thus defining a new position, often called "initial" position (despite of the "initial" term is somewhat confusing). In the actual state (bias + vibration), the current coordinate of the same material point becomes $a_i+w_i+\hat{u}_i$. If there is no dynamic stress acting on the boundary surface of the resonator, the boundary conditions in the reference axes of the natural state express in the following manner:

$$(4)\quad N_i \tilde{P}_{ij} = 0 \text{ on } S_o$$

where S_o is the boundary surface, defined in the natural state, and N_i the corresponding normal unit, outwardly directed. The set of above equations is obtained from the more general equations of Sect II of Ref [8]. It is consistent with equations appearing in Sect II of Ref [3] obtained in a different but straightforward manner applicating a variational principal onto the kinetic and strain energy densities. In addition, we took the liberty of assuming that the static strain itself is sufficiently small to allow a linear expansion of the stress vs. strain relationship. Because the incremental $\hat{C}_{ijkl}$ coefficients are not constant inside the bulk of the resonator, the above equations (1), considered as second order partial differential equations in terms of $\hat{u}_i$ with respect to the coordinates a_i, are governed by locally varying effective coefficients. Generally, those differential equations can not be exactly solved through a direct method. Instead, we adopted the perturbation method proposed by Tiersten in Sect. II of Ref [8]. Then, if $\hat{u}_\alpha$ denotes the associated displacement of a resonant mode of ω_α angular frequency, the frequency shift induced by the bias is simply given by the following formula:

$$(5)\quad \frac{\Delta\omega}{\omega_\alpha} = \frac{1}{2\omega_\alpha^2} \frac{\int V_o \hat{C}_{ijkl}\, \partial\hat{u}_l^\alpha/\partial a_k \cdot \partial\hat{u}_i^\alpha/\partial a_j \, dV_o}{\int V_o \rho_o \hat{u}_i^\alpha \hat{u}_i^\alpha \, dV_o}$$

where the volume V_o of the studied resonator is calculated in the natural state. Because the basic formulas (1-4) derive from a rotationnally invariant formulation [11], the perturbation integral (5) can be evaluated in the reference frame $(O, \vec{a_1}, \vec{a_2}, \vec{a_3})$ defining the studied doubly rotated cut in the natural state. Then, the plate normal direction is given by the natural coordinate a_2 (See Fig.1). Since we deal with essentially thickness modes [9], we are allowed to only retain the subscript values k = i = 2 in the repeated indexes sum of (5), thus defining the approximate basic formula of the present analysis.

II Static Fields

To compute the above integral (5), we need to know the static strain as well as the displacement gradient in a quartz disk submitted to two compression forces acting on a diameter. In this paper, Ψ denotes the azimuth of the forces, counterclokwise counted in the above mentionned doubly rotated frame. For convenience, we assume that the static stresses distribution is not strongly affected by the anisotropy of quartz. Then, the static stresses, in Piola's meaning, can readily by expressed in a new coordinates system (O, a_1', a_2', a_3'), obtained from the doubly rotated system by performing a third rotation of angle Ψ around the plate normal. Then, the static stress Piola tensor in the natural coordinates system is given by [12]:

$$(6)\quad \begin{aligned} T_1' &= \frac{F}{\pi h_o}\left\{\frac{(a_1'+D/2)^3}{[(a_1'+D/2)^2+a_3'^2]^2} + \frac{(D/2-a_1')^3}{[(D/2-a_1')^2+a_3'^2]^2} - \frac{1}{D}\right\} \\ T_3' &= \frac{F}{\pi h_o}\left\{\frac{(a_1'+D/2)a_3'^2}{[(a_1'+D/2)^2+a_3'^2]^2} + \frac{(D/2-a_1')a_3'^2}{[(D/2-a_1')^2+a_3'^2]^2} - \frac{1}{D}\right\} \\ T_5' &= \frac{F}{\pi h_o}\left\{\frac{a_3'(a_1'+D/2)^2}{[(a_1'+D/2)^2+a_3'^2]^2} - \frac{a_3'(D/2-a_1')^2}{[D/2-a_1')^2+a_3'^2]^2}\right\} \\ T_2 &= T_4 = T_6 = 0, \end{aligned}$$

where D is the diameter of the disk, $2h_o$ is the maximum thickness at the plate center. The acting force F is supposed uniform along the thickness, and the radius of contour is sufficiently large to be disregarded in this part of the analysis In these equations and in what follows, the usual contraction of indexes is used (uppercase indexes). Although the compatibility equations of an isotropic medium were needed to obtain the above expressions (6), we derive the static strains from (6) through the proper anisotropic stress vs.strain inverse relations for quartz :

$$(7)\quad E_I' = S_{IJ}\, T_J'$$

$$[S_{IJ}] = [C_{IJ}]^{-1} \quad I,J = 1,2\ldots6$$

Since the material cannot be simultaneously isotropic and anisotropic, such a method is not selfconsistent, although it is a very common practice. Especially, the static displacement cannot be integrated from the whole set of static strains. Nevertheless, disregarding the strain components E_4 and E_6, generated by anisotropic properties of quartz, we can deduce the static displacement derivatives from the restricted set of E_1, E_3 and E_5 strains, with help of an additional condition to avoid the rotation of the plate. Because of the above mentionned inconsistency, the final results slightly depend on the choice of that additional condition. In the results presented in this paper, we considered that the displacement w_1 vanishes at every point of the diametre $a_2'=0$.

Hence, we obtain the following approximate expressions for the static displacement derivatives in the current frame :

$$(8)\quad \begin{aligned} \partial w_1'/\partial a_3' &= \int_0^{a_1'} (\partial E_1'/\partial a_3')\, da_1' = S_{1J} \int_0^{a_1'} (\partial T_J'/\partial a_3')\, da_1' \\ \partial w_1'/\partial a_1' &= E_1' \;;\; \partial w_1'/\partial a_2' = 0 \;; \\ \partial w_2'/\partial a_2' &= E_2' \;;\; \partial w_3'/\partial a_1' = E_5' - \partial w_1'/\partial a_3' \;;\; \partial w_3'/\partial a_3' = E_3' \end{aligned}$$

Using these expressions implies that the remaining derivatives $\partial w_2'/\partial a_1'$ and $\partial w_2'/\partial a_3'$ are essentially odd functions of a_2' and the corresponding terms give no contribution to the integral (5) since they are associated in it with dynamic quantities being even functions along the thickness. This simplification

occurs from only retaining k = i = 2 in expression (5). The partial derivatives of the mechanical displacement constitute a second rank tensor, so that the corresponding quantities can easily be obtained from (8) in the doubly rotated frame in which the basic integral(5)is calculated.

III Dynamic Modelling

To describe the vibration of a doubly rotated plano-convex resonator, we used the model recently proposed by Stevens and Tiersten [4]. This model transforms the component of the mechanical displacement, u_i, to the component $\hat{u}_i$ in the thickness solution eigenvector system :

$$(9)\quad u_i = Q_{ir}\,\hat{u}_r$$

where Q_{ir} is the r.th component of the i.th eigenvector for a plane wave propagating along the thickness direction of the plate. Such a linear transformation leads to a new set of partial differential equations in terms of u_i instead of $\hat{u}_i$. These equations are governed by a new set of elastic coefficients (9x9 matrix on compressing the indexes) we denote as $\tilde{C}_{IJ}$. The uncompressed indexes transformation law is :

$$(10)\quad C_{ijk\ell} = Q_{pi}\,Q_{s\ell}\,C_{ipks}$$

Since ordering the eigenvectors of the thickness propagation problem is somewhat arbitrary, the major transformed component of the mode of interest (slow shear C mode, fast shear B-mode) may always be denoted by u_1 by a proper reordering of the Q_{ij} and $\tilde{C}_{IJ}$ matrixes. Then, according to Stevens and Tiersten's work [4], the unknown u_1 is solution of the following partial differential equation, obtained from the asymptotic dispersion equation at small propagation wavenumbers in the plate plane:

$$(11)\quad M_n \frac{\partial^2 u_1}{\partial a_1^2} + P_n \frac{\partial^2 u_1}{\partial a_3^2} + Q_n \frac{\partial^2 u_1}{\partial a_1 \partial a_3} + \left(\rho\omega^2 - \frac{n^2\pi^2}{4h^2(a_1,a_3)}\tilde{C}_{96}\right) u_1 = 0$$

where M_n, P_n, Q_n are rather complex quantities depending on both the overtone number n and the appropriate $\tilde{C}$ tensor for the studied set of quasi thickness modes. $2h(a_1, a_3)$ is the current thickness at any point of the contoured resonator. The term containing the mixed derivative $\partial^2 u_1/\partial a_1 \partial a_3$ is eliminated by using a new coordinates system (x_1, x_3) obtained from (a_1, a_3) by a rotation of angle $\beta = \frac{1}{2}\tan^{-1}(-Q_n/(M_n - P_n))$ about the a_2 - axis. This rotation angle vanishes in an AT-Cut. In the more general case of a SC-Cut, it depends on the mode (B or C) and on the overtone number and it gives the orientation of the mode pattern. In the latter coordinates system, the dynamic displacement amplitude is given by :

$$(12)\quad u_1^{nmp} = H_m(\sqrt{\alpha_n}\,x_1)\,e^{-\frac{\alpha_n x_1^2}{2}}\,H_p(\sqrt{\beta_m}\,x_3)\,e^{-\frac{\beta_n x_3^2}{2}}\,\sin\left(\frac{n\pi x_2}{2h_o}\right)$$

where H_m and H_p are Hermite polynomials and :

$$(13)\quad \alpha_n = \frac{n\pi}{2h_o}\sqrt{\frac{\tilde{C}_{96}}{2Rh_oM'_n}}\ ;\quad \beta_n = \frac{n\pi}{2h_o}\sqrt{\frac{\tilde{C}_{96}}{2Rh_oP'_n}}$$

R_o is the radius of contour of the plano-convex resonator. The prim quantities M_n and P_n are deduced from M_n, P_n, Q_n [4]and m and p are integer indexes corresponding to the number of nodal lines along the x_1 and x_3 axes [13]. Eventually, we just have to perform the reverse decomposition from the transformed displacement component expressed in the (x_1, x_3) axes to the classical displacement in the (a_1, a_3) axes, since the usual sense is required for the mechanical displacement in the perturbation formalism.

IV Programming Techniques

If the vibration is sufficiently trapped [14], the mode pattern does not depend on the azimuth of the external forces. Hence, there is no band between the symmetries of the static and dynamic displacement fields. Integrating along the plate thickness is very easy, but a fully numerical method is required to integrate over the plane of the disk, since there is no means to separate the integration variables using either cartesian or polar coordinates. The programming task should then be optimized to avoid excessive CPU time We adopted the following principles :

- -the basic integral (5) is evaluated in the usual frame of the doubly rotated cut, thus avoiding any further computation of the third order elastic coefficients whenever the azimuth Ψ is incremented. A polar integration is done after all occuring quantities have been obtained inside the disk, the integration step in polar angle being equal to an integer fraction of $\Delta\Psi$ chosen for the final curves. Taking advantage of the isotropic assumption retained for the static stresses, this feature allows a large improvement of the computing time. The main steps of the calculation are the followings :
- -the static stress values and their derivatives with respect to the a'_3 coordinate of the triply rotated frame associated to the current azimuth of the force are stored in an array at the beginning of the program.
- -each time Ψ is incremented, those array elements are recalled with an appropriate reordering of indexes, thus taking into account the connexion between $\Delta\Psi$ and the step in polar angle on integrating.
- -Thereafter $\partial w'_1/\partial a'_1$, $\partial w'_3/\partial a'_3$, $\partial w'_3/\partial a'_1$ are easily obtained from previously calculated quantities through the relations (8) and the strain-stress relationships (7) in the current frame.
- -eventually, the needed static quantities in (5) are obtained with help of the proper tensorial rotations.
- -the dynamic displacement is also stored in an initial array, already taking into account the third rotation β of Sect.III. That array is also recalled on each new value of Ψ.
- -the results we show at next section have been obtained with help of an integration step of polar angle equal to 5 degrees and a radial step equal to R/66. Then, a single K_f vs. Ψ curve requires about 4 minutes with the minicomputer we used. Some less precise results have also been obtained with a desk-top computer and were presented in a previous paper(15).

V Theoretical and Experimental Results

To facilitate the comparison with results given by other authors, our results are presented in terms of Ratajski's coefficient[1]

$$(13)\quad K_f = \frac{\Delta f}{f}\cdot\frac{n^2}{f_F}$$

where f denotes the resonant frequency.

The figures 2 and 3 are for 5 MHz, 5-th overtone, AT-Cut (θ = 35° 25', 2h = 1 679 mm, R = 150 mm, D = 15 mm). Figure 2 shows the sensitivity of the C, 5, 0, 0 - mode calculated from the relation (5) (solid line) and the sensitivity merely calculated at the center of the resonator (dashed line). As we can see, the curves are very close together. The crosses indicate experimental values. The agreement with calculated values is fairly good, excepted for the values of Ψ in the neighbourhood of zero. As explained by Lee and Tang [10], this occurs from assuming isotropic properties on evaluating the static stresses repartition.

Other results are shown for 5 MHz, 3-rd overtone, SC-Cut resonator (θ = 35° 15', φ = - 22° 25', 2h = 1 089 mm, R = 300 mm, D = 15 mm). Figures 4 and 5 are for C-modes, figures 6 and 7 for B-modes The differences between the calculated sensitivities at the center of the plate and those obtained from the complete integral (5) are small for overtones and more important in the case of anharmonic modes. Experimental values exhibit significant deviations from theoretical values, especially in the case of the B-modes.

CONCLUSION

In the case of the overtone modes (n,o,o) which have a gaussian distribution along the in plane axes, the force sensitivities of the resonant frequencies are very near together, evaluated either by integrating over the whole disk or calculated at its center. This is also valid for doubly rotated cuts. Having a more complex mode shape, anharmonic modes (n, m, p) are expected to behave differently from overtone modes, as well in terms of maximum K_F, as in terms of the azimuth angle value corresponding to a zero sensitivity. This is confirmed by experimental results, although some significant deviations between predicted and measured sensitivities are observed in the case of doubly rotated cut. Further developments could improve the experimental vs. theoretical agreement by properly introducing the anisotropy of the material on determining the static fields.

Aknowledgments

This work was supported by the Direction des Recherches Etudes et Techniques (Contract. Nu. 86/058)

Appendix

We give the complete expressions of the effective elastic coefficients occuring in the integral perturbation in the case of quasi-thickness modes :

$$\frac{\Delta\omega}{\omega_\alpha} = \frac{1}{\omega_\alpha} \cdot \frac{\int_{V_o} q_j \, q_\ell \, \hat{C}_{2j2\ell} \, (u_{1,2})^2 \, dV_o}{\int_{V_o} \rho \, u_1^2 \, dV_o}$$

where (with contracted notation of indices) :

$$\hat{C}_{99} = T_2 + 2(C_{66}W_{1,1} + C_{26}W_{1,2} + C_{46}W_{1,3}) + C_{66K} E_K$$

$$\hat{C}_{44} = T_2 + 2(C_{46}W_{3,1} + C_{24}W_{3,2} + C_{44}W_{3,3}) + C_{44K} E_K$$

$$\hat{C}_{22} = T_2 + 2(C_{26}W_{2,1} + C_{22}W_{2,2} + C_{24}W_{2,3}) + C_{22K} E_K$$

$$C_{49} = \hat{C}_{94} = C_{46}W_{1,1} + C_{24}W_{1,2} + C_{44}W_{1,3} + C_{66}W_{3,1} + C_{26}W_{3,2} + C_{46}W_{3,3} + C_{46K} E_K$$

$$\hat{C}_{24} = \hat{C}_{42} = C_{26}W_{3,1} + C_{22}W_{3,2} + C_{24}W_{3,3} + C_{46}W_{2,1} + C_{24}W_{2,2} + C_{44}W_{2,3} + C_{24K} E_K$$

$$\hat{C}_{29} = \hat{C}_{92} = C_{26}W_{1,1} + C_{22}W_{1,2} + C_{24}W_{1,3} + C_{66}W_{2,1} + C_{26}W_{2,2} + C_{46}W_{2,3} + C_{26K} E_K$$

where $W_{i,j} = \frac{\partial w_i}{\partial a_j}$

REFERENCES

(1) J.M. RATAJSKI "The force sensitivity of A.T. Cut quartz crystals", Proc. 20th A.F.C.S. pp. 39-46, 1966.

(2) J.P. PERDIGUES "Etude de l'influence des forces sur la fréquence d'un résonateur à quartz" Thèse D.I. N° 58, Besançon 1975.

(3) P.C.Y. LEE, Y.S. WANG, X. MARKENSCOFF "High frequency vibrations of crystal plates under initial stresses". J. Acoust. Soc. Am., Vol. 57, N° 1, Jan. 1975.

(4) B.K. SINHA, H.F. TIERSTEN "On the influence of a flexural biasing state on the velocity of piezoelectric surface waves" Wave motion 1, pp 37,51, (1979) North Holland Publishing Company.

(5) R.A. TOUPIN, B. BERNSTEIN "Sound waves in Deformed Perfectly Elastic Materials - Acousto-elastic Effect", J. Acoust. Soc. Am., Vol. 33, N° 2, Feb. 1961.

(6) J.C. BAUMHAUER, H.F. TIERSTEN "Non linear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am, Vol. 54, N° 4, pp 1017, (1973).

(7) R.N. THURSTON "Waves in Solids", Handbuch der Physik VIa/4, Springer Verlag, Berlin 1974.

(8) H.F. TIERSTEN "Perturbation theory for linear electroelastic equations for small fields superposed on a bias", J. Acoust. Soc. Am., Vol. 64, N° 3, pp. 832-837, (1978).

(9) D.S. STEVENS, H.F. TIERSTEN "An Analysis of doubly rotated quartz resonators utilizing essentially thickness modes with transverse variation", J. Acoust. Soc. Am., Vol. 79, N° 6, pp. 1811-1826, (1986).

(10) P.C.Y. Lee, M.S.H. Tang "Initial Stress field and resonance frequencies of incremental vibrations in crystal resonators by finite element method" Proc. of 40th A.F.C.S., pp. 152-160, (1986).

(11) H.F. TIERSTEN "Rotational Invariance, Biasing States and Quartz Resonators", Proc. of the 12th Int. Cong. on Acoustics, Can. Acoust. Assoc., Bauregard Press Ltd, Toronto 1986.

(12) S.P. TIMOSHENKO "Théorie de l'Elasticité" Librairie Polytechnique Ch. Beranger, Paris 1935.

(13) W.J. SPENCER "Observations of resonant vibrations and Defect Structure in Single Crystals by X-Ray Diffraction Topography", in Physical Acoustics, Vol. V, Academic Press, NY, 1968.

(14) B. DULMET, H.F. TIERSTEN "On the Effect of Edge Boundaries on Nearly Trapped Vibrations of Contoured Rectangular Quartz Resonators", Proc. of 1986 Ultrasonics Symposium, IEEE N° 86CH2375-4.

(15) B. DULMET, R. BOURQUIN "Sensibilité d'un résonateur à quartz plan-convexe à la compression diamètrale" First European Forum on Time and Frequency, Besançon, March 1987.

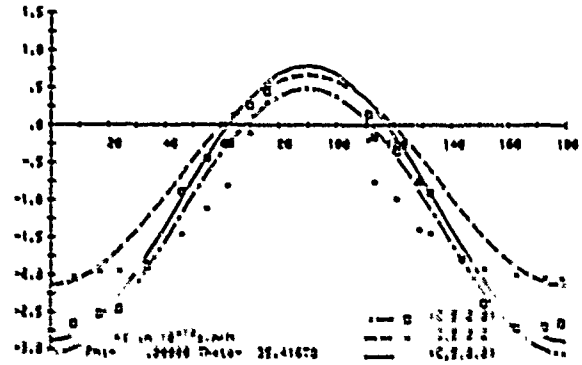


Fig.3

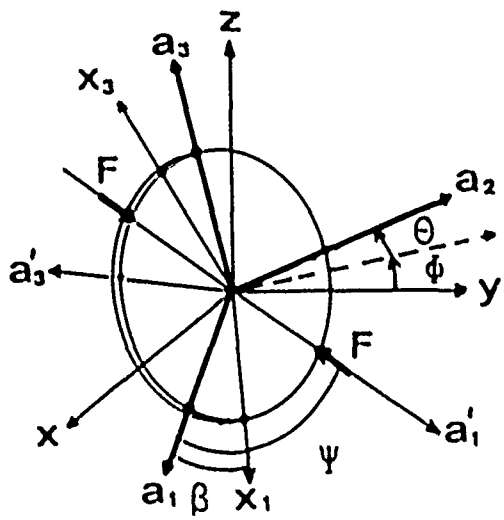


Fig. 1

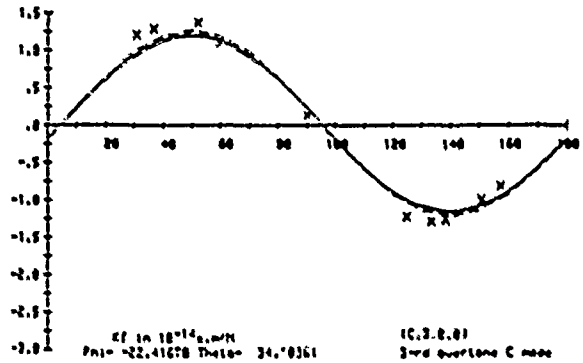


Fig.4

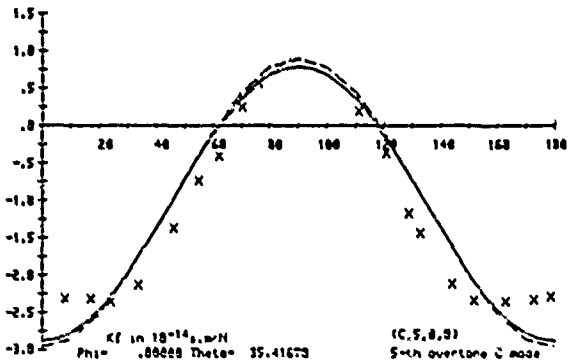


Fig.2

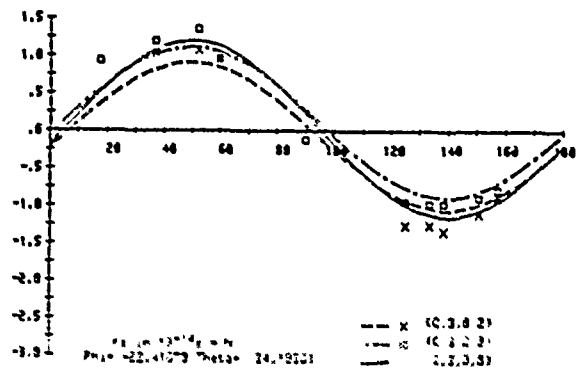


Fig.5

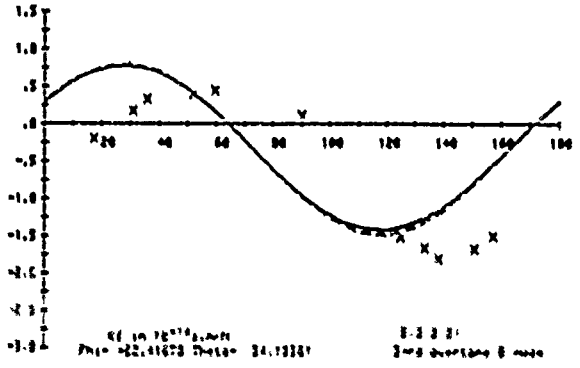


Fig.6

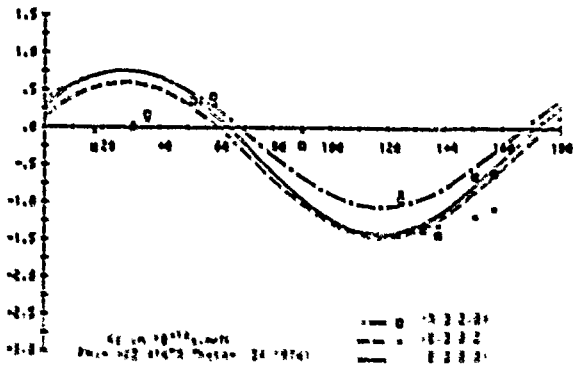


Fig.7

41st Annual Frequency Control Symposium - 1987

AT QUARTZ STRIP RESONATORS

L. N. Dworsky, Motorola Inc.
Schaumburg, Illinois 60196

Summary

An AT-cut quartz strip resonator is a rectangular plate of quartz, typically fully electroded in the (narrower) width direction. Strip resonator analyses show that the X thickness-shear vibration typical of AT quartz resonators can couple strongly to other modes, principally X-Y flexural modes, in these devices. These couplings will, in general, cause deterioration of the motional capacitance and frequency-temperature performance of the resonators. The dimensional tolerances necessary to assure repeatable satisfactory performance of strip resonators has for many years made them a second choice to the more easily fabricated trapped-energy resonators. However, progress in quartz grinding and etching technologies along with the electronics industry's continual push towards smaller products have gradually caused the strip resonator to become the preferred resonator choice in many applications.

This paper reviews analyses of AT strip resonators, including the predominant couplings, and the device performance that may be expected.

Introduction

For many years now most AT quartz resonators manufactured have been trapped energy devices. In these devices, the mass loading of the electrodes creates an effective acoustic cavity which is confined to the quartz region between the electrodes. The principle result of creating this cavity is that the device operation is essentially independent of the nature and exact location of the edges of the plate. The resulting device performance is then easily repeatable and largely free of coupling to unwanted vibration modes (activity dips) while stringent tolerances on plate size and electrode location are not required.

Unfortunately, trapped energy devices are not efficient utilizers of quartz real estate - that is, the unelectroded area does not contribute to the motional capacitance and is in a sense wasted. The trend in electronic equipment today is towards miniaturization. This means that there is a demand for smaller quartz devices. Also, modern photolithographic device manufacturing techniques which take advantage of wafer scale processing relate cost directly to device surface area. The two factors add up to a demand for a reduction of unelectroded quartz surface area.

The obvious approach to reducing unelectroded area is simply to eliminate it - i.e. fully electrode the quartz surface. Maximum utilization of surface area on a quartz wafer is achieved by building rectangular devices. Combining these two factors, we have the current interest in rectangular, fully electroded, (strip) resonators.

The AT strip resonator is not a new device, as the bibliography at the end of this paper clearly shows. However, until recently, lack of convenient computer design capability and lack of photolithographic tolerance capability have made production of these devices an unattractive prospect - especially when compared to the more easily produced trapped energy devices. Now, however, both the needs and the means have matured and many companies are successfully producing AT strip resonators.

Background

The simplest AT resonator to analyze is the one that can't be built - the plate of thickness (Y dimension) 2h which is infinite in extent along the surface (X and Z dimensions). The electrical circuit couples only to X thickness shear (TS1) vibrations and these vibrations in turn do not couple to any other vibrations. The (fundamental) resonance frequency is inversely proportional to h and is given by

$$\omega = \frac{\pi}{2h}\left(\frac{\bar{c}_{66}}{\rho}\right)^{1/2} \qquad [1]$$

where all quartz constants are referred to the AT reference system and are as defined by Tiersten and Smythe (1981).

The ratio of the static to motional capacitance (Co/Cmot) for the infinite AT plate is approximately 160.

If we assume that only the TS1 mode is present for a rectangular finite AT plate, we find that the displacement wave function for the fundamental vibrational mode is given by

$$u_x = \cos\left(\frac{\pi x}{2L_x}\right)\sin\left(\frac{\pi y}{2h}\right) \qquad [2]$$

where 2Lx is the width of the plate in the X direction. Note that the length (2Lz) does not enter into the above equation. The resonance frequency is given by

$$\omega = \left(\omega_0^2 + \frac{\bar{c}_{11}}{\rho}\left(\frac{\pi}{2L_x}\right)^2\right)^{1/2} \qquad [3]$$

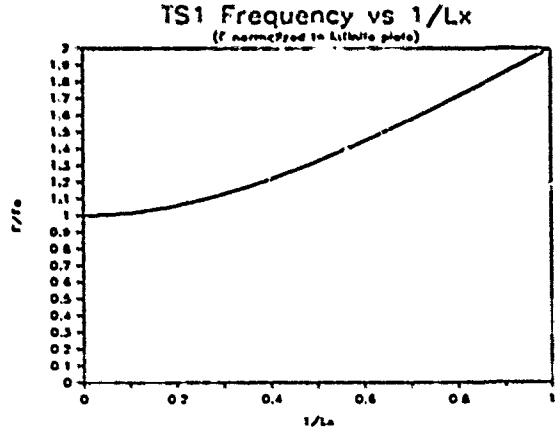


Figure 1

For practical cases, Lx >> h and therefore the frequency is a slowly decreasing function of Lx. Figure 1 shows the functional form of equation [3] vs. 1/Lx. The capacitance ratio for this structure is approximately 200.

Unfortunately, physical strip resonators do not behave as predicted by this simple model. Since the electrodes travel out to the (X) edges and the device width is not as large a multiple of plate thickness as is typical of round plate trapped energy resonators, the edges are a non-negligible part of the vibrating system. At these edges other modes of vibration couple mechanically to the TSI mode and the observed performance is often very different from the simple case. In order to understand the effects of a second mode mechanically coupling to our desired mode, let us postulate a simple coupled resonator system as shown in Figure 2.

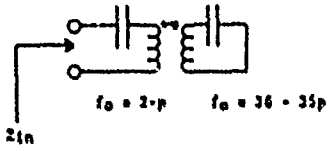


Figure 2. Coupled Resonator Model

The TSI mode is represented by the left hand resonator of figure 2, and some as yet unspecified mode is represented by the right hand resonator. Note that we only have direct access to the left hand resonator – all of our electrical experimental information about the right hand resonator comes from observing the performance of the total system, as seen through our access to the left hand resonator. Anticipating actual results, we will assume that the left hand self-resonance frequency is a slowly decreasing function of a parameter P, while the right hand self resonance frequency is a rapidly decreasing function of the same P, with the two frequencies being the same when P = 1:

$$\omega_{o_1} = 2 - P \qquad [4]$$

$$\omega_{o_2} = 36 - 35P \qquad [5]$$

$$\omega_{o_1}(P = 1) = \omega_{o_2}(P = 1) = 1 \qquad [6]$$

Figure 3 shows these two self-resonance frequencies as a function of P, which is labelled width in figure 3 in further anticipation of actual results.

Due to the coupling between the resonators (modes), the observed resonances are not as shown in figure 3, but instead are as shown in figure 4.

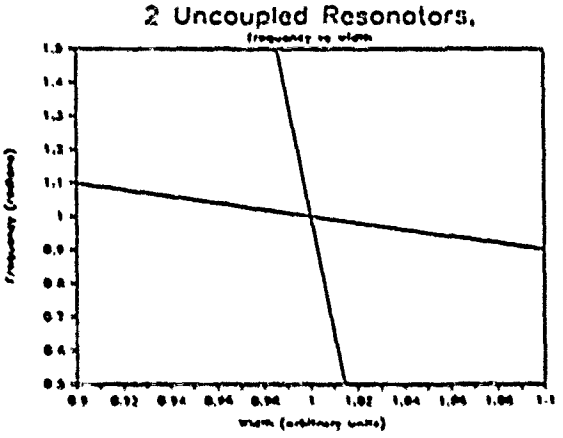


Figure 3

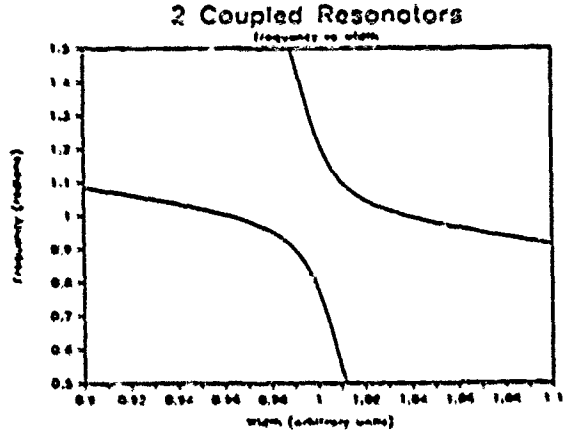


Figure 4

As may be seen, when P is not close to 1, we see the two individual resonances. On the other hand, when P is close to 1 the coupling perturbs the resonances and at P = 1 we do not see any resonance at the original frequency. Note also that when P is close to 1 it is impossible to identify either resonance with a particular resonator – both observed resonances are resonances of the total system.

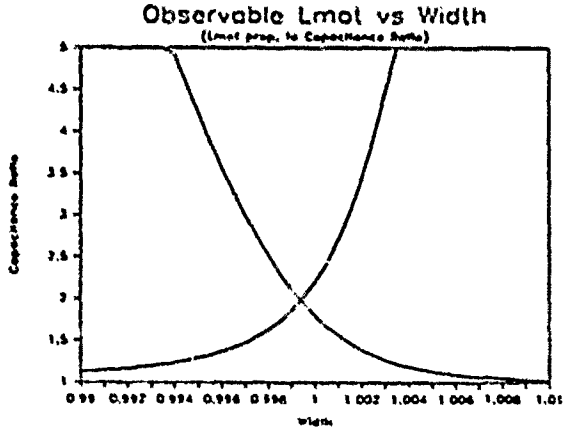


Figure 5

Figure 5 shows the effective inductance seen at the two resonances (normalized to 1) of the coupled system, again as a function of P. This effective inductance is proportional to the capacitance ratio that would be seen for an actual quartz resonator. As would be expected, for P not close to 1, we see essentially only the inductance of the left hand resonator. On the other hand when P is close to 1, we see two inductances (one at each resonance) but neither of them are "as good" as the inductance of the uncoupled left hand resonator.

From the above contrived example we would expect that mechanical coupling of the TS1 mode to other modes is not desirable - when the resonance frequencies of the TS1 and a coupling mode are close we see multiple resonances and deteriorated motional parameters. It is reasonable to predict that frequency-temperature response will also differ from the ideal Bechmann Curve AT resonator response. The principle modelling challenge in the design of AT strip resonators is to predict mode coupling as a function of mechanical parameters - so that we may design to best avoid it.

Device Modelling

A detailed account of the several device analyses which have appeared in the literature is beyond the scope of this treatment. It is of value, however, to summarize 2 of these analyses so that the underlying assumptions may be presented and understood.

Mindlin's analysis (1951) begins with the (6) stress equations for AT quartz:

$$T_1 = T_{11} = C_{11}\frac{\partial u_x}{\partial x} + C_{12}\frac{\partial u_y}{\partial y} + C_{13}\frac{\partial u_z}{\partial z} + C_{14}\left(\frac{\partial u_y}{\partial z} + \frac{\partial u_z}{\partial y}\right) \quad [7]$$

+ 5 others, and the (3) equations of motion:

$$\frac{\partial T_1}{\partial x} + \frac{\partial T_6}{\partial y} + \frac{\partial T_5}{\partial z} = -\rho\omega^2 u_x \quad [8]$$

+ 2 others.

The assumed boundary conditions are that T2 = 0 throughout the (thin) plate, that T4 = T6 = 0 on the major surfaces, and that T1 = T6 = 0 on the X edges.

The X and Z displacement functions are assumed to be a simple linear function of y multiplied by some function of X and Z, while the Y displacement function is assumed to be only a function of X and Z:

$$u_x = y\Psi_x(x,z) \quad [9]$$

$$u_y = \eta(x,z) \quad [10]$$

$$u_z = y\Psi_z(x,z) \quad [11]$$

We now go through a somewhat lengthy algebraic procedure of substituting the functions into the equations and then integrating over Y. Assume that since Lz >> Lx all partial derivatives with respect to Z are very small and hence ignorable. We then substitute into our boundary conditions and get equations of the form

$$\frac{d^2\Psi_x}{dx^2} + (P_1 - P_2)\Psi = P_2\frac{d\eta}{dx} \quad [12]$$

$$\frac{d\Psi_x}{dx} + \frac{d^2\eta}{dx^2} + P_3\eta = 0 \quad [13]$$

where the Pi are known functions.

For the resonator to have a non-zero motional capacitance, Ux must be even in X, therefore let

$$\Psi_x = A_x\cos(\beta x) \quad [14]$$

and from the form of the differential equation immediately that

$$\eta = A_y\sin(\beta x) \quad [15]$$

which is odd in X.

This leads to the homogeneous system

$$\begin{bmatrix} (P_1 - P_2 - \beta^2) & -\beta \\ -\beta & (P_3 - \beta^2) \end{bmatrix} \begin{bmatrix} A_x \\ A_y \end{bmatrix} = 0 \quad [16]$$

Equation [16] has 2 roots, each with a corresponding set Ax:Ay. In general, therefore, the solutions are a sum over these 2 roots,

$$\Psi_x = \sum B_i A_{xi}\cos(\beta_i x) \quad [17]$$

$$\eta = \sum B_i\sin(\beta_i x) \quad [18]$$

and again using the boundary conditions we arrive at the homogeneous system

$$\begin{bmatrix} A_{x1}\beta_1\sin(\beta_1 L_x) & A_{x2}\beta\sin(\beta_2 L_x) \\ (\beta_1 + A_{x1})\cos(\beta_1 L_x) & (\beta_2 + A_{x2})\cos(\beta_2 L_x) \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = 0 \quad [19]$$

Solving equation [19] (that is, setting the determinant of the coefficients = 0) leads to a transcendental equation with an infinite number of roots. These roots correspond to frequencies of resonance, and their values turn out to be a function of the Lx:h ratio only.

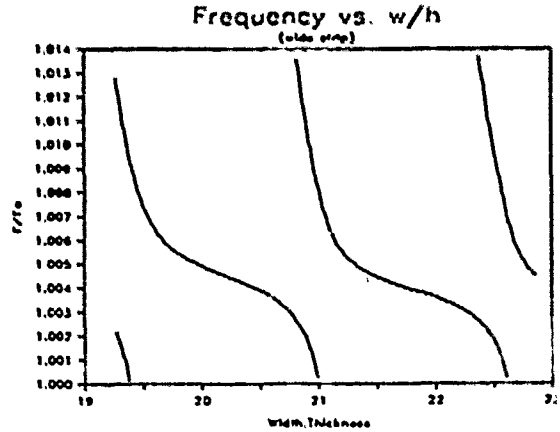


Figure 6

Figure 6 shows those roots (frequencies) near the ideal TS1 strip resonator frequency (equation [3]) versus Lx:h. As may be seen, the frequency response is that of an ideal TS1 strip periodically coupling to a mode with a very steep frequency:width dependance. Examining the displacement functions shows that the TS1 mode is coupling to even overtones of X-Y flexure. If we were to extend the axis of figure 6 over a wider range we would further see that the coupling is very strong for small values of Lx:h, and gets progressively weaker as Lx:h increases. (This "strength" of coupling trend is seen as a gradual transition from coupling regions that look like figure 4 towards those that look like figure 3.

Milsom's analysis (1981) begins with the differential equations obtained by combining the stress equations and the equations of motion, (given here only in X and Y),

$$\rho\omega^2 u_x + C_{11}\frac{\partial^2 u_x}{\partial x^2} + (C_{12}+C_{44})\frac{\partial^2 u_y}{\partial x \partial y} + C_{44}\frac{\partial^2 u_x}{\partial y^2} = 0 \quad [20]$$

$$\rho\omega^2 u_y + (C_{12}+C_{44})\frac{\partial^2 u_x}{\partial x \partial y} + C_{44}\frac{\partial^2 u_y}{\partial x^2} + C_{22}\frac{\partial^2 u_y}{\partial y^2} = 0 \quad [21]$$

Beginning with the observation that Ux is even in X, by to satisfy [20] and [21] we see that Ux and Uy must be of the forms

$$u_x = A_x \cos(\beta_x x)\sin(\beta_y y) \quad [22]$$

$$u_y = A_y \sin(\beta_x x)\cos(\beta_y y) \quad [23]$$

Equations [22] and [23] satisfy [20] and [21] exactly, leading to a set of homogeneous equations. Setting the determinant of the coefficients = 0 leads to a quadratic equation in 3 unknowns. By guessing at two of these unknowns we may solve for the third - actually two values for the third unknown since the homogeneous equation is quadratic:

Guess	Calculate		
ω, β_x	β_{y1}, β_{y2}	$A_y = 1$	A_{x1}, A_{x2}

The general form of the solutions is now

$$u_x = \sum B_n A_n \cos(\beta_x x)\sin(\beta_{yn} y) \quad [24]$$

$$u_y = \sum B_n \sin(\beta_x x)\cos(\beta_{yn} y) \quad [25]$$

The above solutions are now substituted into the boundary conditions at the major surfaces, T2 = T6 = 0. Again we arrive at a set of homogeneous equations, leading to an equation which allows us to reduce the number of unknowns to 1 (again, 2 cases). The general form of the solutions is now

$$u_x = \sum_m C_m \cos(\beta_{xm} x)\sum_n B_{mn} A_n \sin(\beta_{ymn} y) \quad [26]$$

$$u_y = \sum_m C_m \sin(\beta_{xm} x)\sum_n B_{mn} \cos(\beta_{ymn} y) \quad [27]$$

Thus far we are working without any approximations The next step is to substitute equations [26] and [27 into the X edge boundary conditions, T1 = T6 = 0. Thi leads, unfortunately, to 2 equations which cannot b satisfied for all Y, and some approximation will b necessary in order to continue. Noting that T1 is odd i Y and T6 is even in Y, we take weighted integra approximations

$$\int T_{11}\sin(k_n y)dy = 0 \quad [28$$

$$\int T_{12}\cos(k_n y)dy = 0 \quad [29$$

and we get our 3rd set of homogeneous equations Again, we take the determinant of the coefficients an set it = 0. We now have a sufficient number o equations to solve for all unknowns. The system i implicit and numerical procedures must be used to adjus all unknowns until this final determinant = 0.

Once the solution is complete, the motional capacitanc may be found by substituting the resonance frequency Ux, and Uy into

$$C_{mot} = \frac{e_{26}^2}{\rho\omega^2 t_x^2}\,\frac{\left[\int\int u_x(x,\,h,\,z)dxdz\right]^2}{\int\int\int\left(u_x^2 + u_y^2\right)dxdydz} \quad [30$$

Another parameter of interest may be seen by examinin the denominator of equation [30]. This denominator i the sum of the (kinetic) energies of the Ux and U vibrations. Since these energies must be calculated i order to calculate Cmot, it is an easy matter to take th ratio of these energies and look for some insight int the coupled system of resonators.

The resonance frequencies vs Lx:h predicted by thi analysis are essentially identical to those predicted b the Mindlin analysis, (Figure 6).

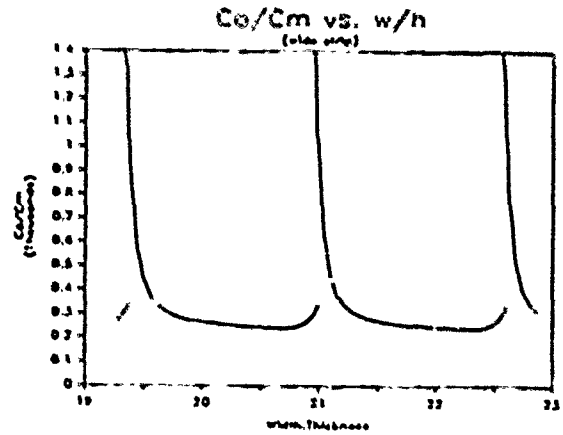


Figure 7

Figure 7 shows the capacitance ratio (Co/Cmot) vs Lx:h As was predicted by the simple circuit model, th capacitance ratio looks best away from regions of stron mode coupling. The lowest value of Co/Cmot i approximately 240. Note, however, that this lowest valu occurs close to a region of mode coupling, an manufacturing tolerances will make designing for thi best capacitance a poor practical decision.

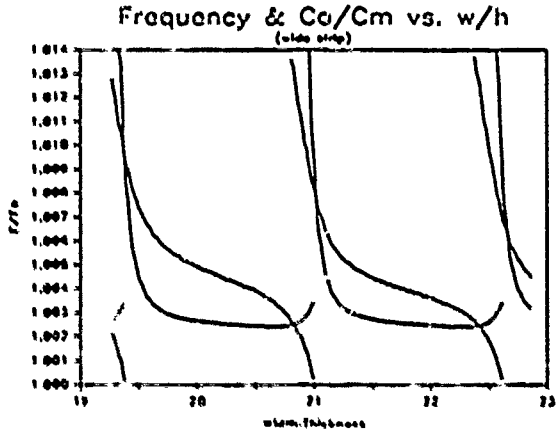


Figure 8

Figure 8 shows a superposition of the capacitance ratio and resonance frequency vs Lx:h curves. Clearly, Cmot is best (and most consistent) away from regions of mode coupling.

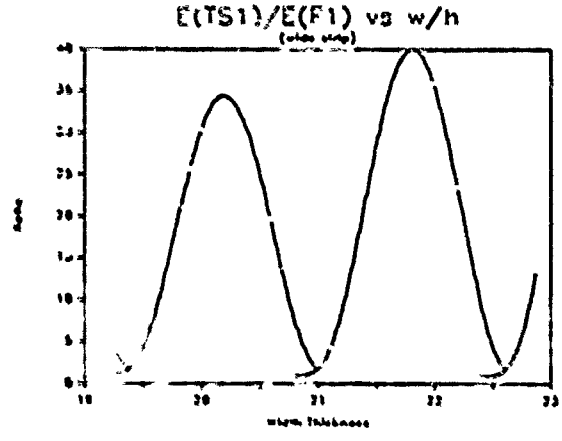


Figure 9

Figure 9 shows the ratio of the Ux energy to the Uy energy (Erat) vs Lx:h. As may be seen, Erat peaks away from regions of mode coupling - i.e. there is very little vibrational energy in Uy away from regions of coupling. Note also that the magnitude of the peaks increases with increasing Lx:h. This means that wider devices, designed for regions of low mode coupling, will behave more like ideal TSI AT devices than will narrower devices.

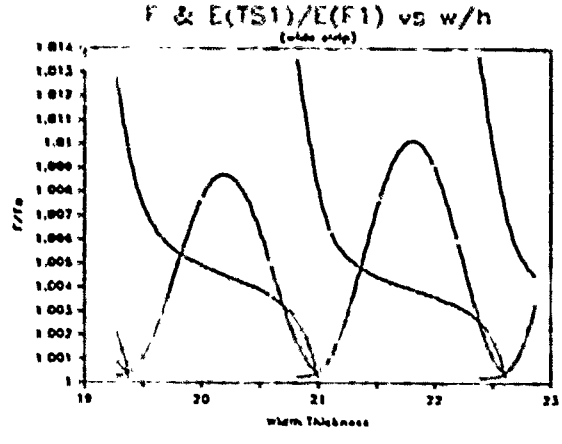


Figure 10

Figure 10 shows Erat superimposed on the resonance frequencies. Note that Erat peaks when Lx/h is essentially mid way between two coupling regions. On the other hand, Erat does not peak where Co/Cmot is lowest (Figure 11).

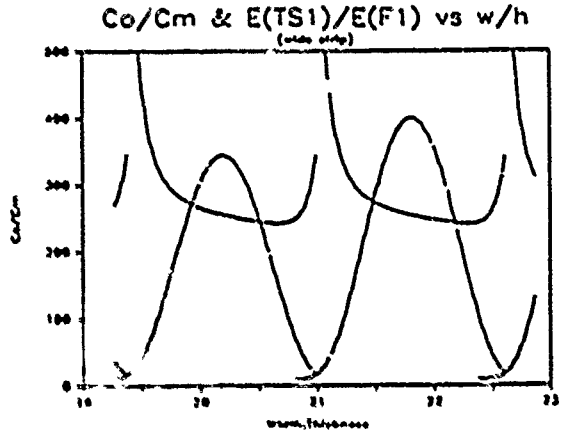


Figure 11

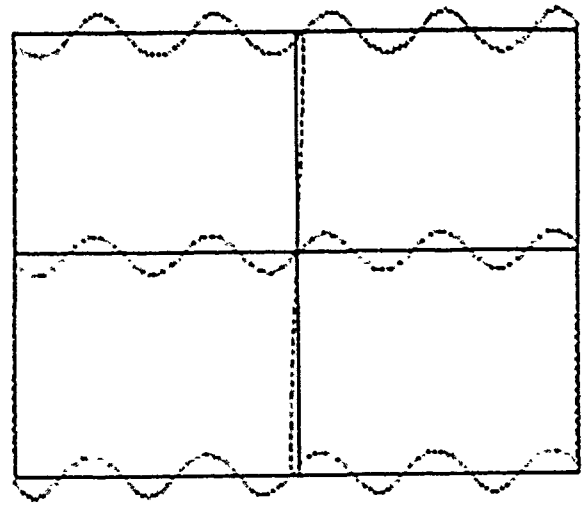


Figure 12. Typical Displacement Functions

Figure 12 shows the displacement functions for a cross section of a device. In this case Lx:h was chosen so that the TS1 mode is coupling to the 10 half-wavelength overtone of F1 (flexure). In figure 12, Lx:h is not shown correctly – the aspect ratio was distorted to give a good perspective for the printed figure. The relative Ux and Uy amplitudes are, however, correct.

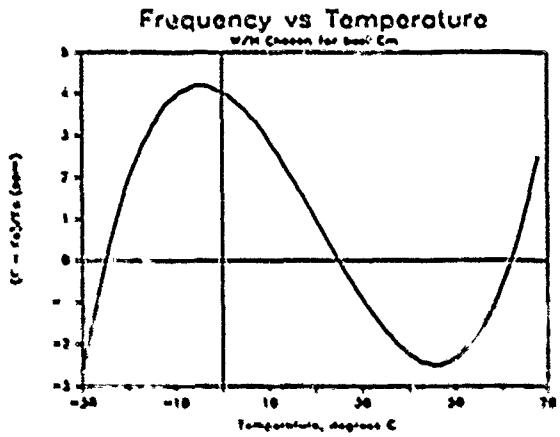


Figure 13

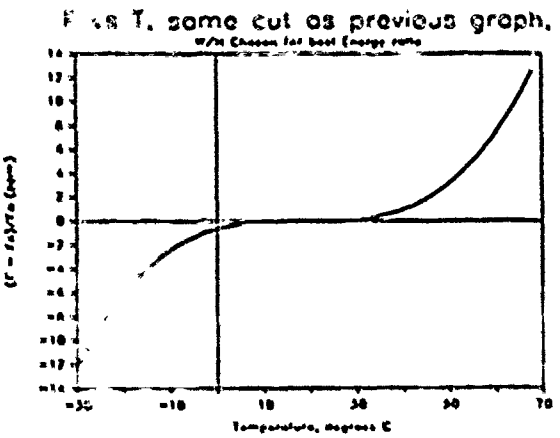


Figure 14

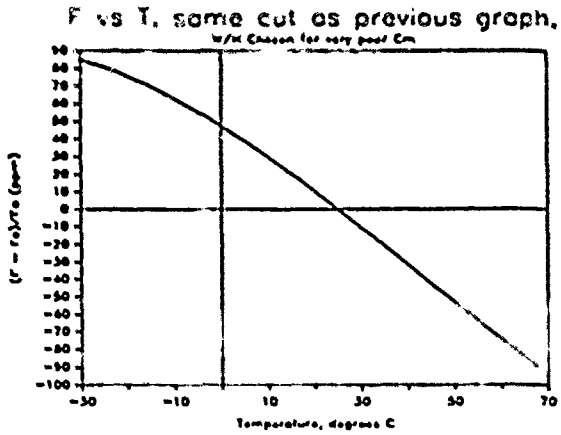


Figure 15

Figures 13, 14 and 15 show the frequency vs temperatu response which can be expected from a strip resona based on the above analysis. The data for figure 13 w obtained by picking a value of Lx:h (approximately 20. which maximizes Cmot and then adjusting the angle the AT cut to get the desired response. This angle tur out to be approximately .25 degrees lower than th required for conventional trapped energy devices. Figu 14 shows the response for a device of the same AT c angle as that of figure 13, but with Lx:h moved to maximum Erat point (approximately 20.2). Note t significant change in response for a very small change device dimensions.

Figure 15 shows the frequency vs temperature respon for a device using the same AT cut angle as the devic of figures 13 and 14, but with Lx:h chosen for a ve poor Cmot (strong mode coupling). In this extreme ca the response no longer resembles a Bechmann curve, a the frequency excursion is almost 200 ppm over the 1 degree temperature range shown.

A comparison of figures 13 and 15 leads to t speculation that different frequency vs temperatu responses might be possible with the same device. F example, figure 16 shows a superposition of figure 13 Bechmann curve, essentially TS1 device) with t response obtained for the same device from the coupli of the TS1 mode to the next lower F1 mode (rememt that figure 6, for example, only shows a portion of t full situation both horizontally and vertically – a vertic line through any point on the axis will cross an infini number of resonances). There are interesti possibilities for a dual mode AT strip resonator with o mode acting as a thermometer.

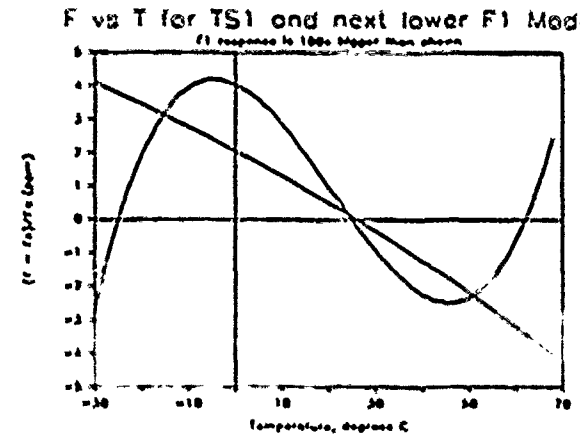


Figure 16

Neither of the above analyses has considered leng effects. If the plate is fully electroded, we may expe a full (infinite) set of anharmonic modes due to t finite device length. Figure 17 shows the resonan frequencies vs Lx:h for a few of these modes. practice there will be anharmonic modes above t operating frequency similar to trapped energy anharmor modes, and also anharmonic modes below the operati frequency. These latter modes are actually higher orc anharmonic modes of the next lower TS1 – F1 couplin

If the electrodes on the device do not extend all t way to the Z edges there will be energy trapping in t Z direction and a corresponding change in the number anharmonic modes. Milsom's (1981) analysis predicts

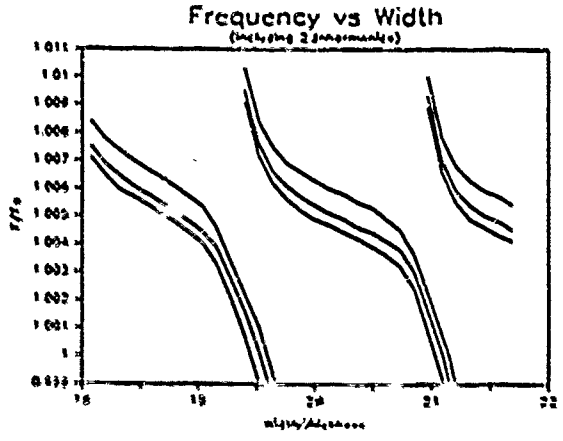


Figure 17

coupling to a TS2 mode which must be taken into account if the energy trapping is to be calculated correctly.

Overtone Operation

Mindlin, in 1983, extended his previous analysis to include various types of mode couplings as well as 3rd harmonic overtone operation. Figure 18 shows some resonance frequency vs Lx:h curves for fundamental mode operation from this newer analysis. As may be seen, in addition to the previously discussed couplings to flexure (Fn) there are couplings to face shear modes (FSn). Since these latter couplings occur in the same region as flexure mode couplings, they are in a region which is avoided based on flexure mode coupling analyses, and lend no new design insights.

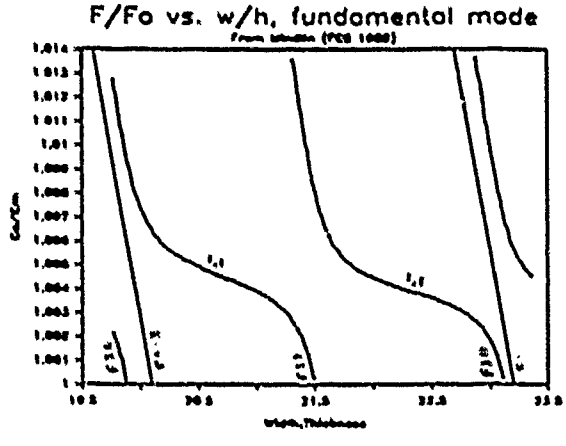


Figure 18

Figure 19 shows some resonance frequency vs Lx:h curves for third harmonic overtone operation. As may be seen, in addition to the flexure (Fn) and face shear (FSn) couplings present in fundamental mode devices, there are also couplings to the anharmonic overtones of both Z thickness shear (TS2) and X thickness shear (TS1) modes. Clearly, overtone design of strip resonators is much more complicated than fundamental mode device design, and tighter tolerance requirements for repeatable device performance are to be expected.

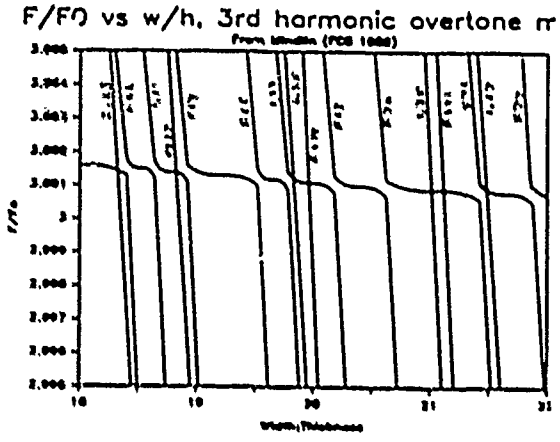


Figure 19

Practical Considerations

Strip resonators may be fabricated either by grinding the rectangular strip from a larger quartz plate or by photolithographically patterning and then etching the strip from a large wafer. The latter approach is to be preferred both because tolerance control is better and because many devices can be produced simultaneously from a large wafer.

Most devices in production today are cantilever mounted on a header which is then encased in a small cylindrical cover. Both epoxy and solder attachment schemes are used for affixing the resonator to the header.

Due probably to the low stress one-ended cantilever mount, strip resonators usually display excellent frequency retrace characteristics over temperature. Due both to this mount and also to their small size, the strips are typically very rugged under shock and vibration. Aging and starting resistance characteristics are very manufacturing process dependant and strip resonators probably do not differ meaningfully from other AT devices insofar as these characteristics are concerned.

Bibliography

J. Balbi & M. Dulmet, "Simple Model for an AT Cut Rectangular Quartz Plate." Proc. 35th AFCS, 1981.

L. Dvorsky, "Dual Frequency, Dual Mode Quartz Resonator." U.S. Patent #4,525,647, 6/85.

Y. Fujiwara & N. Wakatsuki, "LiTaO3 and LiNbO3 Strip-Type Resonators." IEEE Trans. on Ultrasonics Ferroelectrics, and Frequency Control, Vol 34, #1, 1/87.

Y. Fujiwara, et al, "Strip Type Resonator of Lithium Tetraborate." Proc. of the 39th AFCS, 1985.

T. Kato & H. Ueda, "Frequency Temperature Characteristics of Rectangular AT-Cut Quartz Plates." Proc. 33rd AFCS, 1979.

P. Lee, "Extensional, Flexural and Width-Shear Vibrations of Thin Rectangular Crystal Plates." Proc. 25th AFCS, 1971.

P. Lee & C. Lam, "Stresses in Rectangular Cantilever Crystal Plates Under Transverse Loading." Proc. 35th AFCS, 1981.

P. Lee et. al., "Thickness-Shear, Thickness-Twist, and Flexural Vibrations of Rectangular AT-Cut Quartz Plates with Patch Electrodes." Proc. 32nd AFCS, 1978.

T. Meeker, "Extension, Flexure and Shear Modes in Rotated X-Cut Quartz Rectangular Bars." Proc. 33rd AFCS, 1979.

R. Milsom, et. al., "Three-Dimensional Mode-Matching Theory of Rectangular Bar Resonators Using Complex Wavenumbers." Proc. 35th AFCS, 1981.

R. Milsom, "Three-Dimensional Variational Analysis of Small Crystal Resonators." Proc. 33rd AFCS, 1979.

R. Milsom, et al, "Analysis and Design of Coupled Mode Minature Bar Resonators and Monolithic Filters." IEEE Trans. on Sonics & Ultrasonics, Vol 30, #3, 5/83.

R. Mindlin & W. J. Spencer, "Thickness-Twist Overtones of Thickness-Shear and Flexural Vibrations of Rectangular AT-Cut Quartz Plates." Proc. 21st AFCS, 1967.

R. Mindlin, "Third Overtone Quartz Resonator." Proc 36th AFCS, 1982.

R. Mindlin, "Thickness-Twist Vibrations of a Quartz Strip." Proc. 24th AFCS, 1970.

Y. Mochizuki, "Simple Exact Solutions for Thickness Shear Mode of Vibration of a Crystal Strip." Proc. 29th AFCS, 1975.

M. Okazaki & N. Manabe, "AT-Cut Strip Resonators Enclosed in Cylindrical Package." Proc. 38th AFCS, 1984.

M. Onoe et. al., "4 MHz AT-Cut Strip Resonator for Wrist Watch." Proc 31st AFCS, 1977.

M. Onoe & M. Okazaki, "Miniature AT-Cut Strip Resonators with Tilted Edges." Proc. 29th AFCS, 1975.

J. Royer, "Rectangular AT-Cut Resonators." Proc. 27th AFCS, 1973.

H. Tiersten & R. Smythe, "Coupled Thickness Shear and Thickness Twist Resonances in Unelectroded Rectangular and Circular AT-Cut Quartz Plates." Proc. 35th AFCS, 1981.

J. Tomase & L. Dworsky, "Analysis and Design of Coupled Mode AT Rectangular Resonators." Proc. 38th AFCS, 1984.

S. Yamashita, et. al., "A 4.19 MHz Beveled Miniature Rectangular AT-Cut Quartz Resonator." Proc 32nd AFCS, 1978.

S. Yamashita et. al., "New Frequency-Temperature Characteristics of 4.19 MHz Beveled AT-Cut Quartz Resonator. Proc. 33rd AFCS, 1979.

A. Zumsteg & P. Suda, "Properties of a Flat Rectangular Quartz Resonator Vibrating in a Coupled Mode." Proc. 30th AFCS, 1976.

A. Zumsteg et. al., "Energy Trapping of Coupled Modes in Rectangular AT-Cut Resonators." Proc. 32nd AFCS, 1978.

41st Annual Frequency Control Symposium - 1987

A LAGRANGEAN, HIGH FREQUENCY PLATE ELEMENT FOR THE STATIC TEMPERATURE BEHAVIOR OF LOW FREQUENCY QUARTZ RESONATORS.

Y.K Yong
Dept of Civil and Environmental Engineering
Rutgers University
P.O.Box 909
Piscataway, New Jersey 08855-0909
Tel. # (201) 932-3219

SUMMARY

Finite element matrix equations based on the Lagrangean, first order, incremental plate equations of motion superposed on homogenous thermal strains were formulated using virtual work principles. The plate equations are adequate for frequencies up to the fundamental thickness shear and its overtones. Three translational displacements and three rotational displacements making a total of six displacement components were used. A program for an isoparametric, four node quadrilateral element was written and applied to the study of the F-T behavior of flexure mode quartz resonators. Since the coordinate reference frame is fixed, the finite element mesh does not change with temperature. The lumped mass and consistent mass matrices were found to yield practically the same F-T curves. For simple prismatic resonators, the two schemes: reduced/selective integration and incompatible modes produce relatively similar F-T curves. The incompatible modes scheme yielded better results for resonators of more complex shapes such as the tuning fork. For flexure vibrations in a X_i-X_j plane with the length along the X_j axis, the thermal strain component in the X_i direction is predominant. The off-diagonal components of the thermal strain tensor can be neglected. The six degrees of freedom/node element is needed for the F-T behavior of a fully anisotropic flexure mode resonator.

INTRODUCTION

A numerical procedure such as the finite element method may be used to predict with reasonable accuracy the natural frequencies of low-frequency resonators having realistic and complex geometric shapes. Some recent works include the use of a three-dimensional, eight node hexahedral element[1,2] and a plate element with membrane stresses[3] in the study of quartz tuning forks. The tuning fork is one example of a low-frequency resonator possessing a relatively difficult geometric shape. The same finite element program with some minor modifications could be applied to the study of static frequency-temperature behavior, that is, the characteristics of the change in natural frequencies due to a steady and uniform change in temperature. Currently, published results on the various aspects of the application of this numerical method in frequency-temperature predictions seem to be lacking. This paper reports the results of a Langrangean plate element with six degrees of freedom per node employed in a study of the static temperature behavior of flexure mode quartz resonators.

The word "Langrangean" ascribe to a formulation used in reference 4 where all the governing field equations are referred to a fixed reference frame. This approach is theoretically consistent with formulation used by Lee, Wang and Markscenscoff[5] in problems of high frequency vibrations in crystal plates under initial stresses. Since the physical dimensions of the resonator are referred to a fixed reference frame, the finite element mesh does not change with the temperature. While the plate equations are more complicated than the three-dimensional equations, hence resulting in a somewhat more complicated set of finite element equations, there are advantages to its use: (1) The plate element with six degrees of freedom will yield a given degree of accuracy with a smaller number of elements and (2) it is easier to identify and interprete the two-dimensional mode shapes. This is especially true for higher frequency calculations.

The paper is divided into three major sections. The first section details the derivation of the finite element equations using the principle of virtual work. An isoparametric, four node quadrilateral element is chosen. Both the reduced/selective integration[6] and incompatible modes[7] schemes are considered. In the second section, the finite element results are discussed with respect to (1) the preference of the incompatible modes scheme over the reduced/selective integration, (2) the adequacy of the lumped mass matrix when compared with the consistent mass matrix, (3) the role of the thermal expansion coefficient in the thickness direction and (4) the effects of using a three degrees of freedom/node plate element versus that of six degrees of freedom/node element. Some of the current numerical results were compared with Nakazawa et al's[8] experimental results and the three-dimensional finite element results from a previous paper[1]. A brief summary and conclusion are given in the last section.

I. FINITE ELEMENT FORMULATION OF THE TWO-DIMENSIONAL INCREMENTAL PLATE EQUATIONS OF MOTION SUPERPOSED ON HOMOGENOUS THERMAL STRAINS.

a) Constitutive and strain-displacement relations of the first order plate equations.

CH2427-3/87/0000-303 $1.00 C 1987 IEEE

The finite element model is based on the Lagrangean, first order, incremental plate equations of motion superposed on homogenous thermal strains[4]. These equations are adequate for frequencies up to the fundamental thickness shear and its overtones. The interested reader is referred to sections II, V and the first paragraph of section VI in reference 4 for the details in the development of these equations. The constitutive and strain-displacement relations given in Eq.(51)$_{3,4}$ of the same reference are modified slightly and stated here in matrix notation:

$$\text{Constitutive equations, } \underline{t} = \underline{\underline{D}}\,\underline{e} \tag{1}$$

where

$$\underline{t} = [t_1^{(0)}, t_2^{(0)}, t_3^{(0)}, t_4^{(0)}, t_5^{(0)}, t_6^{(0)}, t_1^{(1)}, t_3^{(1)}, t_4^{(1)}, t_5^{(1)}, t_6^{(1)}]^T \tag{2}$$

is an eleven component Lagrangean incremental stress vector, and $\underline{e}$ is an eleven component Lagrangean incremental strain vector:

$$\underline{e} = [e_1^{(0)}, e_2^{(0)}, e_3^{(0)}, e_4^{(0)}, e_5^{(0)}, e_6^{(0)}, e_1^{(1)}, e_3^{(1)}, e_4^{(1)}, e_5^{(1)}, e_6^{(1)}]^T. \tag{3}$$

$\underline{\underline{D}}$ is a matrix of temperature dependent elastic constants referred to the fixed reference frame. The components of $\underline{\underline{D}}$ are given below.

$$\underline{\underline{D}} = \begin{bmatrix} \bar{D}_{11} & \bar{D}_{12} & \bar{D}_{13} & \bar{D}_{14} & \bar{D}_{15} & \bar{D}_{16} & 0 & 0 & 0 & 0 & 0 \\ \bar{D}_{12} & \bar{D}_{22} & \bar{D}_{23} & \bar{D}_{24} & \bar{D}_{25} & \bar{D}_{26} & 0 & 0 & 0 & 0 & 0 \\ \bar{D}_{13} & \bar{D}_{23} & \bar{D}_{33} & \bar{D}_{34} & \bar{D}_{35} & \bar{D}_{36} & 0 & 0 & 0 & 0 & 0 \\ \bar{D}_{14} & \bar{D}_{24} & \bar{D}_{34} & \bar{D}_{44} & \bar{D}_{45} & \bar{D}_{46} & 0 & 0 & 0 & 0 & 0 \\ \bar{D}_{15} & \bar{D}_{25} & \bar{D}_{35} & \bar{D}_{45} & \bar{D}_{55} & \bar{D}_{56} & 0 & 0 & 0 & 0 & 0 \\ \bar{D}_{16} & \bar{D}_{26} & \bar{D}_{36} & \bar{D}_{46} & \bar{D}_{56} & \bar{D}_{66} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{D}_{11} & \bar{D}_{13} & \bar{D}_{14} & \bar{D}_{15} & \bar{D}_{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{D}_{13} & \bar{D}_{33} & \bar{D}_{34} & \bar{D}_{35} & \bar{D}_{36} \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{D}_{14} & \bar{D}_{34} & \bar{D}_{44} & \bar{D}_{45} & \bar{D}_{46} \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{D}_{15} & \bar{D}_{35} & \bar{D}_{45} & \bar{D}_{55} & \bar{D}_{56} \\ 0 & 0 & 0 & 0 & 0 & 0 & \bar{D}_{16} & \bar{D}_{36} & \bar{D}_{46} & \bar{D}_{56} & \bar{D}_{66} \end{bmatrix} \tag{4}$$

where $\bar{D}_{pq} = 2b\, k_{(p)} k_{(q)} D_{pq}$; $k_{(2)}^2 = k_{(4)}^2 = k_{(6)}^2 = \pi^2/12$

$$k_{(1)}^2 = k_{(3)}^2 = k_{(5)}^2 = 1 \tag{5}$$

$$\text{and } \tilde{D}_{pq} = \frac{2b^3}{3}\left(D_{pq} - D_{2p} D_{2q} / D_{22} \right) \tag{6}$$

The term $k_{(p)}$ in Eq.(5) is the usual correction factor for thickness vibrations. The relation for $\tilde{D}_{pq}$ in Eq.(6) is obtained by setting the stress $t_2^{(1)}$ to zero and accommodating the strain $e_2^{(1)}$ in the first order constitutive relations. The elastic constants D_{pq} are dependent on the temperature change $\theta = (T - T_o)^\circ C$ up to the third order:

$$D_{pq} = C_{pq} + D_{pq}^{(1)}.\theta + D_{pq}^{(2)}.\theta^2 + D_{pq}^{(3)}.\theta^3 \tag{7}$$

where T_o is the reference temperature at 25°C, and C_{pq} and $D_{pq}^{(n)}$ are the fundamental elastic stiffness and "Lagrangean" temperature coefficients of elastic stiffness, respectively. The word "Lagrangean" is used to differentiate $D_{pq}^{(n)}$ from the temperature coefficients by Bechmann, Ballato and Lukaszek[10] which are referred to a temperature dependent reference frame. The measured values of C_{pq} are given by Bechmann, Ballato and Lukaszek and the calculated values of $D_{pq}^{(n)}$ are provided in the appendix of reference 4. For completeness, these values are reprinted in Table 1.

There are six displacement components in the strain vector, namely, $u_1^{(0)}$, $u_2^{(0)}$, $u_3^{(0)}$, $u_1^{(1)}$, $u_2^{(1)}$ and $u_3^{(1)}$. The zeroth order components represent translational displacements while those of the first order are dimensionless quantities which measure the rotational displacements. The matrix form of the strain-displacement relations (Eq.(51)$_4$ of ref. 4) is given below.

$$\text{Strain-Displacement Relations, } \underline{e} = \underline{\underline{d}}\,\underline{\underline{b}}\,\underline{u} \tag{8}$$

where $\underline{\underline{d}}$, $\underline{\underline{b}}$ and $\underline{u}$ are, respectively,

$$\underline{\underline{d}} = \begin{bmatrix} .x_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & .x_3 & 0 & 0 & 0 \\ 0 & .x_3 & 0 & 0 & 0 & 1 \\ .x_3 & 0 & .x_1 & 0 & 0 & 0 \\ 0 & .x_1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & .x_1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & .x_3 \\ 0 & 0 & 0 & 0 & .x_3 & 0 \\ 0 & 0 & 0 & .x_3 & 0 & .x_1 \\ 0 & 0 & 0 & 0 & .x_1 & 0 \end{bmatrix} \tag{9}$$

$$\underline{\underline{b}} = \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} & 0 & 0 & 0 \\ \beta_{12} & \beta_{22} & \beta_{23} & 0 & 0 & 0 \\ \beta_{13} & \beta_{23} & \beta_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & \beta_{11} & \beta_{12} & \beta_{13} \\ 0 & 0 & 0 & \beta_{12} & \beta_{22} & \beta_{23} \\ 0 & 0 & 0 & \beta_{13} & \beta_{23} & \beta_{33} \end{bmatrix} \tag{10}$$

$$\text{and } \underline{u} = [u_1^{(0)}, u_2^{(0)}, u_3^{(0)}, u_1^{(1)}, u_2^{(1)}, u_3^{(1)}]^T \tag{11}$$

Matrix $\underline{\underline{d}}$ is a linear differential operator and its elements $.x_i$ (i = 1,2 or 3) signify partial differentiation with respect to the coordinates of the fixed reference frame. The explicit temperature dependency of the strain is reflected in the matrix $\underline{\underline{b}}$ containing elements β_{ij} which are functions of the thermal expansion coefficients $\alpha_{ij}^{(n)}$ and temperature change θ:

$$\beta_{ij} = \delta_{ij} + \alpha_{ij}^{(1)}.\theta + \alpha_{ij}^{(2)}.\theta^2 + \alpha_{ij}^{(3)}.\theta^3 \tag{12}$$

where δ_{ij} is a Kronecker delta. The values of the thermal expansion coefficients were measured by Bechmann, Ballato and Lukaszek[10] and for completeness are provided in Table 2.

b) Finite Element Matrix Equations.

The finite element equations for free vibrations are formulated using the virtual work principle, which is stated as:

$$\delta U_e = \delta W_e \tag{13}$$

where δU_e and δW_e are respectively the element virtual strain energy of internal stresses and the virtual work on the element, that is,

$$\delta U_e = \int_A \delta \underline{q}^T \underline{t}\, dA \quad A \text{ is the plate element area} \qquad (14)$$

$$\delta W_e = \int_A - \delta \underline{y}^T \underline{\underline{m}}\, \ddot{\underline{y}}\, dA \qquad (15)$$

In the expression for the virtual work δW_e, only the inertial action is considered. The surface traction is neglected for free vibration problems. The matrix $\underline{\underline{m}}$ represents the mass of the plate element, namely,

$$\underline{\underline{m}} = \rho \begin{bmatrix} b_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_1 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & b_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & b_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_2 \end{bmatrix} \quad \text{where } \rho \text{ is the density.} \quad b_1 = 2b \quad b_2 = \frac{2b^3}{3} \qquad (16)$$

The plate thickness is 2b.

In the finite element approximation, the generic displacement $\underline{y}$ is represented by the interpolation of a set of displacements $\underline{Q}$ at discrete points called nodes:

$$\underline{y} = \underline{\underline{N}}\, \underline{Q} \qquad (17)$$

where $\underline{\underline{N}}$ is a matrix of interpolation or shape functions and $\underline{Q}$ is a vector of nodal displacements. Hence the virtual displacements in Eq.(15) may be written as,

$$\delta \underline{y}^T = \delta(\underline{\underline{N}}\, \underline{Q})^T = \delta \underline{Q}^T\, \underline{\underline{N}}^T \qquad (18)$$

If the displacement $\underline{y}$ in Eq.(8) is substituted by Eq.(17), the following relation is obtained:

$$\underline{q} = \underline{d}\, \underline{b}\, \underline{\underline{N}}\, \underline{Q} = \underline{\underline{B}}\, \underline{Q} \qquad (19)$$

where $\underline{\underline{B}} = \underline{d}\, \underline{b}\, \underline{\underline{N}}$ is generally known as the strain-displacement matrix which gives strains at any point within the element due to unit values of nodal displacements. Hence the virtual strain $\delta \underline{q}^T$ in Eq.(14) is:

$$\delta \underline{q}^T = \delta(\underline{\underline{B}}\, \underline{Q})^T = \delta \underline{Q}^T\, \underline{\underline{B}}^T \qquad (20)$$

Putting Eq.(19) into the constitutive relations of Eq.(1) yields the stress-displacement matrix equation:

$$\underline{t} = \underline{\underline{D}}\, \underline{q} = \underline{\underline{D}}\, \underline{\underline{B}}\, \underline{Q} \qquad (21)$$

Substitution of Eq.(14) and Eq.(15) into the virtual work principle Eq.(13), using the relevant expressions from Eqs.(17), (18), (20) and (21) yields:

$$\delta \underline{Q}^T \Big[\int_A \underline{\underline{B}}^T \underline{\underline{D}}\, \underline{\underline{B}}\, dA \Big]\, \underline{Q} = \delta \underline{Q}^T \Big[- \int_A \underline{\underline{N}}^T \underline{\underline{m}}\, \underline{\underline{N}}\, dA \Big]\, \ddot{\underline{Q}} \qquad (22)$$

Since $\delta \underline{Q}^T$ is common in Eq.(22), we obtain the following matrix displacement equation for vibration without damping:

$$\underline{\underline{K}}\, \underline{Q} = -\underline{\underline{M}}\, \ddot{\underline{Q}} \qquad (23)$$

where $\underline{\underline{K}} = \int_A \underline{\underline{B}}^T \underline{\underline{D}}\, \underline{\underline{B}}\, dA$ and $\underline{\underline{M}} = \int_A \underline{\underline{N}}^T \underline{\underline{m}}\, \underline{\underline{N}}\, dA$ are the element stiffness matrix and the element consistent mass matrix, respectively.

The aforementioned equation is in the time domain. For harmonic motion,

$$\ddot{\underline{Q}} = -\omega^2\, \underline{Q} \qquad (24)$$

The general eigenvalue problem is obtained when Eq.(23) is written in the frequency domain:

$$\underline{\underline{K}}\, \underline{Q} = \omega^2\, \underline{\underline{M}}\, \underline{Q} \qquad (25)$$

c) Four-Node, Isoparametric Quadrilateral Element.

The 4-node, isoparametric quadrilateral element is one of the simplest element for a plate. The four shape functions[6] are:

$$N_1(\xi,\eta) = \tfrac{1}{4}(1-\xi)(1-\eta)$$
$$N_2(\xi,\eta) = \tfrac{1}{4}(1+\xi)(1-\eta)$$
$$N_3(\xi,\eta) = \tfrac{1}{4}(1+\xi)(1+\eta)$$
$$N_4(\xi,\eta) = \tfrac{1}{4}(1-\xi)(1+\eta) \qquad (26)$$

where ξ and η are the local coordinates. Hence the matrix of shape functions $\underline{\underline{N}}$ in Eq.(17) is,

$$\underline{\underline{N}} = [\underline{\underline{n}}_1,\ \underline{\underline{n}}_2,\ \underline{\underline{n}}_3,\ \underline{\underline{n}}_4] \qquad (27)$$

where $\underline{\underline{n}}_i$ (i=1,2,3 or 4) $= \underline{\underline{I}}\, N_i$. For a plate with M degrees of freedom/node, $\underline{\underline{I}}$ is a MxM identity matrix. The displacement vector $\underline{y}$ can also be written as:

$$\underline{y} = \sum_{i=1}^{4} \underline{\underline{n}}_i \underline{q}_i \qquad (28)$$

where $\underline{q}_i$ is a vector of displacements at node i so that,

$$\underline{Q} = [\underline{q}_1,\ \underline{q}_2,\ \underline{q}_3,\ \underline{q}_4]^T. \qquad (29)$$

In a three degrees of freedom/node plate, similar to the one proposed by Hughes et al.[4], the strains: $e_1^{(0)}$, $e_2^{(0)}$, $e_3^{(0)}$, $e_4^{(0)}$, $e_4^{(1)}$ and $e_4^{(1)}$ along with the respective stresses are neglected. Only three displacement components, namely, one translational displacement $u_3^{(0)}$ and two rotational displacements $u_1^{(1)}$ and $u_2^{(1)}$, are retained. Results of the three degrees of freedom/node plate will be compared with that of the six degrees of freedom/node plate.

There are schemes: incompatible modes[7] or reduced/selective integration[4] which can be introduced to "soften" an otherwise overconstrained quadrilateral plate element. In the incompatible modes scheme, two extra shape functions are included:

$$N_5(\xi,\eta) = 1-\xi^2$$
$$N_6(\xi,\eta) = 1-\eta^2 \qquad (30)$$

These shape functions do not belong to any nodes. The element stiffness terms associated with these functions are removed by static condensation, while those of the element mass terms are reduced by Guyan[9] reduction. If the reduced/selective integration scheme is used, a 1x1 Gauss quadrature integration is performed on the element stiffness terms associated with the shear

strains $e_4^{(0)}$ and $e_5^{(0)}$. A 2x2 Gauss quadrature is employed on the rest of the stiffness terms.

A finite element program which incorporates all the aforementioned options was written. An efficient eigenvalue solver for sparse matrices was added to the program.

II LAGRANGEAN FINITE ELEMENT PLATE ANALYSIS OF THE FREQUENCY-TEMPERATURE BEHAVIOR OF QUARTZ CRYSTAL RESONATORS VIBRATING IN THE FLEXURE MODE.

The resonant frequencies of the fundamental flexure mode of quartz bars or tuning forks and their frequency-temperature behavior are studied with respect to (1) the fineness of the finite element mesh, (2) the type of "softening" scheme, (3) the kind of mass matrix, (4) the role of the thermal expansion coefficient in the thickness direction and (5) the number of degrees of freedom/node in the plate element.

Figure 1 shows all the finite element meshes used in this study. The coordinate system is shown in the upper left corner of the figure, with the X_2 axis going into the plane of the paper. The thickness, width and length are taken respectively in the X_2, X_3 and X_1 direction. Hence the dimensions of a bar are quoted in the same sequence. With the exception of the tuning fork in fig.1c, the flexure mode vibrates predominantly in the X_3-X_1 plane. The tines of the tuning fork vibrate predominantly in the X_3-X_1 plane. Figure 1a gives the finite element meshes in increasing degree of fineness for a 5° X-Cut cant lever. It must be noted at this point that the vibration characteristics of a cantilever are quite different from those of a tuning fork tine. The beam in fig.1b is an NT-Cut bar, (YXwt) +90°/-8.5°/-30°. The dimensions of the tine in fig.1c is the same as those in the cantilever of fig.1a, while the base dimensions are 0.51x0.63x2.55 mm. These crystal cuts and dimensions were chosen so as to facilitate comparisons of some of the present numerical results with the NT-Cut bar and rotated X-Cut tuning forks empirical data provided by Nakazawa et al.

Figure 2 shows the results of the different finite element meshes of fig.1a using either the reduced/selective integration or incompatibe modes scheme. Although the frequencies fluctuate a little, both schemes give turnover temperatures which converge monotonically to the right. For a mesh of 2x20 elements, the two schemes offer reasonably similar frequency-temperature curves. The mesh of 2x20 elements was used in subsequent finite element studies.

The consistent mass matrix is a higher order numerical approximation of the mass continuum of the vibrating body. The lumped mass matrix is the simplest form of representation. Figure 3 shows that while the two mass matrix schemes yield slightly different natural frequencies, they produce frequency-temperature curves which are practically identical. The results are given for two types of "softening" schemes. The mesh of fig.1a was used. The lumped mass matrix is a diagonal matrix; hence it takes up less storage space and is much easier to factorize in eigenvalue calculations.

The fourth order differential equation for a Lagrangean classical plate, derived in previous paper: Eq.(35) of ref.1, display only one component of the tensor β_{ij}, namely, β_{22}. Of interest are the temperature components in β_{22}(see Eq.(12)):

$$\alpha_{22} = \alpha_{22}^{(1)}.\theta + \alpha_{22}^{(2)}.\theta^2 + \alpha_{22}^{(3)}.\theta^3 \qquad (31)$$

α_{22} is physically the component of the homogenou thermal strain in the X_2 direction. The role of α_{22} the frequency-temperature behavior of flexure mod quartz resonators is further explored. Figure exhibits the results of the mesh of fig.1b for fixed-fixed NT-Cut bar. Incompatible modes were used The use of a diagonalized thermal strain tensor, tha is $\alpha_{ij} = 0$ for $i \neq j$, produced a frequency-temperatur curve which was virtually identical the original curve There were relatively small changes in t frequency-temperature curves when all the components except for α_{22}, of the thermal strain tensor were s to zero. Hence, the off-diagonal thermal strai components are insignificant in frequency-temperatu effects and among the diagonal components, t component in the thickness direction is by far t predominant term. This is further illustrated in fig. for a 1° X-Cut tuning fork using the mesh of fig.1c a incompatible modes. The curve with the full therm strain tensor is relatively similar to one in which a the thermal strain components, except for α_{33}, are s to zero. In this case, α_{33} instead of α_{22} predominant because the tuning fork tine flexes in t X_3-X_1 plane. The experimental results from fig.16 Nakazawa et al.[8] is also shown. The numerical curve are quite close to the empirical curve.

One question of particular interest is whether three degrees of freedom/node plate element is adequa for a flexure mode resonator. The result is compar against the present six degrees of freedom/node eleme in fig.6 for a fixed-fixed NT-Cut bar. The mesh fig.1b and either one of the two "softening" sche were employed. The three degrees of freedom/no element yields different frequency-temperature curv because it does not satisfactorily model the flexu modes which are slightly out of the X_3-X_1 plane a which are coupled to a small degree with a torsion mode. This occurs in a fully anisotropic beam, but n in monoclinic cut plates such as a Y-Cut bar.

The results in figures 3 and 6 indicate that t reduced/selective integration and incompatible mod schemes yield relatively similar frequency-temperatu curves. One might then conclude that t reduced/selective integration scheme is the scheme choice since it is computationally cheaper. Howeve there are other factors which may negate t conclusion. Figure 7 shows the results for a 1° X-C tuning fork using the mesh of fig.1c. The incompatil modes scheme gives a curve which is quite close to t empirical curve by Nakazawa et al. T reduced/selective integration scheme produced different frequency-temperature curve. The reason li in the reduced integration of the stiffness te associated with the bending shear strains $e_4^{(0)}$and $e_5^{(0)}$ which implies bending in the X_3-X_1 plane. The tuni fork tine vibrates in the X_2-X_1 plane which perpendicular to the former plane. Any attempt tensor rotate by 90° the field equations so that implied bending occurs in the X_2-X_1 plane is not qui satisfactory because it leads to difficulties in modelling of the crotch of the tuning fork. One mi then be tempted to try uniform reduced integrati Unfortunately, the scheme introduces a larger number zero-energy modes which may lead to some spuri eigenmodes and erroneous results.

Figures 8 and 9 show the comparisons of the pres results using incompatible modes and the meshes fig.1b and fig.1c against the three-dimensional fin element solutions from a previous paper[1] and Nakaz et al.[8] 's data. For the three-dimensional soluti

the off-diagonal components of the thermal strain tensor were neglected and the diagonal components were assumed equal to α_{22}. An eight-node hexahedral element with incompatible modes was employed. The curves by Nakazawa et al. were obtained from their experiments and from their classical beam equation calculations using either Ariga and Bechmann et al.'s constants. In fig.8 the present numerical curve for a free-free NT-Cut bar is observed to compare reasonably good with the empirical data and the three-dimensional finite element solution. Similar comparisons are obtained in fig.9 where the turnover temperatures TZTC's of various rotated X-Cut tuning forks are plotted against the angle of rotation.

It must be noted that the numerically produced frequency-temperature curves changes slightly with the number of elements and the form of finite element mesh employed, that is, the relative size of the elements and the arrangement of the elements in the mesh. Hence it is quite possible by using a certain element mesh to obtain a three-dimensional solution curve which matches closely a plate element curve in either fig.8 or 9.

III CONCLUSIONS

The following conclusions are inferred from the study of a six degrees of freedom per node, Lagrangean, four-node quadrilateral plate element applied to the frequency-temperature problems of flexure mode quartz resonators:

1) Although the lumped mass matrix compared with the consistent mass matrix yields slightly different natural frequencies, it produces frequency-temperature curves which are practically identical. This is true for low frequency vibrations.

2) For simple prismatic resonators, the two "softening" schemes: reduced/selective integration and incompatible modes schemes yield relatively similar frequency-temperature behavior if element meshes of adequate fineness are used. Although the reduced/selective integration scheme is computationally cheaper, it may not be the scheme of choice in resonators with complex shapes, for example a tuning fork, because of flexure vibrations in a plane other than that implicit in the scheme. The incompatible modes scheme has greater generality.

3) For the frequency-temperature problems of flexure mode resonators vibrating in the X_i-X_j plane with the X_j axis along the length of the vibrating element, the thermal strain in the X_i (thickness) direction is by far the predominant component. The off-diagonal components of the thermal strain tensor can be neglected. Reasonably good solutions were obtained even when all the components of the thermal strain tensor, except the one in the thickness direction, were set to zero. This correlates well with the three-dimensional results from a previous paper where the off-diagonal components of the thermal strain were neglected and all the diagonal components were set equal to the component in the thickness direction.

4) In general, a six degrees of freedom/node element is needed for the frequency-temperature predictions of a fully anisotropic flexure mode resonator. This is due to vibrations which are slightly out of the X_i-X_j plane and coupled to a small degree with a torsional mode. The three degrees of freedom/node element is adequate for only certain crystal cuts, for example, monoclinic cuts such as the Y-Cut.

5) Reasonably good comparisons between the present numerical results and empirical data by Nakazawa et al." were obtained for a NT-Cut bar and various rotated X-Cut tuning forks.

These conclusions are deduced from the results of the Lagrangean quadrilateral plate element when used to predict the frequency-temperature characteristics of low frequency flexure in quartz bars and tuning forks. As such, some of the inferences may not be applicable to other modes of vibration, especially those in the higher frequency ranges.

Acknowledgements

The author would like to thank Michael Tang and Peter C.Y. Lee of Princeton University for providing the eigenvalue solver written by David S. Scott of University of Texas at Austin.

REFERENCES

1) Yong, Y.K., "Three-Dimensional Finite Element Solution of The Lagrangean Equations for the Frequency-Temperature Behavior of Y-Cut and NT-Cut Bars," Proceedings of the 40th Annual Symposium on Frequency Control, 1986, pp 179-186.
2) EerNisse, E.P. and Wiggins, R.B., "A Resonator Temperature Transducer with No Activity Dips," Proceedings of the 40th Annual Symposium on Frequency Control, 1986, pp 216-223.
3) Tomikawa,Y., Sato,K., Konno,M. and Johnson,R.A., "Second-Mode Tuning Forks for High Frequencies: Finite Element Analysis and Experiments," IEEE Trans. on Sonics and Ultrasonics, Vol.SU-27, No.5, Sept.1980, pp 253-257.
4) Lee,P.C.Y and Yong,Y.K., "Frequency- Temperature Behavior of Thickness Vibrations of Doubly-Rotated Quartz Plates Affected by Plate Dimensions and Orientations," J. Appl. Phys., 60(7), October 1986, pp 2327-2342.
5) Lee,P.C.Y., Wang,Y.S. and Markenscoff,X., "High Frequency Vibrations of Crystal Plates Under Initial Stresses," J. Acoust. Soc. Am., Vol.57, No.1, Jan 1975, pp 95-105.
6) Hughes, T.J.R., Cohen, M. and Haroun, M., "Reduced and Selective Integration Techniques in the Finite Element Analysis of Plates," Nuclear Engineering and Design, Vol 46, 1978, pp 203-222.
7) Wilson, E.L., Taylor, R.L., Doherty, W.P. and Ghaboussi, J., "Incompatible Dispacement Models," Numerical and Computer Methods in Structural Mechanics, edited by S.J.Fenves, N. Perrone, A.R. Robinson and W.C. Schnobrich, (Academic Press, New York, 1973) pp 42-57.
8) Nakazawa,M., Nakamura,Y and Miyashita,S, "Frequency-Temperature Characteristics of Quartz Crystal Flexure Bars and Quartz Crystal Tuning Forks," IEEE Trans. on Sonics and Ultrasonics, Vol.SU-26, No.5, Sept.1979, pp 369-376.
9) Guyan, R.J., "Reduction of Stiffness and Mass Matrices," AIAA Journal, Vol.3, No.2, Feb. 1965, pp 380.

10) Bechmann,R., Ballato,A.D. and Lukaszek,T.J., "Higher-Order Temperature Coefficients of the Elastic Stiffnesses and Compliances of Alpha-Quartz," Proc. IRE 50, 1962, pp 1812-1822.

TABLE 1: Elastic stiffnesses and Lagrangean temperature coefficients of elastic stiffness of Alfa-Quartz at 25°C

[pq]	$C_{[pq]}$ Pa 10^9	$D^{(1)}_{[pq]}$ Pa/°C 10^4	$D^{(2)}_{[pq]}$ Pa/°C² 10^3	$D^{(3)}_{[pq]}$ Pa/°C³
11	86.74	-5.922	-10.11	-5.370
13	11.91	-8.398	-12.18	-0.6165
14	-17.91	-1.509	0.9746	14.61
33	107.2	-21.18	-16.91	10.04
44	57.94	-10.78	-15.78	-2.643
66	39.88	6.308	6.024	-9.542

TABLE 2: Thermal expansion coefficients of Alfa-Quartz at 25°C (Density, ρ = 2.649 Kg/m³)

ij	$\alpha^{(1)}_{ij}$ 1/°C 10^{-6}	$\alpha^{(2)}_{ij}$ 1/°C² 10^{-9}	$\alpha^{(3)}_{ij}$ 1/°C³ 10^{-12}
11	13.71	6.5	-1.9
13	13.71	6.5	-1.9
14	7.48	2.9	-1.5

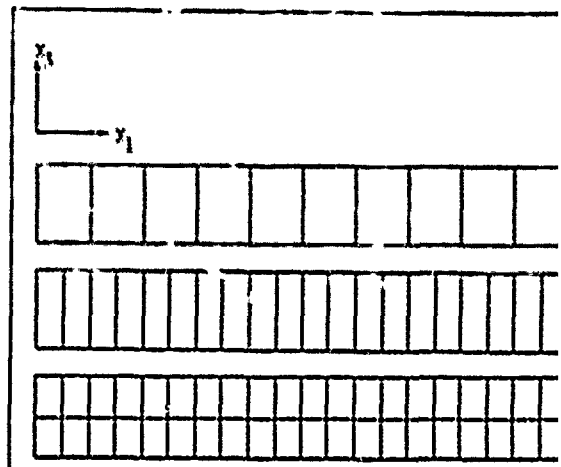


FIG. 1A : FINITE ELEMENT MESH FOR AN AT-CUT CANTILEVER OF A 0.51 x 0.51 x 3.45 mm 5°

FIG. 1B : FINITE ELEMENT MESH FOR AN AT-CUT BAR, DIMENSIONS: 3.6 x 0.8 x 50 mm

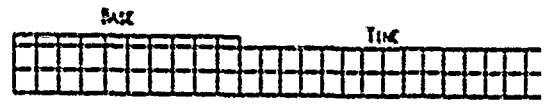
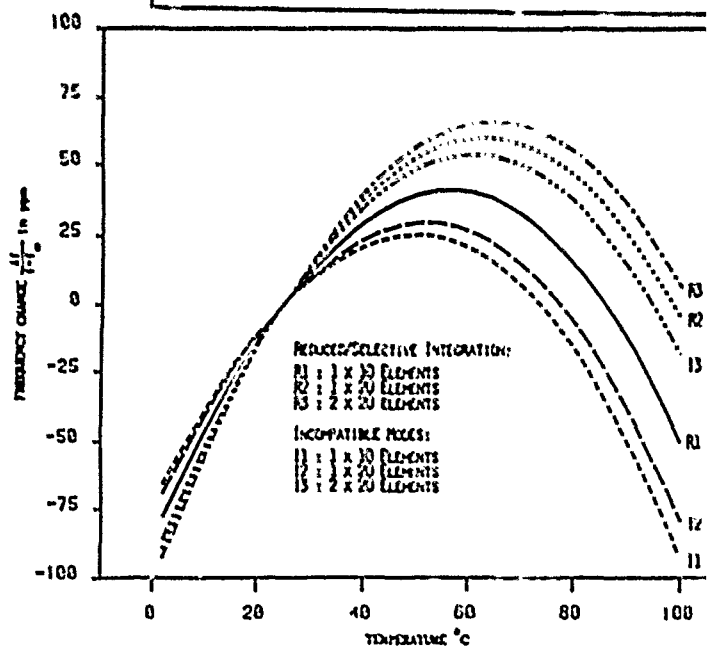
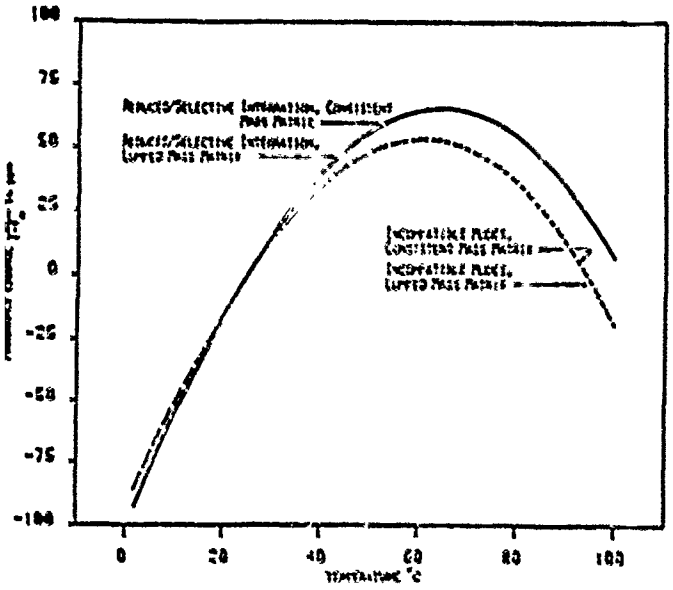


FIG. 1C : FINITE ELEMENT MESH FOR ONE OF THE TINES OF A TUNING FO (ALL DISPLACEMENT COMPONENTS IN THE X_3 DIRECTION ALONG THE CONTOUR AT THE BASE OF THE TUNING FORK WERE SET TO ZERO. THE LEFT END OF THE BASE WAS FIXED.)



Finite Element Mesh	Reduced/Selective Integration		Incompatible Modes	
	Frequency Khz	Turnover Temp °C	Frequency Khz	Turnover Temp °C
1 x 10 elements	37.3657	56.1	37.0719	49.6
1 x 20 elements	37.2884	62.7	37.0521	51.7
2 x 20 elements	37.3075	64.5	37.2398	60.7

FIG.2 : F-T CURVES OF A 5° X-CUT CANTILEVER FOR DIFFERENT ELEMENT MESHES USING REDUCED/SELECTIVE INTEGRATION OR INCOMPATIBLE MODES SCHEME.



Mass Matrix Scheme	Reduced/Selective Integration		Incompatible Modes	
	Frequency Khz	Turnover Temp °C	Frequency Khz	Turnover Temp °C
Consistent Mass	37.3076	64.5	37.3148	68.7
Lumped Mass	37.3046	64.5	37.3154	68.6

FIG.3 : COMPARISON OF F-T CURVES OF A 5" X-CUT CANTILEVER USING CONSISTENT OR LUMPED MASS MATRIX SCHEME.

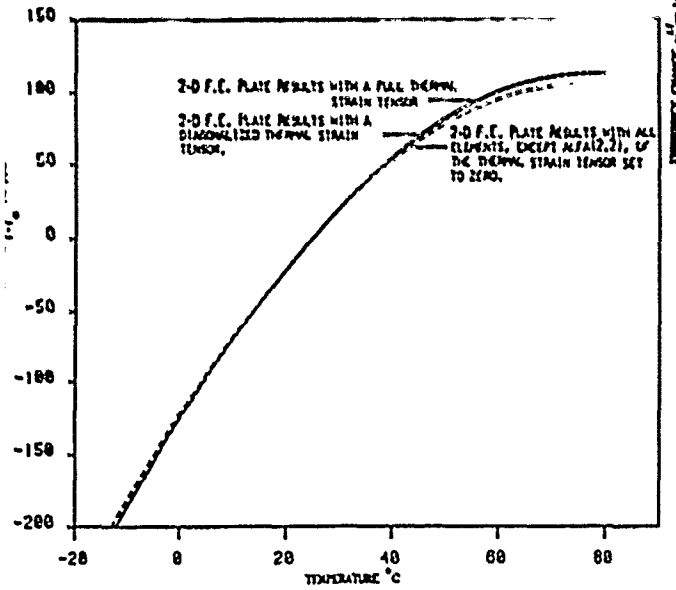


FIG.4 : F-T CURVES OF A FIXED-FIXED NT-CUT BAR USING DIFFERENT REDUCED THERMAL STRAIN TENSORS.

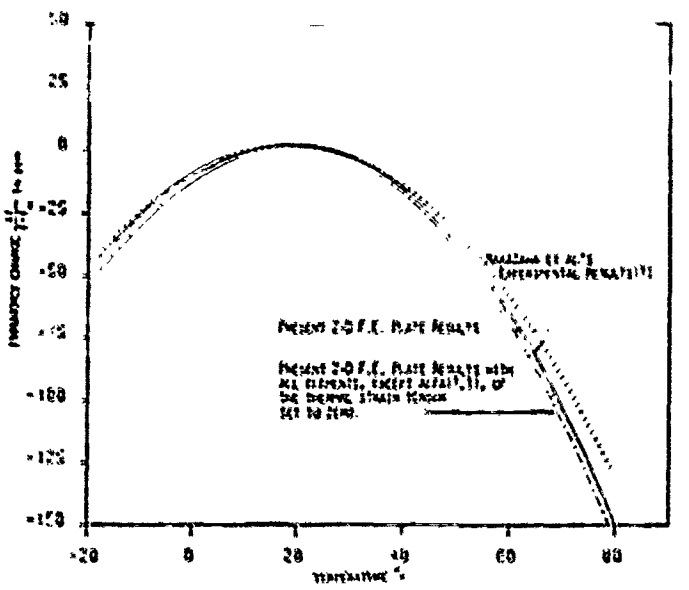
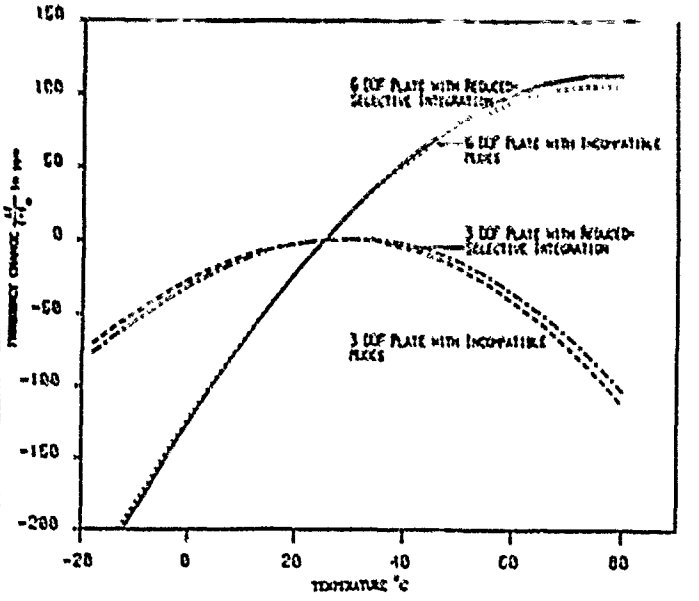


FIG.5 : COMPARISON OF NAKAZAWA ET AL. EXPERIMENTAL F-T CURVE OF A TUNING FORK WITH PRESENT NUMERICAL CURVES



Finite Element Plate	Reduced/Selective Integration		Incompatible Modes	
	Frequency Khz	Turnover Temp °C	Frequency Khz	Turnover Temp °C
6 DOF/Node	21.9762	76.7	21.8167	74.9
3 DOF/Node	22.2528	29.6	22.0947	27.5

FIG.6 : F-T CURVES OF A FIXED-FIXED NT-CUT BAR FOR A 6 DOF/NODE OR 3 DOF/NODE ELEMENT.

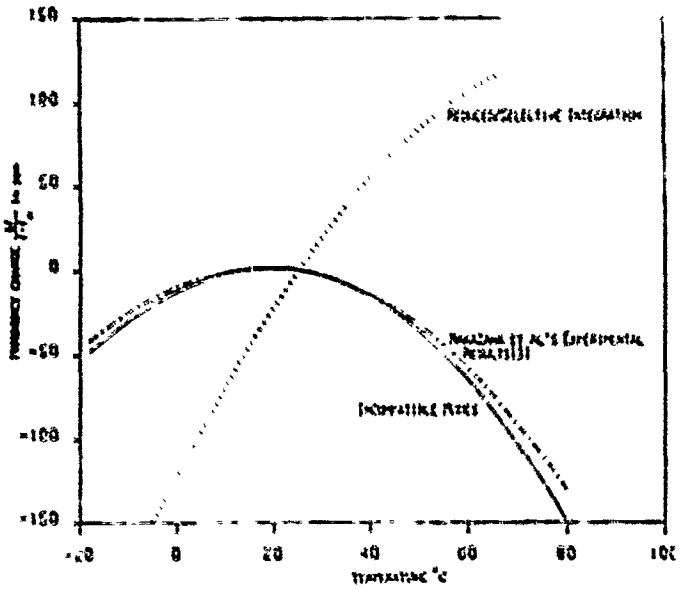


FIG.7 : COMPARISON OF NAKAZAWA ET AL. EXPERIMENTAL F-T CURVE OF A 1° X-CUT TUNING FORK WITH PRESENT NUMERICAL CURVES.

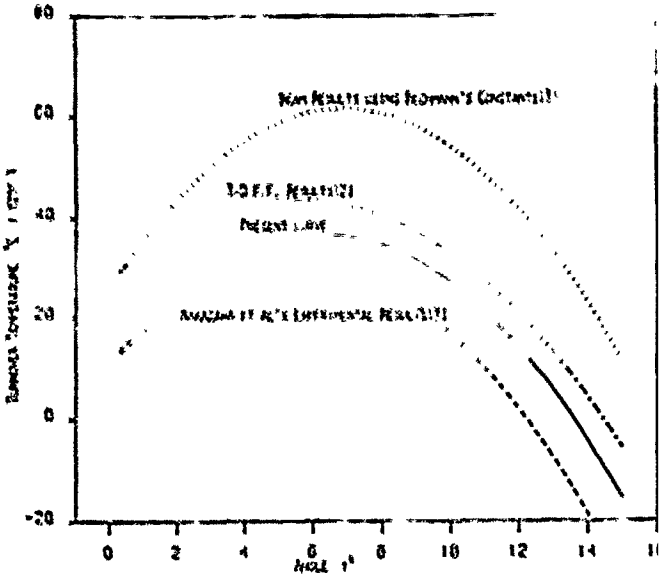


FIG.9 : COMPARISON OF PRESENT RESULTS FOR A TUNING FORK WITH THE 3-D FINITE ELEMENT CURVE AND NAKAZAWA ET AL.'S RESULTS.

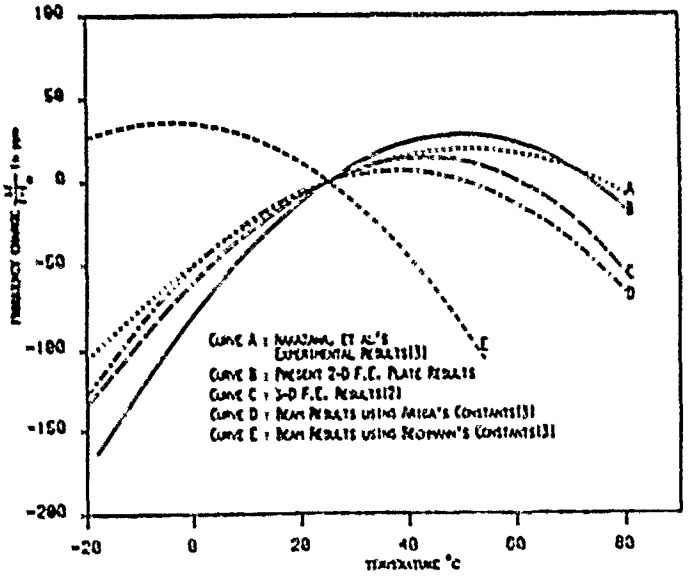


FIG.8 : COMPARISON OF PRESENT F-T CURVE FOR A FREE-FREE NT-CUT BAR WITH THE 3-D FINITE ELEMENT CURVE AND NAKAZAWA ET AL.'S RESULTS.

AN APPROXIMATE EXPRESSION FOR THE MOTIONAL CAPACITANCE OF A LATERAL FIELD RESONATOR

R.C. Smythe
Piezo Technology Inc.
Orlando, Florida 32804

H.F. Tiersten
Rensselaer Polytechnic Institute
Troy, N.Y. 12180

Abstract

A simple, approximate expression is obtained for the motional capacitance of a lateral-field quartz resonator. Comparison with measured values for fundamental mode and third overtone SC-cut resonators shows agreement within 10% to 50%.

Lateral field excitation of SC-cut resonators is of current interest, since either the B-mode or the C-mode may be preferentially excited, and by a suitable arrangement of electrodes it might be possible to excite the two separately. In this letter some simple, approximate, but useful expressions will be developed for the motional capacitance and inductance of lateral field resonators.

Figure 1 shows a rotated Y-cut quartz thickness-shear resonator having the desired electrode configuration. Opposing top and bottom electrodes are connected together. Depending upon the azimuthal orientation of the electrodes with respect to the rotated crystallographic axes of the plate, either the B-mode or the C-mode or both may be excited by the lateral field thus produced.[1] A resonator of this type is called a lateral field resonator (LFR).

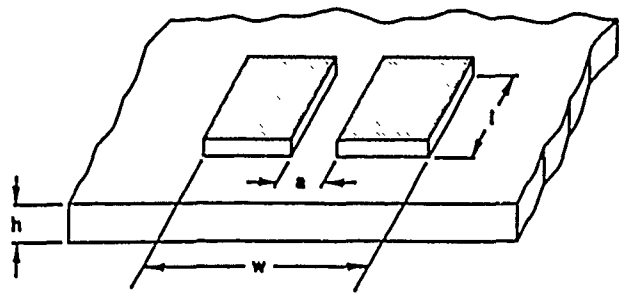


FIG. 1. LFR ELECTRODE CONFIGURATION

Consider first the semi-infinite strip model of Figure 2, in which the mode is confined by the electrodes. Calculation of the motional capacitance requires knowledge of the mode shape and the quasi-electrostatic electric field distribution. In this paper the simplest approximations will be used for both.

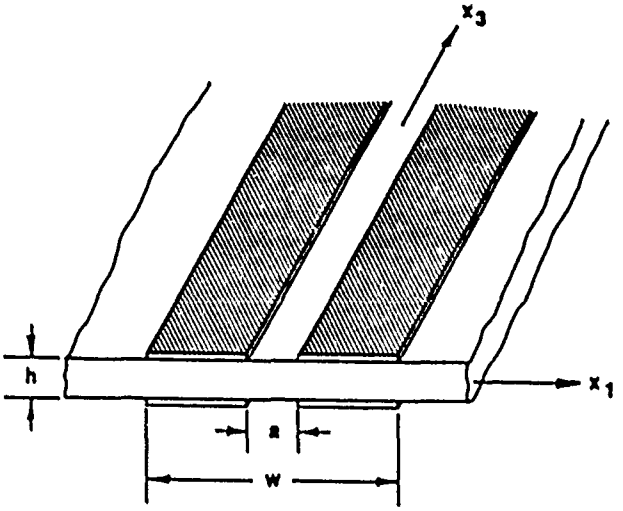


FIG. 2. TRAPPED-ENERGY STRIP ELECTRODE MODEL OF LFR

1) Mode Shape: The mode amplitude is assumed to be uniform over the entire width, w, of the electrode array, and zero elsewhere.

This is not an unreasonable approximation to the actual mode amplitude[2,3], which varies cosinusoidally in the electroded region, has a hyperbolic cosine behavior in the inter-electrode gap, and decays exponentially outside the electroded region.

2) Electric Field: The actual field has both lateral and transverse components. Because of symmetry, only the lateral components contribute to the excitation of the mode. In this elementary treatment we take the simplest approximation for the lateral field, which is that it is of uniform intensity in the gap and zero elsewhere.

For this semi-infinite strip model there is no x_3 dependence of either electric field or mode amplitude.

If this mode shape were excited by a lateral field uniform on w, the motional capacitance would be that for the simple thickness case[1]:

$$C_1 = (8k^2/n^2\pi^2)C \qquad (1)$$

CH2427-3/87/0000-311 $1.00 C 1987 IEEE

where k is the appropriate lateral-field piezoelectric coupling coefficient, n is the overtone, and C is the static capacitance, neglecting fringing, which would be produced by a pair of electrodes through the plate, normal to x_1, and separated by w; i.e., for a uniform lateral field. For a length, l, in the x_3-direction, C is given by

$$C = \epsilon\, l h/w \qquad (2)$$

where ϵ is the dielectric constant in the x_1-direction.*

Substituting in Equation (1), we have:

$$C_1 = (8\epsilon k^2/n^2\pi^2) l h/w \qquad (3)$$

However, we have assumed a field uniform on a and zero elsewhere. To the accuracy of this approximation, only the central portion of the mode, within the gap width a, is excited, but the field intensity is increased by the same proportion, w/a, that the excited region is reduced. As a consequence of the uniform mode assumption, the two offset one another, so that C_1 is still given by Equation (3). To the accuracy of the approximations, then, C_1 is independent of a.

Equation (3) may be rewritten to eliminate the plate thickness, given by

$$h = nN/f \qquad (4)$$

where f is the frequency and N the frequency-thickness constant. Substituting in Equation (3),

$$C_1 = (8\epsilon k^2 N/\pi^2)\ (l/w)\ (1/nf) \qquad (5)$$

Equations (3) and (5) may be applied to the more realistic configuration of Figure 1 by assuming the mode shape in the x_3-direction to be uniform on l and zero elsewhere (although the correct mode shape could readily be included) and ignoring the effect of finite l on the electrostatic field distribution.

The development so far has not been restricted to any particular cut. For the SC-cut, N = 1797 Hz-m, and $\epsilon = 40 \times 10^{-12}$ F/m. If the x_1-direction is chosen to maximize C-mode excitation, then k = 9.38%.[1] Substituting these values in Equation (5) gives

$$C_1 \doteq 0.5 \cdot (l/w)/nf \qquad (6)$$

where f is in MHz and C_1 is in femtofarads.

For a square array (l = w) Equation (5) further reduces to

$$C_1 \doteq 1/2nf \qquad (7)$$

The corresponding expression for the motional inductance, in henries, is

$$L_1 \doteq 50n/f \qquad (8)$$

* Note that C is not the actual static capacitance, which, to the accuracy of approximation 2, is $C_o = \epsilon\, lh/a$.

Equation (5) may also be applied to the contoured resonator. The actual mode shape in this case is essentially determined by the contour rather than the electrodes, and is gaussian in x_1 and x_3 (Reference[4], Equation (3.13)). It is well approximated by a rectangular mode enclosing the inflection points. The corresponding value of l/w is

$$l/w = (M_n/P_n)^{1/4} \qquad (9)$$

where M_n and P_n are effective elastic constants.[4,5] For third and fifth overtone SC-cut contoured resonators, this ratio is approximately .96 and .97, respectively; hence, the coefficient in equations (7) and (8) is essentially unchanged.

The functional dependence shown in Equation (8) was first observed by Warner and Goldfrank[6] for contoured resonators. Their empirical expression for SC-cut LFR motional inductance is

$$L_1 = 17\ h \qquad (10)$$

where h is in millimeters and L_1 in henries. Using Equation (4), this reduces to

$$L_1 \doteq 30\ n/f \qquad (11)$$

or about 60% of the calculated value.

To test the validity of Equation (8), fundamental mode and third-overtone SC-cut resonators were fabricated at 10 MHz. Quartz wafers for the fundamental mode units were uncontoured and polished, with a diameter of 10 mm. The third-overtone units were plano-convex with a radius of curvature of 0.53 m (1 diopter), and a wafer diameter of 14 mm. The electrodes were circular with a bisecting gap, oriented with an azimuth angle, psi, of -22 degrees[1], corresponding to maximum C-mode coupling. Motional inductances were measured using a H-P 4191A RF Impedance Analyzer[7]. In Table 1, the measured values are compared with values calculated from Equation (8). For the uncontoured resonators, agreement is within 15%. For the contoured resonators, the measured values are about two-thirds of calculated, but within 10% of Warner's empirical value. We consider the agreement with calculation to be good, in view of the nature of the approximations. For this azimuth angle, the calculated B-mode coupling is 1.4%. Motional inductance of the B-mode was not measured, but the impedance at resonance was extremely high, indicating that the B-mode is effectively suppressed.

TABLE 1. MEASURED AND CALCULATED MOTIONAL INDUCTANCES

OVERTONE	ELEC. DIM. (mm)		MOTIONAL INDUCTANCE (H)		
			Calculated	Measured	
	Diameter	Gap		Mean	Std. Dev.
1	3.3	0.25	5.0	5.74	0.34
3	7.5	0.5	15.0	9.71	0.39
3	7.5	1.0	15.0	10.04	0.50

Acknowledgments

The authors wish to acknowledge helpful dicussions with Arthur Ballato, U.S. Army LABCOM. They would like to thank A.W. Warner for calling Equation (10) to their attention, and Peter Morley and John Hunt, Piezo Technology, Inc., for assistance in resonator fabrication.

A portion of this work was supported by U.S. Army LABCOM under contract DAAL01-86-C-0006.

References

[1] A. Ballato, E.R. Hatch, M. Mizan, & T. Lukaszek, "Lateral Field Equivalent Networks and Piezocoupling Factors of Quartz Plates Driven in Simple Thickness Plate Modes," IEEE Trans. on Ultrasonics, Ferroelectrics, and Frequency Control, v. UFFC-33, no. 4, July, 1986, pp. 385-393.

[2] H.F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of Thickness-Shear," Proc. 28th Annual Symposium on Frequency Control, 1974, pp. 44-48.

[3] H.F. Tiersten, "Analysis of Trapped-Energy Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness Twist," J. Acoust. Soc. Am., v. 59, no. 4, April, 1976, pp. 879-888.

[4] H.F. Tiersten & R.C. Smythe, "An Analysis of Contoured Resonators Operating in Overtones of Coupled Thickness-Shear and Thickness Twist," J. Acoust. Soc. Am., v. 65, no. 6, June, 1979, pp. 1455-1460.

[5] D.S. Stevens & H.F. Tiersten, "An Analysis of Doubly Rotated Quartz Resonators Utilizing Essentially Thickness Modes with Transverse Variation," J. Acoust. Soc. Am., v. 79, no. 6, June, 1986, pp. 1811-1826.

[6] A.W. Warner, "Lateral Field Resonators," Proc. 39th Annual Symposium on Frequency Control, 1985, pp. 473-474.

[7] R.C. Smythe, "An Automated Resonator Measurement System Using a Reflection Coefficient Bridge," Proc. 35th Annual Symposium on Frequency Control, 1981, pp. 280-285.

Energy Trapping in Plane and Corrugated Resonators : Application to Quartz and Berlinite

J. Détaint, J. Schwartzel, C. Joly, E. Philippot*

Centre National d'Etudes des Télécommunications PAB/BAG/MCT Bagneux 92220 France
* Laboratoire de Physicochimie des Matériaux Inorganiques Université de Montpellier 34000 France

ABSTRACT

Trapped energy resonators of plane and bevelled geometry are widely used for frequency generation and filtering. A model using a solution of the approximate equation governing the lateral dependance of thickness modes is established for plane resonators having nearly arbitrary electrode shape. This model uses solutions in the form of serie of products of Bessel and trigonometric functions and considers the expression of the approximate continuity conditions at a discrete number of points of the electrode edge. This method of solution is applied to study the influence of the shape of the electrodes on the properties of the resonators and to demonstrate on a few examples that the usual circular geometry is not optimal with respect to several criterions. Resonators with electrodes respecting approximatively the lateral anisotropy of the plate (rectangular and elliptical) are considered in these examples. An extension of the model is made to consider plane resonators having shallow grooves outside the electrodes to improve the energy trapping.

I - INTRODUCTION

The thickness mode resonators are the most commonly used since their performances and their frequencies correspond well to the specifications of modern electronic systems. Mathematical models of these resonators having more and more precision were elaborated in recent years. Most of these models were derived from the theories of the thickness vibrations of piezoelectric plates established by Professors Mindlin, Lee and Tiersten. Two principal types of theory can be distinguished :

The first type is based on two dimensional equations(1) of successively higher order of approximation that are obtained from the tridimensional equations of linear piezoelectricity by expansion of the different variables in series of power of the thickness coordinate x_2(2) (or of trigonometric functions of x_2(3)) and integration over the thickness. These theories can consider various modes coupled to the thickness ones.

In the second type of theory, a dispersion relation valid near the overtones of a thickness mode having essentially a thickness dependance similar to the corresponding one dimensional mode and having slow lateral variations is established. Then, a scalar equation governing the lateral variations of the mode is derived from the dispersion relation. The way used to established the dispersion relation ensure the verification of the boundary conditions on the surfaces normal to the thickness, to the second order of approximation.

The scalar equation governing the lateral dependance of thickness shear established by Tiersten and co-worters has been used to obtain several models of the plane and planoconvex resonators (4) (5) (6), of resonators with cylindrical bevels (7) (8), of resonators with unparallel surfaces (9) and of monolithic filters (10).

In this paper solutions of the scalar equation for the case of plane and corrugated resonators are proposed. Some examples of comparison of experimental and computed results are given in the last part of the paper.

II - MODEL OF PLANE RESONATORS WITH ELECTRODES OF ARBITRARY SHAPE

As basis of this model, we use the scalar equation gouverning the lateral dependance of the anharmonics in the viccinity of one overtone (number n) of a given thickness model (noted(1)). This equation was first established for monoclinic plates (4) then for plates of an arbitrary orientation (5). In this case the equation has, in a coordinate system with x_2 normal to the thickness and x_1 and x_3 choosen to eliminate the mixed derivative, (5) the following form :

$$M_n' \frac{\partial^2 \tilde{u}_1^n}{\partial x_1^2} + P_n' \frac{\partial^2 \tilde{u}_1^n}{\partial x_3^2} - \frac{n^2\pi^2 c^{*(1)}}{4h^2} \tilde{u}_1^n + \rho\omega^2 \tilde{u}_1^n =$$

$$\rho\omega^2 (-1)^{(n-1)/2} \frac{e_{26}}{c^{(1)}} \frac{4 V_0 e^{j\omega t}}{n^2 \pi^2}$$

CH2427-3/87/0000-314 $1.00 C 1987 IEEE

In this equation, $\tilde{u}_1^n(x_1, x_3, t)$ is the lateral dependance of $\hat{u}_1^n(x_1, x_2, x_3, t) = \tilde{u}_1^n \sin(n\pi x_2/2h)$.

$\tilde{u}_1^n$ results of the transformation of the conventional component of the displacement u_1 and of the potential ϕ, make to replace the inhomogeneous boundary conditions $\phi = V_e e^{j\omega t}/2$ at $x_2 = h$ by the homogeneous conditions $\phi = 0$. The inhomogeneous term of the equation results of this transformation.

$$u_1^n = \tilde{u}_1^n - \frac{e_{26}}{c^{*(1)}} \frac{V_e e^{j\omega t}}{2h} \cdot x_2$$

$c^{*(1)}$ is either $c^{(1)}$ for the unelectroded part of the resonator or $\bar{c}^{(1)}$ for the electroded part. $\bar{c}^{(1)}$ is the stiffened elastic constant relative to the corresponding one dimensional mode (Eigen-value of the Christoffel Matrix) :

$$\bar{c}^{(1)} = c^{(1)}(1 - 8\,k^{(1)2}/n^2\pi^2 - 2\,\bar{R})$$

is a constant that includes the electrical and mechanical effects of the metallization. $4k^{(1)2}/n^2\pi^2$ is the relative frequency lowering due to the electrical effect of the metallization ; $\bar{R} = 2\rho'h'/\rho h$ is the relative frequency lowering due to the inertia of the electrodes (figure 1). $c^{(1)} = \bar{c}^{(1)}(1-k^{(1)2})$ is a pseudo ordinary elastic constant. $k^{(1)}$ is the coupling coefficient of the corresponding one dimensional mode.

M'_n, P'_n are intricate functions of the material constants and of the plate orientation (5).

To simplify the notations and also to be able to compare the computed results to experimental ones obtained with Y rotated plates, we consider in the following only the case of plates of class 32 crystals with monoclinic symetries. This does not involve any losses of generality in the calculations since it is sufficient to replace M_n, C_{55}, C_{66}, k_{26} by the values of M'_n, P'_n, $c^{(1)}$, $k^{(1)}$ computed for an arbitrary orientation and a particular thickness mode to consider again the general case (11).

As, most frequently, we make also the assumption that the lateral dimensions of the plate, are sufficient to have, for the considered modes, a negligible vibration amplitude at the plate edges (i.e. we consider only perfectly trapped modes).

For the electroded part of the plate the equation is now for an anharmonic mode near the n^{th}. overtone ($n=1,3,5\ldots$) :

$$M_n \tilde{u}_{1,11}^n + C_{55} \tilde{u}_{1,33}^n - \frac{n^2\pi^2}{4h^2} \bar{C}_{66} \tilde{u}_1^n + \rho\omega^2 \tilde{u}_1^n = \rho\omega^2 (-1)^{(n-1)/2} (e_{26}/C_{66}) 4 \frac{V_0 e^{j\omega t}}{n^2\pi^2} \qquad [1]$$

where : $\hat{u}_1 = \tilde{u}_1(x_1, x_3, t) \sin\left(\frac{n\pi x_2}{2h}\right)$

For the unelectroded part we have :

$$M_n u_{1,11}^n + C_{55} u_{1,33}^n - \frac{n^2\pi^2}{4h^2} \bar{C}_{66} u_1^n + \rho\omega^2 u_1 = 0 \qquad [2]$$

Eigen Modes at V=0 ($\tilde{u}_1 \equiv u_1$)

In a first step we calculate the eigen modes at $V_0=0$ of the resonator, the equations are normalized by division by C_{66}, they take the form of Helmotz equations :

For the electroded region :

$$\frac{M_n}{C_{66}} \tilde{u}_{1,11}^n + \frac{C_{55}}{C_{66}} \tilde{u}_{1,33}^n + k_e^2 \tilde{u}_1^n = 0 \qquad [3]$$

For the unelectroded region :

$$\frac{M_n}{C_{66}} \tilde{u}_{1,11}^n + \frac{C_{55}}{C_{66}} \tilde{u}_{1,33}^n + k_1^2 \tilde{u}_1^n = 0 \qquad [4]$$

where :

$$k_e^2 = \frac{\pi^2}{4h^2} \frac{(f^2 - f_{ce}^2)}{f_{66}^2} \qquad [5]$$

$$k_1^2 = \frac{\pi^2}{4h^2} \frac{(f_{c1}^2 - f^2)}{f_{66}^2} \qquad [6]$$

$$f_{ce} = \frac{n}{4h} \sqrt{\frac{\bar{C}_{66}}{\rho}} \qquad [7]$$

(cut-off frequency for the electroded part)

$$f_{c1} = \frac{n}{4h} \sqrt{\frac{\bar{C}_{66}}{\rho}} \qquad [8]$$

(cut-off frequency unelectroded part)

$$f_{66} = \frac{1}{4h} \sqrt{\frac{\bar{C}_{66}}{\rho}} \qquad [9]$$

$$\alpha^2 = \frac{C_{66}}{M_n} \; ; \; \beta^2 = \frac{C_{66}}{C_{55}} \qquad [10]$$

Equations [3] and [4] are transformed in isotropic equations by the transformation $x'_1 = \sqrt{C_{66}/M_n}\,x_1$; $x'_3 = \sqrt{C_{66}/C_{55}}\,x_3$ and then separated in polar coordinates r and τ ($r\cos\tau = x'_1$; $r\sin\tau = x'_3$).

$$\tilde{u}_1^n = R(r).\,T(\tau) \qquad [11]$$

For the electroded part, the following ordinary differential equations are obtained :

$$\left| r^2 \frac{d^2R}{dr^2} + r \frac{dR}{dr} + R\,(r^2k_1^2 - \nu^2) = 0 \right. \qquad [12]$$

$$\left| \frac{d^2T}{d\tau^2} + \nu^2 T = 0 \right. \qquad [13]$$

The separation constant ν must be an integer since the solution $T(\nu)$ must be periodical with $2\pi\nu$ as period ; $\nu = m$.

$$T = C_m^n \cos m\tau + D_m^n \sin m\tau \qquad [14]$$

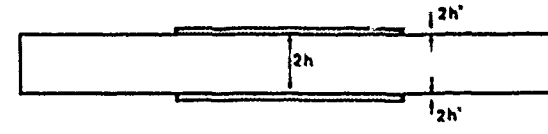


Figure 1 : The plane resonator.

Equation [12] is a Bessel equation which has for solutions, finite at r = 0, the Bessel functions of first kind

$$R(r) = L_m^n \; J_m(r.k_e) \qquad [15]$$

We choose to express the general solution of [3] under the form :

$$\tilde{u}_1^n = \sum_{m=0}^{m=\infty} J_m(r\,k_e).(C_m^n \cos m\tau + D_m^n \sin m\tau) \qquad [16]$$

In a similar maner, for the unelectroded part, the general solution contains modified Bessel functions of the second kind, which are the solutions bounded at infinity of the modified Bessel equation resulting of the separation of [4].

$$\tilde{u}_1^n = \sum_{m=0}^{m=\infty} K_m(r.k_1).(E_m^n \cos m\tau + F_m^n \sin m\tau) \quad [17]$$

Generally, the electrodes and the modes that can be excited in plate with parallel faces have symetries that permit to reduce expressions [16] and [17]. For example with electrodes symetrical in x_1 and x_3 only modes symetrical in x_1 and x_3 are excited, then the coefficients D_m and F_m vanishe. However, even in this case, it may be usefull to consider also the antisymetric modes that can be excited, in real resonators, as the result of minor defects (slightly unparallel plates, electrode misaligment etc...). An example of such modes will be given below.

Boundary and Continuity Conditions

Due to the sin ($n\pi x_2/2h$) dependance of u_1 and to the method used to establish the equations [1] and [2], the boundary conditions on the major surfaces normal to x_2 are approximatively verified (5). At the electrode edge we have to express the continuity of u_1 and of $\partial u_1/\partial n$. This can be approximated by expressing these conditions at only a discrete number of points p. (12) (13). In the case of a symetrical electrode and of a symetrical mode these conditions at the point $M^q(r^q, \tau^q)$ take the form ($D_m = 0$; $F_m = 0$; $m = 2i$) :

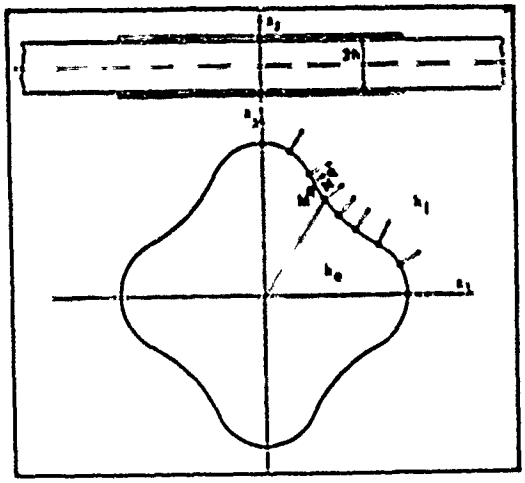


Figure 2 : Discretization of the continuity conditions at the electrode edge.

$$\sum_0^{\infty} C_m^n J_m(r^q.k_e)\cos m\tau^q = \sum_0^{\infty} E_m^n K_m(r^q k_1)\cos m\tau^q \quad [18]$$

$$\sum_0^{\infty} C_m^n r^q k_e J_m' \cos m\tau^q (\alpha N_1^q \cos\tau^q + \beta N_3^q \sin\tau^q)$$
$$+ \sum_0^{\infty} C_m^n.m.J_m \sin m\tau^q (\alpha N_1^q \sin\tau^q - \beta N_3^q \cos\tau^q) \quad [19]$$
$$= \sum_0^{\infty} E_m^n r^q.k_1 K_m' \cos m\tau^q (\alpha N_1^q \cos\tau^q + \beta N_3^q \sin\tau^q)$$
$$+ \sum_0^{\infty} E_m^n.m.K_m \sin m\tau^q (\alpha N_1^q \sin\tau^q - \beta N_3^q \cos\tau^q)$$

where : the index m has only even values m = 2 i

$\alpha = \sqrt{C_{66}/\rho}$; $\beta = \sqrt{C_{66}/C_{55}}$

$\vec{N}^q(N_1^q, N_3^q)$ is the normale at point M^q

J_m, J_m', K_m, K_m' are respectively :

$J_m(r^q.k_e)$, $\dfrac{d\,J_m(r^q.k_e)}{d(r^q.k_e)}$, $K_m(r^q.k_1)$ etc..

For the p points at the electrode boundary, 2p series equalities are obtained, to have an eigen solution, the serie are truncated to p terms ; this gives an homogeneous system with 2 p coefficients. To have non zero solutions in the coefficients, the determinant of the homogeneous linear system must vanish. This constitute an equation to the eigen frequencies $f_{n\mu}$ [μ = 1, 2, 3...) which is numerically solved in the interval $]f_{ce}, f_{ci}[$.

A matrix expression of the truncated continuity conditions of u_1 [18] can be found for the p points, as :

$$t_{mq}^{I}\ C_m^n = t_{mq}^{II}\ E_m^n \qquad m, q = 1, \ldots p$$

where : $t_{mq}^{I} = J_m(r^q.k_e)\cos(m\tau^q)$

$t_{mq}^{II} = K_m(r^q.k_1)\cos(m\tau^q)$

Similarly the continuity of $\partial u_1/\partial n$ at the p points can be expressed after truncation by :

$$s_{mq}^{I}\ C_m^n = s_{mq}^{II}\ E_m^n$$

The homogeneous system is :

$$\begin{pmatrix} t^I & -t^{II} \\ s^I & -s^{II} \end{pmatrix} \begin{pmatrix} C \\ E \end{pmatrix} = 0$$

This formulation will be used for the case of the corrugated resonators.

The eigen modes are obtained by the solution of the homogeneous system for each of the eigen frequencies.

On the whole, the eigen modes for the electroded region are given by :

$$\tilde{u}_{1e}^{n\mu} = \sin(n\pi x_2/2h)\ \tilde{u}_{1e}^{n\mu}$$
$$\tilde{u}_{1e}^{n\mu} = e^{j\omega t} \sum_{i=0}^{i=p-1} \mathcal{A}_{2i}^{n\mu} J_{2i}(r.k_e^{n\mu})\cos(2i\tau) \quad [20]$$

and for the unelectroded region :

$$\tilde{u}_{1l}^{n\mu} = \sin(n\pi x_2/2h)\ \tilde{u}_{1e}^{n\mu}$$
$$\tilde{u}_{1l}^{n\mu} = e^{j\omega t} \sum_{i=0}^{i=p-1} \mathcal{B}_{2i}^{n\mu} K_{2i}(r.k_1^{n\mu})\cos(2i\tau) \quad [21]$$

where : the $\mathcal{A}_i$ and $\mathcal{B}_i$ are of the form C_i/C_s and E_i/C_s

$$k_e^{n\mu} = \left[\frac{\pi^2}{4h^2}(f_{n\mu}^2 - f_{ce}^2)/f_{66}^2\right]^{1/2}$$
$$k_1^{n\mu} = \left[\frac{\pi^2}{4h^2}(f_{ci}^2 - f_{n\mu}^2)/f_{66}^2\right]^{1/2}$$

The corresponding transformed potentiel is :

$$\hat{\varphi}_e^{n\mu} = \frac{e_{26}}{c_{22}}\,\tilde{u}_1^{n\mu}\left(\sin\left(\frac{n\pi x_2}{2h}\right) - (-1)^{(n-1)/2}\frac{x_2}{h}\right) \quad [22]$$

These eigen solutions must verify the orthogonality relations :

$$\iiint_{Vt} \hat{u}_1^{n\mu}.\hat{u}_1^{k\lambda}\,dV = \delta_{nk}\,\delta_{\mu\lambda}\,N^{(n)(\mu)} \quad [23]$$

with :

$$\iiint_{V_t} \tilde{u}_1^{n\mu}\,\tilde{u}_1^{k\lambda}\,dV = \iiint_{V_e} \tilde{u}_{1e}^{n\mu}\,\tilde{u}_{1e}^{k\lambda}\,dV + \iiint_{V_1} \tilde{u}_{11}^{n\mu}\,\tilde{u}_{11}^{k\lambda}\,dV$$

V_t = total volume of the plate

V_e = electroded volume

$V_1 = V_t - V_e$

Forced vibrations :

The modes forced by the potential $V_0 e^{j\omega t}$ can be obtained by a linear combination of the eigen modes at $V_0 = 0$

For the potential a suplementary term must be added (14)

$$\tilde{u}_1 = \sum_n \sum_\mu H^{n\mu}\,\tilde{u}_1^{n\mu} \sin(n\pi x_3/2h) \qquad [24]$$

$$\tilde{\Phi} = \sum_n \sum_\mu H^{n\mu}\,\tilde{\Phi}^{n\mu} + \frac{x_3 V_0}{2h} e^{j\omega t} \qquad [25]$$

where : As in [23] $\tilde{D}_1^{n\mu}$ is either $\tilde{D}_{1e}^{n\mu}$ or $\tilde{D}_{1i}^{n\mu}$

$u_1 = \hat{u}_1 - \frac{e_{26}\, x_3\, V_0}{2h\, C_{66}} e^{j\omega t}$

u_1 being the forced solution.

Accounting for the fact that the orthogonality of $\sin(n\pi x_3/2h)$ has already been expressed in [1] and [2] we substitute in these equations the lateral dependance of u_1.

Then we multiply these relations by $\tilde{u}_1^{k\lambda}$ and integrate respectively on V_e and V_1. The summation of these two relations gives :

$$\sum_n \sum_\mu H^{n\mu} \frac{\pi^2}{4h^2 f^2_{66}} (f^2 - f^2_{n\mu}) \iiint_{V_t} \tilde{u}_1^{n\mu}\,\tilde{u}_1^{k\lambda}\,dV = (-1)^{(n-1)/2} \frac{\rho\omega^2 e_{26}\, 4V_0\, e^{j\omega t}}{C^2_{66}\, n^2\pi^2} \iiint_{V_e} \tilde{u}_1^{k\lambda}\,dV$$

Using the orthogonality relation :

$$\iiint_{V_t} \tilde{u}_1^{n\mu}\,\tilde{u}_1^{k\lambda} dV = 2h \iint_{S_t} \tilde{u}_1^{n\mu}\,\tilde{u}_1^{k\lambda}\,dS = 2\,\delta_{nk}\,\delta_{\mu\lambda}\,N^{(n)(\mu)}$$

We obtain :

$$H^{n\mu} = (-1)^{(n-1)/2} \frac{f^2\, h\, e_{26}}{(f^2 - f^2_{n\mu})\, n^2\pi^2 C_{66}} \frac{4V_0\, e^{j\omega t} \iint_{S_e} \tilde{u}_1^{n\mu} dS}{2N^{(n)(\mu)}} \qquad [26]$$

where : S_n = electroded surface

S_t = total surface

Electrical response :

$$D_2 = e_{26}\, u_{1,2} - \varepsilon_{22}\, \Phi_{,2}$$

$$D_2 = \frac{\varepsilon_{22}}{2h} V_0 e^{j\omega t}(k^2_{26} + 1) + \sum_n \sum_\mu (-1)^{(n-1)/2} H^{n\mu}\,\tilde{u}_1^{n\mu} \frac{e_{26}}{h}$$

and the relation $Y = -\frac{1}{V_0 e^{j\omega t}} \iint_{S_e} D_2^*\,dS$

we obtain the expression of the admittance

$$Y = j\omega \left[\frac{S_e \varepsilon_{22}}{2h} (k^2_{26} + 1) - \sum_n \sum_\mu \frac{f^2\, 4\, k^2_{26}\, \varepsilon_{22}\, h \left| \iint_{S_e} \tilde{u}_1^{n\mu}\,dS \right|^2}{(f^2 - f^2_{n\mu})\, 2n^2\pi^2\, N^{(n)(\mu)}} \right]$$

The identification with the admittance of the classical equivalent scheme leads to :

$C_Z = \frac{S_e \varepsilon^S_{22}}{2h} (k^2_{26} + 1)$ is the capacitance [28] at $f = 0$

$$C_{n\mu} = \left[\frac{8\, k^2_{26}\, \varepsilon_{22}}{n^2\pi^2\, 2\, h} \right] \frac{h \left| \iint_{S_e} \tilde{u}_1^{n\mu}\,dS \right|^2}{N^{(n)(\mu)}} \qquad [29]$$

As expected for stiffened mode resonators the parallel capacitance contains C_Z and an infinite sum of the opposite of the capacitances of all the modes :

$$C_0 = C_Z - \sum_n \sum_\mu C_{n\mu} \qquad [30]$$

In the case of one dimensional modes of an infinite plate, the sum $-\sum_n \sum_\mu C_{n\mu}$ is equal to $-k^2 S_e \varepsilon_{22}/2h$ which reduce the parallel capacitance at high frequency to the usual value of $C_0 = S_e \varepsilon^S_{22}/2h$ whereas $C_Z = S_e \varepsilon^S_{22}/2h$ is the low frequency capacitance.

III - MODEL FOR THE CORRUGATED RESONATORS

Resonators having a groove outside the electrode were proposed by Nakazawa et al (15). Similarly, plane resonators having an inverse mesa structure were presented to control the energy trapping by Lukaszek (16). This method of using discrete lateral variations of the thickness to control the energy trapping in plane resonators can be extended to confer to this type of resonators the energy trapping behaviour of bevelled resonators or of plano-convex resonators. This can be of large interest for the case of the high frequency resonators for which it is very difficult to realize the contoured shape by the conventional methods using the lapping and polishing techniques. In this part we propose a model to analyse the energy trapping behaviour of resonators having, outside the electrode, several flane zones with different thicknesses.

The figure 3 indicate that the considered resonators have N different zones with different thicknesses, N-1 being situated outside the electrode. We make the simplifing hypothesis that the cut-off frequencies for theese regions are all higher than that of the electroded region. We make also the hypothesis that the thicknesses of all the regions are not very different so that no boundary conditions are to be considered on the different steps appearing across the resonator.

III.1 - Eigen modes at V = 0 :

The partial derivative equations governing the lateral variations of the displacement in the different regions of these resonators are equations [1] and [2]. For the central electroded region, the eigen solution is still



Figure 3 : Geometry of grooved resonators.

given by the expression [16]. For the external regions, except the last, without any particular hypothesis concerning the thickness distribution, it is necessary to include in the solution given by expression [17] the modified Bessel functions of first kind since the eigen solution has not now to be bounded nor at zero neither at infinity.

For such a region M, the eigen solution (symetrical mode) is thus :

$$\tilde{u}^{n}_{1M} = \sum_{2\Gamma=0}^{\infty} \cos(2\Gamma\tau)\left[E^{n}_{2\Gamma} K_{2\Gamma}(r\cdot k_M) + G^{n}_{2\Gamma} I_{2\Gamma}(r\cdot k_M)\right] \quad [31]$$

where :
$$k_M^2 = \frac{\pi^2(f^2 - f_{CM}^2)}{4h_M^2 f_{66}}$$
$$f_{CM} = \frac{n}{4h_M}\sqrt{\frac{c_{66}}{\rho}}$$

$2h_M$ is the thickness in this region, f_{CM} is the corresponding cut-off frequency.

To simplify the formalism, in the following, we use for this eigen solution the form :

$$\tilde{u}^{n}_{1M} = \sum_{0=m}^{\infty} A^{M}_{m} C_{m}(r\cdot k_M)\cos(i(m)\cdot\tau) \quad [32]$$

where : C is either K or I

For the most external region, since we consider only trapped modes, the solution only contains the modified functions K.

Continuity conditions : As for the case of the plane resonators we use the approximation of expressing the continuity of u_1 and of $\partial u_1/\partial n$ at only a discrete number of points (p) at each region boundary (figure 4). After truncation of the series to p terms, these conditions can take the matricial form given in paragraph II.

For the p points at the limit of region M and M + 1 the continuity of u_1 can take the form :

$$t^{M}_{mj} A^{M}_{m} = t^{M+1}_{mj} A^{M+1}_{m} \quad [33]$$

where :
$$t^{M}_{mj} = C_m(r^j\cdot k_M)\cos(i(m)\cdot\tau^j)$$
$$t^{M+1}_{mj} = C_m(r^j\cdot k_{M+1})\cos(i(m)\cdot\tau^j)$$

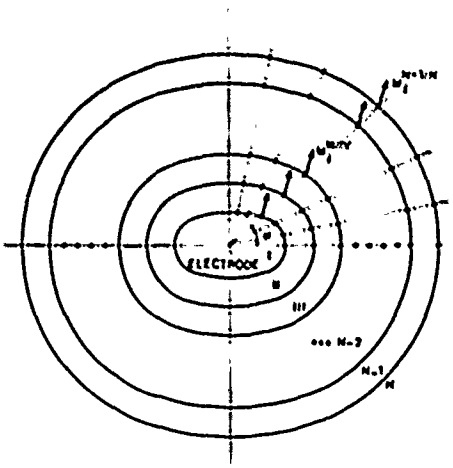


Figure 4 : Discretization of the continuity conditions for the grooved resonators.

Similarly the continuity of the normal derivative can be written :

$$s^{M}_{mj} A^{M}_{m} = s^{M+1}_{mj} A^{M+1}_{m} \quad [34]$$

For all the region boundaries we have :

$$\begin{vmatrix} t^{I}A^{I} = t^{II}A^{II} \\ t^{\cdot II}A^{II} = t^{III}A^{III} \\ \cdots\cdots\cdots\cdots \\ t^{\cdot N}A^{N-1} = t^{N}A^{N} \end{vmatrix} \quad \begin{vmatrix} s^{I}A^{I} = s^{II}A^{II} \\ s^{\cdot II}A^{II} = s^{III}A^{III} \\ \cdots\cdots\cdots\cdots \\ s^{\cdot N-1}A^{N-1} = s^{N}A^{N} \end{vmatrix} \quad [35]$$

from which one obtains :

$$t^{I}A^{I} = t^{II}.(t^{\cdot II})^{-1}.t^{III} \quad \ldots.(t^{\cdot N-1})^{-1}t^{N}.A^{N} \quad [36]$$

$$s^{I}A^{I} = s^{II}.(s^{\cdot II})^{-1}.s^{III} \quad \ldots.(s^{\cdot N-1})^{-1}s^{N}.A^{N} \quad [37]$$

or :

$$\begin{matrix} t^{I}A^{I} = V^{N}A^{N} \\ s^{I}A^{I} = W^{N}A^{N} \end{matrix} \quad \text{or} \quad \begin{pmatrix} t^{I} - V^{N} \\ s^{I} - W^{N} \end{pmatrix}\begin{pmatrix} A^{I} \\ A^{N} \end{pmatrix} = 0 \quad [38]$$

The problem admits non trivial solutions in (A^I, A^N) if the determinant of the homogeneous sytem [38] vanishes. This constitutes the equation to the eigen frequencies which must be numerically solved for the eigen frequencies f_{n_μ}. The coefficients of the eigen modes in region I and N are obtained as the solutions of the homogeneous linear system for the eigen frequencies. The coefficients for the other regions are obtained from back substitution in the relations [35] and resolution.

The forced modes can then be obtained by a method which is formally identical to those given in § II, u_1 being now defined by piece in the unelectroded region.

IV - PROPERTIES OF THE RESOLUTION METHOD

Convergence For the resonators with circular electrodes, the convergence of the eigen frequency is very fast, when the number of points at which the continuity conditions at the electrode edge are expressed, is increased. In figure 5 this is represented for the case of the fundamental and the first anharmonic of an AT resonator (17) (2h = 200 μm ; $\hat{R}$ = 1,92 %; re = 2.8 mm).

In the case of a resonator with an elliptical electrode respecting the lateral anisotropy of the plate (i.e. with an axis ratio $\frac{a}{b} = \left(\frac{M_n}{c_{55}}\right)^{\frac{1}{2}}$ a totally analytical solution can be obtained in r, τ coordinates. With this solution it is sufficient to express the continuity conditions for the (unique) corresponding value of r. To test the numerical model we have computed the eigen frequencies for an AT plate with 4π ab = 25 π mm²; $\frac{a}{b} = \left(\frac{M_n}{c_{55}}\right)^{\frac{1}{2}}$; $\hat{R}$ = 0.95 %, as a function of the number of points at the electrode edge. To the precision at which the eigen frequencies are computed, they are identical for p = 4 to p = 9.
f_{11}=10692794,886 f_{12}=10773750,813 f_{13}=10785592,817 Hz.

In figure 6 are represented the convergence properties for an elliptical electrode with 4πab = 25π mm² ; a/b = 2.0 and $\hat{R}$ = 0.95 %. This demonstrate that, as the electrode shape becomes more and more different of the

case corresponding to the lateral anisotropy, more and more points are required at the electrode edge to have the same precision. However due to the limited absolute precision on the material constants and the resonator parameters, in all cases, for symetrical electrodes, 6 to 10 points give more than sufficient precision on the eigen frequencies and the eigen modes, even for the case ($\frac{1}{2}$Y) of resonators having polygonal electrodes.

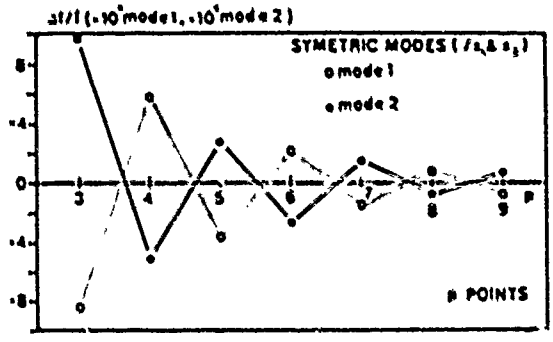


Figure 5 : Convergence of the eigen frequencies for circular electrodes.

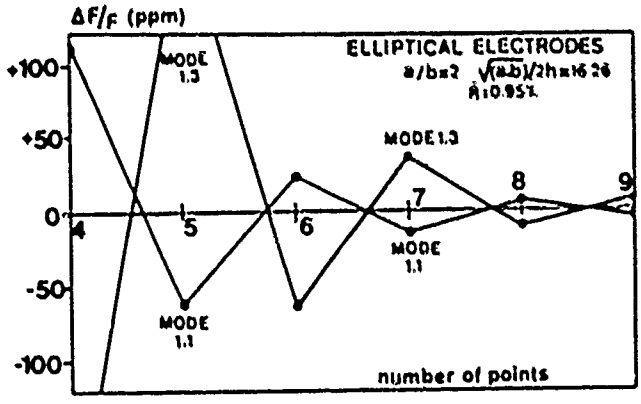


Figure 6 : Convergence of the eigen frequencies for elliptical electrodes.

Unsymetrical electrodes. Unsymetrical modes

In the case of unsymetrical electrodes (in x_1 or/and in x_3) more general solutions of equations [1] and [2] must be considered so that more terms must be retained in expressions [16] and [17]. In this case, the antisymetrical modes in x_1 and x_3 can be excited since the integral $\iint_{Se} \tilde{u}_{1e}\, dS$ does not vanish.

The case of unsymetrical modes excited in plates with symetrical electrodes presenting sligh imperfections (18) (electrode misaligment, tab not exactly in x_1 or x_3 direction etc...) can be covered by this theory.

The case of unsymetrical modes resulting from defects of parallelism of the plate can be considered by introducing in equations [1] and [2], $h(x_1, x_3)$. This has already been done by Dworsky (9) with the use of a finite difference methode of resolution. A semi analytical method with a discretization of the continuity conditions can also be used since, in this case the wave equations are separable in Airy equations whose solutions can be expressed in terms of Bessel functions of fractional orders.

The experience, in the case of sligh defects of parallelism, shows that the eigen frequencies of the antisymetrical modes computed for the corresponding parallel case, give a very good approximation of the observed frequencies and modes.
In figure 7 the 3 first eigen frequencies and modes of a Y resonator using $AlPO_4$ are displayed (17). The second and the third modes are respectively the first antisymetrical modes in x_3 and in x_1.

Effects of the electrode tabs : The method of solution permits to consider the effect of infinite electrode tabs on the resonance frequencies and modes provide that they have the shape of circular sectors (figure 8). This is obtained in considering in the tab zone a solution in the form :

$$\tilde{u}^n_{1m} = \sum_{2i=0}^{\infty} G^n_{2i} K_{2i}(r.k_m) \cos(2i\tau)$$

with :
$$k_m^2 = \frac{(f_{cm}^2 - f^2)}{4\, h^2\, f_{66}^2}$$

f_{cm} = cut off frequency of the semi electroded tab region

The continuity conditions are then expressed as previously described at the electrode edge. Doing this, we neglect the continuity conditions along the radii limiting the tab, but since u_1 is small in this region and has generally a quasi circular symetry this is of small importance.

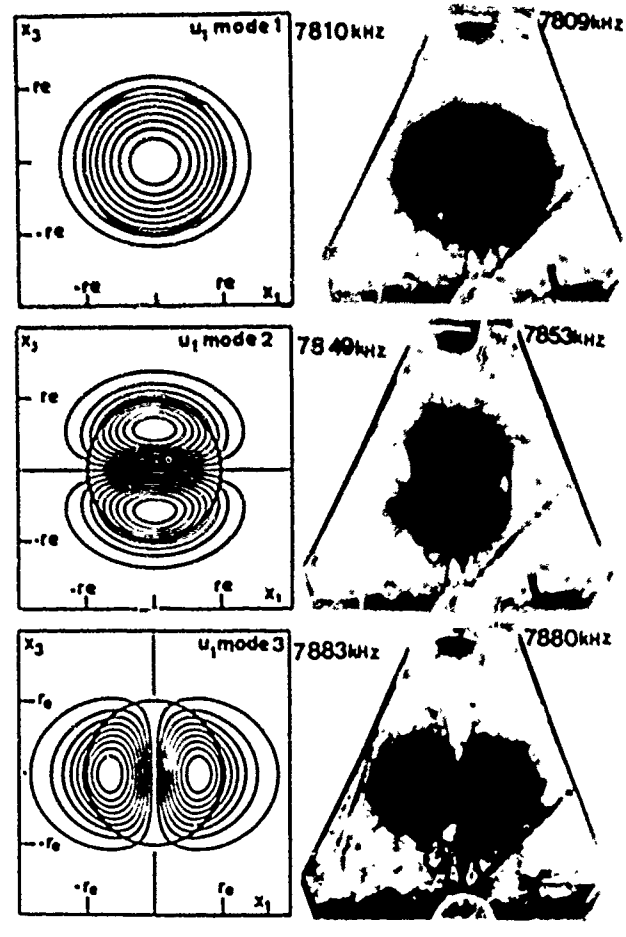


Figure 7 : Antisymetrical anharmonics.

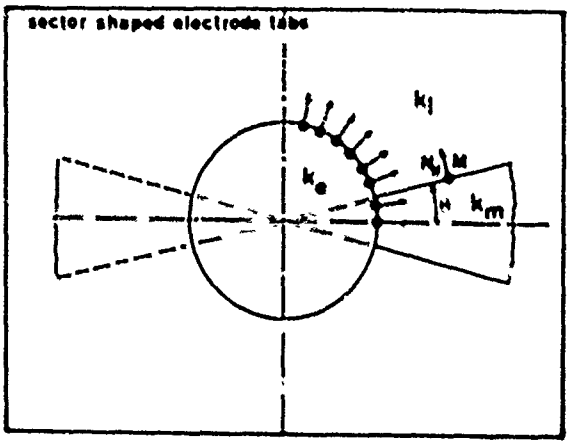


Figure 8 : Model for the electrode tabs.

Table 1 effect of electrode tabs

	computed without tabs	computed with tabs	experiment
Mode 1	10712.	10711.	10714.
Mode 2	10785.	10754	10786
Mode 3	10812	___	10807

For the computed results, the hypothesis of infinite lateral dimensions of plate and tabs, has the consequence that all the modes having frequencies above fcm (figure 9) become untrapped when the tabs are considered.

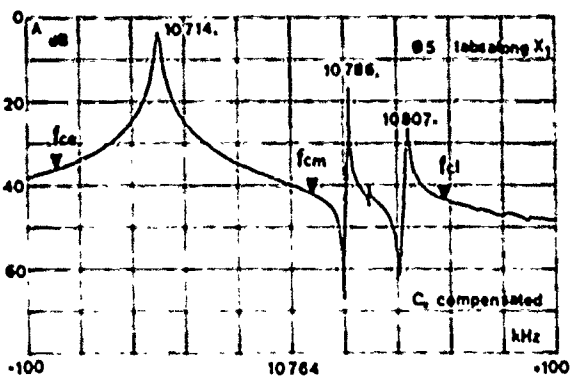


Figure 9 : Frequency response of the resonator.

A comparison with experimental results was made in several cases. A typical example is given in table 1 for an AT resonator having the following characteristics :

AT quartz θ =-35°15'

2h = 153.5 μm $\bar{R}$ = .93 %
f_{ce} = 10688 kHz
f_{cl} = 10823 kHz
f_{cm} = 10772 kHz

sumit angle of the tabs 20°
tabs in the x_1 direction.

The experimental results indicate that the effect of the tabs computed for the fundamental mode is small and is probably found of the right order of magnitude. For the 2nd and the 3 rd modes, they indicate, that the major effect is that they become lossy because they reach the fixation of the resonator. This is confirmed by the examination of the experimental modes on figure 10.

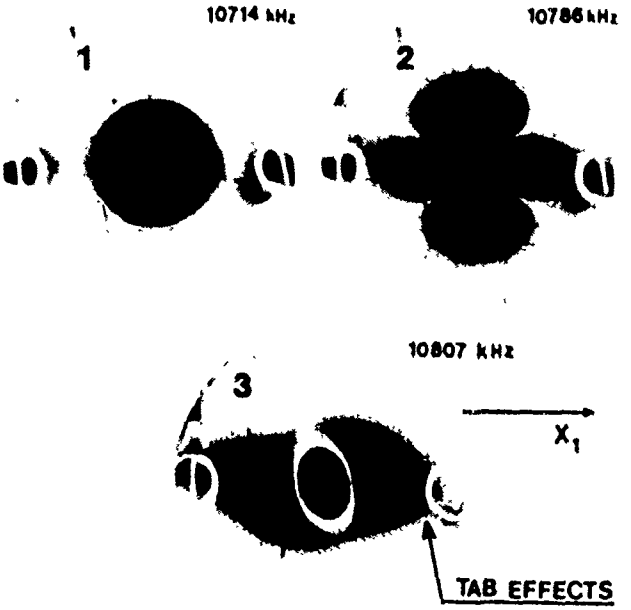


Figure 10 : Experimental modes.

On the whole, the proposed method can describe the small effect of the electrode tabs on modes with eigen frequencies lower than f_{cm}. For modes with frequencies higher than fcm it is necessary to take in account the finite dimensions of the plate (19) and the properties of the fixation. In all cases, the defects of parallelism of the plate near its edges play an important role.

V. Comparison with experimental results : effect of the electrode geometry

Resonators with circular, rectangular, hexagonal electrodes : To evaluate the numerical model, fundamental mode resonators with various electrode shapes were investigated using AT quartz plates having a thickness near 153,75 μm and a mass loading generally of 0,95 %.

On figure 11 are represented the computed and observed fundamental modes and frequencies for a resonator with circular electrodes (2 re = 5 mm).

On figure 12 the case of a resonator with rectangular electrodes is considered (ax1 = 5,15 ax3 = 4,15 mm). One, can observe that even with electrodes having angular points the computed eigen mode is an accurate prediction of the observed one. For this resonator the mass loading was 0,95 %.

On Figure 13 we have the case of a resonator with hexagonal electrodes a_{hex} = 2,748 mm $\bar{R}$ = 0,95 %. Here again, there is a good agreement between the observed and the computed frequencies and modes.

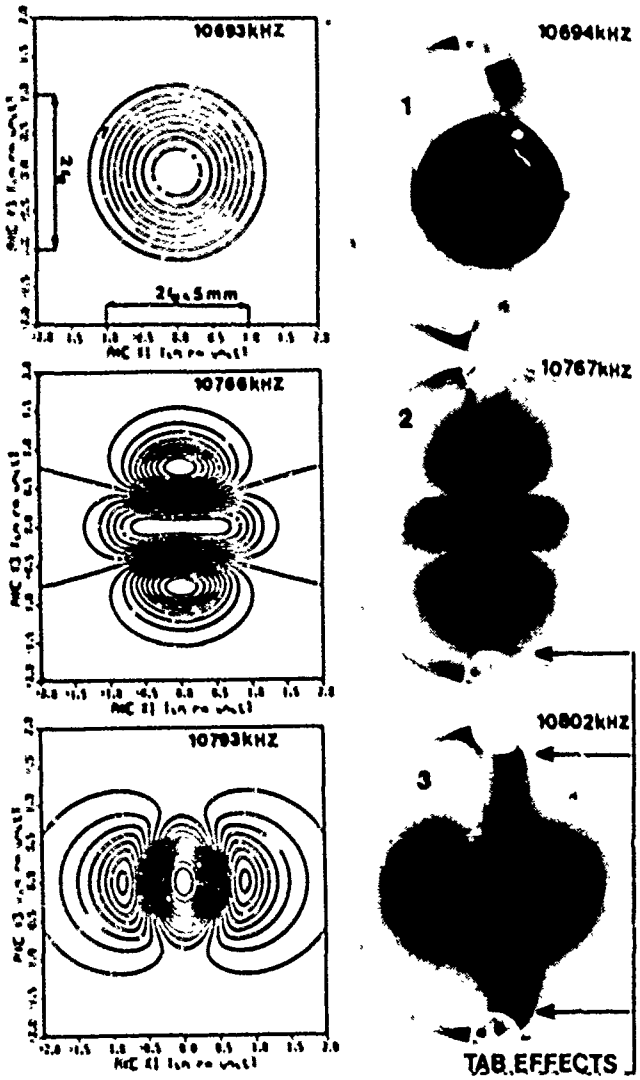


Figure 11 : Resonator with circular electrodes.

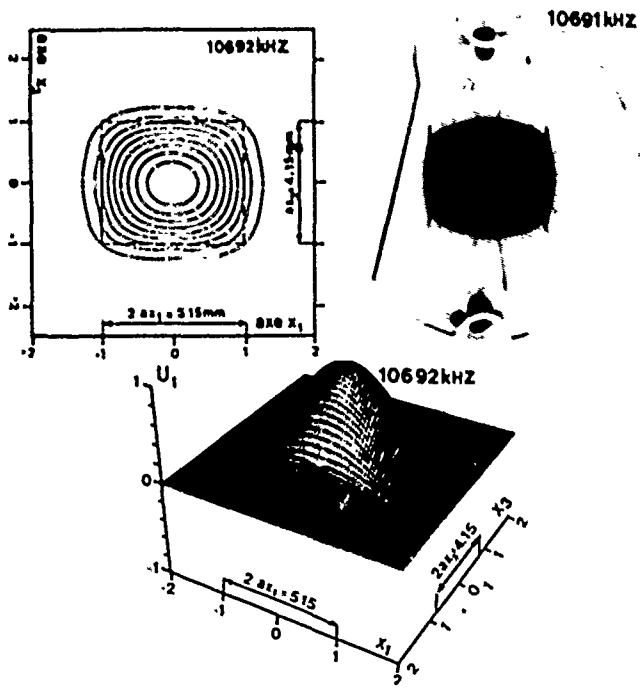


Figure 12 : Resonator with rectangular electrodes.

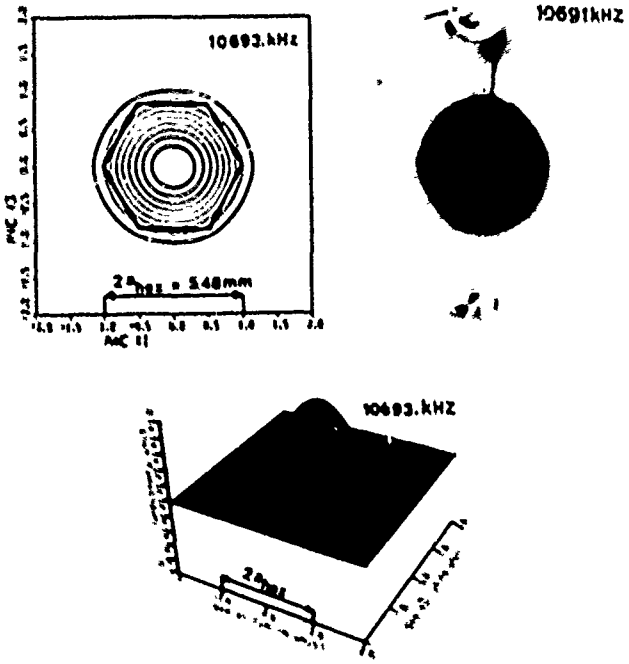


Figure 13 : Resonator with hexagonal electrodes.

Effect of the electrode geometry on the anharmonics. AT plates, θ = - 35°15' with a diameter of 13 mm, thickness near 153,75 µm were metallized with electrodes having the geometries given in figure 14. The mass loading (0,95 %) was identical in all cases and the surfaces nearly identical (S ≈ 20 mm²). The observed and computed frequencies are given in table 2. Since the relative differences between the air gap frequencies of the different plates were lower than 2.10^{-4} and the differences between the mass loading lower than a few 10^{-4}, we can say that the effect (experimental or computed) of the electrode shape is small on the fundamental frequency.

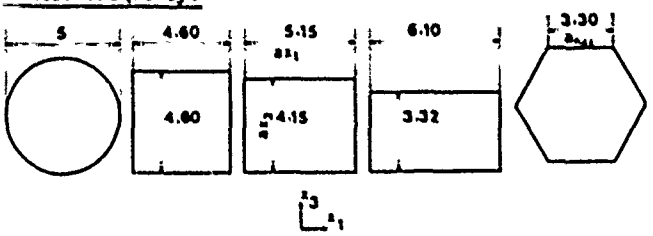


Figure 14 : Considered electrode geometries.

Table 2 : Computed and observed frequencies

		CIRCUL. ELECTR. 2r=5mm	SQUARE ELECTR. 4.6 × 4.6	RECTANG. ELECTR. 5.15 × 4.15	RECTAN ELECTR. 6.10 × 3.32	HEXAGO ELECTR. r=1.75
COMPUTED RES.	MODE 1	10 693.	10 692.	10 692.	10 694.	10 693.
	MODE 2	10 766.	10 772.	10 779.	10 764.	10 766.
	MODE 3	10 793.	10 793.	10 781.	10 804.	10 793.
EXPERIMENTAL	MODE 1	10 694.	10 691.	10 691.	10 694.	10 691.
	MODE 2	10 767.	10 770.	10782.	10 763.	10 763.
	MODE 3	10 802.	10 801.	10785.	10 810.	10 796.

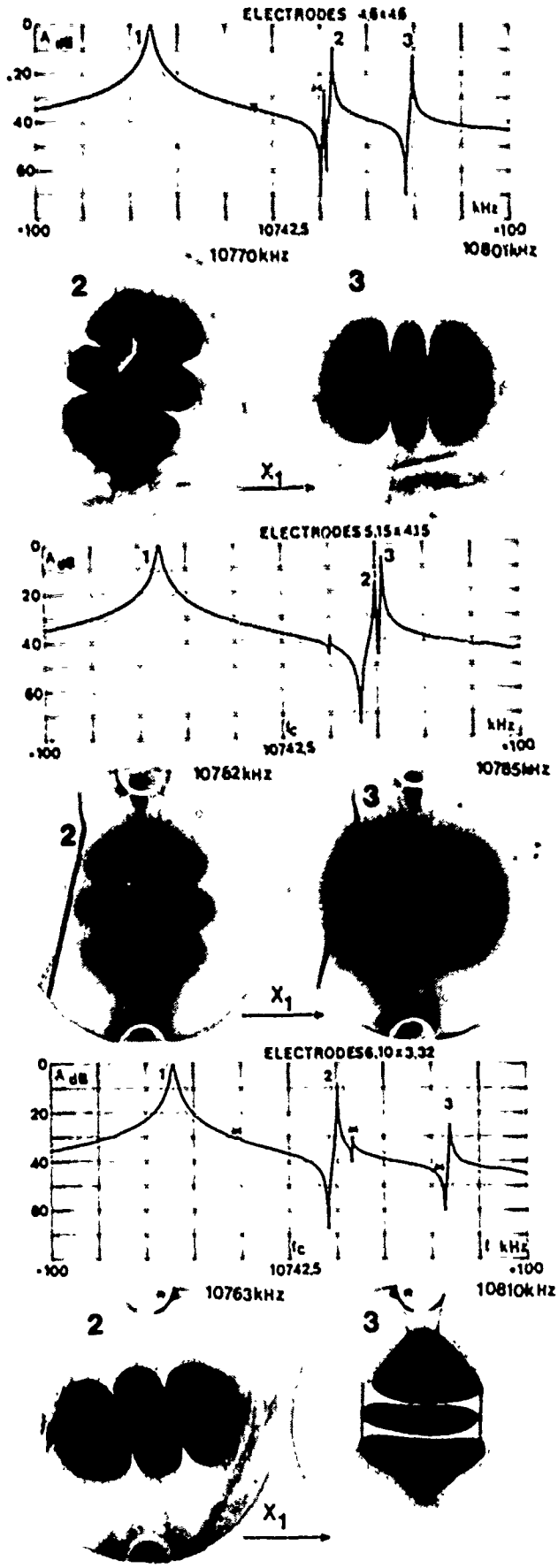


Figure 15 : Response curve and experimental modes for rectangular electrodes.

To the contrary, we can observe, in table 2, that the effect of the electrode geometry can be very important for the anharmonic modes. First, we must notice that there is no fundamental differences between the distributions of the anharmonic frequencies for the cases of circular, square and hexagonal electrodes. But for the rectangular electrodes with different ax_1/ax_3 ratios, large differences can be observed. The resonator having ax_1 = 5,15 mm and ax_3 = 4,15 mm is particular since these dimensions were choosen so that $ax_1/ax_3 \simeq \sqrt{Mn/C_{55}}$. In this case, we can observe that the 2nd and the 3rd anharmonic modes have nearly equal frequencies. It is also in this case that these anharmonics are the more distant of the fundamental frequency. This signify that, if the surface or the mass loading is reduced, it is for this particular rectangular shape (with $ax_1/ax_3 \simeq \sqrt{Mn/C_{55}}$) that the anharmonics will first desappear. This has also the signification, that, it is for this type of resonator (among all considered here) that is is possible to have the maximum dynamic capacitance with a single mode response (for mass loading near or higher than 1 %, the dynamic capacitance of the fundamental mode has small (and favourable) variations as ax_1/ax_3 is varied at constant surface). With the electrode surface and the mass loading used here we have 2 symetrical anharmonics.

For the rectangular electrodes, as ax_1/ax_3 is varied, the order of the symetrical mode having respectively 2 nodal lines along x_1 and along x_3 change for a ratio close to $\sqrt{Mn/C_{55}}$ (figure 15).

Some of these properties, were already signaled by several authors, they have the consequence that it is of large interest to make use of the lateral anisotropy of the plate in the design of the resonators particularly for their applications to filters and VCXO.

Incidentally, we can notice in figure 15 that, in some cases, the nodal lines appears to have direction making an angle with the x_1 or the x_3 direction. This is the result of a forced mode effect due to the coupling of an antisymetrical anharmonic mode to the symetrical one. Then, at least two terms in expression [24] have non negligible values in this frequency range and the forced mode is practically a linear combination of the two nearest eigen-modes.

Elliptical electrodes. The elliptical electrodes with an axis ratio $a/b = ax_1/ax_3 = \sqrt{Mn/C_{55}}$ have exactly the same anisotropy as the crystal plate. To investigate the effect of the electrode anisotropy on anharmonic modes, we have computed, at equal surface ($4\pi ab$ = 25 mm) the effect of the variation of a/b. The considered mass loading were 0.5 %, 1 % and 2 %, the thickness was such that $(ab)^{1/2}/2h$ = 16.26.

The results are given in reduced frequency ($\Omega = \sqrt{(f_r^2 - f_c^2)}/f_{66}^2$ as a function of a/b in figure 15 and figure 16. For R = 0.5 % (figure 17) we can observe, as previously discussed, that there is no important variations of the frequency of the fundamental mode with a/b. To the contrary, the unique anharmonic mode has large frequency variations and it vanishes when a/b approaches $\sqrt{Mn/C_{55}}$. On figure 17 we can observe, for a/b near $\sqrt{Mn/C_{55}}$, the exchange of the 2nd and 3rd anharmonics as noticed for the rectangular electrodes, and also, for R = 2 %, the fact that the 4th anharmonic mode desappears for a/b near $\sqrt{Mn/C_{55}}$.

On both figures, we can observe that the case of the circular electrode (a/b = 1) is less favourable than the case with $a/b \neq \sqrt{Mn/C_{55}}$. The design diagram for AT Quartz with circular electrodes is recalled on figure 18 where the reduced dimension corresponding to $(a.b)^{1/2}/2h = re/2h = 16,26$ is noted.

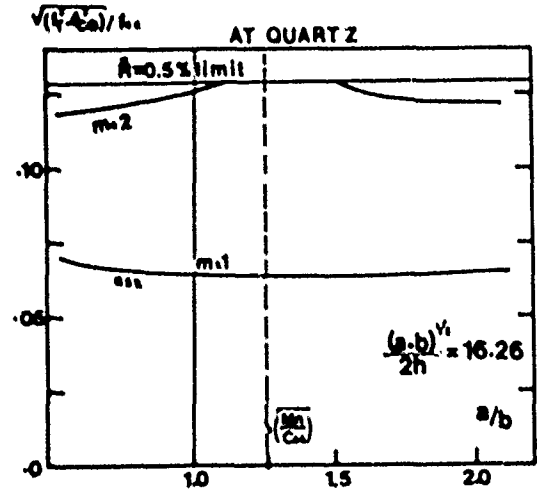


Figure 16 : Reduced frequencies of eigen modes excited by elliptical electrodes ($\bar{R}$ = 0,0050)

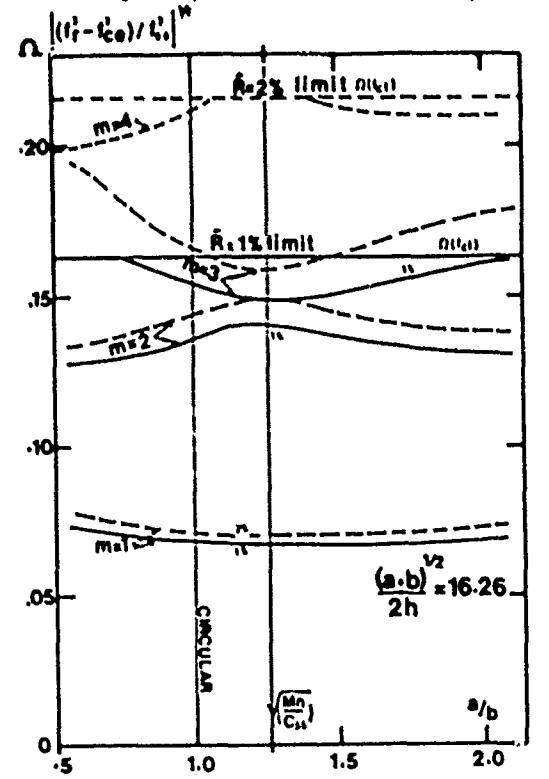


Figure 17 : Reduced frequencies of eigen modes excited by elliptical electrodes ($\bar{R}$ = 0,01 and 0,02)

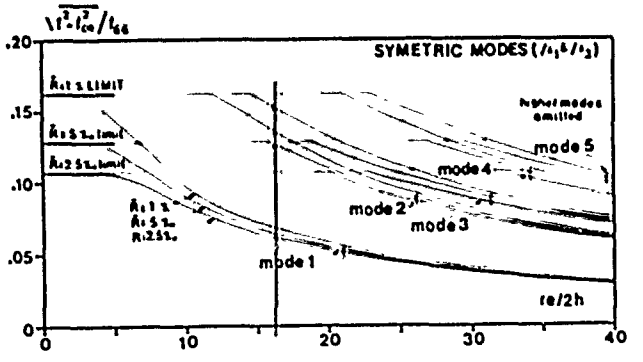


Figure 18 : Design diagram for AT resonators with circular electrodes.

circular and elliptical electrodes :
S = 25π /4 mm² 2 h = 153.75μ R̄ = .95 %

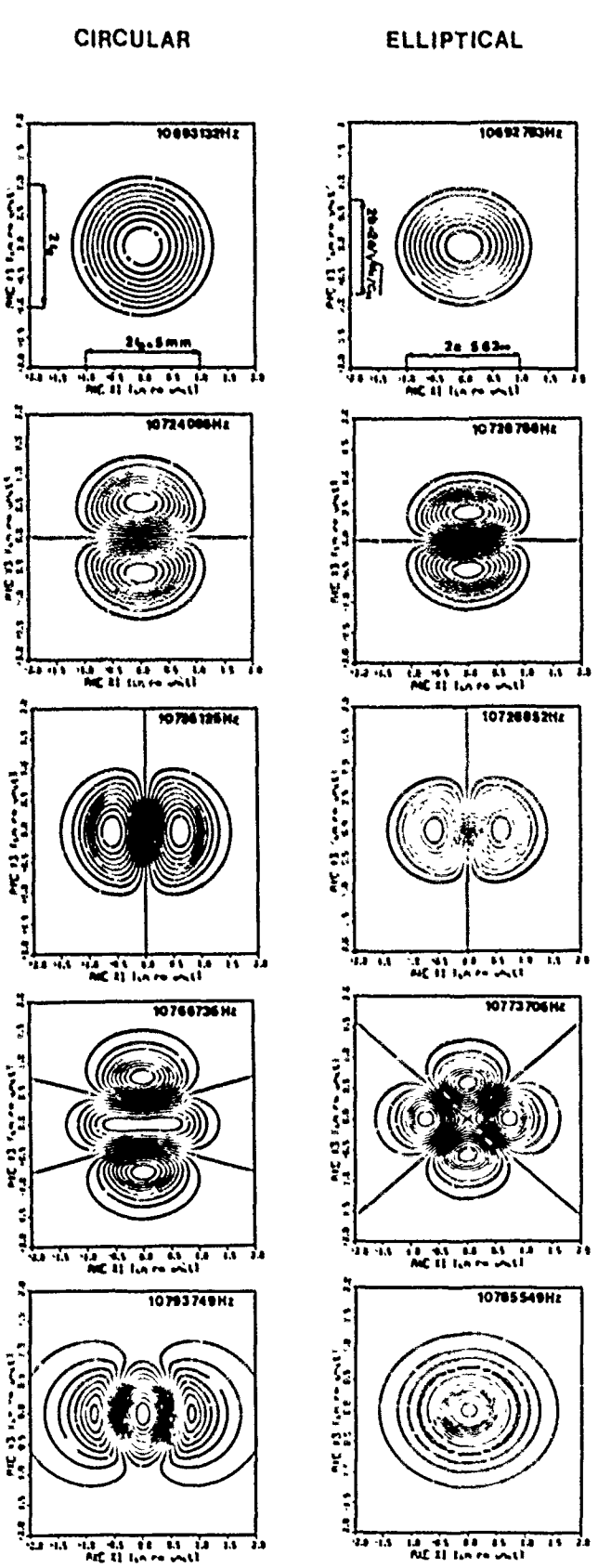


Figure 19 : Comparison of computed modes excited by circular and elliptical electrodes.

On figure 19 we compare the computed modes of resonators with circular of the same area and elliptical electrodes of the same area. On this figure, we can observe that the resonant frequencies of the first two antisymetric modes excited by elliptical electrodes are nearly equal and also that with elliptical electrodes, we dont need to use plates of larger surface than for circular ones.

VI. Conclusion

The use of the approximate equation established by Professor Tiersten and his associates together with a semi analytical method of solution using a discretization of the continuity conditions at the electrode edge permit to construct precise models for the plane and corrugated resonators. Many other possibilities of this method still exist such as those to take in account the effects of the finite dimensions of plates of arbitrary geometry.

Several examples prooving the interest of making resonators with electrode shapes respecting approximatively the lateral anisotropy of the crystal plates were given. Since this is very easy to implement, and permits to improve the performances of the resonators, especially for filter and VCXO applications, there is no reasons, to continue to use circular electrodes in all cases.

Acknowledgments : The authors whish to aknowledge M. H. Carru for valuables indications during the course of this work and the skilful assistance of R. Chenebault, A. Daniel and G. Le Thuault for the elaboration of the many samples used during this study. Special thanks are due to M.P. Louis for preparing this manuscript.

REFERENCES

(1) R.D. Mindlin
Int. J. Solids Structures vol 8 p 895 (1972)

(2) P.C.Y. Lee, W.J. Spencer
J. Acoust Soc. Am. vol. 45 n° 3 p 637 (1969)

(3) P.C.Y Lee, S. Syngellakis, J.P. Hou
Proc. 1986 IEEE Ultra Sonics Symposium (1986)

(4) H.F. Tiersten
J. Acoust. Soc. Am. vol 59 p 879 (1976)

(5) D.S. Stevens, H.F. Tiersten
J. Acoust. Soc. Am. vol 79 n° 6 p 1811 (1986)

(6) B. Dulmet ; H.F. Tiersten
Proc 1986 IEEE Ultrasonics Symposium

(7) B.K. Sinha, D.S. Stevens
J. of Acoust. Soc. Am. vol 66 n° 1 p 192 (1979)

(8) D.C.L. VangheLuwe
Proc. 34 ann. Freq. Control Symp. p 412 (1980)

(9) L.N. Dworsky
Proc. 37th. Ann. Freq. Control Symp. p 232 (1983)

(10) L.N. Dworsky
Proc. 38th Ann. Freq. Control Symp. p 141 (1984)

(11) H.F. Tiersten, R.C. Smythe
J. Acoust. Soc. Am. vol 65 n° 6 p 1455 - 1979

(12) K. Nakamura ; H. Shimizu
Proc. IEEE Ultrasonics Symp p 606 (1976)

(13) H. Seikimoto
IEEE. Trans. Sonics Ultrason. vol SU 31 p 664-669 (1984)

(14) R. Holland, E.P. Eernisse
Design of resonant piezoelectric Devices
MIT Press 1969

(15) M. Nakazawa, T. Lukaszek, A. Ballato
Proc. 35th. Ann. Symp Frequency Control p 71 (1981)

(16) T. Lukaszek, A. Ballato
IEEE Trans. Sonics. Ultrasonics vol Sa 18 n° 4 p 238 (1971)

(17) J. Détaint, A. Zarka, B. Capelle, Y. Toudic, J. Schwartzel, E. Philippot, J.C. Jumas, A. Goiffon, J.C. Doukhan
Proc. 40th. Ann. Frequency Control Symp. p 101 (1986)

(18) A. Glowinski, R. Lançon, R. Lefèvre
Proc. 27th Ann. Frequency Control Symp. p 233 (1973)

(19) H. Seikimoto
Proc. 39th Ann. Frequency Control Symp. p 386 (1985)

(20) J. Détaint, J. Schwartzel, C. Joly, J.C. Jumas
Proc. 1er Forum Européen Temps-Fréquence Besançon (1987)

41st Annual Frequency Control Symposium - 1987

LATERAL- AND THICKNESS-FIELD COUPLING IN ZINCBLENDE STRUCTURES

Arthur Ballato, Theodore Lukaszek,

Muhammad Mizan, and John Kosinski

US Army Electronics Technology & Devices Laboratory
LABCOM, Fort Monmouth, NJ 07703-5000

ABSTRACT

Most binary semiconductors of type III-V and II-VI crystallize in the cubic zincblende structure, class 4-bar 3m, and are therefore piezoelectric. This leads to the possibility of acousto-electronic signal processing devices and technology, a field with great opportunities for growth and development.

It might be thought that, because of the cubic symmetry, the acoustic and piezoelectric properties would not vary greatly with orientation. Such is not the case. Using semi-insulating gallium arsenide plates as an example, graphs are given of propagation velocity and piezoelectric coupling for both lateral- and thickness-field excitation of simple thickness modes. It is found that there is considerable acoustic anisotropy in GaAs, as well as a number of orientations having interesting device properties. The maximum levels of piezocoupling are about 5%, or about one-half those of quartz, for both forms of excitation. An equivalent electrical transmission line network for the plate vibrator is also given that includes losses due to conduction.

INTRODUCTION

Piezoelectric resonators are fashioned today nearly exclusively of insulating materials such as quartz and lithium tantalate. Most semiconductors have moderate piezoelectric coupling coefficients, and this fact gives rise to the possibility of their use in integrated analog-digital signal processing devices and frequency control sources.[1] In this paper some of the pertinent properties of this class of materials are examined. Composite resonators, consisting of insulating and semiconducting layers, are a further application of some promise.[2] Table I provides an overview of properties for some crystalline materials. Each has presently, or potentially, a niche in the marketplace. Quartz was one of the first crystals in which piezoelectricity was discovered, and remains still the most important material for classical frequency control. Newer crystals such as lithium niobate and tantalate are now firmly established as key materials. The newest substances shown: berlinite, lithium tetraborate, and piezo-semiconductors such as GaAs, have yet to prove themselves.

Progression in materials synthesis relating to our subject is shown in Fig. 1 as function of year, starting with the discovery of piezoelectricity in 1880. Two main branches are distinguished: hydrothermal synthesis, begun around 1908 for quartz by Spezia, and Czochralski melt growth originating in 1918. Under the latter method come the unary, non-piezoelectric semiconductors Ge and Si in class m3m (diamond structure), the very high piezocoupling refractories $LiNbO_3$, and $LiTaO_3$, and the binary semiconducting compounds such as GaAs.

Figure 2 shows the growth of IC density that has revolutionized the electronics industry; similar hopes are held for MMIC technology using binary compounds. The added feature of piezoelectricity exhibited by these materials permits greater versatility of function.

TABLE I.

PIEZOELECTRIC COUPLING & FREQUENCY-TEMPERATURE COEFFICIENTS

CRYSTAL	CLASS	ORIENTATION	MODE	PIEZOELECTRIC COUPLING (%)	TEMPERATURE COEFFICIENT (10^{-6}/°C)
GaAs	$\bar{4}3m$	(100) (111)	SHEAR EXTENSION	0 4	-30 +72
$Li_2B_4O_7$	mm2	(90°, 47°, 0°) Z	SHEAR EXTENSION	20 10	0 +141
$LiTaO_3$	3m	(36°, 90°, 0°) Z	SHEAR EXTENSION	5 10	0 -46
$LiNbO_3$	3m	Z Z	SHEAR EXTENSION	70 17	-85 -85
$AlPO_4$	32	(90°, 21°, 0°) X	SHEAR EXTENSION	11 5	0 -86
SiO_2	32	(90°, 36°, 0°) X	SHEAR EXTENSION	0 14	0 +20

MATERIALS PROGRESSION IN ACOUSTICS

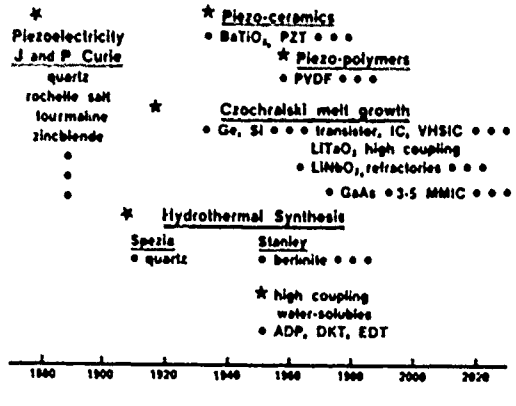


Figure 1. Materials progression in acoustics.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

BINARY SEMICONDUCTORS

The binary semiconductors shown in Fig. 3 crystallize in either the sphalerite or wurtzite form.[3] The latter is hexagonal, class 6mm (and is piezoelectric); the former is cubic, class $\bar{4}3m$, and is also known as the zincblende structure. Zincblende (β - ZnS) was among the crystals in which the brothers Curie found the piezoeffect (cf. Table I). Because of the simplicity of their structures, these binaries appear not to have any representatives that possess a zero temperature coefficient of frequency; a resonator proximate to microelectronic elements, however, certainly can be compensated by them.

TRANSISTORS ON A SILICON CHIP

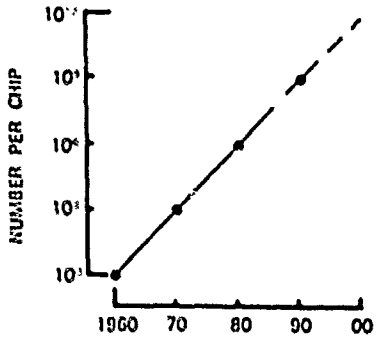


Figure 2. Transistors on a silicon chip.

ZINCBLENDE STRUCTURES

Gallium arsenide is presently the most important representative of class $\bar{4}3m$. Its symmetry elements are shown in Fig. 4, along with the elastopiezodielectric matrix elements. The zincblende-structure materials have but one dielectric permittivity, and are therefore optically isotropic. They have three piezo matrix elements; these are identical. Of the nine linear elastic constants, three are independent. The elastopiezodielectric matrix shown in the figure is identical to that of crystals in class 23. When other effects (e.g., photoelasticity, optical activity) are considered, the difference between the classes is manifested. Symmetry elements for each class are:

$\bar{4}3m$: 6 mirror planes; 6 $\bar{2}$-axes; 3 $\bar{4}$-axes; & 3 (inherent) 2-axes.
23: 3 mirror planes; & 3 2-axes.

Further differences emerge when the numbers of independent, higher-order elastic constants are compared, as seen in Table II.

BINARY SEMICONDUCTORS

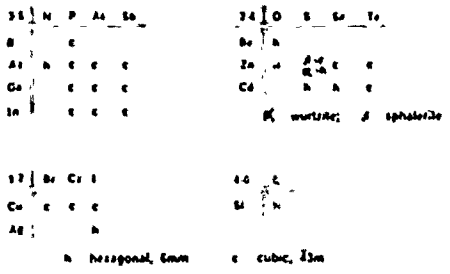
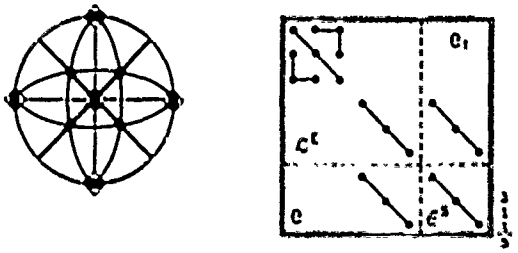


Figure 3. Binary semiconductors and their crystal classes.

GALLIUM ARSENIDE



CUBIC SYSTEM VIIb CRYSTAL CLASS $\bar{4}3m$ [T_d]

Figure 4. Gallium arsenide, symmetry elements and term scheme.

PRIMITIVE REGION

Orientations considered below are described by the usual Euler angles $\mathscr{X}$ and θ for the axial double rotation given by $(YXw\mathscr{L})\,\mathscr{X}/\theta$. The first rotation about X_3 (Z), by angle $\mathscr{X}$, is followed by a second rotation about X_1' (X'), by angle θ. The primitive region for quartz (class 32) is customarily given as $0° \leq \mathscr{X} \leq 30°$, $-90° \leq \theta \leq 90°$, or as $0° \leq \mathscr{X} \leq 60°$, $0° \leq \theta \leq 90°$, for a total of $\pi/3$ steradians. For cubic materials, the primitive region is bounded by the great circle paths passing through the [100], [110], and [111] directions. These boundaries are described in Refs. 4 and 5 as follows: Path I, [100] to [110], $0° \leq \mathscr{X} \leq 45°$, $\theta = 0°$; Path II, [110] to [111], $\mathscr{X} = 45°$, $0° \leq \theta \leq \sin^{-1}(1/\sqrt{3})$; Path III, [100] to [111], $0° \leq \mathscr{X} \leq 45°$, $\tan\theta = \sin\mathscr{X}$. The total solid angle is $\pi/12$ steradians. It is convenient to employ the space $0° \leq \mathscr{X} \leq 45°$, $0° \leq \theta \leq 90°$; this consists of three primitive regions.

TABLE II. NUMBER OF INDEPENDENT CONSTANTS.

Elastic Constant Type	Crystal Class: $\bar{4}3m$ & m3m	Crystal Class: 23	Order of Magnitude*
(Hooke) c_2	3	3	10^2
c_3	6	8	10^3
c_4	11	14	10^4
c_5	18	26	10^5
c_6	32	48	10^6
c_7	48	76	10^7

* (0.3 to 3.0) GPa times number indicated for main elements.

MATERIAL CONSTANTS[6-15]

In Table III are listed the material constants for indium phosphide (43m)[6], the ternary alloy[10] $A\ell_xGa_{1-x}As$, (43m) and silicon[3] (m3m). For aluminum gallium arsenide, the permittivity given is

$$\epsilon_{11} = \epsilon_{11}^T - e_{14}d_{14} = \epsilon_{11}^T - e_{14}^2/c_{44},$$

to account for the piezoelectric contribution to ϵ^S. Figure 5 plots the relative positions of a number of piezoelectric and/or semiconductor materials as function of unit cell size and bandgap energy, that is, the energy required to break a bond. Larger bandgaps in general indicate higher temperature capability. For $A\ell_xGa_{1-x}As$, both coordinates are functions of the alloy ratio X, as are the material values in Table III.[10]

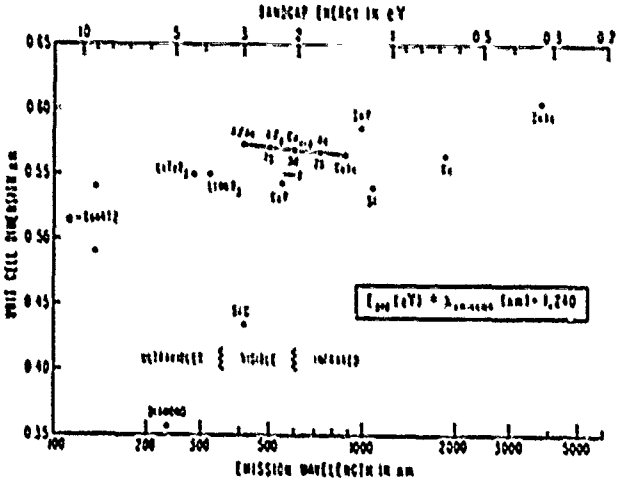


Figure 5. Bond energy versus cell dimensions for acousto-electronic materials.

GALLIUM ARSENIDE[4]

The (100) cut, corresponding to the [100] direction, has effective elastic constants of $c_a = c_{11}$ and $c_b = c_c = c_{44}$. All modes are elastically uncoupled and piezoinactive for thickness-directed fields. The frequency constants are $N_a = 2.363$ and $N_b = N_c = 1.672$ MHz-mm. For lateral field excitation the shear modes may be driven, with coupling factors shown in Fig. 6. The displacement for the extensional mode is along (100); the other displacements are perpendicular to (100), but are otherwise arbitrary.

For the (110) cut, effective stiffnesses are $c_a = c_{44} + (c_{11} + c_{12})/2$, $\bar{c}_b = \bar{c}_{44} = c_{44} + e_{14}^2/\epsilon_{11}$, and $c_c = (c_{11} - c_{12})/2$. Frequency constants are $N_a = 2.617$, $N_b = 1.675$, and $N_c = 1.237$ MHz-mm. Thickness coupling factors are $k_a = k_c = 0$, and $k_b = 6.06\%$; k_b (thickness excitation) for this cut has the same magnitude as $\underline{k}_b$ and $\underline{k}_c$ (lateral excitation) for the (100) cut. The "b" mode is polarized along X_3 (Z), and the "a" mode is polarized along [110]. Figure 7 gives $\underline{k}_a$ versus psi, the azimuthal angle; the shear modes are inert to lateral fields. The (111) cut has modal elastic stiffnesses $\bar{c}_a = (c_{11} + 2c_{12} + 4c_{44} + 4e_{14}^2/\epsilon_{11})/3$, $c_b = c_c = (c_{11} + c_{44} - c_{12})/3$. The frequency constants are $N_a = 2.699$, $N_b = N_c = 1.397$ MHz-mm. The shear modes are piezoinactive for thickness fields, while $k_a = 4.34\%$. For lateral fields, $\underline{k}_a = 0$, and $\underline{k}_b = 4.20\%$, independent of psi. Mode "a" is along [11$\bar{1}$], while the shear modes are perpendicular to [111], but otherwise arbitrary.

Passing now from individual cuts to the general family of doubly rotated GaAs cuts, we show, in Fig. 8, the frequency constants N_m ($\phi\theta$) for m = a, b, and c modes, and 0°(9°)45° for ϕ. Corresponding thickness-directed coupling factors for m = a, b, and c are given in Figs. 9, 10, and 11, respectively. As mentioned previously, Fig. 8 covers three primitive regions. Accordingly, a number of points may be identified

TABLE III. MATERIAL CONSTANTS

QUANTITY	UNIT	Indium Phosphide In P (43m)	Gallium Aluminum Arsenide $A\ell_x Ga_{1-x} As$ (43m)	Silicon Si (m3m)
Elastic stiffness				
c_{11}	10^9 Pa	102.2	118.8 + 1.4X	165
c_{12}		57.6	53.8 + 3.2X	64
c_{44}		46.0	59.4 - 0.5X	79.2
Piezoelectric constant				
e_{14}	C/m²	0.11	-0.16 - 0.065X	—
Permittivity				
ϵ_{11}	pF/m	109.35	116.27 - 28.06X	105.36
Mass density				
ρ	10^3kg/m³	4.787	5.36 - 1.6X	2.328

with equivalent cuts. N_a is maximum at $\phi = 45°$, $\theta = \sin^{-1}(1/\sqrt{3}) \cong 35.26°$; this is (111). At this same angle-pair, N_b may be seen to equal N_c. At $\theta = 45°$, and $\phi = 0°$, N_c is a minimum; this is (110). For $\theta = 90°$, the (001) cut, equivalent to (100), is reached.

The great circle paths I, II, & III, described above, may readily be traced on the curves of Figs. 8-11. For example, path III in Fig. 8 runs along the $\phi = 45°$ line for each mode from $\theta = \sin^{-1}(1/\sqrt{3})$ to $\theta = 90°$. The lateral field coupling factors for these paths were given in Refs. 4 and 5.

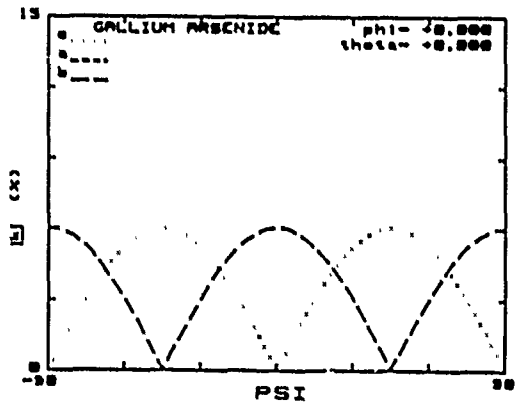


Figure 6. Lateral coupling versus azimuth angle for X-cut GaAs.

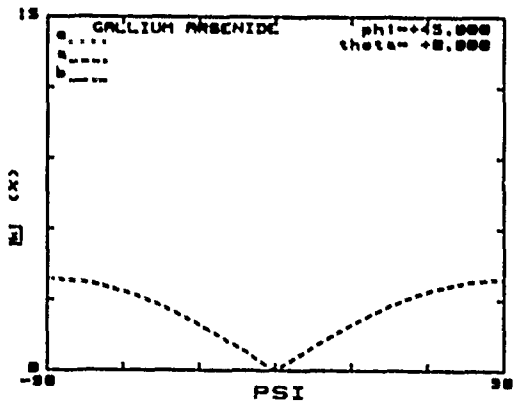


Figure 7. Lateral coupling versus azimuth angle for (110) cut GaAs.

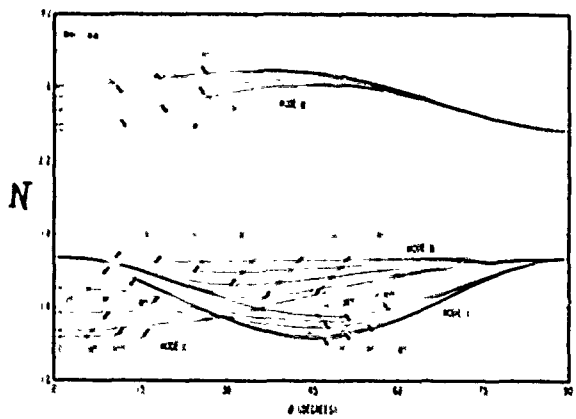


Figure 8. Frequency constants Nm for doubly rotated cuts of GaAs.

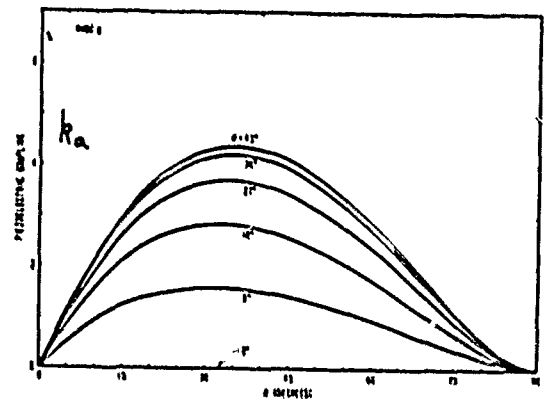


Figure 9. Thickness coupling for mode a of doubly rotated cuts of GaAs.

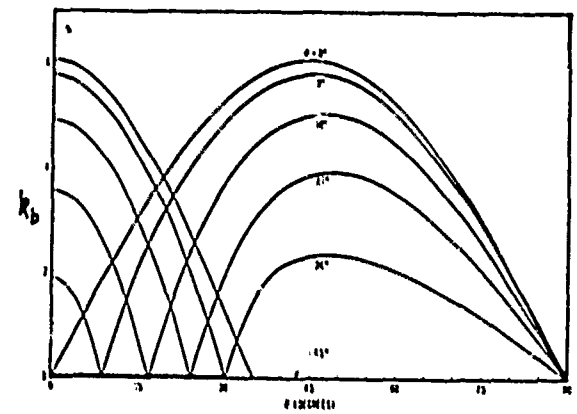


Figure 10. Thickness coupling for mode b of doubly rotated cuts of GaAs.

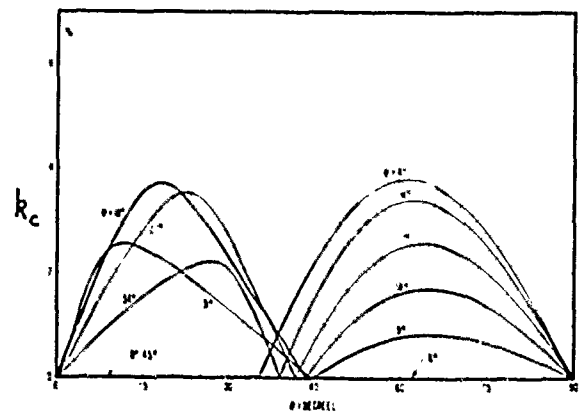


Figure 11. Thickness coupling for mode c of doubly rotated cuts of GaAs.

ACOUSTIC CONE

Another point of interest in Fig. 8 is that for which $\theta \approx 28.7°$ for N_C. This point has all N_C (β) values very nearly equal, corresponding to a cone of propagation directions. Figures 9-11 disclose that the thickness coupling factors k_c (β) vary appreciably with β around the cone. This locus of equality of mode c velocities is not particular to GaAs. As will be seen in Figs. 12, 14, and 16, on enlarged scales, for GaAs, AℓAs ($A\ell_x Ga_{1-x}$ As with x = 1), and InP, respectively, the same phenomenon occurs. Figures 13, 15, and 17 provide the corresponding lateral field coupling factors $\underline{k}_c$ (β). Although the wave speeds on the acoustic cones are nearly the same, the excitation strengths for both thickness- and lateral-directed fields are strong functions of angle.

EQUIVALENT NETWORK

Reference 16 describes the transmission line and lumped element equivalent network representations for lateral field excitation of plates comprised of insulating piezoelectric media; the transmission line network for thickness excitation differs by the presence of a negative capacitance, and is shown in Ref. 17.

With the addition of electrical conduction the equivalent networks take on a modified form. In semiconductors, the current terms other than displacement and piezoelectric flows arise from carrier drift and gradients. Charge carrier gradients contribute a diffusion term, while the drift component appears as an ohmic term. In either or both cases, the conductivity in effect makes the effective dielectric permittivity complex. The piezoelectric stiffening term is therefore also complex, so that acoustic velocity and admittance are complex numbers as well. Table IV provides the relations for the equivalent network parameters for insulating and semiconducting crystals. The quantity F given there is for conduction arising solely from charge carrier drift; inclusion of diffusion complicates the form of F but does not change the forms of the equivalent networks in Fig. 18 and 19.

Figure 18 is the exact network for a single, simple thickness mode of an unbounded plate with electronic conduction, driven by a thickness-directed field. Lateral field excitation would be represented by short circuiting the negative C_0 and G_0 elements in series with the piezo transformer. The network element values are given in Table IV.

In Fig. 18 another source of loss could be incorporated without changing the form of the circuit. Admitting the presence of acoustic viscous loss adds another imaginary term to the stiffened elastic constant.

Losses due to elastic viscosity and electronic (and/or hole) conductivity are most simply expressed in the forms of the Butterworth-Van Dyke lumped equivalent circuits given in Fig. 19. These are derivable from the transmission line form of Fig. 18 by means described in Ref. 16. The effect of elastic viscosity shows up in the presence of R_1, the motional resistance, which is in series with L_1 and C_1. In the absence of this loss, $R_1 \rightarrow 0$. The effect of electronic conductivity appears in the guise of a resistance R_0, the static resistance, which is in parallel with the static capacitance C_0. This is identical to the equivalent circuit of a piezoelectric ceramic resonator.[18] As the semiconductor becomes more like an insulator $R_0^{-1} \rightarrow 0$.

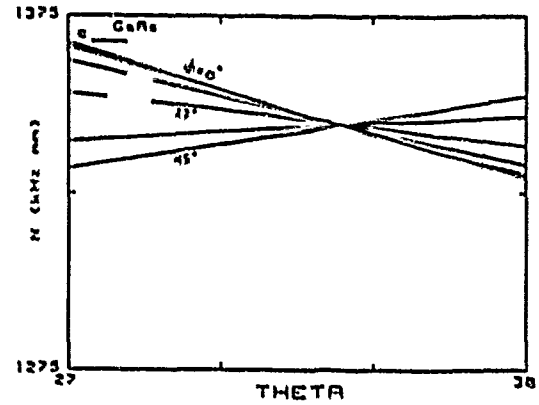


Figure 12. Frequency constants, N_C, for GaAs near (YXwℓ)β/28°.

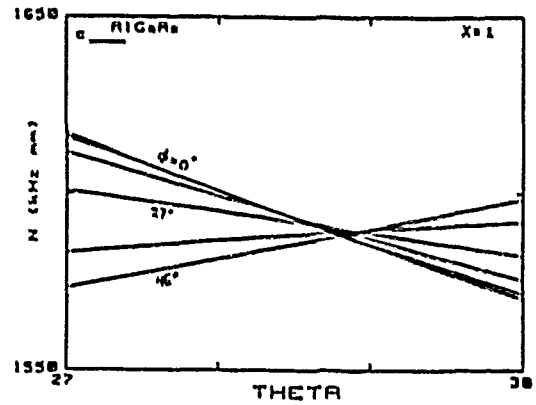


Figure 14. Frequency constants, N_C, for AℓAs near (YXwℓ)β/28°.

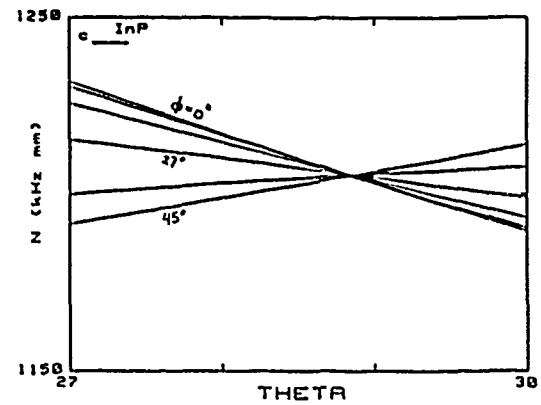


Figure 16. Frequency constants, N_C, for InP near (YXwℓ)β/28°.

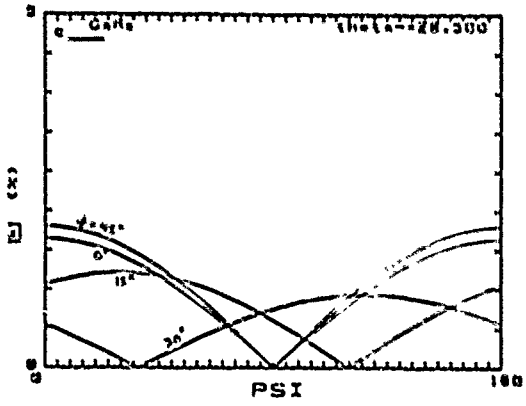


Figure 13. Lateral couplings, k_c, for GaAs near (YXwℓ)β/28°.

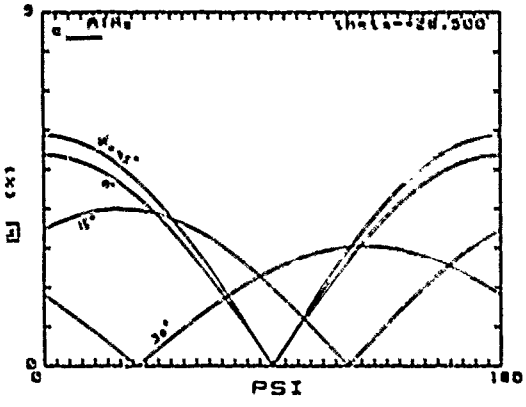


Figure 15. Lateral couplings, k_c, for AℓAs near (YXwℓ)β/28°.

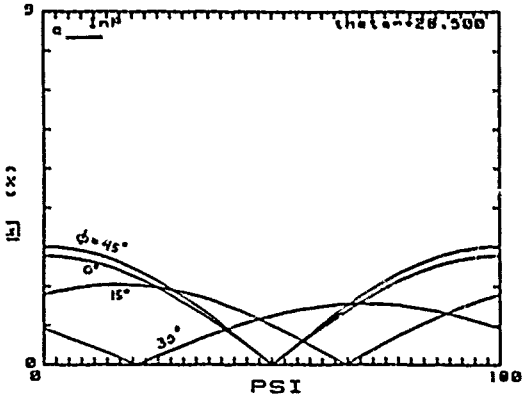


Figure 17. Lateral couplings, k_c, for InP near (YXwℓ)β/28°.

For semiconductors, the temperature coefficient of R_0 (for each mode and cut) is apt to be very large and a strong function of mode, cut, temperature range of measurement, and material-dependent features such as doping and treatment. It is a potential source of error in the determination of the temperature coefficients of the elastic constants unless provided for in the experiments.

Table V gives a comparison of the static and motional time constants for insulating quartz,[12,19,20] for the nonpiezoelectric semiconductor silicon, and for the piezoelectric semiconductor gallium arsenide. It is seen that GaAs, as regards conductivity and static time constant at room temperature, is approximately at the geometric means of the quartz-silicon values. This is approximately true for many other binaries also. The elastic viscosities of GaAs are not known at present, but are thought to be roughly the same order of magnitude as those for quartz and silicon. This would yield the same estimate for its motional time constants, approximately ten femtoseconds.

TABLE IV.

ELECTRICAL NETWORK PARAMETERS FOR INSULATING AND SEMICONDUCTING CRYSTALS

Quantity	Unit	Insulator		Semiconductor	
		Symbol	Relation	Symbol	Relation
Static Capacitance	F	C_0	$\epsilon A/2h$	C_0^*	$C_0 \cdot C_0/1-$
Static Conductance	S			G_0	$\epsilon A/2h$
Piezoelectric Transformer Ratio	C/m	n	$Ae/2h$	n*	n
Acoustic Admittance	s/Ω	Y_0	$1/A \cdot v$	Y_0^*	$Y_0 \cdot F^{-1/2}$
Normalized Frequency Variable	-	x	ωh	x*	$x \cdot F^{-1/2}$

$F = (1-k^2/(1+j\omega\tau))$; $\tau = \tau_0 = C_0/G_0$

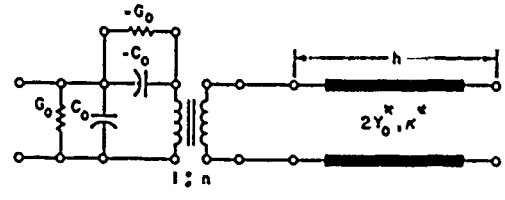


Figure 18. Exact, single mode, transmission-line network for the simple thickness modes of a piezoelectric semiconducting plate.

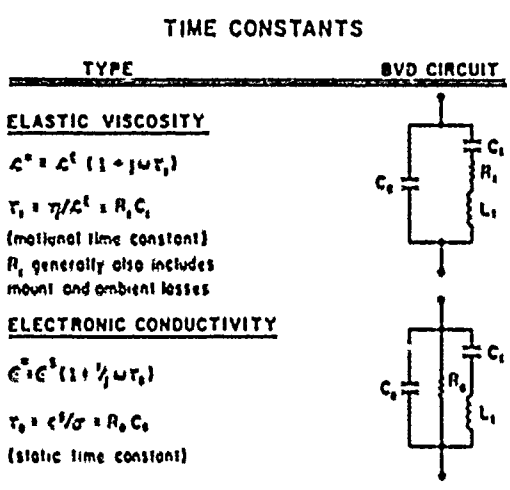


Figure 19. Time constants and the Butterworth-Van Dyke representation of Fig. 18.

REFERENCES

1. E.A. Gerber, T. Lukaszek, and A. Ballato, "Advances in microwave acoustic frequency sources," IEEE Trans. Microwave Theory Tech., Vol. MTT-34, No. 10, October 1986, pp. 1002-1015.

2. D.V. Shick, D.S. Stevens, and H.F. Tiersten, "Quality factor of the piezoelectric thin film on semiconductor composite resonator resulting from radiation into the semiconductor wafer," J. Appl. Phys., Vol. 60, No. 7, October 1986, pp. 2238-2249.

3. K.-H. Hellwege, ed., Landolt-Börnstein, New Series, Group III, Volume 11, Springer, Berlin, 1979.

4. A. Ballato and C. Bosco, "Acoustic waves in cubic crystals: networks for semiconducting vibrators, and applications to gallium arsenide," Technical report DELET-TR-79-9, April 1979, U.S. Army ERADCOM, Fort Monmouth, NJ 07703-5000.

5. A. Ballato, E.R. Hatch, T. Lukaszek, and M. Mizan, "Lateral-field coupling of rotated BAW plates with 3m, 4mm, & $\bar{4}3m$ symmetries," IEEE Ultrasonics Symposium Proceedings, 1986, pp. 339-342.

6. J. Henaff, M. Feldmann, M. Carel, and R. Dubois, "Acoustic wave propagation on indium phosphide surfaces," Appl. Phys. Lett., Vol. 41, No. 1, July 1982, pp. 22-24.

7. F.S. Hickernell and W.R. Gayton, "Elastic constants of single-crystal indium phosphide," J. Appl. Phys., Vol. 37, No. 1, January 1966, p. 462.

8. F. Hickernell, "The electroacoustic gain interaction in III-V compounds: gallium arsenide," IEEE Trans. Sonics Ultrason., Vol. SU-13, No. 2, July 1966, pp. 73-77.

9. J.K. Poe, "Measuring the temperature coefficients of the elastic constants of indium phosphide," Master's Thesis, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, December 1983. Document AFIT/GE/EE-83D-59.

TABLE V.

MATERIAL CONSTANTS

QUANTITY	SYMBOL	UNIT	QUARTZ (32)	GALLIUM ARSENIDE ($\bar{4}3m$)	SILICON (m3m)
Elastic stiffness	c	10^9 Pa	29 (AT) slow shear	59 (110) fast shear	165 (c_{11}); 79 (c_{44})
Elastic viscosity	η	10^{-6} N.s/m^2	344 (AT)	—	1505 (η_{11}); 553 (η_{44})
Motional time constant	$\tau_1 = \eta/c$	10^{-15}s	12 (AT)	—	9.1 (long.); 8.6 (shear)
Permittivity	ϵ	10^{-12} F/m	40	116	105
Conductivity	σ	10^{-12} S/m	1 (Z); 0 005 (X,Y)	$\sim 10^{+5}$	$4.3 \cdot 10^{+8}$
Static time constant	$\tau_0 = \epsilon/\sigma$	s	40 ; 8000	$\sim 10^{-3}$	$2.4 \cdot 10^{-7}$
Ratio	τ_0/τ_1	—	$4 \cdot 10^{15}$; $6 \cdot 10^{17}$	—	$2.7 \cdot 10^{+7}$

10. S. Adachi, "GaAs, AlAs, and $Al_x Ga_{1-x} As$: Material parameters for use in research and device applications," J. Appl. Phys., Vol. 58, No. 3, August 1985, pp. R1-R29.

11. J.S. Cohen and A.G. Schlijper, "Long-range order in $Al_x Ga_{1-x} As$," Phys. Rev. B, scheduled for 15 July 1987.

12. J. Lamb and J. Richter, "Anisotropic acoustic attenuation with new measurements for quartz at room temperatures," Proc. Roy. Soc. A, Vol. 293, 1966, pp. 479-492.

13. H.J. McSkimin and P. Andreatch, Jr., "Elastic moduli of silicon vs hydrostatic pressure at 25.0°C and -195.8°C," J. Appl. Phys., Vol. 35, No. 7, July 1964, pp. 2161-2165.

14. J.J. Hall, "Electronic effects in the elastic constants of n-type silicon," Phys. Rev., Vol. 161, No. 3, September 1967, pp. 756-761.

15. M. Ezz-El-Arab, B. Galperin, J. Brielles, and B. Vodar, "Variation des vitesses de propagation des ultrasons dans le silicium monocristallin entre 25° et 830°C," Solid State Commun., Vol. 6, 1968, pp. 387-390.

16. A. Ballato, E.R. Hatch, M. Mizan, and T.J. Lukaszek, "Lateral field equivalent networks and piezocoupling factors of quartz plates driven in simple thickness modes," IEEE Trans. Ultrason., Ferroelec., & Frequency Control, Vol. UFFC-33, No. 4, July 1986, pp. 385-392.

17. T.R. Meeker, "Bulk Acoustic Waves and Resonators," in Precision Frequency Control, Vol. 1, (E.A. Gerber and A. Ballato, eds.), Chap. 2.1, p. 105, 1985. Academic Press, Orlando, FL.

18. B. Parzen, Design of Crystal and Other Harmonic Oscillators, Chapter 3, Wiley, New York, 1983.

19. W.G. Cady, Piezoelectricity, McGraw-Hill, New York, 1946; Dover, New York, 1964.

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

41st Annual Frequency Control Symposium - 1987

NEW PROSPECTS FOR ACOUSTIC SENSORS: AN OVERVIEW

Richard M. White
Department of Electrical Engineering and Computer Sciences
and the
Berkeley Integrated Sensor Center
University of California
Berkeley, CA 94720

Summary

Sensors employing elastic waves have been devised for detecting a wide range of non-acoustical measurands. Examples include: biological substances (antigens in a liquid); chemicals, such as styrene vapor and hydrogen sulfide; electrical and magnetic fields; temperature; and a host of mechanical measurands, such as position, force and pressure, acceleration, and material properties such as viscosity. The reason for using an acoustical sensor varies with the application. High sensitivity is attractive in the case of chemical, acceleration and temperature sensing. Near-linear response and wide dynamic range are also important considerations. Low cost and simplicity characterize several acoustical position sensing systems for use with robots and computer displays. In many of these applications, the output variable is a frequency, which is a convenient input variable for digital storage and processing circuitry.

Most of the newer acoustic sensing devices employ resonators or delay-line bulk- or surface-wave oscillators whose frequency is affected by the measurand. Resonant elements that are used include conventional piezoelectric crystal resonators, such as plates and tuning forks, SAW resonators formed on crystalline piezoelectric substrates or in sputtered piezoelectric thin films, and tiny micromechanical beams and bridges formed by integrated-circuit fabrication techniques. Other sensors of current interest, particularly for position sensing and materials characterization, use measurement of timing and reflection, transmission, or attenuation made on ultrasonic waves propagating in air or in a solid.

Device principles, operating characteristics, and possible limitations of a number of these acoustic sensors will be described.

****

Acoustical techniques have been used in sensors for many different measurands (see Table I).

Table I. List of measurands for which sensors might be of interest. Bullets (•) indicate measurands for which acoustical sensors have been reported (Original table of measurands is from Reference 1).

A1. Acoustic		
A1.1	•	Wave amplitude, phase, polarization, spectrum
A1.2	•	Wave velocity
A1.3		Other (specify)
A2. Biological		
A2.1	•	Biomass (identities, concentrations, states)
A2.2		Other (specify)
A3. Chemical		
A3.1	•	Components (identities, concentrations, states)
A3.2		Other (specify)
A4. Electric		
A4.1		Charge, current
A4.2		Potential, potential difference
A4.3	•	Electric field (amplitude, phase, polarization, spectrum)
A4.4	•	Conductivity
A4.5	•	Permittivity
A4.6		Other (specify)
A5. Magnetic		
A5.1	•	Magnetic field (amplitude, phase, polarization, spectrum)
A5.2		Magnetic flux
A5.3		Permeability
A5.4		Other (specify)
A6. Mechanical		
A6.1	•	Position (linear, angular)
A6.2	•	Velocity
A6.3	•	Acceleration
A6.4	•	Force
A6.5	•	Stress, pressure
A6.6	•	Strain
A6.7	•	Mass, density
A6.8		Moment, torque
A6.9	•	Speed of flow, rate of mass transport
A6.10	•	Shape, roughness, orientation
A6.11	•	Stiffness, compliance
A6.12	•	Viscosity
A6.13	•	Crystallinity, structural integrity
A6.14		Other (specify)
A7. Optical		
A7.1		Wave amplitude, phase, polarization, spectrum
A7.2		Wave velocity
A7.3		Other (specify)
A8. Radiation		
A8.1		Type
A8.2		Energy
A8.3		Intensity
A8.4		Other (specify)
A9. Thermal		
A9.1	•	Temperature
A9.2		Flux
A9.3		Specific heat
A9.4		Thermal conductivity
A9.5		Other (specify)

Principles long-used in ultrasonic sensors include (1) measurement of the characteristics of reflected ultrasonic waves (used in non-destructive materials evaluation, surface characterization, position sensing, etc.), and (2) study of sound emitted from structures under stress (for example, acoustic emission measurements to assess structural integrity). More recently, ultrasonic microsensors have been based on two different principles:

- Shift of resonant frequency of a resonator (used to sense biological substances, vapors, pressure, acceleration, pressure, etc.)
- Shift of operating frequency of a delay-line oscillator (used to sense biological substances, vapors, electric and magnetic field, pressure, acceleration, etc.)

CH2427-3/87/0000-333 $1.00 C 1987 IEEE

These devices detect a change in ultrasonic wave velocity caused by the measurand of interest. Surface, rather than bulk, waves have been utilized in the newest sensors for higher sensitivity. Increasingly, techniques developed for integrated circuit manufacture are being used to fabricate miniature sensor structures to support acoustic wave propagation. Examples are membranes and beams, formed by etching silicon wafers, that may be only a few microns thick and have lateral dimensions ranging from tens of microns to many millimeters. For wave transduction on silicon, which is not piezoelectric, one may use sputtered zinc-oxide thin films.

Four effects that may cause the ultrasonic wave velocity to change have been used in these sensors:

- Mass loading of the surface
- Change of material stiffness
- Change of electrical conductivity at the surface
- Change of permittivity at the surface

The first two occur with non-piezoelectric and piezoelectric media, while the last two occur only with piezoelectrics. In general, the directions of the first two changes are consistent with the observation that wave velocity v_p depends on density ρ and stiffness c as $v_p \propto (c/\rho)^{1/2}$.

Ultrasonic techniques are used in sensors in part because of their high sensitivity and large dynamic range. Furthermore, ultrasonic sensors whose output is a variable frequency are attractive because their quasi-digital nature permits them to interface easily with digital systems for data storage, display and interpretation.

SAW Sensor Characteristics

Table II lists sensitivities of some recently described SAW sensors. The biological and chemical sensors listed all are based on the delay-line oscillator, which is described in more detail below. The position sensor is based on the attenuation of a SAW probe beam caused when an object such as a human finger presses against a glass display screen. Resonator frequency shifts are the basis of the acceleration, pressure, and temperature sensors listed in Table II. The viscosity sensor involves measuring the loss caused when a viscous liquid contacts a piezoelectric crystal in which a horizontally-polarized shear wave is propagating. Finally, the mass sensor result listed is for a photoresist-coated SAW delay-line oscillator fabricated in lithium niobate.

SAW Delay-Line Oscillator Sensors

Fig. 1 shows schematically a SAW delay-line-oscillator chemical vapor sensor that could be realized with a piezoelectric single crystal, such as lithium niobate or quartz, or with zinc oxide on a silicon substrate. Oscillation at a frequency f within the common passbands of the two interdigital transducers will occur if (1) the amplifier provides sufficient gain to overcome losses in the acoustical path and the transducers, and (2) the following phase condition is satisfied:

$$(2\pi f/v_p)L+\phi_E=2\pi N \qquad [1]$$

where L is the center-to-center transducer spacing, ϕ_E is the phase shift in the amplifier, and N is an integer. From this condition, which specifies that the total phase shift around the loop must be an integer times 2π radians, one finds that the fractional frequency shift is equal in magnitude and opposite in sign to the fractional velocity change caused by a measurand. Analysis based on a perturbation approach shows that in the case where absorption of a chemical substance causes a change in mass loading of the surface but not a change in stiffness, the fractional frequency shift of the delay-line oscillator can be written as

$$\delta(f/f_0)=-Cf_0\delta(M/A) \qquad [2]$$

where f_0 is the original oscillation frequency, C is a coefficient roughly equal to $2\times10^{-7}m^2s/kg$, and M and A are the mass and area of the surface where material has been added or removed.[2] Such oscillator sensors have been operated at frequencies from 10 to 300 MHz, and have stabilities as high as 1 part in 10^7. The coefficient C depends upon the crystal and orientation used, but it is nearly the same as the corresponding coefficient for the familiar bulk resonator vapor sensor, first described by Sauerbrey.[3] Because of the presence of the f_0 term and the fact that SAW devices can be made at frequencies much higher than are practical for bulk resonators, the sensitivity of the SAW sensor can be much higher than that of the bulk wave sensor. (Table II shows the computed sensitivity of the SAW sensor at the highest practical frequency of 3GHz. This sensitivity corresponds to detection of about one thousandth of a monolayer of added material in the acoustic path.)

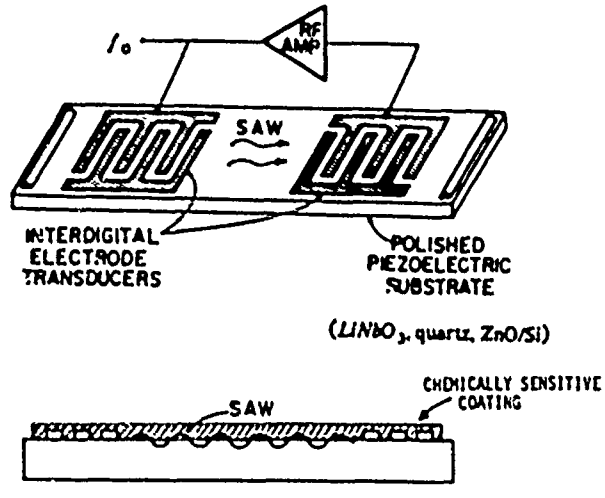


Fig.1 Surface-acoustic-wave oscillator for chemical sensing.

Some techniques used to deposit coatings on delay-line gas or vapor sensors are listed in Table III. In addition, techniques used routinely in immunoassay studies for immobilizing biological substances, such as antibodies and enzymes, may be useful for making ultrasonic delay-line biosensors.

Achieving Selective Response in SAW Sensors

As Table II shows, many measurands can cause a change of ultrasonic wave velocity. Hence, achieving selective response to a desired measurand must be addressed. In accelerometers, particular crystallographic orientations that minimize cross-coupling have been investigated. Because of the effects of temperature upon velocity, many investigators have used temperature-

compensated cuts such as ST-quartz as the sensor substrate. In addition, most reported SAW sensors have employed a reference and an active device to reduce interference due to measurands other than that of primary interest. Change of the difference in frequency of the reference and active device is then measured. Automatic control of the gains of the two amplifiers is also found to improve sensor performance. Recently Venema et al.[4] have described an all-silicon SAW delay-line sensor utilizing active and reference ZnO-on-Si delay lines, and a companion integrated circuit containing amplifiers and mixer to derive the difference frequency.

Table II. SAW Sensor Characteristics

MEASURAND	SENSITIVITY	REFERENCES
Biological		
Antigen	Exp. 13 micrograms 10-20 MHz	5
Chemical		
Vapor	Theor. to 3×10^{-11} g/e m^2 at 3 GHz	2
Dimethyl acetamide	Exp. 358 Hz/ppm at 294 MHz; selective	6
H_2S	Exp. to 10 ppb claimed with activated WO_3	7
NO_2	Exp. to 0.5 ppm, 100 Hz/ppm at 40 MHz with pthalocyanine; selective; mass loading and conductivity change	4
Mechanical		
Position	SAW on glass, 4-10MHz, attenuation from contact on display screen	7
Acceleration	Exp. 1400-10,000 Hz/g at 105 MHz Exp. response to 10^{-5}g	9,10,11
Pressure	Exp. 400 Hz/kPa at 105 MHz, temp. compensated -40 to 80°C	9
Mass, density	Exp. photoresist etching, 75MHz osc., 690kHz increase for 1 micron thickness reduction	12
Viscosity	Exp. 159MHz HPSW (SSBW), attenuation $\propto (\eta)^{1/2}$	13
Thermal		
Temperature	Exp. 2-3 kHz/K, 0.1mC min. resolvable	14

With chemical and biological sensors particularly, the following techniques for increasing selectivity have been considered:

- Filtering (e.g., using zeolites that admit molecules smaller than a pore diameter)
- Separating molecules by rate of diffusion in a porous absorber
- Exploit solubility parameter differences (match absorber coating to target vapor)
- Utilize a chemical reaction

Table III. Coating techniques for ultrasonic delay-line gas and vapor sensors.

Technique	Comments
1. Subliming	Limited applicability.
2. Painting	Thickness uniformity difficult to achieve.
3. Spraying	Uniformity an issue; droplets noted in deposit.
4. Ink-jet printing	Control good; process can be automated; suitable for localized deposits
5. Dipping	Usually effective; ultimate is Langmuir-Blodgett technique for depositing mono-molecular layers.
6. Solvent casting	Provides thickness control via dilution used; improved uniformity with two-component solvent.

- Exploit specificity of certain biochemical reactions (e.g., antibody-antigen reaction)
- Employ spectroscopy (e.g., thermal desorption spectroscopy)
- Employ pattern recognition (seek pattern in responses of many absorbers whose responses are not completely selective)

We discuss below some of these techniques for obtaining selective response.

SAW Resonator with Porous ZnO Coating

Martin et al.[15] at Sandia Laboratories showed that with their ZnO-on-Si SAW resonator oscillator one could obtain a correlation between molecular size and rate of sensor response. They found in work with a 109MHz oscillator that the frequency dropped quickly upon admission of a gas, and then continued to decrease at a much lower rate (Fig. 2a). They interpret the sharp initial drop as being due to adsorption on the ZnO surface, and the slower fall as a result of the gradual permeation of the porous ZnO by gas molecules. The time constant for this second process correlated well with molecular diameter (Fig. 2b): The larger molecules diffused in rapidly, while the smaller ones diffused more slowly into the film as they tended to follow the convolutions of the pore walls. Thus measurements of this time constant would permit discriminating among gases on the basis of their molecular sizes.

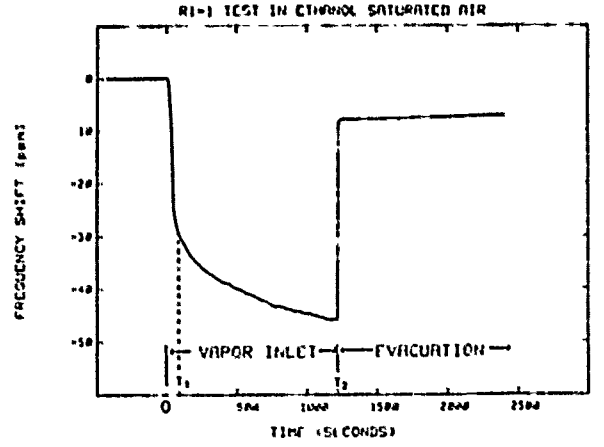


Fig.2a Frequency shift induced in the SAW resonator sensor by vapor exposure and re-evacuation (From Reference 15).

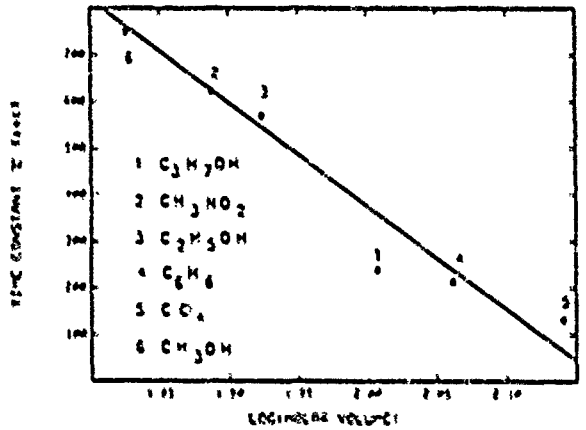


Fig.2b Long-term equilibration time constant vs. size of the vapor molecule, suggesting size restricted permeation into the ZnO layer (From Reference 15).

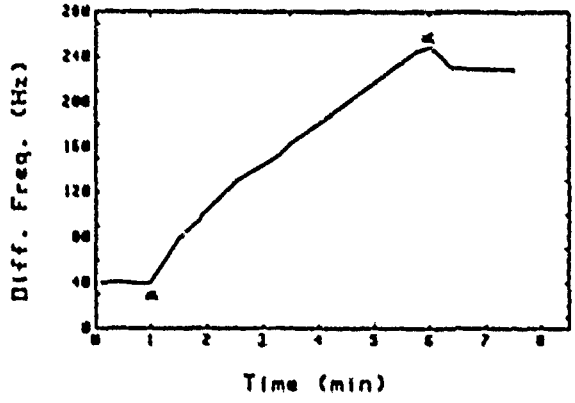


Fig.3b Response of dual-oscillator SAW sensor exposed t 50ppm of styrene.

SAW Sensor Exploiting a Chemical Reaction

Zellers[16] has described a SAW delay-line-oscillator sensor for styrene vapor that achieves its selectivity by employing a chemical substitution reaction. This sensor is also reversible.

Fig. 3 shows schematically a cross-section of the sensor substrate, polymer coating, and vapor flow outside the sensor. The polymer, operating above its glass transition temperature to ensure rapid diffusion of absorbed molecules, contains solid particles of a platinum-based compound that contains an ethylene group. Absorption of styrene vapor permits a reaction to occur in which styrene substitutes for the ethylene group, releasing the ethylene to diffuse out of the polymer into the surrounding atmosphere. The mass per unit area of the polymer increases as a result, since the styrene molecule is heavier than the ethylene molecule (molecular weights are 104 and 28 respectively), and the oscillator frequency falls. When tested with styrene at 50ppm together with a typical co-contaminant such as methyl-ethyl-ketone (MEK) at 300ppm, the sensor did not respond to the co-contaminant. Furthermore, the polymer coating could be regenerated after prolonged exposure to styrene by flowing ethylene over it. This 30MHz, ST-quartz device could be used to measure both integrated dose (total frequency shift over an exposure time) and current concentration (rate of change of oscillator frequency). The sensor frequency changes at a rate of 25Hz/min at 50 ppm styrene concentration. The estimated minimum measurable concentration in this device was 3ppm.

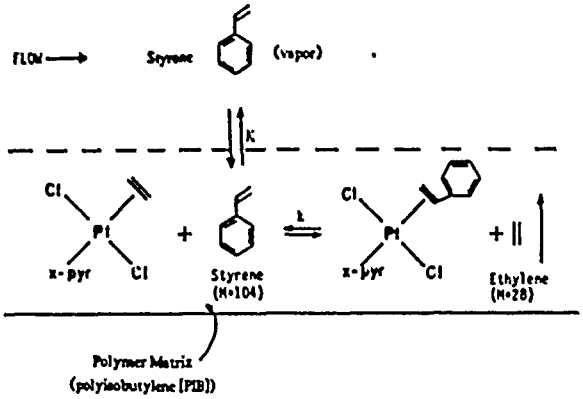


Fig.3a Schematic cross section through polymer coating on SAW styrene sensor.

Immersible SAW Sensors

Sensors for detecting biological substances, as well as cer tain chemicals, may have to function while immersed in a liquic Workers familiar with SAW devices would anticipate difficult since the propagating surface waves will tend to radiate energ into the liquid and hence suffer attenuation. Viktorov's[17] studie suggest that losses around 4 dB/MHz-cm will arise. The experi mental biosensor results of Roederer and Bastiaans[18] and of Bas tiaans[5] show that the losses are acceptable at frequencies as lo as 10MHz, however. In spite of the presence of a liquid colum and an O-ring that sealed the serum container to the SAW dev ice, the authors showed a clear frequency shift when a antigen i the serum was exposed to the corresponding antibody which ha been immobilized on the ST-quartz SAW substrate.

At higher frequencies, serious loss is anticipated, so othe designs for low-loss immersed ultrasonic sensors are of interes We should note first that two conditions are necessary for energ radiation into the liquid: the wave on the solid must have a di placement component normal to the surface, and the wave velc city along the surface must exceed the velocity of sound in th liquid. Two approaches have been suggested for violating thes conditions and hence eliminating the radiation loss. The first i to use a horizontally-polarized wave, and the second is to use low-velocity Lamb wave.

Horizontally-Polarized Shear Device

Workers at Sandia[13] have used interdigital transducers o an ST-quartz plate to generate horizontally-polarized she waves. When a liquid layer was present on the device shown i Fig. 4, viscous coupling of the vibrating surface to the liquid w the only source of loss -- no radiative loss was observed. I fact, in tests with pure water, no loss at all was observed until small quantity of surfactant was added to permit the water couple to the quartz. Study showed that bulk shear modes rad ated down into the plate from the input transducer, and that th device was sensitive to viscous liquids placed either above below the quartz crystal.

Analysis shows that the attenuation in this device, measur in dB, is proportional to the square root of the viscosity. T thickness of the layer of liquid that takes part in the vibrato motion is $\delta = (2\eta/\rho\omega)^{1/2}$, where η is the viscosity, ρ the densi and ω the angular frequency of the wave. This layer can quite thin, permitting one to measure viscosities of very sm liquid samples; for example, the thickness for water at

160MHz operating frequency is only about 500 angstroms. The device can also be used in a delay-line oscillator mode to measure changes of mass loading at the surface, as in the study of corrosion of a deposited metal film.

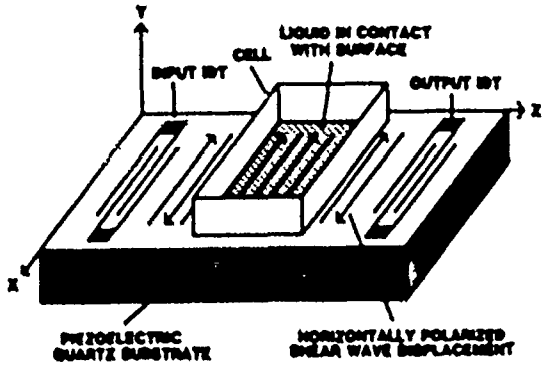


Fig.4a Schematic diagram of micro-viscometer using a horizontally polarized shear wave (From Reference 12).

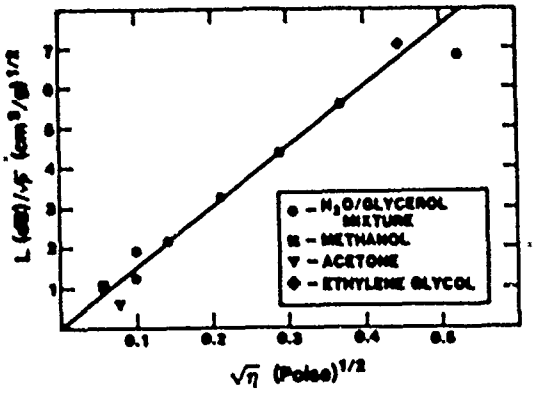


Fig.4b Relative power loss between transducers due to viscous damping of the horizontally polarized shear wave (From Reference 12).

Lamb-Wave Membrane Sensor

The phase velocity of the lowest-order antisymmetric Lamb wave approaches zero as the membrane supporting the wave is made arbitrarily thin. Since velocities of only a few hundred meters per second are possible with practical membrane dimensions and operating frequencies,[19] an immersible membrane sensor (Fig. 5) appears feasible. Computer studies have shown that a zinc oxide film on such a structure will have a significantly higher piezoelectric coupling coefficient than on a silicon wafer of standard thickness. More importantly, the computed sensitivity of oscillator frequency to added mass or thickness is typically an order of magnitude higher than that of a SAW device at the same wavelength[16] (a wavelength of 140 microns was chosen, corresponding to 31 MHz operation of the SAW device and 1.5MHz operation of the 2-micron-thick membrane device).

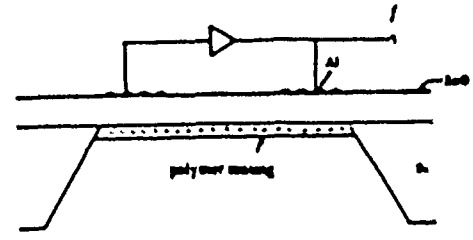


Fig.5 Cross section of ZnO-coated plate-mode oscillator sensor. With a 2-micron-thick silicon nitride membrane, the lowest antisymmetric mode had, at 2.9MHz (140 micron wavelength), a measured phase velocity of only 400m/s.

Conclusions

Sensors based on ultrasonic techniques typically have high sensitivity and large dynamic range. The resonator and delay-line oscillator acoustic sensors also have a convenient quasi-digital output. Acoustic sensors for many different measurands have been reported, as indicated in Table I. The use of piezoelectric thin films on a semiconductor permits combining acoustic sensors with integrated circuitry for signal conditioning, linearization, and other purposes. In addition to their promise for conventional sensing applications, these devices are also opening new avenues for making measurements in scientific research.

Acknowledgement

This research was supported in part by the Berkeley Integrated Sensor Center, an NSF/Industry/University Cooperative Research Center.

References

1. R. M. White, "A sensor classification scheme," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 125-6, March 1987.
2. H. Wohltjen, "Mechanism of operation and design considerations for surface acoustic wave device vapour sensors," *Sensors and Actuators*, vol. 5, pp. 307-325, June, 1984.
3. G. Sauerbrey, "Verwendung von Schwingquarzen zur Wagung duenner Schichten und zur Mikrowaegung," *Z. Phys.*, vol. 155, pp. 206-222, 1959.
4. A. Venema and others, "NO2 gas-concentration measurement with a SAW-chemosensor," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 149-155, March 1987.
5. G. J. Bastiaans and C. M. Good, "Development of a surface acoustic wave biosensor," *Transducers '87 Digest*, June 1987.
6. H. Wohltjen and others, "Trace chemical vapor detection using SAW delay line oscillators," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 172-8, March 1987.
7. J. F. Vetelino and others, "Hydrogen sulfide surface acoustic wave gas detector," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 157-161, March 1987.

8. R. Adler and P. J. Desmares, "An economical touch panel using SAW absorption," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 195-201, March 1987.

9. D. Hauden and others, "SAW cantilever-beam accelerometer sensitivities," *Proc. 1985 IEEE Ultrasonic Symp.*, pp. 486-9, 1985.

10. D. Hauden, "Miniaturized bulk and surface acoustic wave quartz oscillators used as sensors," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 253-8, March 1987.

11. M. E. Motamedi, "Acoustic accelerometers," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 237-242, March 1987.

12. S. G. Joshi, "Use of a surface-acoustic-wave (SAW) device to monitor etching of thin films," *Proc. IEEE*, to be published 1987.

13. A. J. Ricco and S. J. Martin, "Acoustic wave viscosity sensor," *Appl. Phys. Lett.*, vol. 50 (21), pp. 1474-6, 25 May 1987.

14. D. Hauden and others, "Pressure and temperature measurements with SAW sensors," *Proc. 36th Ann. Freq. Cont. Symp.*, pp. 284-9, 1982.

15. S. J. Martin and others, "Gas sensing with surface acoustic wave devices," *Transducers '85 Digest*, pp. 71-73, 1985.

16. R. M. White and others, "Plate-mode ultrasonic oscillator sensors," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. UFFC-34 (2), pp. 162-171, March 1987.

17. I. A. Viktorov, *Rayleigh and Lamb Waves*, Plenum, New York, 1967.

18. J. E. Roederer and G. J. Bastiaans, "Microgravimetric immunoassay with piezoelectric crystals," *Analytical Chemistry*, vol. 55, pp. 2333-2336, December 1983.

19. K. Uozumi, K. Ohsone, and R. M. White, "Generation and detection of ultrasonic Lamb waves in a thin deposited film by using interdigital transducers," *Appl. Phys. Letters*, vol. 43, pp. 917-919, 15 Nov. 1983.

THEORETICAL MODELING OF QUARTZ RESONATOR PRESSURE TRANSDUCERS

Copyright 1987 by IEEE.
Presented at 41st Annual Symposium on Frequency Control, May 27-29, 1987, Philadelphia, PA.

E. P. EerNisse

Quartztronics, Inc.
1020 Atherton Drive, Bldg.C
Salt Lake City, UT 84123

Summary

A new concept for a high pressure range, high temperature, quartz resonator pressure transducer is presented. A thickness shear mode resonator is surrounded by a non-cylindrical, integral shell structure that transmits forces to the resonator from externally applied pressure. The forces lack circular symmetry because of the shell geometry, which is a circular shell with diametrically opposed flats. This non-symmetric force distribution creates non-symmetric stress in the resonator that shifts the resonant frequency proportional to applied pressure.

Finite element modeling has been used to calculate the stress distribution in the resonator due to the non-symmetric shell. A theoretical model is presented that simulates the non-symmetric stress with superposition of a uniform radial force and diametric force pairs. This simulation is used to predict two key features of the pressure transducer: the scale factor and the temperature dependence of the scale factor. The model predicts that proper crystallographic orientation and dimensions of the flats can be combined to increase the scale factor over that obtained with a circularly symmetric shell and to improve the temperature dependence of the scale factor. The new degrees of freedom (flat orientation and dimensions) increase the choices for design of a pressure transducer.

Introduction

The resonant frequency of a thickness shear mode quartz resonator changes when a static stress, called initial stress, is present in the resonator blank.[1-3] The effect, studied initially in order to eliminate it in frequency control applications, can be used as a transducer mechanism. The most successful transducer application has been the pressure transducer offered commercially by Hewlett Packard.[4-8] Their device is a BT-cut resonator surrounded by a unitary quartz shell that isolates the active region of the resonator from the pressure medium (usually a liquid) but has no joints in the high stress region near the resonator. The shell is circular, which results in a uniform radial force, or a uniform in-plane initial stress. The only degree of freedom available to the transducer designer using the circular shell design is the crystallographic orientation of the resonator/shell unit. Thus, only one operating characteristic of the transducer can be optimized; the other characteristics must be accepted without modification for a given crystallographic orientation. Some of the transducer characteristics that might be desired for optimization are frequency vs. temperature behavior, scale factor (frequency vs. pressure), temperature dependence of scale factor, temperature transient effects, size, and cost. With only one degree of freedom to work with, Hewlett Packard[6] decided to optimize the temperature dependence of scale factor while obtaining reasonable frequency vs. temperature behavior. The BT-cut is the only choice available for this criteria when a circular shell is used. Since the BT-cut requires a relatively large diameter to thickness ratio for proper resonator electrical performance free of activity dips, the overall size of the transducer is large. The BT-cut also has a strong dependence of frequency vs. temperature behavior with small changes in contouring of the convex surface, which raises fabrication costs.

The present work was initiated in order to add more freedom to the choices for transducer design, particularly in terms of using other cuts of quartz to obtain the key advantage of small size for lower cost and faster thermal response. The method taken was to break up the circular symmetry used previously. This idea has been proposed by others using slots or bridges to transfer the forces, or stresses, from the shell to the resonator without joints. However, stress concentrations in these structures near the slots preclude use at higher pressures because of brittle fracture of the quartz. The proposed approach leaves the interface between shell and resonator intact around the entire resonator perimeter and obtains the desired symmetry effects by using a non-circular shell.

Shell Shape Effects

Figure 1 shows the circular shell concept used by the earlier workers. Figure 2 shows the non-circular shell used in the present work. The end caps are left off in Fig. 2 for viewing purposes. Both structures have the advantage that the joints to the end caps where a glue substance must be used are isolated from the resonator by a length of shell structure. Since quartz is single crystal, there are no gross hysteresis mechanisms. The circular shell can only generate a uniform radial compressive stress in the resonator disc as outside pressure exceeds inside pressure. The non-circular shell, on the other hand, causes a non-uniform stress in the resonator. The subject of this section is what that stress pattern looks like and how it depends on the shell shape.

CH2427-3/87/0000-339 $1.00 C 1987 IEEE

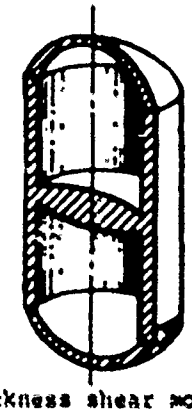


Figure 1: Thickness shear mode disc resonator with circular integral shell.

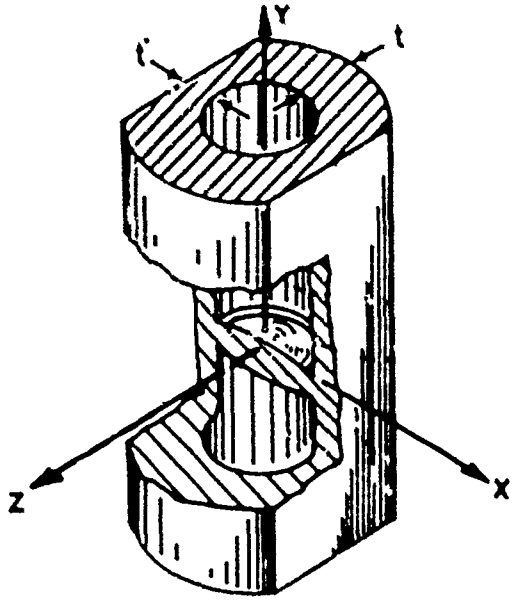


Figure 2: Thickness shear mode disc resonator with non-circular integral shell and the axes system used.

An Aluminum mock-up of Fig. 2 fitted with a strain gauge on the center of one face of the disc was used initially to study the nonuniform stress distribution. A 45° strain gauge Rosette provided strain measurements in three directions. The two principle directions, x and z are shown in Fig. 2. Knowledge of S_{xx} and S_{zz} provides completeness; the 45° answer is the average of the other two.

The Al mockup was placed in a pressure bomb. The strain gauge resistance was monitored as the pressure was increased. After characterizing the slope of strain vs. pressure, the Al mockup was removed from the pressure bomb, mounted onto the bed of a mill, and the flats were machined deeper. The mockup was then placed back into the pressure bomb and strain vs. pressure was again monitored. The process was repeated a second time. Figure 3 shows the strain ratio S_{xx}/S_{zz} (see Fig. 2) vs. the ratio of shell thickness at the flats and away from the flats (see Fig. 2 for t' and t). Note in Fig. 3 that large changes in the S_{xx} and S_{zz} ratio can be obtained.

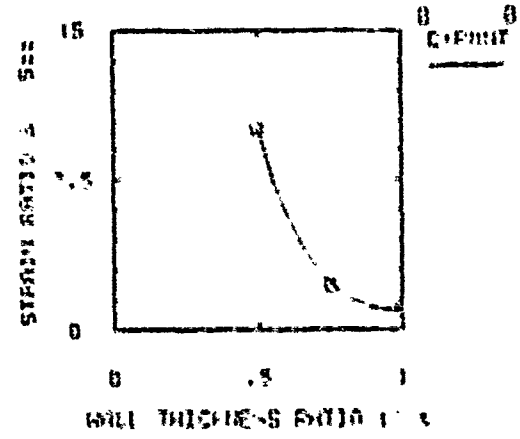


Figure 3: Strain ratio S_{xx}/S_{zz} for the experimental Al mockup found using a strain gauge.

A finite element model was used for analytical investigations. Shell elements were used, which handle both bending stresses and membrane stresses. The program used ran on a desk top computer. The results showed the same type of stress behavior seen in Fig. 3 except that there is a complex tradeoff between t'/t and t_c/t, where t_c is the resonator disc thickness. There is an endless range of the three t variables, so the only quantitative results presented here will be those related to actual quartz devices tested. Qualitatively, the amount of symmetry in the stresses in the disc depends on whether the circular symmetry of the disc or the non-circular symmetry of the shell predominates. Thick discs with thin shells tend to have stresses closer to circular symmetry while thin discs with thick shells tend to have stresses non-circular in symmetry.

Figure 3 shows one new degree of freedom, namely, the depth of the flats and the resulting S_{xx}/S_{zz} ratio. The other new degree of freedom is the crystallographic orientation of the flats. The choice of crystallographic orientation depends on which characteristics of the transducer one wishes to optimize and the interplay between the initial stress pattern and the thickness shear vibration.

Frequency Shift Effects

The effect of initial stress on the resonant frequency of a thickness shear mode has been treated here empirically. It is possible to handle the frequency shift effect analytically at room temperature with available theory incorporating the third-order elastic constants.[4] However, no data exists for the temperature dependence of the third-order elastic constants, so a purely analytical approach is impractical if we are to include the temperature dependence of scale factor in the characteristics of interest of the transducer.

We can take advantage of past experimental data in formulating an empirical model.[9] Figure 4 shows the frequency shift coefficient for the AT-cut due to a diametric force pair as previously published for temperatures of 25 and 78°C.[9] The fractional change in frequency, df/f_o (f_o is the frequency for zero initial stress) is given by

$$df/f_o = K_p N_o F/(d_r t_r) \qquad (1)$$

where N_o is the frequency constant (1660 m.Hz for the AT-cut), F is a positive force acting inwardly, d_r and t_r are the resonator diameter and thickness, respectively, and K_p is the coefficient of frequency shift at azimuthal angle p.

Figure 5 shows dK_p/dT for the AT-cut, which is the key to being able to optimize the temperature dependence of scale factor.

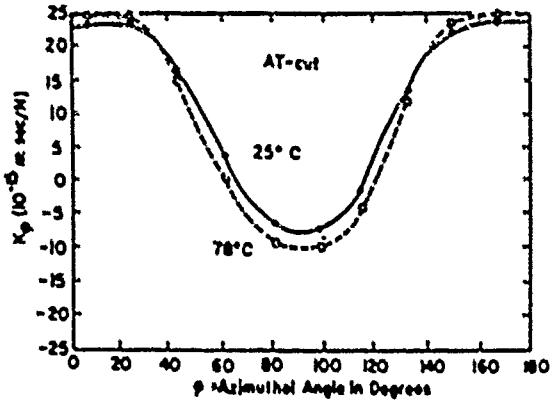


Figure 4: Force sensitivity coefficient K_p of the AT-cut vs. azimuthal angle for 25°C and 78°C.

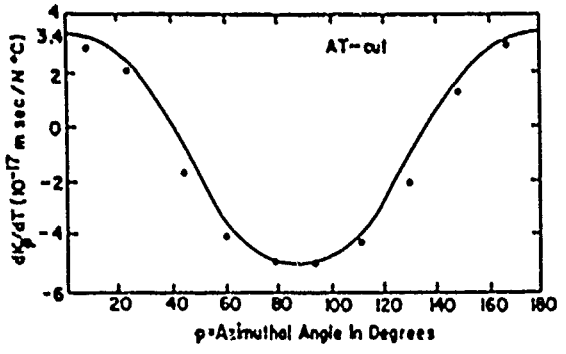


Figure 5: Temperature derivative of the force sensitivity coefficient K_p vs. azimuthal angle p for the AT-cut.

Our empirical model[9,10] assumes isotropic behavior of the quartz structure for obtaining an analytical solution of the initial stress problem. We first recognize that by superposition, the df/f_o for circular symmetry is given by integration of Fig. 4,

$$df/f_o = k_{oo} T_{oo}, \qquad (2)$$

where T_{oo} is radial stress (positive for tension) and,

$$k_{oo} = -(N_o/2)\int_0^{\pi} K_p \, dp. \qquad (3)$$

where p is azimuthal angle as measured from the x crystallographic axis. The minus sign in Eq. (3) converts from the positive force inwards convention adopted years ago to the conventional definition of stress (tension positive). For the AT-cut, $k_{oo} = -2.7\times10^{-11} m^2/N$.

We now use superposition to represent the initial stress as the sum of a uniform radial compression T_{oo} and two diametric force pairs, one (F_x) aligned along the flat normals, the other (F_z) aligned normal to F_x. Figure 6 shows this situation. T_{xx} and T_{zz} are the stresses at the disc center. We know that

$$T_{oo} = (T_{xx} + T_{zz})/2. \qquad (4)$$

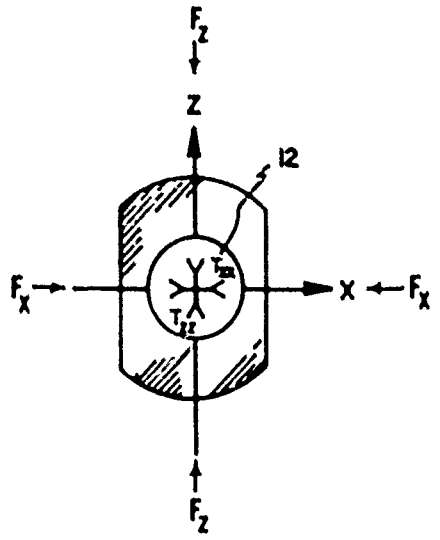


Figure 6: The two pairs of diametric force pairs used to simulate the non-circular symmetry of T_{xx} and T_{zz}.

We let the stresses T_{xx} and T_{zz} be represented by T_{oo} and incremental stresses dT_{xx} and dT_{zz}:

$$T_{xx} = T_{oo} + dT_{xx} \qquad (5)$$

$$T_{zz} = T_{oo} + dT_{zz}. \qquad (6)$$

These incremental stresses arise from the applied forces F_x and F_z. An applied force F causes a stress at the disc center of $-6F/(\pi t_r d_r)$ along the direction of force application and $2F/(\pi t_r d_r)$ stress at right angles. Thus,

$$(\pi d_r t_r) dT_{xx} = -6F_x + 2F_z \qquad (7)$$

and

$$(\pi d_r t_r) dT_{zz} = 2F_x - 6F_z. \qquad (8)$$

Inverting Eqs. 7 and 8, we have

$$F_x = -\pi d_r t_r (dT_{zz} + 3dT_{xx})/16 \qquad (9)$$

and

$$F_z = -\pi d_r t_r (dT_{xx} + 3dT_{zz})/16. \qquad (10)$$

Equations 4, 9, and 10 give us T_{oo}, F_x, and F_z if we know T_{xx} and T_{zz}. Also, in terms of the stress ratio

$$R = T_{xx}/T_{zz} \qquad (11)$$

we find that

$$F_x = -2\pi d_r t_r T_{oo}(R-1)/[16(R+1)] \qquad (12)$$

and

$$F_z = -F_x. \qquad (13)$$

We can determine T_{xx}/T_{zz} from finite element modeling or strain gauges on metal mockups. Given T_{xx}/T_{zz}, we know T_{oo}, F_x and F_z, so we can predict df/f_o.

$$df/f_o = T_{oo}\,(k_{oo} - 2\pi N_o(R-1)^* (K_p - K_{p+90^\circ})/[16(R+1)]), \qquad (14)$$

where K_{p+90° is the K_p value at p+90°

Abbreviated,

$$df/f_o = kT_{oo}. \qquad (15)$$

We can differentiate Eqs. 14-15 with respect to temperature if we assume T_{oo} and R are independent of temperature.

$$dk/dT = dk_{oo}/dk - 2\pi N_o(R-1)^* (dK_p/dT - dK_{p+90^\circ}/dT)/[16(R+1)] \qquad (16)$$

where we find dk/dT_{oo} from Fig. 5 by integration:

$$dk_{oo}/dT = -(N_o/2)\int_o^\pi (dK_p/dT)dp. \qquad (17)$$

For example, we find $(dk_{oo}/dT)/k_{oo}$ at room temperature to have a value of -1000ppm/°C from integration of Figure 5.

There is an amplification of the stresses due to the shell structure, so we find that for a pressure P outside the shell,

$$T_{oo} = -CP \qquad (18)$$

Where C=2.6 from finite element analysis of our structure and C=2.4 experimentally to best fit our data with the empirical model. Defining K for pressure scale factor and (dK/dT)/(K) for temperature sensitivity of scale factor, we find that

$$df/f_o = KP \qquad (19)$$

where

$$K = -C\,(k_{oo} - 2\pi N_o(R-1)^* (K_p - K_{p+90^\circ})/[16(R+1)]) \qquad (20)$$

and

$$(1/K)(dK/dT) = ((dk_{oo}/dT)/k_{oo}) - (2\pi N_o^* (R-1)(dK_p/dT - dK_{p+90^\circ}/dT)/[16k_{oo}(R+1)]) \qquad (21)$$

Experiment vs. Theory

Quartz devices were built of AT-cut orientation with different flat depths (different t'/t ratios)[11] The details of the structure are found elsewhere. The outside diameter was 0.720 inches. Flat to flat dimensions varied from 0.72 to 0.595 inches. Finite element analysis gave T_{xx}/T_{zz} values. Then, use of Eqs. 19-21 and Figs. 4 and 5 allow calculation of expected results. Actual results were found by testing of the quartz devices over pressure and temperature. The frequency, pressure, and temperature data were analyzed to extract experimental values for K and (1/K)(dK/dT). The results are shown in Fig. 7 for K and Fig. 8 for (1/K)(dK/dT). A value of 2.4 was used for C.

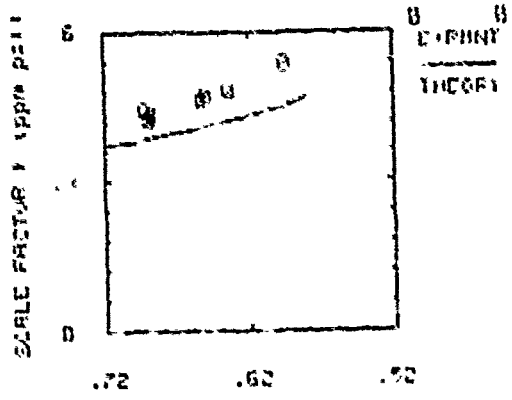


Figure 7: Comparison of empirical model and experimental results for the pressure coefficient of frequency (scale factor), K, vs. flat depth.

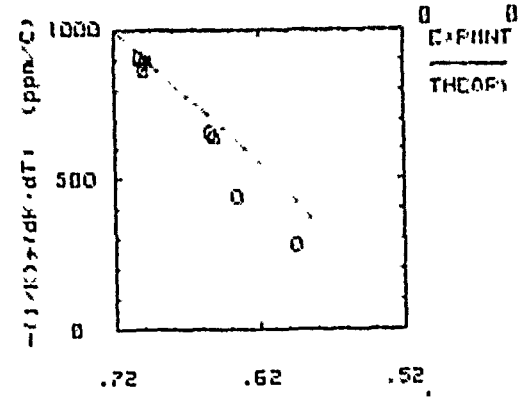


Figure 8: Comparison of empirical model and experimental results for the temperature derivative of the pressure coefficient of frequency, (1/K)(dK/dT), vs. flat depth.

The agreement is excellent. Our experimental data extrapolate to the empirical model result of -1000 ppm/°C for (1/K)(dK/dT). The horizontal axis in Figs. 7 and 8 is flat depth (flat-to-flat dimension) in Figs. 6 and 7. For reference, finite element analysis shows T_{xx}/T_{zz} of 1.47 and 1.20 for flat to flat dimensions of 0.595 and 0.652 inches, respectively. Inside diameter was 0.450 inches, so t'/t varied from 1.0 to 0.54. The T_{xx}/T_{zz} ratios did not vary as much as one might infer from Fig. 2 for the range of t'/t, but the disc thickness was larger relative to the shell wall thickness in the quartz devices compared to the early Al metal mockup.

Examination of Figs. 4 and 5 and Eqs. 20 and 21 shows that, since $k_{\infty} < 0$, flats aligned with the normal along p=0 (X crystallographic axis) enhance the K over the k_{∞} of a circular symmetry because both K_p and $K_{p+90°}$ contribute constructively. A similar examination for (1/K)(dK/dT) shows that both dK_p/dT and $(dK_{p+90°})/dT$ combine to reduce dk_{∞}/dT. Thus, some improvement in both transducer characteristics occurs for deeper flats, as seen in Figs. 7 and 8. This effect reverses if the flats are moved to p=90°, i.e., the transducer becomes less sensitive and the temperature dependence of scale factor gets worse than for a circular geometry.

Conclusions

The new empirical model has allowed the use of previously published experimental data on the diametric force-frequency effect to predict more complicated force geometries encountered in a new pressure transducer concept. This new transducer concept uses a non-circular shell integral with a circular resonator to create non-symmetric initial stress patterns in the resonator. The agreement between the results of the empirical model and experimental data is excellent. The overall results show that the lack of circular symmetry allows two new degrees of freedom for the transducer designer to use in optimizing transducer performance.

Acknowledgement

The Al mockup experiment was performed by Larry D. Clayton.

REFERENCES

[1] A.D. Ballato and R. Bechmann, "Effect of Initial Stress in Vibrating Quartz Plates," Proc. IRE, vol. 48, pp. 261-262, 1960.

[2] J.M. Ratajski, "Force-Frequency Coefficient of Singly Rotated Vibrating Quartz Crystals," IBM Journal, vol. 12, pp. 92-99, 1968.

[3] C.R. Dauwalter, "The Temperature Dependence of the Force Sensitivity of AT-Cut Quartz Crystals," Proc. 26th Annual Freq. Control Symp., pp. 108-112, 1972.

[4] A. Ballato, E.P. EerNisse, and T. Lukaszek, "The Force-Frequency Effect in Doubly Rotated Quartz Resonators," Proc. 31st Annual Freq. Control Symp., pp. 8-16, 1977.

[5] E.P. EerNisse, "Temperature Dependence of the Force-Frequency Effect for the AT-, FC-, SC- and Rotated X-cuts," Proc. 34th Annual Freq. Control Symp., pp. 426-430, 1980.

[6] H.E. Karrer and J. Leach, "A Quartz Resonator Pressure Transducer," IEEE Trans. of Industrial Electronics and Control Instr., vol. IECI-16, pp. 44-50, 1969.

[7] H.E. Karrer and J.G. Leach, "Quartz Resonator Pressure Transducer," U.S. Patent #3,561,832, February 9, 1971.

[8] A. Benjaminson and D.L. Hammond, "Piezoelectric Transducer and Method for Mounting Same," U.S. Patent #3,617,780, November 2, 1971.

[9] E.P. EerNisse, "Resonator Pressure Transducer," U.S. Patent #4,550,610, November 5, 1985.

[10] E.P. EerNisse, "AT-Cut Crystal Resonator Pressure Transducer," U.S. Patent #4,660,420, April 28, 1987.

[11] R.W. Ward and E.P. EerNisse, "A Reduced Hysteresis, Extended Range Quartz Pressure Transducer," This Proceedings.

A REDUCED HYSTERESIS, EXTENDED RANGE QUARTZ PRESSURE TRANSDUCER

Copyright 1987 by IEEE.
Presented at 41st Annual Symposium on Frequency Control, May 27-29, 1987, Philadelphia, PA.

Roger W. Ward
Errol P. EerNisse

Quartztronics, Inc.
1020 Atherton Drive
Building C
Salt Lake City, Utah 84123

SUMMARY

Experimental AT-cut pressure transducers, based upon the theory presented elsewhere in this Symposium ("Theoretical Modeling of Quartz Resonator Pressure Transducers"), have been fabricated and tested. These tests have been used to verify that the use of diametrically opposed flats to create a non-symmetric stress in the resonator of an AT-cut crystal can be used as two additional degrees of freedom (location with respect to crystallographic axes and depth of the flats) to reduce the temperature coefficient of scale factor and to increase the scale factor (ppm/psi) of the pressure transducer[1,2]. The transducer has been made as small as reasonably possible. This reduces cost and improves the thermal transient response of the transducer. Tests to 16,000 psi and 175°C show deviations of less than 1 psi, including hysteresis, for steady-state conditions. Thermal transients of 3°C result in 2-3 psi errors which dissipate in 3 minutes. Tests of units having increased flat depth to 20,000 psi and 200°C resulted in the twinning of some of the resonators, probably due to ferroblelastic effects. Units having a compromise flat depth survive these tests without twinning and show no hysteresis. A torsional-mode tuning fork temperature sensor and an SC-cut reference crystal are mounted adjacent to the pressure transducer, enabling real-time temperature compensation and the mix-down of the temperature and pressure frequency to a suitable frequency for digital transmission and/or storage; the accuracy of the system is better than 5 psi during any combination of pressure or temperature transients likely to be encountered in the field. Recovery time to steady-state conditions is typically less than 3 minutes.

BACKGROUND

The force-frequency effect on an AT-cut has been known for many years [3]. This effect has been used for temperature compensating AT-cut crystals over wide temperature ranges [4], and for pressure measurement [5,6]. However, these applications all used the crystal to measure a force. Since pressure is a force per unit area, it was necessary to provide a conversion means to allow the crystal to measure a pressure (i.e., a diaphragm, a lever mechanism, etc.). All of these applications were troubled with frequency drift caused by creep in the materials attached to the quartz.

In 1967, Benjaminson and Hammond [7] at Hewlett Packard filed a U.S. Patent on an all quartz pressure transducer consisting of a quartz resonator surrounded by a hermetically sealed cylindrical housing which permits the resonator to be squeezed uniformly around its circumference. This allows the crystal to respond directly to the pressure applied to its periphery, without going through a pressure-to-force conversion, with its inherent error and hysteresis producing mechanisms, Figure 1. The only joints in the HP design occur away from the force sensitive resonator, and are positioned at the zero shear point along the length of the cylinder, and are located such that any creep in the joint will not cause frequency shifts in the resonator.



Figure 1. Hewlett Packard's BT-cut pressure transducer, circa 1970[8].

The Hewlett Packard quartz pressure transducer has been reported on several times at this symposium [8-10]. Kaitz has reported on hysteresis and resistance problems with the HP BT-cut design which they associate with high pressures and temperatures, especially with certain types of quartz source material. In addition, the large size of the HP design makes it expensive (because of the raw material cost). But more importantly, the large size causes a rather large thermal transient restabilization phenomenon, which is aggravated by the remote location of their reference crystal.

To overcome some of these problems, it was proposed to make the quartz transducer as small as practical (lower thermal mass, faster thermal stabilization, lower raw material cost). The theoretical development of this transducer is detailed in the previous paper in this Proceeding [11]. This paper shall concentrate on the experimental development of the transducer.

DESIGN

The first design attempt of this small pressure transducer used a 0.375 inch resonator diameter, which proved to be too small to permit the generation of the bi-convex resonator, and the thin, high frequency resonator fractured where it joined the bore-wall during the generation of the bore.

The second (and final) design uses a 0.450 inch diameter resonator which is 0.055 inch thick (3.610 Mhz, 3rd overtone). The final frequency was determined by the lack of activity dips over the pressure and temperature test range for the final transducer design (0-20,000 psi, 0-200°C). This design is shown in Figure 2. The sensor has a sensitivity of 0.5×10^{-7} ppm/psi (1.8 Hz/psi).



Figure 2. An At-cut quartz pressure transducer with X-axis flats.

FREQUENCY-TEMPERATURE

The frequency-temperature (f-T) response of the transducer is a function of the applied pressure -- Figure 3. The apparent angle of the AT-cut increases approximately 3 arc-minutes per 1000 psi of applied pressure. The optimum AT-angle was selected to provide a small f-T slope at elevated temperatures and pressures.

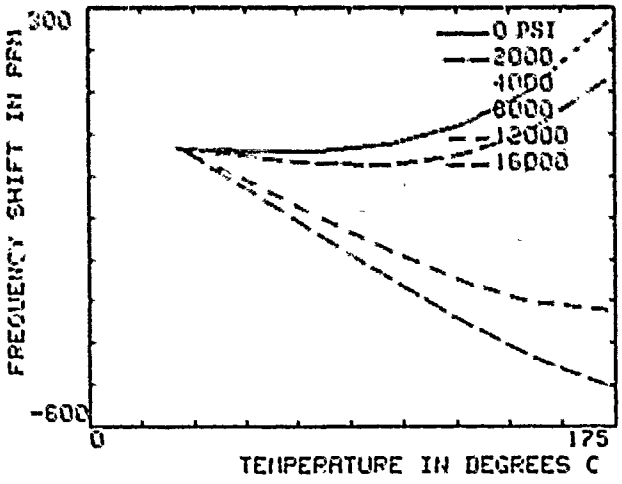


Figure 3. Frequency vs. temperature curves for an AT-cut pressure transducer with X-axis flats at various pressures. Apparent angle shift: 3 arc-min/1000 psi.

The converse effect is that the slope of the frequency-pressure response is a function of temperature. Put another way, the temperature coefficient of the scale factor is non-zero. This is shown in Figure 4, as evidenced by the spread of the frequency vs. pressure curves at different temperatures. The temperature sensitivity of scale factor, as well as the pressure dependence of the f-T response, can be made to be zero by adjusting the depths of the flats on the transducer, as described by EerNisse [11]. However, the resulting transducer has rather thin walls, resulting in a reduced full scale pressure capability.

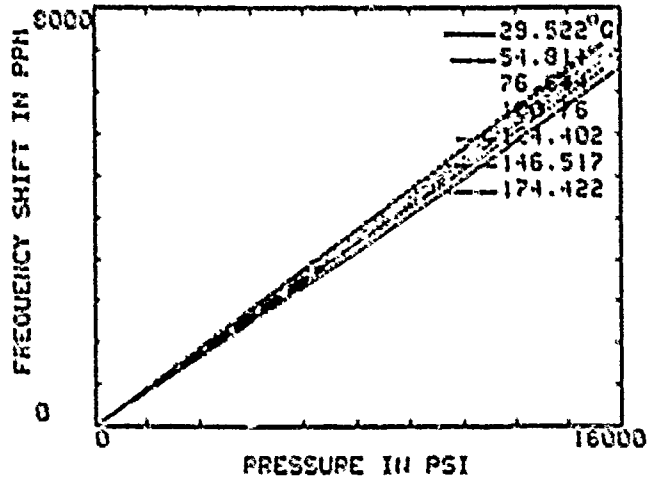


Figure 4. Temperature coefficient of scale factor for an AT-cut pressure transducer with X-axis flats.

TEMPERATURE COMPENSATION

Due to the temperature behavior of the pressure transducer, it is obvious that some form of temperature compensation is required if one is to achieve a high precision pressure transducer over an extended temperature range. The worst case slope of Figure 3 is equivalent to about 8 psi/°C. Hence, to have an overall accuracy of better than ± 1 psi over the "specified" 0-16,000 psi and 0-175°C range, a temperature correction of better than 0.1°C is required.

Due to the mass and physical size of the pressure sensing crystal, there exists a thermal time lag effect upon the output frequency. To help reduce errors from the frequency excursion due to thermal transients, a torsional-mode quartz tuning fork (TTF) temperature sensor [12,14] is mounted in a pressure-proof quartz package designed to simulate the thermal response of the pressure crystal. (Note: the stress created by temperature gradients along the pressure crystal's length and radius cannot be compensated by this technique.) The TTF assembly is mounted adjacent to the pressure crystal in an oil-filled pressure housing. It has a nominal frequency of 180 kHz and a nominal temperature sensitivity of 50 ppm/°C. In Figure 5, a 3000 psig to 0 psig rapid decompression is seen to create a 0.6°C adiabatic cooli.g effect at the TTF, which

causes a -1.4 psi pressure error in the pressure crystal. This same decompression causes a 2-3°C abrupt change in the temperature of the surrounding oil when measured with a low-mass temperature sensor.

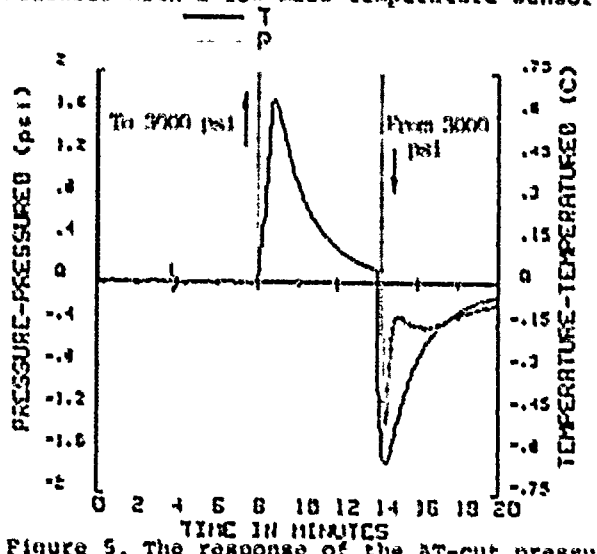


Figure 5. The response of the AT-cut pressure transducer and the TTF temperature sensor to a sudden 3000 psi pressure step.

With this active temperature compensation, the pressure crystal's output is within 1 psi of its final value within 45 seconds after the thermal transient occured. The temperature sensor took 3.8 minutes to recover to within 0.15°C of its final temperature.

During rapid (1.5°C/minute) thermal ramps, the pressure crystal shows a worst-case error of -4 psi, which dissipates rapidly as the temperature stabilizes, Figure 6. This behavior indicates a slight difference in the thermal conduction paths for the temperature compensation vs. the pressure sensor resonator. (The pressure spikes in Figure 6 were caused by the manual adjustments required to maintain the dead-weight-tester's weights in their "float" range as the dead-weight-tester's oil expanded due to the temperature rise.)

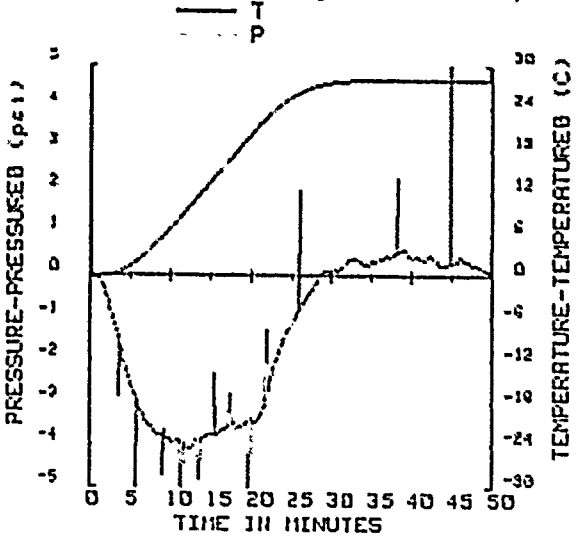
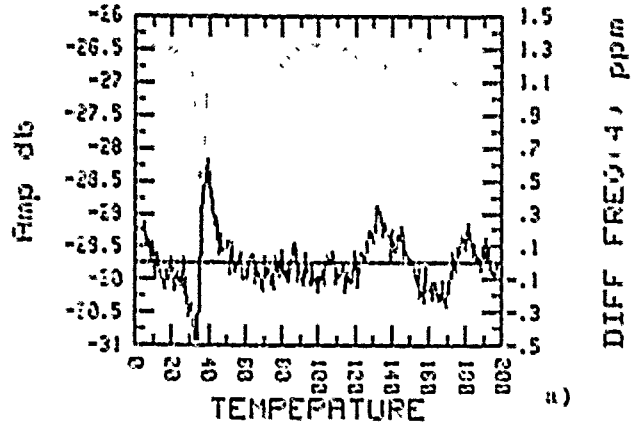


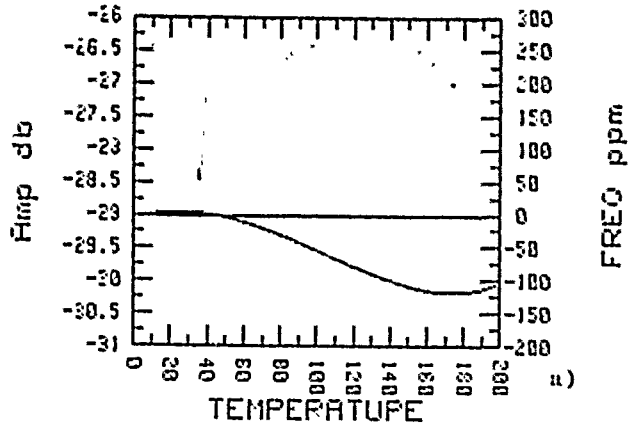
Figure 6. 25°C temperature ramp at 10,000 psi.

REFERENCE CRYSTAL

One requirement for an oil and gas service down-hole frequency-based transducer is that its output frequency must be lower than 50 kHz to transmit the frequency up a 20,000 ft cable. This was accomplished by including a 7.200 MHz fundamental-mode SC-cut reference crystal in a pressure-proof quartz package adjacent to the pressure crystal. Upon dividing the reference frequency by two, and mixing with the pressure frequency, a 10 kHz output frequency results, with a nominal 1.8 Hz/psi sensitivity. At 20,000 psi, a 46 kHz output is obtained.

Several researchers have commented that the SC-cut is free from activity dips. This is not universally true, as seen in Figure 7. An activity dip in a 7.2 MHz fundamental, plano-convex SC, having a ±0.5 ppm frequency deviation is shown, with a corresponding 2dB activity decrease. This data was taken in a lag-block in an air oven. Dry ice was used to obtain temperatures below room temperature. Also shown is an SC with no dips (greater than perhaps 1x10⁷). In these plots, the general shape of the residuals is due to temperature scan rate variations, since each crystal in a run will exhibit similar general perturbations. The noise is due to frequency measurement errors associated with the HP3585 spectrum analyzer used to make the frequency and amplitude measurements.





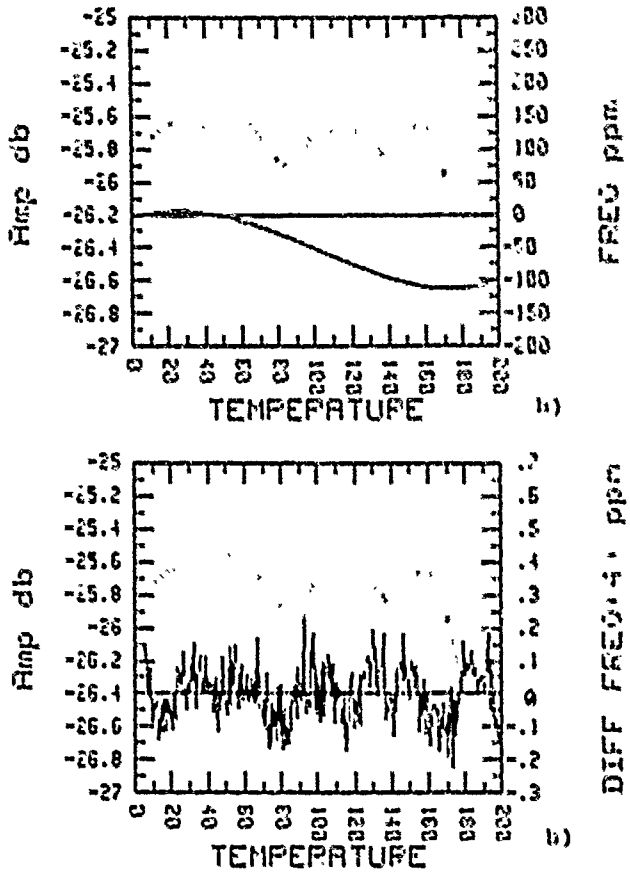
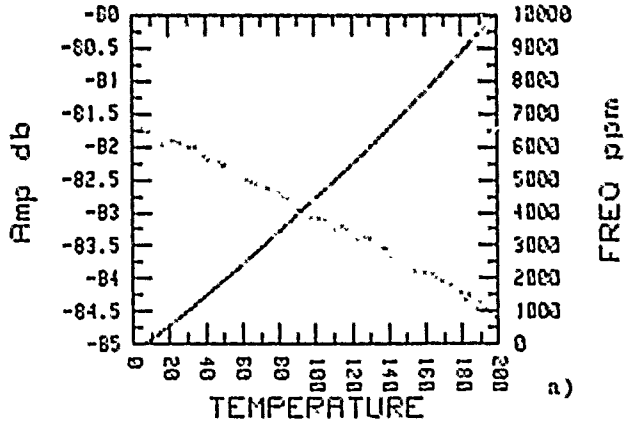


Figure 7. Two SC-cut reference crystals: a) with activity dip; b) no activity dips.

RESULTS

Figure 8 shows a temperature scan on one of the TTF temperature sensors from 0-200°C showing no activity dips [12,13]. The tail on the residual curve below room temperature is due to time lag effects during cool-down effects in the lag-block. Correlation between two sensors in a common lag-block is better than 5 ppm, or 0.1°C.



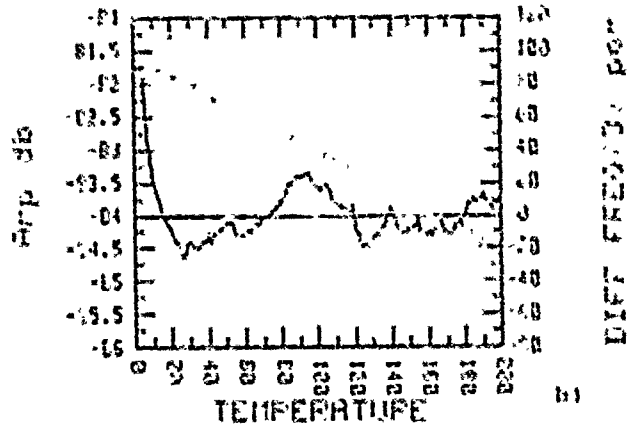


Figure 8. A TTF temperature sensor having no activity dips.

The pressure-temperature performance of the three crystal combination [14] is shown in Figures 9-11. Figure 9 illustrates that the TTF frequency is independent of pressure. Figure 10 shows a temperature accuracy of better than ± 0.03°C from ambient to 175°C. Figure 11 shows the calculated pressure residual versus pressure at seven temperatures from 20°C to 174°C. The residual is less than ± 0.8 psi for any combination of pressure and temperature.

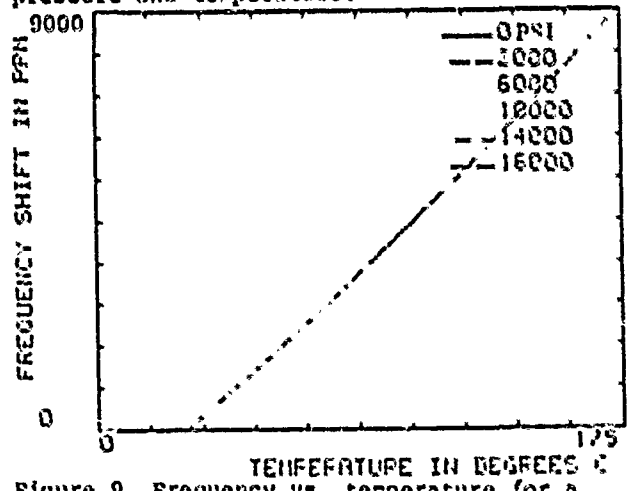


Figure 9. Frequency vs. temperature for a packaged TTF temperature sensor with no pressure sensitivity.

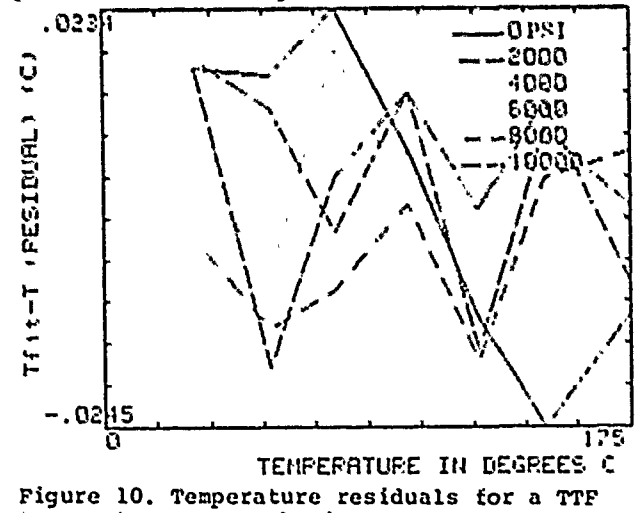


Figure 10. Temperature residuals for a TTF temperature sensor having no pressure sensitivity or activity dips.

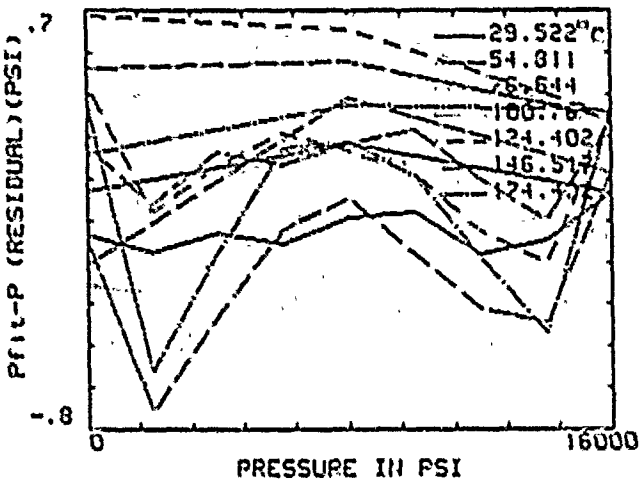


Figure 11. Pressure residuals for the AT-cut pressure transducer with TTF temperature compensation.

The pressure crystal's raw data is curve-fit using a fourth-order fit on frequency vs. temperature and a third-order fit on pressure; a fourth-order fit is used on the TTF's frequency vs. temperature. In total, 20 coefficients are used.

Figure 12 demonstrates the resolution and noise in the pressure transducer system. Three successive 0.1 psi pressure steps are shown, with a noise floor of less than 0.01 psi.

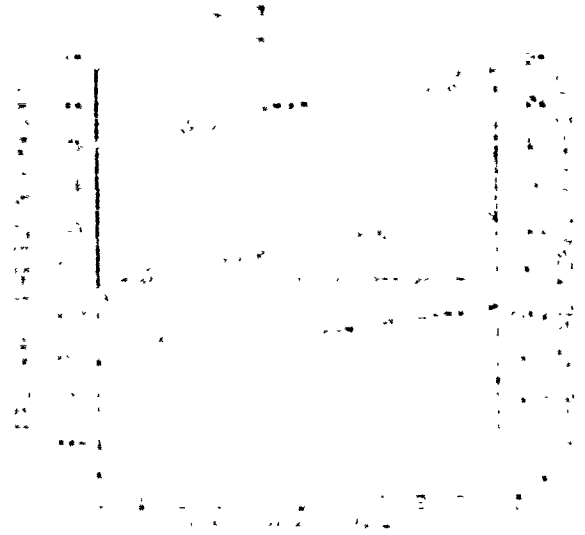


Figure 12. 0.1 psi pressure steps. Noise floor is below 0.01 psi.

Pressure is generated with a DHI dead-weight tester which is NBS traceable to ±0.01% of reading. Temperature is measured using a platinum resistance thermometer and a DVM.

FERROBIELASTIC TWINNING

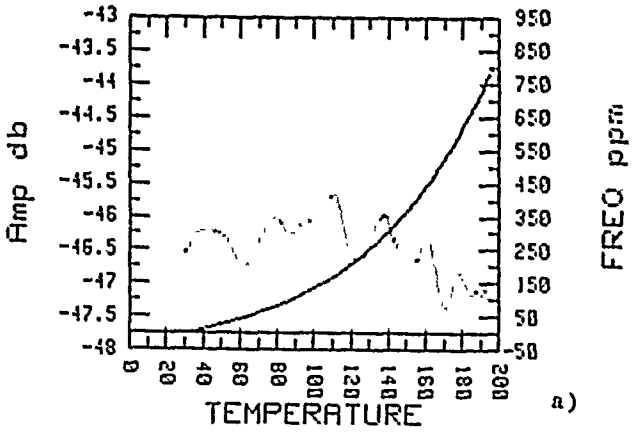
During early experimental testing to determine the effects of flat depth and orientation on the temperature sensitivity of scale factor [11], it was found that as the flat depth increased, the temperature sensitivity decreased. The minimum distance across the flats tested was 0.590 in. In two separate tests of units with 0.590 inch flats, both units quit oscillating at 20,000 psi as the temperature was ramped from 175°C to 200°C. The units were disassembled and etched, revealing twinning in the resonator of each unit. One unit twinned along the X-diameter of the resonator. The second had a small, circular twin region near the center of the resonator. Inspection under a microscope revealed a microscopic inclusion (possibly acmite in the cultured quartz) in the center of the twin region. Units with shallower flat depths have not shown this susceptibility to twinning. Anderson, et. al. [15] demonstrated that as the temperature of a quartz sample increased, the stress level required to twin the sample decreased; no external stress is required at 573°C (the Curie temperature for natural quartz), while a stress level around 6×10^{4} psi is required at room temperature. This phenomenon is called ferrobielastic twinning, and probably accounts for the twinning observed on the two deep-flat units, i.e., at 20,000 psi the resonator has about 60,000 psi stress along its X-axis.

HYSTERESIS

Early units were fabricated from Sawyer pure-Z optical grade cultured quartz. Results on those units showed a worst-case hysteresis of ±3 psi to 20,000 psi and 200°C.

Subsequent units, using improved manufacturing techniques, were fabricated from natural quartz. Any hysteresis present in these natural quartz pressure transducers has been unmeasurable up to the design limits of 20,000 psi and 200°C. Data on over 20 natural quartz units to 16,000 psi and 175°C shows consistent performance, for both increasing and decreasing pressures, to better than ± 1.2 psi. In addition resistance changes are typically around ± 10% to 200°C, Fig. 13.

Several units have been fabricated, using current manufacturing techniques, from Sawyer cultured quartz. These units are presently being tested, using current testing and data reduction techniques, to see if similarly fabricated cultured units exhibit different characteristics from the current natural quartz units. Early results are inconclusive, but cultured units (pure-Z) definitely show more hysteresis than similar natural units.



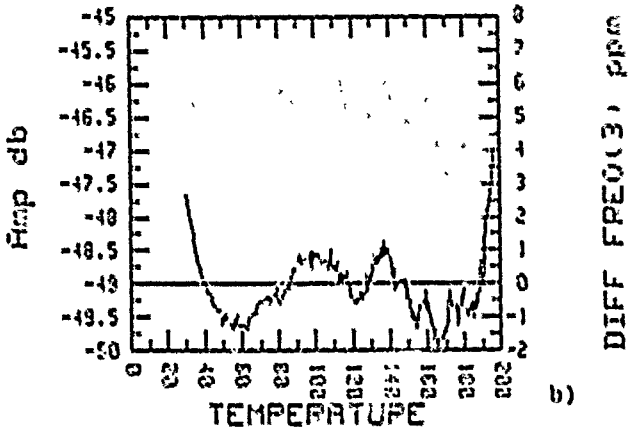


Figure 13. Relative resistance changes of ± 10% on a natural quartz AT-cut pressure transducer to 200°C at 0 psi.

The earlier results on cultured material (no swept-natural material has been studied) appear to not confirm Hewlett Packard's hysteresis results [9,10] Two explanations are possible: 1) The AT-cut is less susceptible to material induced hysteretic effects than is the BT-cut; or 2) there exists slip or glide planes in the BT-cut which cause larger hysteresis effects than in the AT-cut.

Further studies on the differences between these two high performance transducers are planned. The results could lead to even higher performance transducers in the future, through a better understanding of the raw material--crystalline quartz, both natural and cultured.

CONCLUSIONS

A 0.01%, 20,000 psi, 200°C quartz pressure transducer has been built, based upon a theoretical model utilizing x-axis flats on an AT-cut cylindrical housing to provide two additional degrees of freedom to simultaneously minimize the transducer's temperature coefficient of scale factor and maximize the magnitude of its scale factor. The transducer, due to its small size and active temperature compensation, exhibits small thermal transient effects that dissipate rapidly. Hysteresis is unmeasurable. This transducer system will find use in down-hole oil and gas service and logging operations.

REFERENCES

[1] E.P. EerNisse, "Resonator Pressure Transducer," U.S. Patent #4,550,610, Nov. 5, 1985.

[2] E.P. EerNisse, "AT-cut Crystal Resonator Pressure Transducer," U.S. Patent #4,660,420, April 28, 1987.

[3] V.E. Bottom, "An Anomalous Thermal Effect in Quartz Oscillator-Plates," American Mineralogist, vol. 32, pp. 137-140, 1947.

[4] E.A. Gerber and M.H. Miles, "Reduction of the Frequency-Temperature Shift of Piezoelectric Resonators by Mechanical Stress," Proc. IRE, pp. 1650-1654, 1961.

[5] J.P. Corbett, "Oscillating Crystal Force Transducer System," U.S. Patent #3,541,849, Nov. 24, 1970.

[6] E. Karrer and R. Ward, "A Low Range Quartz Pressure Transducer," ISA Trans., vol. 16, pp. 90-98, 1977.

[7] A. Benjaminson and D.L. Hammond, "Piezoelectric Transducer and Method for Mounting Same," U.S. Patent #3,617,780, November 2, 1971.

[8] J.G. Leach, "5 MHz BT-Cut Resonators", Proc. of the 24th Annual Freq. Control Symp., pp. 117-125, 1970.

[9] G.S. Kaitz, "Extended Pressure and Temperature Operation of BT-Cut Pressure Transducers," Proc. of the 38th Annual Freq. Control Symp., pp. 245-250, 1984.

[10] 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 Application," Proc. of the 39th Annual Freq. Control Symp., pp. 223-233, 1985.

[11] E.P. EerNisse, "Theoretical Modeling of Quartz Resonator Pressure Transducers", Proc. of the 41st Annual Freq. Control Symp., This Proceedings, 1987.

[12] E.P. EerNisse and R. B. Wiggins, "A Resonator Temperature Transducer With No Activity Dips," Proc. 40th Annual Freq. Control Symp., pp. 216-223, 1986.

[13] E.P. EerNisse and R.B. Wiggins, "Resonator Temperature Transducer," U.S. Patent #4,592,663, Jun. 3, 1986.

[14] E.P. EerNisse and R.W. Ward, U.S. Patent Pending.

[15] T.L. Anderson, R.E. Newnham, and L.E. Cross, "Coercive Stress for Ferrobielastic Twinning in Quartz," Proc. 31st Annual Freq. Control Symp., pp. 171-177, 1971.

STUDY OF LIQUIDS IN SHEAR USING A QUARTZ RESONATOR

K. Keiji Kanazawa and C.E. Reed
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099

Several recent developments have driven our efforts to extend the pioneering work of Mason[1] and use the simple quartz resonator, directly interfaced to the liquid, as a tool for studying the fluid behavior at large cylic shear rates. These developments include the quantitative description of the compound resonator formed between the resonator and contacting liquid, the availability of instrumentation for accurate impedance analysis and the resurgence of interest in the rheological behaviors of polymeric liquids[2]. With this technique we have shown that simple liquids behave as purely viscous fluids, at least up to frequencies of 5 MHz. Initial studies on polyperfluorinated fluids suggest an elastic component which increases with increasing molecular weight.

When a viscoelastic material contacts one face of a quartz crystal operating in the thickness shear mode, a compound resonator is formed and the mechanical resonant frequency is altered. This change can be described in terms of the material's thickness t, density ρ, shear modulus μ and viscosity η. In the limiting case of a classical liquid for which $\mu = 0$, the shear wave penetrates only a short distance into the fluid. In spite of this limited penetration, the quality factor (Q) of the compound resonator is markedly decreased due to viscous losses in the liquid. This loss in Q is reflected as a change in the electrical equivalent circuit of the resonator. This equivalent circuit consists of two parallel arms: the "static" arm representing the dielectric capacitance C_p of the quartz and the "motional" arm consisting of a series L, C_s, R circuit representing the motional behavior of the resonator. The inductance represents the energy stored in inertial or kinetic form, the series capacitance represents energy stored in elastic form and of course the resistance reflect the viscous damping. One can show that the quality factor for the shear wave in the liquid is unity. Coupling this with the physical arguments above, an expression for the resistance ΔR upon interfacing to a liquid is given by:

$$\Delta R = \frac{1}{\pi C_s f_o^{1/2}} \frac{\sqrt{\rho\eta}}{\sqrt{\pi \rho_Q \mu_Q}} \qquad [1]$$

where ρ_Q and μ_Q are the density and effective shear modulus for the quartz resonator.

This relationship was tested for a number of fluids by studying the resonance behavior using an impedance analyzer and is seen to be satisfied for simple liquids. For polymeric liquids, the observed resistances were found to be much less than that given by the equation, and in addition the frequency changes were also found to be less than predicted by the mechanical model. Formally, one could account for this by assigning a relaxation time τ to the to the fluid such that for time scales well in excess of τ the fluid behaves viscously and for time scales much less than that time, the fluid behaves elastically, as first described by Maxwell. Fitting τ to the frequency changes for a series of perfluoropolyethers showed that the fitted τ increases with molecular weight.

1. W.P. Mason, W.O. Baker, H.J. McSkimin and J.H. Heiss, Phys.Rev. 75, 936 (1949)

2. Daniel D. Joseph, J. Non-Newt. Fluid Mech. 19, 237 (1986)

CH2427-3/87/0000-350 $1.00 C 1987 IEEE

41st Annual Frequency Control Symposium - 1987

A Novel Technique for Trimming the Frequency of a Sealed Surface Acoustic Wave Resonator

J. A. Greer and T. E. Parker
Raytheon Research Division
131 Spring St.
Lexington, MA 02173

M. Rothschild and D. J. Ehrlich
Lincoln Laboratory, Massachusetts Institute of Technology
244 Wood Street
Lexington, MA 02173-0073

ABSTRACT: An important requirement in many applications is the need to accurately set a SAW oscillator's frequency to a predetermined value. In some cases the oscillator's frequency may have to be set to within ±1 PPM of a specified frequency which usually requires considerable time in external circuit adjustment. Recently, a novel post-seal frequency trimming technique has been developed for use with SAW resonators sealed in an all quartz package (AQP). By taking advantage of the transparent nature of the AQP, we have developed a post-seal frequency trimming technique which utilizes a laser to transfer a thin dielectric film from the inner surface of the package down onto the active region of the SAW device. The oscillator's frequency can be monitored in real time, and by focusing the laser beam down to a width of a few microns the trimming resolution can be less than ±1 PPM for devices operating at 402 MHz. Experimental results indicate that trimming down in frequency as much as -35 PPM is easily obtainable with minimal changes in the resonator phase noise, insertion loss, and unloaded Q values. Preliminary results on aging of laser trimmed devices also look encouraging.

INTRODUCTION: An important requirement in many applications is the need to accurately set a SAW oscillator's frequency to a predetermined value. In some cases the oscillator's frequency may have to be set to within ±1 part per million (PPM) of a specified frequency. If this specified frequency falls within the 1 dB bandwidth of the SAW resonator, the frequency adjustment can usually be accomplished by using a simple adjustable phase shift circuit to compensate for the SAW frequency error as well as phase shift variations in the other oscillator components. A number of frequency trimming techniques have been developed to accurately set the SAW resonator frequency prior to sealing. The technique which has received the most attention is called reactive ion etching[1,2] (RIE). RIE can be used to adjust the resonator frequency with an accuracy of ±1 PPM prior to sealing the device with no degradation in device performance. Unfortunately, all hermetic sealing techniques for SAW resonators which are compatible with long-term frequency stability induce frequency shifts which are variable in nature. This variability in the frequency shift produced by sealing a device may cause the oscillator's frequency to lie outside the 1 dB bandwidth of the resonator. Also, some applications, such as voltage tunable oscillators, may require that the resonant frequency of the SAW device be set within a few PPM of a specified center frequency in order to maximize the available tuning range. Recently, an all quartz package[3] (AQP) was developed which provides a hermetic seal and is compatible with long-term frequency stability. The AQP is particularly attractive for hybrid circuit oscillators because of its small size. Another advantage of this hermetic packaging technique is that it has reduced both the magnitude and spread in sealing-induced frequency shifts. Data from 53 AQP SAW resonator devices operating at 350 MHz indicate that the sealing-induced frequency shift has an average of +54 PPM with a standard deviation of only ±15 PPM.

CH2427-3/87/0000-351 $1.00 C 1987 IEEE

While the AQP is superior to the TO-8 package in terms of frequency settability, there are still many applications which require a sealed resonator's frequency to be more accurately set. Up to the present time, frequency adjustment of a sealed resona could only be accomplished by the time consuming and expensive process of pulling the resonant frequency with external circuit components.[4]

By taking advantage of the transparent nature of the AQP we have developed a pos.-seal frequency trimming technique which utilizes pulsed laser energy to transfer a thin film from the inner surface of the quartz cover down onto the active region of the SAW device. Results indicate that trimming down in frequency as much as -35 PPM is easily obtainable with minimal change in oscillator flicker noise and turnover temperature, as well as device insertion loss and unloaded Q value. Initial results on the long-term frequency stability of laser trimmed devices are encouraging.

MATERIALS AND METHODS: Two-port SAW resonator devices operating between 402 MHz and 950 MHz were fabricated in the center of rotated Y-cut quartz substrates using standard photolithographic techniques. Small trim pads consisting of a thin aluminum film sandwiched between two layers of aluminum oxide were e-beam evaporated onto the center of the matching quartz covers of each AQP as shown in Figure 1. The location of this pad was such that it shadowed the region between the interdigital transducers (IDT's) when the devices were hermetically sealed under high vacuum (2×10^{-7} Torr) with a glass frit as shown in Figure 2. Sealed SAW devices were characterized for resonant frequency, insertion loss, and loaded and unloaded Q. Also, open- and closed-loop flicker noise levels as well as turnover temperatures were obtained for some of the devices prior to laser trimming. The devices were then mounted on small test fixtures and placed into one of four optical setups incorporating either an excimer or Nd:YAG laser and appropriate optics. The optical setups were used to focus the incident laser radiation down through the polished quartz cover and onto the trim pad. The thin Al layer sandwiched between the two oxide layers readily absorbs the laser radiation, converting it to heat which is then conducted to the Al_2O_3 films. Depending on the laser type and incident power density, the oxide film can be either evaporated or ablated off the quartz cover and down onto the acoustically active region of the SAW device. The redeposited film then changes the local acoustic wave velocity and thereby alters the device's resonant frequency.

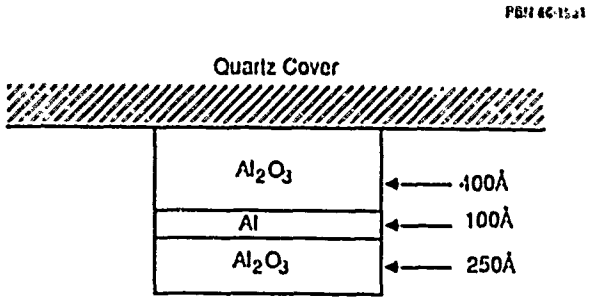


FIGURE 1. Schematic view of the laser trim pad located on the inner surface of the AQP cover.

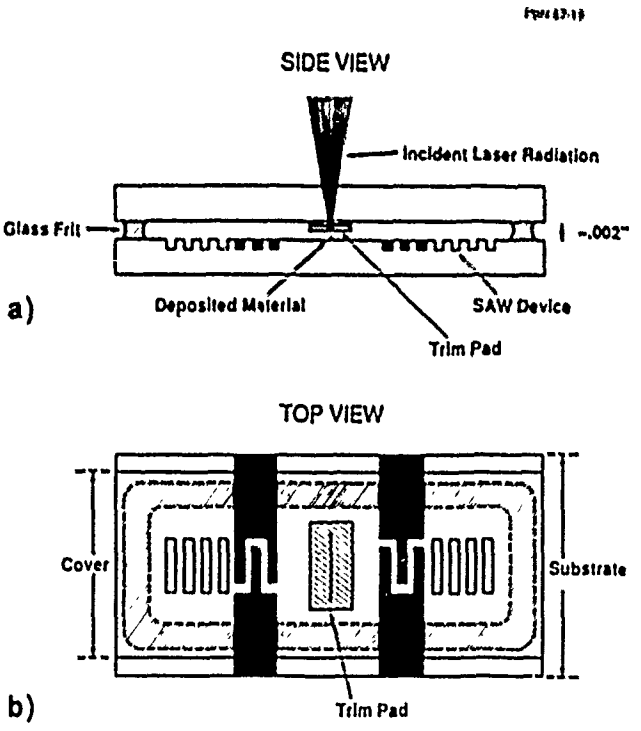


FIGURE 2. a) Side view of an AQP depicting the location of the laser trim pad. b) Top view of the AQP showing a line drawn by a laser across the trim pad running the length of the acoustic aperture.

As noted, two laser types and four different optics systems were utilized for post-seal frequency trimming of the AQP SAW devices. Two excimer laser setups, operating at 193 nm (ArF), were used to trim devices. The first was an experimental system designed for pattern projection in microlithographic applications in a step-and-repeat configuration. It consisted of a 10x objective and a computer controlled X-Y translation stage with 0.25 μm accuracy. It provided about 7 J/cm^2 per pulse at the device with a 15 nsec pulse duration. Typically this system operated at 1 pulse per second (PPS) and a line 7 μm wide by 450 μm long was projected onto the trim pad. A complete line could be drawn across the acoustic aperture (1 mm wide at 402 MHz) by simply moving the X-Y stage in between laser pulses. The second excimer system consisted of a commercially available laser/microscope setup. This system could deliver up to 10 J/cm^2 per pulse at the device trim pad. It was also run at 1 PPS except where noted below. Lines with dimensions of 1 μm by 135 μm were easily projected with this system. Again, a line could be drawn across the acoustic aperture by manually moving the microscope stage.

The Nd:YAG laser systems (commercially available laser trim systems) operated with quite different parameters. They provided an output at 1064 nm at a repetition rate of 1000 PPS with either a 70 ns or 200 ns pulse duration. Fluences up to 20 J/cm^2 per pulse could be produced with circular spot sizes ranging from 12 μm to 80 μm. The devices were scrolled in front of the laser beam at a rate of 5 mm/sec. Thus, a line running the length of the IDT's could be drawn as depicted in Figure 2(b). With either laser system a number of lines could be drawn across the acoustic aperture to obtain the desired frequency shift.

During the laser trimming process the device was connected to an oscillator circuit whose frequency was monitored with a frequency counter. Data was then recorded for each oxide line deposited on the SAW devices for all of the laser systems used.

After post-seal laser trimming was completed, the SAW devices were characterized on a network analyzer for center frequency, minimum insertion loss, as well as loaded and unloaded Q values. Also, open- and closed-loop noise measurements as well as turnover temperature measurements were obtained on several of the devices after laser trimming. Three of the devices operating at 402 MHz were then placed into aging chambers to study the effect of laser trimming on the long-term frequency stability of the SAW oscillator.

For the purpose of comparison, two other sets of SAW resonators were fabricated with operating frequencies at 402 MHz. Two devices were fabricated to serve as aging controls, and two devices were fabricated with thin pads of Al_2O_3 deposited by electron beam evaporation onto a fraction of the area between the two IDT's of the SAW device before the device was sealed. These pads ran the length of and were parallel to the transducer fingers. The shifts in frequency due to the deposition of the oxide pads were obtained as well as their effect on the long-term frequency stability of the SAW devices.

In order to study the laser deposited film morphologies several AQP's were fabricated without a SAW device but with a large area trim pad. These AQP's were cleaved open after laser irradiation to inspect the quality of the deposited films produced by altering laser parameters.

RESULTS: Results of post-seal frequency shifts expressed in PPM for devices trimmed with either a Nd:YAG or excimer laser are shown in Table 1. Also shown in the table are the laser fluence, percent area of the space between the IDT's covered by the laser-deposited oxide material, change in insertion loss, per cent change in the unloaded Q value, and change in the closed-loop phase noise at 10 Hz off the carrier. The average 402 MHz device had a closed-loop flicker noise level of about -76 dBc/Hz at 10 Hz off the carrier, an insertion loss of about 7 dB, and an unloaded Q value of 23,000.

The effects of e-beam evaporating Al_2O_3 pads between the IDT's on the unsealed SAW resonators are shown in Table 2. The width of the deposited film is expressed as a percentage of the length between the two

TABLE 1

POST-SEAL AQP LASER TRIMMING RESULTS

Device ID Frequency	Laser Type	Fluence (J/cm^2)	% Area	ΔF/F (PPM)	ΔIL (dB)	ΔQ_u %	Δ∠(10Hz) (dBc/Hz)
Q2281A 402 MHz	1	7.0	2.2	-35	0.20	-1	-0.2
Q2281B 402 MHz	1	7.0	1.8	-30	0.25	-2	0.5
E2054 945 MHz	2	8.0	0.9	+40	0.00	-7	N.A.
L-4 402 MHz	3	2.4	19.2	- 5	0.78	-16	1.2
A3-2 402 MHz	3	4.9	48.0	-30	0.36	-8	1.5
A4-2 402 MHz	3	5.8	34.6	-42	1.31	-18	0.1
L-3 402 MHz	3	12.7	7.7	-36	0.52	-4	N.A.
I-6 810 MHz	4	2.3	4.1	-67	0.73	+5	-0.2

N.A.= not available
1 = Laboratory excimer projection stepper
2 = Commercial excimer system
3 = Nd:YAG 200 ns pulse length
4 = Nd:YAG 70 ns pulse length

IDT's and the obtained frequency shift is expressed in PPM. Room temperature aging data for three laser-trimmed, two e-beam evaporated, and two control devices all operating at 402 MHz are plotted as normalized frequency shift versus time in Figure 3. Typical laser-deposited film morphologies for both laser types are shown in Figure 4.

DISCUSSION: It is of interest to note that when Al_2O_3 is electron beam evaporated onto an unsealed SAW resonator, the frequency will increase[5] as noted in Table 2. This indicates that the acoustic velocity of the

TABLE 2

PRE-SEAL Al_2O_3 PAD RESULTS

DEVICE ID	PAD WIDTH (%)	PAD THICKNESS (Å)	ΔF/F (PPM)	ΔIL (dB)
Q2284	44	115	+351	-0.85
Q2287	16	270	+164	-0.45

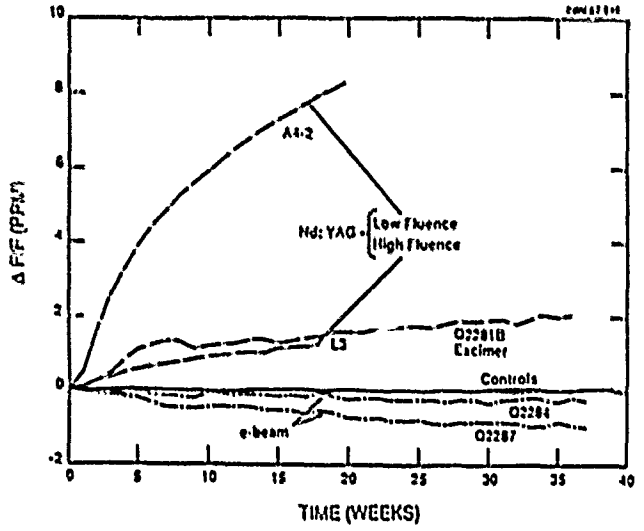


FIGURE 3. Room temperature aging data for several AQP resonator oscillators operating at 402 MHz.

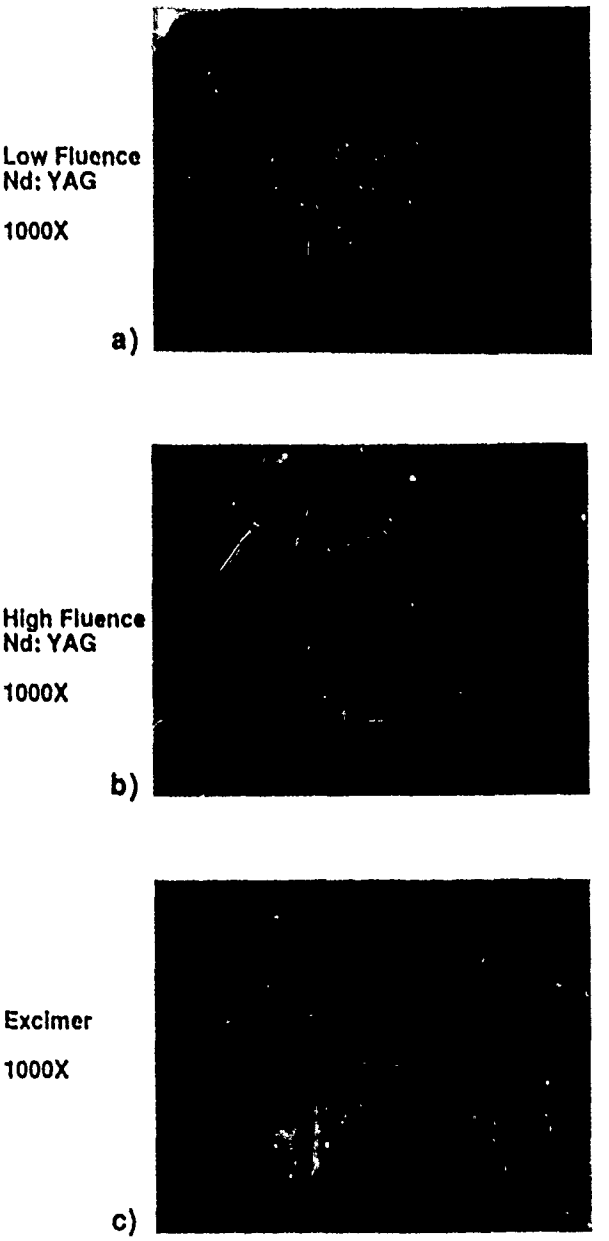


FIGURE 4. Typical film morphologies produced by both the Nd:YAG and excimer lasers.

e-beam evaporated Al_2O_3 pad is greater than that of quartz. Yet in all cases but one (discussed below) the post-seal laser-deposited oxide reduced the oscillator's frequency. Inspection of the films produced by laser deposition (Figure 4) indicated a variety of film morphologies depending on laser type and incident power density (or fluence). These differences in film morphologies have a large effect on the magnitude of the frequency shift produced per unit area by the laser-deposited material. The Nd:YAG laser operating at low fluence levels was capable of producing oxide deposits which had the appearance of an electron beam evaporated film with no evidence of metallic Al at a magnification of 1000 X. The width variations of the deposited line shown in Figure 4(a) are due to fluctuations in output power of the laser when operating at low power levels. Also, inspection of the remaining trim pad on the matching quartz cover after laser irradiation indicated that only the outermost oxide layer and portions of the Al layer were removed with the Nd:YAG laser operating at low fluences. It was hoped that films of this type would produce both an upward frequency shift as well as aging characteristics similar to those obtained with the e-beam evaporated specimens noted in Figure 3. The fact that the frequency of devices trimmed at low fluence levels with the Nd:YAG laser decreased indicates that the quality of the laser-deposited film (stoichiometry, density, adherence, and/or acoustic velocity) is not identical to that of e-beam evaporated material. Also, as noted in Table 1, a larger area between the IDT's needed to be covered with the laser-deposited films produced by low fluences in order to obtain a

given frequency shift. Competition between mass loading (reducing the frequency) and a film with qualities closer to those produced by e-beam evaporation (with an acoustic velocity slightly higher than that of quartz, thus increasing the frequency) may account for the increase in area required to obtain a given frequency shift, as well as the large changes in insertion loss associated with devices trimmed at low fluence levels with the Nd:YAG laser. Since only the outer Al_2O_3 layer and thin Al layer are removed at low fluences, less material is deposited on the active region of the device per unit line width, which also reduces the magnitude of the mass loading effect. However, consider the data from devices Q2281B and A3-2 which were both trimmed down -30 PPM. The excimer-trimmed device required that only 1.8% of the area between the IDT's be covered by the laser-deposited oxide material to produce the noted frequency shift. This is in contrast to 48% coverage of the inter-IDT space required for device A3-2 in order to produce the same frequency shift when trimmed with Nd:YAG operating at low fluence levels. Assuming that only the outer most layer of oxide (250Å) is deposited with the low fluence Nd:YAG and all three layers (750Å) are deposited with the excimer, then one would expect the ratio of the frequency shift produced per unit area due to mass loading alone to be only a factor of three. The actual ratio of frequency shift per unit area covered by the laser-deposited films for these two devices is about a factor of 25. This provides some evidence that films produced by the Nd:YAG at low fluence levels probably have properties closer to the quality of e-beam evaporated films.

At high fluence levels both the excimer and Nd:YAG lasers produced films which had a rough surface morphology indicative of ablation of all three trim pad layers as noted in Figure 4(b) and 4(c). Inspection of the trim pad film on the matching quartz cover from these two deposits indicates that all three layers of the trim pad were removed. Close inspection of the laser-deposited material indicated that submicron islands of metallic Al were present in these films. With film morphologies of this type it is not surprising that the frequency of the SAW oscillator decreases since the surface roughness of the laser-deposited films had features which are smaller than t acoustic wavelength of the SAW device. Thu the deposits act mostly as an effective mas load on the propagating surface wave. By reducing the fluence of the excimer laser w could not produce film deposits similar to those formed with the low fluence Nd:YAG as seen in Figure 4(a), and below fluences of about 6 J/cm^2 the excimer laser was not abl to remove material from the trim pad. This may be related to both differences in laser pulse lengths as well as the fact that the absorption coefficient of Al is over three times smaller at 193 nm than at 1064 nm.[6] Again, the difference in film morphologies produced by the two lasers operating at hig fluence levels produced differences in the magnitude of frequency shift obtained per unit line width. Device L-3 trimmed with Nd:YAG at high fluence required 7.7% covera of the inter-IDT space to produce a shift o -36 PPM, which compares with that of device Q2281A trimmed -35 PPM with a coverage of only 2.2%. An important difference between the two types of laser and optics systems used is that of minimum line width. The excimer-trimmed device had lines drawn with width of 7 μm, while the device trimmed wit the Nd:YAG had line widths of about 80 μm. These dimensions compare with the acoustic wavelength of 7.86 μm at 402 MHz. Thus, th frequency shift produced per unit line widt may vary with the ratio of the acoustic wavelength to that of the width of the lase deposited line in a nonlinear way. Another obvious difference between the laser types i that of peak power density. The pulse lengt for the excimer is about 15 ns, yielding pea power densities of about 500 MW/cm^2. The pulse length for the Nd:YAG laser used to trim both A3-2 and L-3 was 200 ns, yielding peak power densities of 25 MW/cm^2 and 64 MW/cm^2, respectively. These differences in peak power density may produce differences i film qualities such as stoichiometry, etc. Also, tests indicated that with both laser types each individual pulse removed the exposed trim pad, and that at very low fluences (<6 J/cm^2 for the excimer and <1

J/cm^2 for the Nd:YAG) multiple pulses at high repetition rates (200 PPS for the excimer and 8000 PPS for the Nd:YAG) could not remove the trim pad material.

As noted in Table 1, frequency shifts of -35 PPM are readily obtainable with both laser types for devices operating at 402 MHz. This trimming range is more than sufficient to compensate for the spread due to sealing-induced frequency shifts for a SAW AQP operating at 402 MHz. The oscillator frequency may also be trimmed upward, as noted in Table 1 for device E2054 operating at 945 MHz. This upward shift was produced by using the commercial excimer laser/microscope system to first irradiate the three level structure and deposit two 3 um wide oxide lines as described above. This produced a downward frequency shift of about 10 PPM. Then, by refocusing the laser beam onto the freshly deposited oxide, the material was annealed by 1000 pulses at 8 J/cm^2 at a repetition rate of 100 PPS. Similar annealing effects could not be produced with the Nd:YAG laser. While this effect is not well understood at this time it may be due to photochemical effects initiated by UV photons with energies high enough to alter the molecular bond structure of these oxygen-deficient laser-deposited films. The excimer laser operating at 193 nm produces photons with an energy of 6.4 eV/photon, in contrast to only 1.2 eV/photon for the 1064 nm line of the Nd:YAG laser. There is strong evidence that one can produce solid state photo-chemical reactions within an Al/O cermet film[7] as well as remove Al_2O_3 molecules from the bulk by photochemically induced desorption[8] from excimer laser radiation. Also, this upward shift in frequency cannot be explained by a heating effect of the quartz substrate due to absorption of the laser energy. Using standard values for the heat capacity of quartz and the size of the AQP substrate, we calculated the temperature rise per pulse from the excimer to be 0.16 m°C at a fluence of 8 J/cm^2. Even assuming 100% absorption (the absorption coefficient for both Al_2O_3 and quartz are less than 15%) and no cooling between pulses, the temperature rise from 1000 pulses would not be more than 0.16°C. Thus, the observed frequency shift due to laser-induced heating would not be greater than 1 PPM for the rotated Y cuts of quartz used which had temperature coefficients of less than 3 PPM/°C at the trimming temperature. However, transient variations in the oscillator frequency due to localized thermal gradients produced in the quartz substrate by both laser types are seen during laser operation, but their time constants are observed to be on the order of a few seconds.

Frequency trimming with either laser produced no significant changes in oscillator phase noise as noted in Table 1, the measured changes all being less than the resolution of our measuring equipment. The open-loop phase noise levels of most of the devices indicated that the SAW device was the dominant source of noise in the oscillator loop. Increases in device insertion loss of less than 15 mdB/PPM at 402 MHz were obtained with both laser systems operating at high power levels, which is acceptable for most SAW oscillator applications. No significant changes in turnover temperature were obtained for devices undergoing the laser trimming process.

The aging data for several SAW devices operating at 402 MHz are shown in Figure 3. The control resonator oscillators have not deviated from their initial values by more than 0.1 PPM for the duration of the test, which is consistent with aging of SAW devices undergoing a high temperature bake-out during the sealing cycle[9]. The data from the devices with pre-seal e-beam evaporated Al_2O_3 pads (which produced upward frequency shifts) are both drifting down in frequency with time. It is interesting to note that the aging rates of these two devices indicate that a thin oxide pad covering a larger area (which also produced a higher frequency shift) has a lower aging rate than a thicker pad with a narrower profile. Device Q2287 appears to be aging at a rate of 0.25 PPM/year, which is acceptable for most device applications. The data from the pre-seal e-beam evaporated Al_2O_3 devices contrasts that of the post-seal laser-trimmed devices, which are all drifting up in frequency with time. Device A4-2 trimmed with the Nd:YAG laser at low fluence has drifted up over +8 PPM in 20

weeks and would be unacceptable for any applications requiring an accurately set frequency. The cause of the poor aging characteristics of this device is not understood at this time. Again, it may be due to the differences in the oxide film qualities mentioned above or the fact that the thin Al film (used to absorb the laser radiation) is diluted in a smaller volume of oxide, thereby leaving a more oxygen deficient and chemically active laser-deposited film. Also, the chemical bonding of these laser-deposited films may be altered by stress levels of the order of 1×10^{-7} N/m^2 induced by the surface wave, which might increase both the elastic constants and acoustic velocity of the film. Devices L-3 (high power Nd:YAG) and Q2281A (excimer) are aging more gracefully than the low power Nd:YAG trimmed device. Device Q2281A has aged just over +2 PPM in 36 weeks, and if one neglects a short burn in period of about 10 weeks, this device is aging less than +2 PPM per year. Device L-3 appears to be aging at a similar rate as the excimer-trimmed device. While these aging rates are certainly higher than that of both the control and e-beam evaporated devices, they are encouraging considering that both devices were trimmed more than the maximum required to compensate for the spread in sealing-induced frequency shifts for SAW AQP's operating at 402 MHz. Since it has been demonstrated that excimer radiation at 193 nm can increase the device frequency by annealing the laser-deposited film, it may also be capable of annealing out part of this upward frequency drift associated with the laser trimming process, thus yielding improvements in long-term frequency stability. Also, if the aging rates of devices trimmed up in frequency are acceptable (<0.5 PPM/year), then the maximum amount of laser trimming required would be cut in half, as the specified oscillator frequency could then be the target frequency for the sealed SAW resonator. Thus, less material would need to be deposited to trim the device to frequency, with an associated improvement in the aging rate of laser trimmed devices.

CONCLUSIONS: A novel technique for accurately adjusting the frequency of sealed AQP SAW resonator devices has been demonstrated. This new technique may be used to readily compensate for the spread of sealing-induced frequency shifts obtained for the AQP with minimal effect in other device characteristics. With modest improvements in the long-term stability of laser-trimmed devices this technique should be useful for producing large numbers of accurately set SAW stabilized oscillators at lower cost than is now possible.

ACKNOWLEDGMENTS: Thanks to Dr. Joe Callerame and Dr. Gary Montress for their helpful discussions through the course of this project. We would like to also thank Ms. Jo Columbus, Mr. Ernest Sabatino, and Mr. John Lang for fabricating the SAW devices and Steven Fiorillo for assistance in operating the excimer stepper system. Again, a special thanks to Ms. Jo Columbus for patiently characterizing the electrical properties of all the devices. The Lincoln Laboratory portion of this work was supported by the Department of the Air Force, in part under a specific program supported by the Air Force Office of Scientific Research, by the Defense Advanced Research Project Agency, and by the Army Research Office.

1. W. J. Tanski, "Surface Acoustic Wave Frequency Trimming of Resonant and Traveling-wave Devices on Quartz," Appl. Phys. Lett., Vol. 39(1), pp. 40-42, July 1, 1981.

2. P. S. Cross, W. R. Shreve, "Frequency Trimming of SAW Resonators," in IEEE Trans. Sonics Ultrason., Vol. SU-29, No. 6, pp. 231-234, July 1982.

3. 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. 519-526.

4. T. E. Parker, "Precision Surface Acoustic Wave Oscillators," in Proceedings of the IEEE Ultrasonics Symposium, Vol. 1, 1982, pp. 268-274.

5. C. N. Helmick and D. J. White, "Observations of Aging and Temperature Effects on Dielectric-Coated SAW Devices," in Proceedings of the IEEE Ultrasonics Symposium, 1978, pp. 580-585.

6. Handbook of Optics, McGraw-Hill Book Company, New York, 1978, pp. 7-145.

7. D. J. Ehrlich, J. Y. Tsao, and C. O. Bozler, "Submicrometer patterning by Projected Excimer-Laser-Beam Induced Chemistry," J. Vac. Sci. Technol. B, Vol. 3 (1), pp. 1-8, Jan/Feb 1985.

8. R. W. Dreyfus, F. Kelly, R. E. Walkup, "Laser-induced Fluorescence Studies of Excimer Laser Ablation of Al_2O_3," Appl. Phys. Lett., Vol. 49 (21), pp. 1478-1480, Nov 24, 1986.

9. T. E. Parker, "Random Walk Frequency Fluctuations in SAW Oscillators," in Proceedings of 40th Annual Frequency Control Symposium, May, 1986, pp. 241-251.

41st Annual Frequency Control Symposium - 1987

PARAMETRIC FAILURE RATE MODEL FOR QUARTZ CRYSTAL DEVICE AGING WITH APPLICATION TO SURFACE ACOUSTIC WAVE FILTERS

A. A. Feinberg
AT&T Bell Laboratories
1600 Osgood Street
North Andover, MA 01845

Abstract

A theoretical model for the predictions of parametric failure rates of quartz crystal devices is obtained. The model is based on a commonly used aging law for quartz acoustic wave devices. To use the model, it is only necessary to statistically characterize the parameter distribution for the devices at two different times and define a parametric failure limit. An example is given in which a representative sample lot of 80 surface acoustic wave filters is used to characterize the frequency and phase of the population at two different times under accelerated stress conditions. Using these results for the representative lot, the model predicts the cumulative percentage of the population which would fail as a function of time in the field and the instantaneous failure rate in time at use conditions. These predictions are based on defined limits for parametric stability. Such requirements are generally defined by the customer's needs. General results of the model show that when a characteristic parameter of the population being investigated is distributed in Gaussian form, and ages according to the assumed crystal aging law, then the failure rate has a lognormal form in time. The concepts of the model are straightforward, and the methodology may be applied whenever a device aging law is known.

Introduction

The prediction of failure rates for passive and active electronic components is a major undertaking in reliability studies. Accelerated life tests are generally used to help make predictions. The assumption is that a failure mechanism that may occur under normal operating conditions can be made to occur faster under abnormal stress conditions such as high temperature. Often, because of experimental limitations, small sample sizes are used and assumed to be representative of the products' statistical universe. This means that it can be difficult to design an efficient experiment that will yield enough failures in a given time frame so that reliability predictions can be confidently made. If the failures are catastrophic (sudden failures), then a pass/fail analysis is used along with the appropriate reliability model. However, when the device performance degrades gradually in time and eventually drifts out of specification past a parametric limit, causing a parametric failure, then using a pass/fail methodology at the parametric limit may be very inefficient. This is especially true if the stress acceleration factor is small enough so that only a small fraction of the sample population has reached the parametric limit in a reasonable experimental time frame. However, if the kinetics of the parametric degradation can be modeled by an appropriate aging law, then a parametric failure rate model can be obtained and used instead of the pass/fail data analysis. This treatment leads to a more efficient experimental design both in sample size and time requirements. If the model is accurate, the experimental time requirements can be greatly reduced, and the parametric limits need not actually be reached during aging in order to predict the failure rate.

In this memorandum a parametric failure rate model is developed for quartz crystal devices and is used to predict the time-dependent reliability quantities for a 295.6-MHz surface acoustic wave filter (SAWF). The concepts used in obtaining the model may be applied to other types of devices which exhibit parametric degradation according to a known aging law.

The Aging Law

The aging law equation[1] for quartz bulk acoustic wave resonators[2,3] and quartz SAWFs[4,5,6] appears well established and has the form

$$\frac{\Delta P}{P} = C \exp(-E_a/k_B T) \log(1 + at) \qquad (1)$$

where $\Delta P/P$ is the fractional change in a given parameter P of the devices, t is the aging time, C and a are constants (which may differ from device to device), and the expression $\exp(-E_a/k_B T)$ is the familiar Arrhenius thermal activation law. In the case where $at \gg 1$, the aging law may be written as

$$\frac{\Delta P}{P} = A + D \log t \qquad (2)$$

where $D = C \exp(-E_a/k_B T)$ and $A = D \log a$. For example, when $\Delta P/P$ represent synchronous frequency drift, each device's frequency drift ages according to equation (2), with a slope D and intercept A unique for each device.

Recent accelerated aging studies[4] on SAWFs at AT&T Bell Laboratories (AT&T-BL) and elsewhere[5,6] have substantiated the validity of equation (1) for certain device parameters of SAWFs, and its application in the form of equation (2) is discussed in more detail in the last section. Furthermore, the data indicates that those device parameters are distributed quite normally. Examples that demonstrate this normality over most of the population are given in Fig. 1 and 2. In these figures, the synchronous frequency and phase drift distributions of 80 SAWFs are characterized at various measurement times during an accelerated aging study at 120°C. The populations are characterized using normal statistics, and the results are plotted on normal probability graphs. The dashed line through each

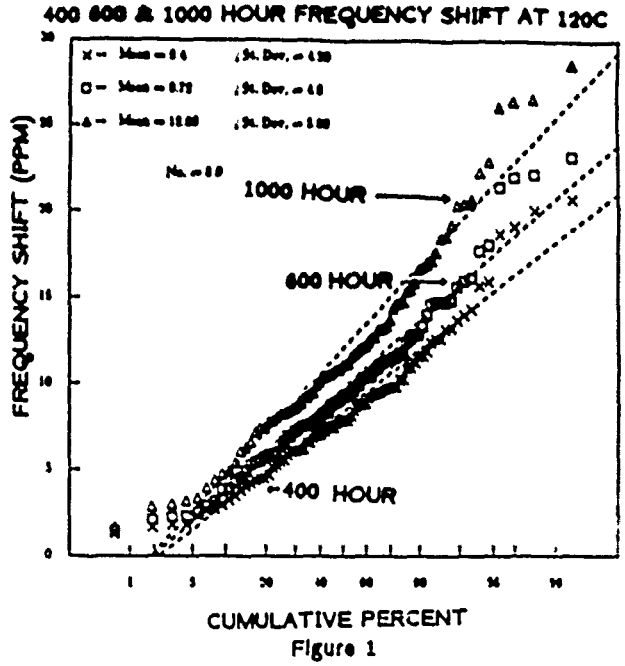


Figure 1

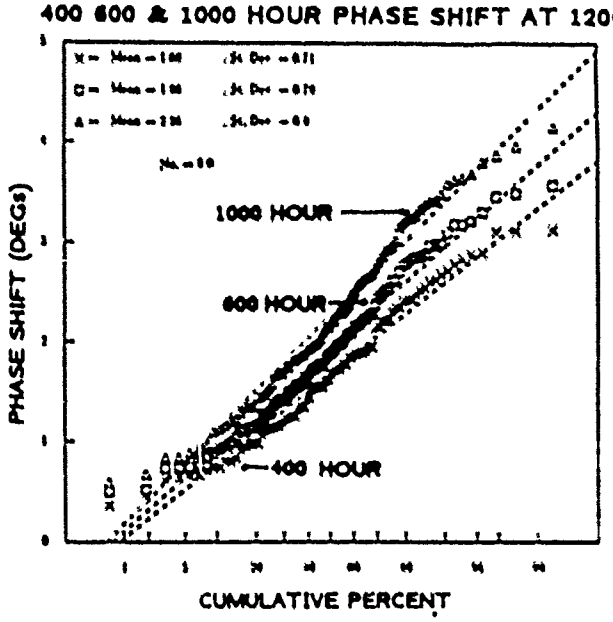


Figure 2

population indicates the best fit for a perfectly normal distribution. A more detailed discussion of these figures is given in the last section.

With these facts in mind, it is possible to construct a parametric failure rate model for SAWF aging. The basic background on parametric failure as it relates to aging of thin-film devices has been previously described by J. S. Fisher[7] and has recently been studied further by P. L. Scarff[8] and A. S Jordan.[9] A similar approach is pursued here. It is worth noting that besides these references, there is very little information on parametric failure in the literature, and apparently, no existing models based on the crystal aging law.[2] Therefore, this paper aims to establish a model based on equation (1).

Predicting Cumulative Percent Failure for Parametric Degradation

To obtain an appropriate failure rate model, one may start with the probability density function f_ϵ for crystal drift parameters in a normal distribution. As previously described, many crystal parameters are normally distributed, so that f_ϵ can be written in Gaussian form

$$f_\epsilon\left(\frac{\Delta P}{P}\right) = \frac{1}{\sigma_p\sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma_p^2}\left(\frac{\Delta P}{P} - \left\langle\frac{\Delta P}{P}\right\rangle\right)^2\right] \quad (3)$$

where $\langle\Delta P/P\rangle$ represents the mean value for a given set of conditions and σ_p is the standard error for the distribution.

Parametric failure can be defined by the maximum permissible shift in $\Delta P/P$ given by

$$\Delta P/P \geq \Delta P/P|_F \quad (4)$$

where $\frac{\Delta P}{P}|_F$ is the threshold of failure.

Figure 3 illustrates the important concepts of this parametric model. The distribution is depicted as it ages, with the fraction of the population that exceeds the allowed drift increasing in time. The figure shows three snapshots in time, showing the distribution before and after the time-to-median failure (τ) and at time $t = \tau$. For large time ($t >> 1$), each device ages according to equation (2) and the mean of the distribution also changes in time with respect to $\Delta P/P|_F$ as

$$\left\langle\frac{\Delta P}{P}\right\rangle = m = A + B \log t \quad (5)$$

For example, $\left\langle\frac{\Delta P}{P}\right\rangle$ above is given by $\frac{1}{N}\sum_{i=1}^{N}\left[\frac{\Delta P}{P}\right]_i$, where N is the number of devices. At time $t = \tau$ the mean value of the distribution is equal to the failure threshold (see Fig. 3b), so that

$$\frac{\Delta P}{P}|_F = A + B \log \tau \quad (6)$$

The cumulative failure function,[10] Q, which expresses the probability that the device will fail according to the criterion in equation (4), is

$$Q\left(\left\langle\frac{\Delta P}{P}\right\rangle\right) = \int_{\frac{\Delta P}{P}|_F}^{\infty} f_\epsilon\left(\frac{\Delta P}{P}\right) d\left(\frac{\Delta P}{P}\right) \quad (7)$$

where we sum over the portion of the distribution that has drifted past the failure threshold

Using equation (3), equation (7) can be rewritten

$$Q = \int_{\frac{\Delta P}{P}|_F}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_p} \exp\left\{-\frac{1}{2\sigma_p^2}\left[\frac{\Delta P}{P} - \left\langle\frac{\Delta P}{P}\right\rangle\right]^2\right\} d\left(\frac{\Delta P}{P}\right) \quad (8)$$

Equation (8) can be expressed conveniently by the complementary error function.

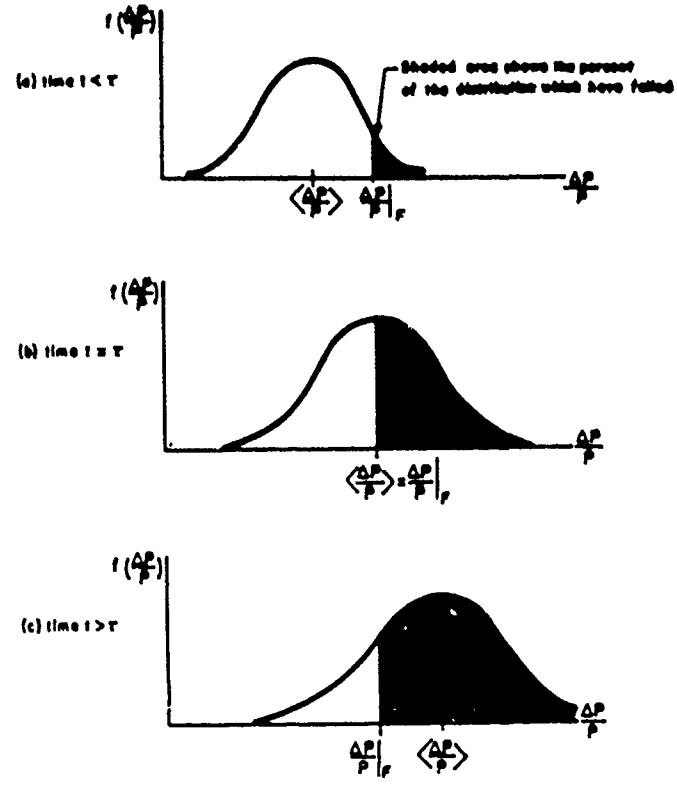


FIGURE 3 SNAPSHOTS OF THE NORMAL DISTRIBUTION BROADENING AND AGING UP IN TIME WITH RESPECT TO $\frac{\Delta P}{P}|_F$

$$Q = \frac{1}{2}\,\mathrm{erfc}\left[\frac{1}{\sqrt{2}\,\sigma_p}\left(\frac{\Delta P}{P}\Big|_P - \left\langle\frac{\Delta P}{P}\right\rangle\right)\right] \qquad (9)$$

or

$$Q = \frac{1}{2}\,\mathrm{erfc}\left[\frac{\langle D\rangle}{\sqrt{2}\,\sigma_p}\log\left(\frac{r}{t}\right)\right], \qquad (10)$$

and the cumulative failure rate (CFR) follows in FITs as $\mathrm{CFR} = \frac{Q}{t}\times 10^9$

Letting $\sigma_p/\langle D\rangle = \sigma_t$, equation (10) reads

$$Q(t) = \frac{1}{2}\left[1 + \mathrm{erf}\left[\frac{1}{\sqrt{2}\,\sigma_t}\log\left(\frac{t}{r}\right)\right]\right]. \qquad (11)$$

This shows that Q(t) has a lognormal form in time (see, for example, Jordan[9]) and that the standard deviation σ_t can be written with temperature dependence, i.e.,

$$\sigma_t = \sigma_p/C\exp\left(-E_a/K_BT\right). \qquad (12)$$

Note that equation (11) is not strictly lognormal, since σ_p is time-dependent as discussed below.

A Time-Dependent Standard Deviation

In the above equations, it is important to note that σ_p is time dependent. To show this, it is also convenient to define the standard deviation for a symmetric distribution in terms of its lower or upper sigma value. Using the lower form,

$$\sigma_p = \left\langle\frac{\Delta P}{P}\right\rangle - \frac{\Delta P}{P}\Big|_{16} \qquad (13)$$

where the subscript, 16, indicates the value of $\frac{\Delta P}{P}$ at the 16th percentile of the distribution. With this definition and equation (2), it is easy to see that σ_p is time dependent and can be written

$$\sigma_p(t) = \langle A\rangle - A_{16} + \left(\langle D\rangle - D_{16}\right)\log t. \qquad (14)$$

The fact that σ_p is time dependent (depicted in Fig 3) is an unusual feature for a parametric-aging model. This feature is, of course, a direct consequence of the aging law's time-dependent form.

Predicting the Instantaneous Failure Rate for Parametric Degradation

Finally, it is of interest to calculate the probability density function for failure, f(t), so that the instantaneous failure rate λ(t) can be obtained. This is obtained through the total derivative of Q(t) as[10]

$$f(t) = \frac{dQ}{dt} = \frac{\partial Q\left(\left\langle\frac{\Delta P}{P}\right\rangle\right)}{\partial\left\langle\frac{\Delta P}{P}\right\rangle}\frac{d\langle\Delta P/P\rangle}{dt} + \frac{\partial Q}{\partial\sigma_p}\frac{d\sigma_p}{dt} \qquad (15)$$

where f(t) ($\neq f_c$) is the probability density function for failure.

Differentiating under the integral and performing the integration yields the desired results. For example, the approach used for the first term reads

$$\frac{\partial Q}{\partial\left\langle\frac{\Delta P}{P}\right\rangle} = -\int_{\frac{\Delta P}{P}\Big|_P}^{\infty}\frac{\partial}{\partial\left\langle\frac{\Delta P}{P}\right\rangle}f_c\left(\sigma_p, \left\langle\frac{\Delta P}{P}\right\rangle, \frac{\Delta P}{P}\right)d\left(\frac{\Delta P}{P}\right), \qquad (16)$$

with $\frac{d\langle\Delta P/P\rangle}{dt} = \frac{\langle D\rangle}{t}$. The integrals, while cumbersome, are straightforward. The result reads

$$f(t) = \left[\langle D\rangle + \left(\langle D\rangle - D_{16}\right)\left(\frac{\frac{\Delta P}{P}\Big|_P - \left\langle\frac{\Delta P}{P}\right\rangle}{\sigma_p}\right)\right]\times \left(\frac{1}{t\sqrt{2\pi}\sigma_p}\right)\exp\left[\frac{-1}{2\sigma_p^2}\left(\frac{\Delta P}{P}\Big|_P - \left\langle\frac{\Delta P}{P}\right\rangle\right)^2\right] \qquad (17)$$

and is valid when $at \gg 1$. Note that the second term in the first square bracket results from the time-dependent sigma term in equation (15).

The instantaneous failure rate can now be obtained in terms of the parameters once we insert equations (9) and (17) into λ(t), which is given in FITs by[10]

$$\lambda(t) = \frac{f(t)}{1 - Q(t)}\times 10^9. \qquad (18)$$

In terms of the time to median failure r,

$$\lambda(t) = \frac{\left[1 + \frac{(\langle D\rangle - D_{16})}{\sigma_p}\log\left(\frac{r}{t}\right)\right]\frac{\langle D\rangle}{t}\exp\left[-\frac{\langle D\rangle^2}{2\sigma_p^2}\left(\log\frac{r}{t}\right)^2\right]}{\sigma_p\sqrt{2\pi}\,t\left\{1 - \frac{1}{2}\,\mathrm{erfc}\left[\frac{\langle D\rangle}{2\sigma_p}\log\frac{r}{t}\right]\right\}} \qquad (19)$$

Equations (11) and (19) indicate that if the characteristic parameter for the devices being investigated is distributed normally and ages according to equation (2), then the CFR and the instantaneous failure rate have a lognormal form in time. It is therefore possible to reliably predict failure rates of SAWFs and other quartz devices using equations (11) and (19).

Furthermore, such predictions can be based on small sample sizes which can be estimated from the characteristics of typical previous distributions and the desired confidence limits. Also, it is not necessary to find the time-to-median failure. The distribution only needs to be characterized at constant temperature for two different times for the model. Once $\frac{\Delta P}{P}\Big|_P$ is defined and $\left\langle\frac{\Delta P}{P}\right\rangle$ and σ_p have been statistically determined for a given sample population at two different times, the reliability quantities may be obtained at any other time. Furthermore these values may be obtained as a function of temperature provided the activation energy is known [see equation (1)].

Using the Model to Predict the Parametric Failure Rate for SAWFs

In this section, the model is demonstrated on a sample population of 80 SAWFs having their nominal frequency at 295.6 MHz. The filters have been aged at 120°C for 600 hours. The population for synchronous frequency and phase drift have been characterized at three different times, and the results are graphed in Fig. 1 and 2, respectively. To characterize the data, 14 measurements were made on each device over the 600-hour aging period. Then the

resulting 14 data points were fitted by method of least squares fit using the aging law as given in equation (2). The resulting least squares fit to the data was then calculated at 400, 600, and 1000 hours for each device. The resulting fitted distribution of the 80 SAWFs is what is shown in Fig. 1 and 2. Fitting the data in this way utilizes all the data. However, fitting all of the data is not necessary to use the proposed model. It is only necessary to characterize the raw data at two different times.

It should be noted that the confidence in the fitted data using the aging law [see equation (2)] is quite good, demonstrating the applicability of the aging law. For example, the 90-percent confidence limit about the mean at 600 hours is 2 ppm for the frequency distribution and 0.8-phase degrees for the phase distribution.

The 295.6-MHz SAWF data is representative of high-performance filter products. While this particular device is used at 10°C at the ocean bottom, many terrestrial SAWFs operate at bay temperatures of 50°C with temperature excursions up to 75°C. The model is used to extrapolate the failure rate in the extreme case at 75°C. However, it should be noted that even at 50°C, the failure rates up to 25 years prove to be negligible.

Figure 1 shows the characteristics of the frequency shift distribution (in ppm) of 80 SAWFs after 400, 600, and 1000 hours at 120°C, while Fig. 2 shows the distribution for phase (in degrees) at these times. The populations are characterized using normal statistics, and the results are plotted on normal probability graphs. The dashed line through each population indicates the best fit for a perfectly normal distribution. The figures help to demonstrate the normality for each distribution, with each distribution having a correlation coefficient to the best-fit dashed line greater than 0.98. This observation lends support for using the Gaussian form to represent the parameter distribution in equation (3). It should be noted that with the abscissa plotted in standard normal deviates, the slope of the dashed line yields the standard deviation for the population.[10] From the figures, then, one can easily observe the mean and the slope (or standard deviation value) increasing with aging time as discussed earlier [see equations (5) and (14)].

Using the means and the standard deviations noted in the figures and defining rather stringent failure criteria of 32 ppm for frequency stability and 1 degrees for phase stability throughout the product life of 25 years and a previously determined

PARAMETRIC FAILURE ANALYSIS AT 75C
FOR FREQUENCY STABILITY

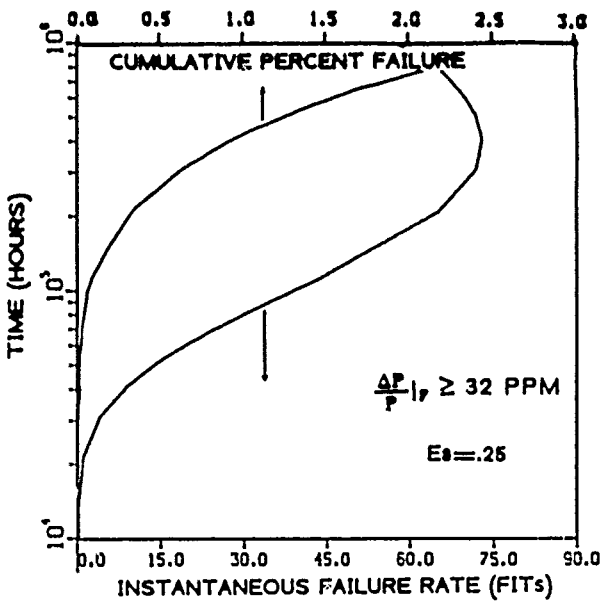


Figure 4

PARAMETRIC FAILURE ANALYSIS AT 75C
FOR PHASE STABILITY

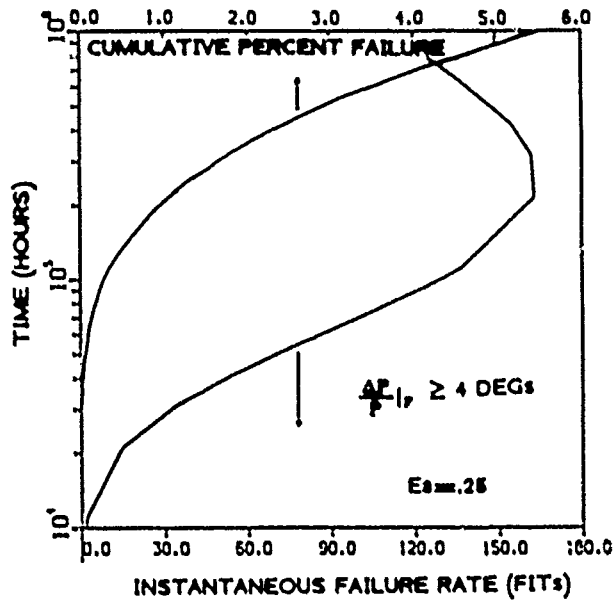


Figure 5

activation energy, the computed reliability quantities are given for the operating temperature of 75°C in Fig. 4 and 5 for synchronous frequency and phase stability, respectively. These figures plot the cumulative percent failure (upper abscissa) and the instantaneous failure rates (lower abscissa) versus time with the arrows in the figures indicating their respective axis. From Fig. 4, the instantaneous failure rate for frequency stability would be less than 5 FITs at 75°C for times up to 30,000 hours (approximately 3.5 years) and goes through a peak of 74 FITs at 25 years with less than 0.5 percent of the devices exceeding the parametric limit of 32 ppm. Figure 5 shows that the instantaneous failure rate for phase stability would be less than 35 FITs for 3.5 years and less than 163 FITS for 25 years at 75°C with less than 1 percent of the devices failing when the parametric phase limit is set at 4° (phase degrees). This calculation is valid when $t \gg 1/a$ [see equation (1)], which has been estimated from aging data on this filter to be greater than 50 hours.

The calculated results will be highly dependent on the activated energy value, since the results have been extrapolated down from the 120° aging data to an operating temperature of 75°C [see equation (1)]. The nominal value, $E_a = 0.25$ ev, has been previously determined from extensive isothermal aging studies, with a 90-percent lower bound confidence limit for the activated energy of $E_a = 0.17$ ev. The failure rates quoted above at 75°C with $E_a = 0.25$ ev would apply at 58°C if this worst-case lower 90-percent confidence value $E_a = 0.17$ ev were used in the calculation for the activation energy, since the 120° to 75°C acceleration factor with $E_a = 0.25$ ev is equivalent to the 120° to 58°C acceleration factor with $E_a = 0.17$ ev. This implies that the 90-percent lower temperature bound for this reliability prediction is at 58°C.

Summary

A theoretical model for the prediction of the parametric failure rate of quartz crystal devices was obtained based on the aging law and a representative failure threshold. The model shows that when the devices age according to this law and parameters are distributed in a Gaussian form, the failure rates prove to have a lognormal form in time. The parametric failure rate for the 295.6-MHz SAWF was obtained at 75°C using the model based on aging data taken on sample population of 80 SAWFs. The model

is easy to use, straightforward, and requires a minimum of input. The methodology may be applied to other types of devices when a device aging law is known.

Acknowledgement

The author would like to thank P. L. Scarff for interesting and friendly discussions, R. L. Rosenberg for helpful suggestions, and T. R. Meeker for enlightening discussions and for making available computer software routines that helped make this study feasible.

REFERENCES

[1] P. T. Landsberg, "On the Logarithmic Rate Law in Chemisorption and Oxidation," *J. Chem. Phys.*, Vol. 23:6, 1955, pp. 1079-1087.

[2] A. W. Warner, D. B. Fraser, and C. D. Stockbridge, "Fundamental Studies of Aging in Quartz Resonators," *IEEE Trans. on Sonics and Ultrasonics*, June 12, 1965, pp. 52-59.

[3] S. H. Olster, *et al.* "A6 Monolithic Crystal Filter Design for Manufacture and Device Quality," *Proc. IEEE on Sonics and Ultrasonics*, 1975, pp. 105-112.

[4] T. R. Meeker, "Reliability of Surface Acoustic Wave Filters for TAT-8," AT&T Bell Laboratories, Internal Memorandum.

[5] D. T. Bell, Jr., "Aging Processes in SAW Resonators," *Ultrasonics Symposium Proceedings, IEEE*, 1977, pp. 851-856.

[6] W. R. Shreve, "Aging in Quartz SAW Resonators," *Ultrasonics Symposium Proceedings, IEEE*, 1977 pp. 857-861.

[7] J. S. Fisher, "Aging Characteristics of Tantalum Nitride Thin-Film Resistors," Proceedings of Electronic Components Conference, 1968, pp. 299-302.

[8] P. L. Scarff, AT&T Bell Laboratories, private communications.

[9] A. S. Jordan "A Comprehensive Review of the Lognormal Failure Distribution with Applications to LED Reliability," *Microelectronics Reliability*, Vol. 18, 1978, pp. 267-279.

[10] Bazovsky, Igor, *Reliability Theory and Practice*, Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

41st Annual Frequency Control Symposium - 1987

APPROXIMATION METHOD FOR PLATE MODES IN SURFACE ACOUSTIC WAVE DEVICES

Janpu Hou

Electronic Materials and Devices Laboratory
Allied-Signal Inc., Morristown, NJ 07960

Abstract - Surface acoustic wave devices produce plate modes which degrade the devices' performance. Experiments also show that SAW chemical sensors in liquid media do not operate as in the gas phase.[1] It is known that the dominant mode of energy propagation is not due to Rayleigh waves in thin substrate devices. The classical analysis procedure for freely propagating plate modes is to seek partial wave solutions to the equations of motion and then to require that sums of these solutions give the prescribed behavior at the plate boundaries. Instead of solving this boundary value problem a two-dimensional theory for high-frequency vibrations of peizoelectric crystal plates is employed to study the plate mode propagation in SAW devices.[2] Dispersion relations are obtained in explicit algebric form and solved for a given thickness of the substrate. Plate mode spectra are obtained for SAW delay lines. It is demonstrated that with this approximation method the spectra are much simpler to compute and the coupling between the various anharmonic modes is easier to determine.

Introduction

The interdigital transducers (IDT) of the surface acoustic waves (SAW) devices are effective for launching not only Rayleigh waves, but also slow quasi-shear waves, fast quasi-shear waves, quasi-longitudinal waves and their anharmonic overtones. The plate mode coupling in SAW devices is well known[3,4] and their influence on the terminal properties of SAW filter were examined by Wagers[5,6]. The efforts were put toward eliminating the plate modes for SAW filter designers. Recent studies on the plate mode sensor devices[7] and chemical sensors in liquids[8] show that plate modes are important in devices on both thin and thick substrates. A plate mode sensor designer needs to understand the plate modes' characteristics in order to select one against the other.

The classical analysis procedure is to seek partial wave solutions to the equations of motion and then to require the sums of these solutions satisfy the plate boundaries' conditions.

In the present paper, two-dimensional equations of motion of successively higher orders of approximation are applied to study the plate modes in SAW delay lines. Because either the face potential or the face charge might be specified as surface conditions, the plate mode frequency can be easily found without iteration for a given substrate thickness.

To examine the accuracy of these equations, one needs to find the dispersion relations between frequency and wavelength/thickness ratio. Dispersion curves of straight-crested wave propagating in the x1 direction in a ST-cut quartz plate are obtained by both the classical method and the approximation method. The comparison shows good agreement.

To demonstrate that these approximate equations can be directly applied to the SAW delay line, we calculate the plate mode frequency of a 50 MHz delay line on eight wavelength thick ST-quartz. The frequencies of the calculation and the device testing results agree well with each other.

Classical Analysis Method

The geometry under consideration is illustrated in Fig. 1. A piezoelectric plate has its length along the x_1-axis and thickness 2b along the x_2-axis, and the modal variations are taken to be independent of x_3. For comparison purpose, we review the classical method comparison purpose, we review the classical method here. The equations of motions are

$$c_{ijkl}u_{k,il} + e_{kij}\phi_{,ki} = \rho u_{j,tt}$$

$$e_{ikl}u_{k,il} - \epsilon_{ik}\phi_{,ki} = 0 \qquad (1)$$

$i,j,k,l = 1,2,3$

The plate is assumed to be stress free, and the boundary conditions are at $x_2 = \pm b$

$$T_{2j} = c_{2jkl}u_{k,l} + e_{k2j}\phi_{,k} = 0$$

open $\quad D_2 = e_{2kl}u_{k,l} - \epsilon_{2k}\phi_{,k} = 0$

or

short $\quad \phi = 0 \qquad (2)$

We assume

$$u_j = a_j e^{i\omega t} e^{ik(x_1 + \ell_2 x_2)}$$

$$\phi = a_4 e^{i\omega t} e^{ik(x_1 + \ell_2 x_2)} \qquad (3)$$

where the transverse propagation constant along thickness is normalized to be a fraction ℓ_2 of that along the substrate length propagation constant $k \cdot k$ is always real, but ℓ_2 may be a complex number.

Substitution of (3) into (1) yields an eighth order polynomials in ℓ_2 with coefficient as function of material constants c_{ij}, e_{ip}, ϵ_{ij} and phase velocity V,

CH2427-3/87/0000-365 $1.00 C 1987 IEEE

$$a_8 t_2^8 + a_7 t_2^7 + a_6 t_2^6 + a_5 t_2^5 + a_4 t_2^4 + a_3 t_2^3 + a_2 t_2^2 + a_1 t_2 + a_0 = 0 \qquad (4)$$

where $a_j = f(C_{ij}, e_{ip}, \epsilon_{ij}, V)$

for a given value of V there are eight roots of t_2 as $t_2^{(n)}$. n = 1, 2, ...8.

The mechanical displacement and electric potential can be expressed as

$$u_j = e^{ick\sum_{m=1}^{4} C_m a_j^{(n)} \cos t_2^{(n)} x_2} ,$$

$$\phi = e^{ick\sum_{m=1}^{4} C_m a_4^{(n)} \cos t_2^{(n)} x_2} . \qquad (5)$$

where $a_j^{(n)}$ (j = 1,2,3,4) are amplitude ratio. By substituting each $t_2^{(n)}$ back into (1), one may solve $a_j^{(n)}$.

An 8 x 8 boundary value determinant can thus be formulated by substituting (5) into the boundary conditions (2).

$$| B_{mn} | = 0 \qquad (6)$$

Solutions to Eq.(6) are found iteratively, first, a value of phase velocity is chosen, then Eq.(4) is solved for the eight $t_2^{(n)}$, the amplitude ration $a_j^{(n)}$ are calculated and the boundary value determinant is evaluated. If Eq.(6) is satisfied, then a plate mode has been found, if not a new velocity is chosen and the process is repeated.

Approximate Analysis Method

In the two-dimensional approximate theory, the components of mechanical displacement and electric potential are expanded in an infinite series with the thickness dependence expressed by trigonometric functions

$$u_i = \sum_{n=0}^{\infty} u_i^{(n)}(x_1,t)\cos\frac{n\pi}{2}\left(1 - \frac{x_2}{b}\right) ,$$

$$\phi = \sum_{n=0}^{\infty} \phi^{(n)}(x_1,t)\cos\frac{n\pi}{2}\left(1 - \frac{x_2}{b}\right) .$$

For clearity and ease of use, equations of zero- and first-order approximation which contain correction factors d_1 and d_2 and general equations of nth order approximation $1 < n < N$ are recapitulated and displayed in detail for the rotated Y-cut crystal with class 32 symmetry.

Zero order equations

$$c_{11}u_{1,11}^{(0)} + e_{11}\phi_{,11}^{(0)} + \frac{a_1}{b}\left(c_{12}u_{2,1}^{(1)} + c_{14}u_{3,1}^{(1)}\right) + \frac{\pi}{4b}\sum_{m=3}^{N} mB_{mo}\left(c_{12}u_{2,1}^{(m)} + c_{14}u_{3,1}^{(m)}\right) + \left(\frac{1}{2b}\right) F_1^{(0)} = \rho u_{1,tt}^{(0)}$$

$$c_{56}u_{3,1}^{(0)} + c_{66}u_{2,1}^{(0)} + \frac{a_1}{b}\left(c_{66}u_{1,1}^{(1)} + e_{26}\phi_{,1}^{(1)}\right) + \frac{\pi}{4b}\sum_{m=3}^{N} mB_{mo}\left(c_{66}u_{1,1}^{(m)} + e_{26}\phi_{,1}^{(m)}\right) + \left(\frac{1}{2b}\right) F_2^{(0)} = \rho a_2^{-2} u_{2,tt}^{(0)}$$

$$c_{55}u_{3,11}^{(0)} + c_{56}u_{2,11}^{(0)} + \frac{a_1}{b}\left(c_{56}u_{1,1}^{(1)} + e_{25}\phi_{,1}^{(1)}\right) + \frac{\pi}{4b}\sum_{m=3}^{N} mB_{mo}\left(c_{56}u_{1,1}^{(m)} + e_{25}\phi_{,1}^{(m)}\right) + \left(\frac{1}{2b}\right) F_3^{(0)} = \rho u_{3,tt}^{(0)}$$

$$e_{11}u_{1,11}^{(0)} - E_{11}\phi_{,11}^{(0)} + \frac{a_1}{b}\left(e_{12}u_{2,1}^{(1)} + e_{14}u_{3,1}^{(1)}\right) + \frac{\pi}{4b}\sum_{m=3}^{N} mB_{mo}\left(e_{12}u_{2,1}^{(m)} + e_{14}u_{3,1}^{(m)}\right) + \left(\frac{1}{2b}\right) D^{(0)} = 0$$

where

$$B_{mo} = \frac{4}{m\pi} \qquad m = 1,3,5...$$

First order equations

$$c_{11}u_{1,11}^{(1)} + e_{11}\phi_{,11}^{(1)} - \frac{2a_1}{b}\left(c_{66}u_{2,1}^{(0)} + c_{56}u_{3,1}^{(0)}\right) - \left(\frac{\pi}{2b}\right)^2\left(c_{66}u_1^{(1)} + e_{26}\phi^{(1)}\right) + \frac{\pi}{2b}\sum_{m=1}^{N}\left\{mB_{m1}\left(c_{12}u_{2,1}^{(m)} + c_{14}u_{3,1}^{(m)}\right) - B_{m1}\left(c_{56}u_{3,1}^{(m)} + c_{66}u_{2,1}^{(m)}\right)\right\} + \left(\frac{1}{b}\right) F_1^{(1)} = \rho u_{1,tt}^{(1)}$$

$$a_2 c_{56}u_{3,11}^{(1)} + a_2^2 c_{66}u_{2,11}^{(1)} - \left(\frac{2a_1}{b}\right)\left(c_{12}u_{1,1}^{(0)} + e_{12}\phi_{,1}^{(0)}\right) - \left(\frac{\pi}{2b}\right)^2\left(c_{22}u_2^{(1)} + c_{24}u_3^{(1)}\right) + \frac{\pi}{2b}\sum_{m=1}^{N}\left\{mB_{m1}\left(c_{66}u_{1,1}^{(m)} + e_{26}\phi_{,1}^{(m)}\right) - B_{m1}\left(c_{12}u_{1,1}^{(m)} + e_{12}\phi_{,1}^{(m)}\right)\right\} + \left(\frac{1}{b}\right) F_2^{(1)} = \rho u_{2,tt}^{(1)}$$

$$c_{55}u_{3,11}^{(1)} + \alpha_2 c_{56}u_{2,11}^{(1)} - (\frac{2\alpha_1}{b})(c_{14}u_{1,1}^{(0)} + e_{14}\phi_{,1}^{(0)}) - (\frac{\pi}{2b})^2 (c_{24}u_2^{(1)} + c_{44}u_3^{(1)}) + \frac{\pi}{2b}\sum_{m=1}^{N}\{mB_{m1}(c_{56}u_{1,1}^{(m)} + e_{25}\phi_{,1}^{(m)}) - B_{m1}(c_{14}u_{1,1}^{(m)} + e_{14}\phi_{,1}^{(m)})\} + (\frac{1}{b})F_3^{(1)} = \rho u_{3,tt}^{(1)}$$

$$c_{11}u_{1,11}^{(1)} - c_{11}\phi_{,11}^{(1)} - \alpha_1(\frac{2}{b})(c_{25}u_{3,1}^{(0)} + c_{2b}u_{2,1}^{(0)}) - (\frac{\pi}{2b})^2(c_{2b}u_1^{(1)} - c_{22}\phi^{(1)}) + \frac{\pi}{2b}\sum_{m=1}^{N}\{mB_{m1}(c_{12}u_{2,1}^{(m)} + c_{14}u_{3,1}^{(m)}) - B_{m1}(c_{2b}u_{2,1}^{(m)} + c_{25}u_{3,1}^{(m)})\} + (\frac{1}{b})D^{(1)} = 0$$

n-th order ($1 < n \leq N$)

$$c_{11}u_{1,11}^{(n)} + c_{11}\phi_{,11}^{(n)} - (\frac{n\pi}{2b})^2 (c_{66}u_1^{(n)} + c_{26}\phi^{(n)}) + \frac{\pi}{2b}\sum_{m=0}^{N}\{mB_{mn}(c_{12}u_{2,1}^{(m)} + c_{14}u_{3,1}^{(m)}) - nB_{nm}(c_{56}u_{3,1}^{(m)} + c_{66}u_{2,1}^{(m)})\} + (\frac{1}{b})F_1^{(n)} = \rho u_{1,tt}^{(n)}$$

$$c_{56}u_{3,11}^{(n)} + c_{66}u_{2,11}^{(n)} - (\frac{n\pi}{2b})^2(c_{22}u_2^{(n)} + c_{24}u_3^{(n)}) + (\frac{\pi}{2b})\sum_{m=0}^{N}\{mB_{mn}(c_{66}u_{1,1}^{(m)} + c_{26}\phi_{,1}^{(m)}) - nB_{nm}(c_{12}u_{1,1}^{(m)} + c_{12}\phi_{,1}^{(m)})\} + (\frac{1}{b})F_2^{(n)} = \rho u_{2,tt}^{(n)}$$

$$c_{55}u_{3,11}^{(n)} + c_{56}u_{2,11}^{(n)} - (\frac{n\pi}{b})^2 (c_{24}u_2^{(n)} + c_{44}u_3^{(n)}) + \frac{\pi}{2b}\sum_{m=0}^{N}\{mB_{mn}(c_{56}u_{1,1}^{(m)} + c_{25}\phi_{,1}^{(m)}) - nB_{nm}(c_{14}u_{1,1}^{(m)} + c_{14}\phi_{,1}^{(m)})\} + (\frac{1}{b})F_3^{(1)} = \rho u_{3,tt}^{(1)}$$

$$c_{11}u_{1,11}^{(n)} - c_{11}\phi_{,11}^{(n)} - (\frac{n\pi}{b})^2 (c_{26}u_1^{(n)} - c_{22}\phi^{(n)}) + \frac{\pi}{2b}\sum_{m=0}^{N}\{mB_{mn}(c_{12}u_{2,1}^{(m)} + c_{14}u_{3,1}^{(m)}) - nB_m(c_{25}u_{3,1}^{(m)} + c_{26}u_{2,1}^{(m)})\} + (\frac{1}{b})D^{(n)} = 0$$

where

$$\alpha_1 = \frac{\pi}{\sqrt{8}}$$

$$\alpha_2 = \sqrt{\frac{(v_s^2 - \bar{c}_{55})}{(\bar{c}_{66}v_s^2 + \bar{c}_{56}^2 - \bar{c}_{55}\bar{c}_{66})}}\, v_s$$

$$v_s^2 = \frac{\rho v_s^2}{c_{66}} \qquad \bar{c}_{66} = c_{66} + \frac{e_{26}^2}{c_{22}}$$

$$\frac{}{c_{pq}} = \frac{c_{pq}}{c_{66}} \qquad v_s = \text{surface wave velocity}$$

For ST-quartz

$$v_s = 0.931051$$

$$\alpha_2 = 0.9554$$

In two-dimensional theory, the boundary conditions on the top and bottom of the plate will be specified in different ways. By specifying the face tractions and face charges, the top and bottom surface boundary conditions are already part of the equation. SAW delay line sensor normally has a thin layer of selective coating on the top of the device. The mechanical mass loading effect can be taken into account as

$$F_j^{(n)} = f_j^{(n)} - 2b'\rho' u_{j,tt}$$

where $2b'$ is the film thickness and ρ' is the film mass density. The electrical effect on the surface can be specified for open conditions as

$$D^{(n)} = 0$$

or for short condition as

$$D^{(n)} = \frac{2}{b}(e_{26}u_1^{(n)} + e_{22}u_2^{(n)} + e_{24}u_3^{(n)}$$

$$n = 1, 3, 5\ldots$$

Solutions of Approximation Method

The relationship between normalized frequency and normalized wavenumber can be found by substituting the displacement and potential in the equations of motion. For a given value of plate thickness, or a given value of normalized wavenumber, we can find the corresponding normalized frequency, or phase velocity of plate modes. We don't need iteration when we use the approximation method.

The matrices of the dispersion relations are symmetric and their elements may be classified into four groups: Amn, Bmn for elements on the main digaonal; Cmn and Dmn for the elements off main diagonal.[8] In terms of these groups the idspersion relations of the N-th order theory for the essentially symmetric families of modes, respectively, can be written as:

$$
\begin{array}{c|cccccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & N \\
0 & A_{00} & C_{01} & 0 & C_{03} & 0 & C_{05} & \cdots & 0 \\
1 & & B_{11} & D_{12} & 0 & D_{12} & 0 & \cdots & D_{1n} \\
2 & & & A_{22} & C_{23} & 0 & C_{25} & \cdots & 0 \\
3 & & & & B_{33} & D_{34} & 0 & \cdots & D_{3n} \\
4 & & & & & A_{44} & C_{45} & \cdots & 0 \\
5 & & & & & & B_{55} & \cdots & D_{5n} \\
\vdots & & & & & & & \cdots & \\
N & & & & & & & & A_{nn}
\end{array}
\Bigg|_{n} = 0
$$

and

$$
\begin{array}{c|cccccccc}
 & 0 & 1 & 2 & 3 & 4 & 5 & \cdots & N \\
0 & B_{00} & D_{01} & 0 & D_{03} & 0 & D_{05} & \cdots & D_{0n} \\
1 & & A_{11} & C_{12} & 0 & C_{14} & 0 & \cdots & 0 \\
2 & & & B_{22} & D_{23} & 0 & D_{25} & \cdots & D_{2n} \\
3 & & & & A_{33} & C_{34} & 0 & \cdots & 0 \\
4 & & & & & B_{44} & D_{45} & \cdots & D_{4n} \\
5 & & & & & & A_{55} & \cdots & 0 \\
\vdots & & & & & & & \cdots & \\
N & & & & & & & & A_{nn}
\end{array}
\Bigg|_{n} = 0
$$

The terms in the groups are defined as follows ($0 < m, n < N$, for any $N > 0$):

$$
A_{nn} = \begin{vmatrix} \bar{c}_{11}z^2 + n^2\bar{c}_{66} - \Omega^2 & \bar{c}_{11}z^2 + n^2\bar{c}_{26} \\ \bar{c}_{11}z^2 + n^2\bar{c}_{26} & -\bar{c}_{11}z^2 - n^2\bar{c}_{22} \end{vmatrix}
$$

$$
B_{nn} = \begin{vmatrix} \bar{c}_{66}z^2 + n^2\bar{c}_{22} - \Omega^2 & \bar{c}_{56}z^2 + n^2\bar{c}_{24} \\ \bar{c}_{56}z^2 + n^2\bar{c}_{24} & \bar{c}_{55}z^2 + n^2\bar{c}_{44} - \Omega^2 \end{vmatrix}
$$

$$
C_{nm} = \pm\frac{16z^2}{(n^2-m^2)^2\pi^2} \begin{vmatrix} (n^2\bar{c}_{66}+m^2\bar{c}_{12})(n^2\bar{c}_{56}+m^2\bar{c}_{14}) \\ (m^2\bar{c}_{12}+n^2\bar{c}_{26})(m^2\bar{c}_{14}+n^2\bar{c}_{25}) \end{vmatrix}
$$

$$
D_{nm} = \pm\frac{16z^2}{(m^2-n^2)^2\pi^2} \begin{vmatrix} (m^2\bar{c}_{66}+n^2\bar{c}_{12})(m^2\bar{c}_{26}+n^2\bar{c}_{12}) \\ (m^2\bar{c}_{56}+n^2\bar{c}_{14})(m^2\bar{c}_{25}+n^2\bar{c}_{14}) \end{vmatrix}
$$

where

$$
z = \frac{c}{(\pi/2b)}, \quad \Omega = \frac{\omega}{\omega_1}, \quad \omega_1^2 = \frac{\pi^2}{4b^2}\frac{\bar{c}_{66}}{\rho},
$$

$$
\bar{c}_{pq} = \frac{c_{pq}}{c_{66}}, \quad \bar{c}_{1q} = \frac{c_{1q}}{(c_{66}c_{22})^{1/2}}, \quad \bar{c}_{ij} = \frac{c_{ij}}{c_{22}}.
$$

The upper sign in the last two groups is applied to the essentially symmetric family of modes and the lower sign is applied to the essentially antisymmetric family of modes. It shall be noted that the correction coefficients a_1 and a_2 are only applied in terms B_{00}, B_{11}, C_{01}, D_{01}.

Theoretical and Experimental Results

The dispersion relations for the fifth, sixth order theories are generated by the approximation method. The dispersion curves for the essentially antisymmetric (fifth) and essentially symmetric (sixth) families of modes are computed for ST-cut quartz SAW delay line and compared with the curves obtained from the classical method. The two sets of curves match very well (Fig. 2-5). To demonstrate that the approximation method is applicable directly to the analysis of SAW delay lines, we use it to calculate the frequency of plate modes in an eight wavelength thick 50 MHz delay line. Figure 6 is the prediction of the location of the plate modes. Figure 7 is the test result of the actual SAW device. The agreement between the approximation theory prediction and the experimental results are very close.

Conclusions

It is concluded that the approximation method provides an alternative way to analyze the plate modes in SAW device. For plate mode sensors application, it has several advantages over the classical method.

First, it is simpler to calculate numerically, because no iteration is needed.

Second, for sensor application mass loading is very important, and the approximate method can take the mass loading into account by specifying the face traction in the equations of motion.

Third, when we study the temperature effects, pressure effects for sensor application or the acceleration effects for accelerometer, we need to find the appropriate cuts. Using the approximation method not only can save time on the calculation, but also give us a better feeling of the individual mode we are looking for.

It is also of interest to note that this approximation method is applicable to the analysis of surface skimming bulk wave devices and reflected bulk wave devices[8][9].

Acknowledgements

The author gratefully acknowledges fruitful discussions with C.S. Lam of the AT&T Bell Laboratory and P.C.Y. Lee of Princeton University. Thanks are also extended to D. Badding of Allied-Signal Inc. for his part in obtaining the experimental data and F. Schnatter in preparing the manuscript.

References

1. Calabrese, G.S., Wohltjen, H., Manas, K.R.: "A study of SAW delay line behavior in liquids," Proceedings of the 1986 IEEE Ultrasonics symposium, pp. 607-610.

2. Lee, P.C.Y., Syngellakis, S., Hou, J.: "A two-dimensional theory for high frequency vibrations of piezoelectric crystal pates with or without electrodes," J. Appl. Phys. 61 (4), 15 February 1987, pp 1249-1262.

3. Ristic, V.M., "Principles of Acoustic Devices," Wiley-Interscience: New York, 1983, 260.

4. Auld, B.A., "Acoustic Fields and Waves in Solids," Wiley, New York, 1973.

5. Wager, R.C. "Plate mode resonances in surface wave delay lines," Proceeding of the 1973 IEEE Ultrasonics symposium, pp 372-273.

6. Wagers, R.C., "Analysis of acoustic bulk mode excitation by interdigital transducers," Applied Physics Letters, Vol. 24, No. 9, pp 401-403, May 1974.

7. White, R.M., Wicher, P.J., Wenzel, S.W., Zellers, E.T., "Plate-mode ultrasonic oscillator sensors," IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, Vo. UFFC-34, No. 2, March 1987.

8. Nikodem, Z., Lee, P.C.Y., "Approximate theory of vibration of crystal plates at high frequencies." Int. J. Solids Structures, vol. 10, 1974, pp 177-196.

9. Yen, K.H., Wagn, K.L., Kagiwada, R.S., Lau, K.F., "Interdigital transducers - a means of effect bulk wave excitation," Proc. 31st Annual Symposium on frequency control, June 1977, pp 266-270.

10. Lewis, M.F., "High frequency acoustic plate mode devices employing interdigital transducers," Electronics Letters, Vol. 17, No. 21, Oct. 1981, pp. 819-821.

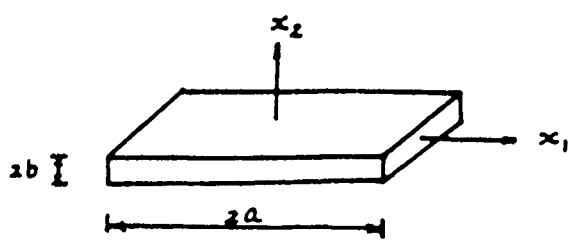


Fig. 1 Plate coordinates and boundaries.

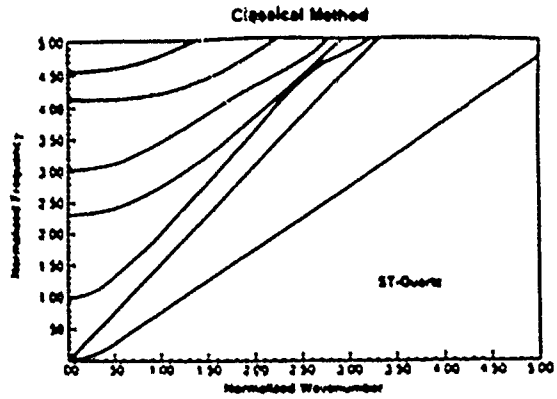


Fig. 2 Dispersion curves of the essentially antisymmetric modes from classical method for waves in x_1-direction in a ST-cut quartz plate.

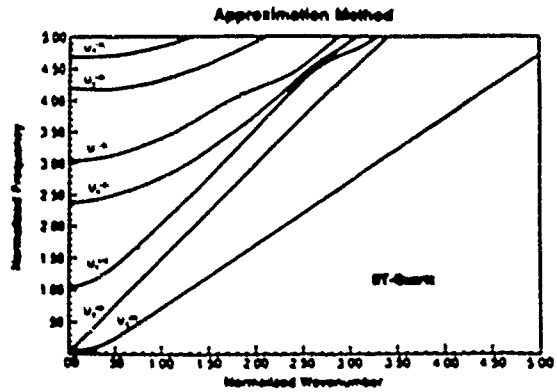


Fig. 3 Dispersion curves of the essentially antisymmetric modes from approximation method for waves in x_1-direction in a ST-cut quartz plate.

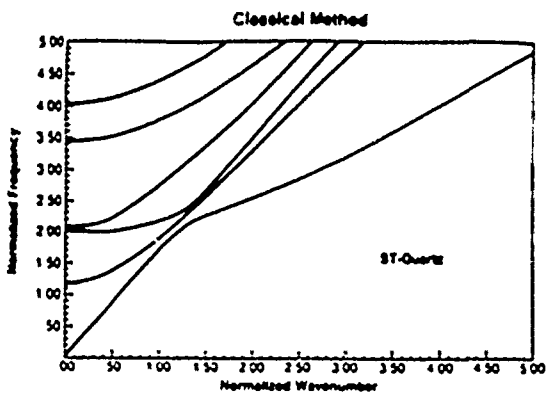


Fig. 4 Dispersion curves of the essentially symmetric modes from classical method for waves in x_1-direction in a ST-cut quartz plate.

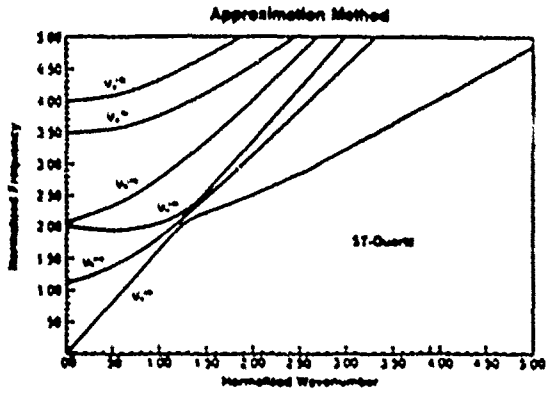


Fig. 5 Dispersion curves of the esentially symmetric modes from approximation method for waves in x_1-direction in a ST-cut quartz plate.

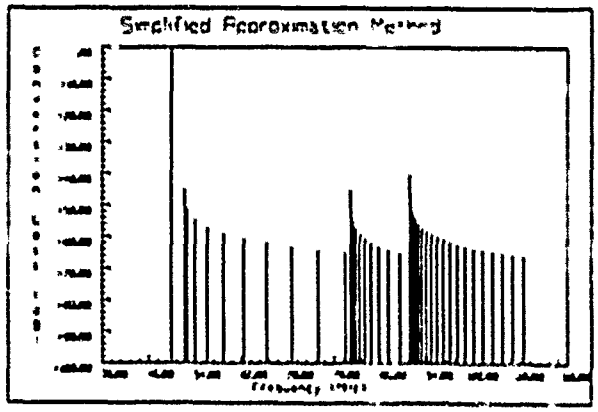


Fig. 6 Prediction of the frequencies of the plate modes in an eight wavelength thick 50 MHz SAW delay line.

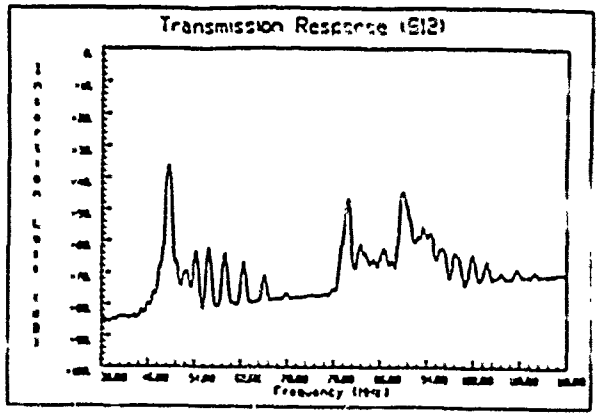


Fig. 7 Transmission response of the plate modes in an eight wavelength thick 50 MHz SAW delay line

THIN FILM RESONATOR TECHNOLOGY

K.M. Lakin, G.R. Kline, R.S. Ketcham,
A.R. Landin, W.A. Burkland, K.T. McCarron
S.D. Braymen, and S.G Burns

Microelectronics Research Center
Iowa State University
1925 Scholl Road
Ames, Iowa 50011

Abstract

This paper will present an overview of the thin film resonator technology discussing issues of materials, materials processing, device physics, and potential applications in oscillators and bandpass filters including performance characteristics achieved to date.

The thin film resonator (TFR) was initially demonstrated in the form of a composite membrane structure on a silicon substrate and later on GaAs. The minimum requirements for an acoustic resonator are reflecting surfaces, a means of exciting the acoustic wave (piezoelectric films), and a set of electrodes (photolithographically defined metal film patterns). With these basic principles in mind, a number of variations of the basic structure have been studied.

From a technological standpoint the impact of the TFR technology stems from two main features; first, the thin film approach allows fabrication of fundamental mode resonators to 2GHz and possibly beyond, and second, the resonators can be fabricated on semiconductor substrates and integrated with active devices.

The fruition of the TFR into an established technology involves solving complex problems in thin film materials growth, materials processing (including IC process compatability), device physics, and systems integration. These issues will be discussed and progress made on these problems will be cited with specific examples.

Introduction

The thin film resonator (TFR) is a high frequency analog of low frequency (primarily quartz based) electromechanical devices.[1-5] The TFR is based, however, on a processing technology that is fundamentally different from the low frequency bulk wave or conventional surface acoustic wave (SAW) technology. As a practical operating definition the TFR is a device whose structure is formed by an additive film deposition process rather than the substractive process of the mechanical extraction or removal of material from bulk crystal ingots. By growing materials on one another, the TFR technology can synthesize complex material heterostructures to fabricate electromechanical devices having unique properties. The functional value-added nature of the TFR technology and the processing techniques employed immediately lends itself to integrated circuits.

Materials processing for the TFR technology belongs to the sophisticated realm of microelectronics since thin film processes are involved and TFR fabrication is done on semiconductor substrates. In carrying the process further, we have fabricated an integrated oscillator which will be reviewed in this paper.

Compared with other acoustic based technologies, the TFR represents a new technology that is more complementary than competitive. For example, conventional acoustic devices are unlikely to compare favorably to the TFR in UHF to 5GHz bandpass filter applications. SAWs have their unique capability for transversal filtering, and the long term stability of low frequency quartz clocks is well established. In a complementary sense, the TFR will find applications in high frequency low insertion loss filtering and in fully integrated applications specific integrated circuits (ASIC) from VHF through 5GHz.

This paper will present a report on the status of the TFR technology as it exists to date in this laboratory and elsewhere when known. Efforts on theoretical modeling that this group has reported elsewhere will not be repeated here in the interest of emphasizing actual experimental results on resonators, filters, and integrated oscillators.

Basic Thin Film Resonator (TFR)

As shown in Fig. 1, the Thin Film Resonator (TFR) is composed of three basic elements common to all piezoelectric resonators; a piezoelectric dielectric for generating waves, electrodes for applying the electric fields, and finally two reflecting surfaces for establishing a standing wave.

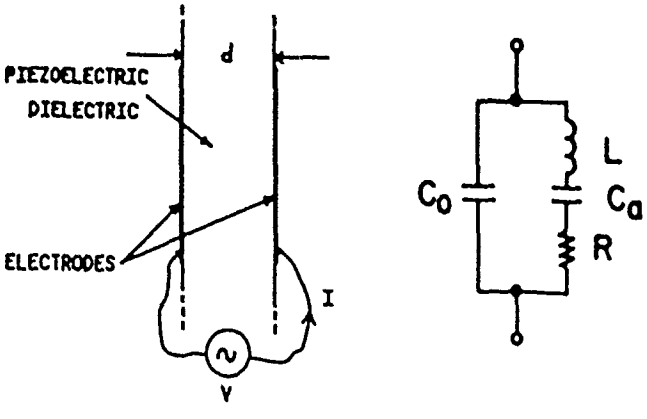


Fig. 1. Schematic representation of a piezoelectric resonator and its approximate Butterworth-Van Dyke equivalent circuit.

Resonators whose extent is large in all lateral directions compared to the thickness can be modeled analytically by a one dimensional resonator. The electrical impedance in this case is given by

$$Z = (1/j\omega C_T)\,[1 - K^2(\tan\theta)/\theta] \qquad (1a)$$

CH2427-3/87/0000-371 $1.00 C 1987 IEEE

where $\theta = kd/2$, half phase across the resonator,

or $= \frac{1}{4}\pi\,(f/fp)$

where $fp = v/2d$, parallel resonant frequency,

and $K^2 = \theta_s/\tan\theta_s$ (1b)

where $\theta_s = \frac{1}{4}\pi\,(fs/fp)$

with fs = series resonant frequency.

A circuit model often used by designers is the Butterworth-VanDyke (BVD) model shown in Fig. 1b. By comparison, its terminal impedance is

$$Z = (1/j\omega C_T)\, Zs/Zp$$

where

$$Zs = 1 - (f/fs)^2 + j(f/fs)\, Qs^{-1}$$

$$Zp = 1 - (f/fp)^2 + j(f/fp)\, Qp^{-1}$$

$$fs = \tfrac{1}{4}\pi\,(LCa)^{-\frac{1}{2}}$$

$$fp = fs\,(1 + Ca/Co)^{-\frac{1}{2}}$$

$$Qs = 2\pi fsCaR$$

$$Qp = Qs\,(1 + Ca/Co)^{\frac{1}{2}}$$

Co/Ca = traditional "capacitance ratio".

For resonators of technological interest the factor $K^2Q > 10$ and therefore the series and parallel resonances appear to occur independently as frequency is changed from fs to fp. The respective Q's around resonance can be shown to be given by

$$Qr = \frac{fr}{2}\left.\frac{d\theta}{df}\right|_{fr} \qquad (2)$$

where fr is either fs or fp. For experimental purposes it is useful to use Q, as defined above, as an implicit indicator of resonance. Thus series resonance is defined as the maximum Q near minimum impedance and parallel resonance defined as the condition of maximum Q near maximum impedance.

The TFR composite structures that have been reported are shown in Fig. 2. In Fig. 2a a silicon substrate is employed with a boron doped p^+ layer used as an etch stop for a selective chemical etch. The resultant p^+ membrane is used as a platform or substrate for subsequent film deposition of piezoelectric layers and eventual metalization for the electrodes. Since p^+ silicon is a good conductor it can act as one electrode itself with a possible backside metalization for purposes of establishing uniform current distribution in the p^+ membrane.

The structure of Fig. 2b is more closely related to a conventional free plate resonator because the p^+ membrane has been removed by a selective etch (reactive ion or plasma etching) process. This configuration is useful for materials evaluation purposes since the resonator is composed primarily of the piezoelectric film itself. However, at microwave frequencies, the finite metal thickness must be accounted for and in that sense all such resonators are considered as composites.

Top side processing techniques are illustrated in Figs. 2c,d. In Fig. 2c an undercutting etch is used through a hole in the piezoelectric to establish a free reflecting surface under the piezoelectric film.[6] In Fig. 2d the process is similar to that used in establishing cross-over air bridge metalization patterns in high frequency IC's.[7] Again the procedure establishes reflecting surfaces for the resonator. At microwave frequencies low Q bridge materials, such as SiO_2, should be avoided.

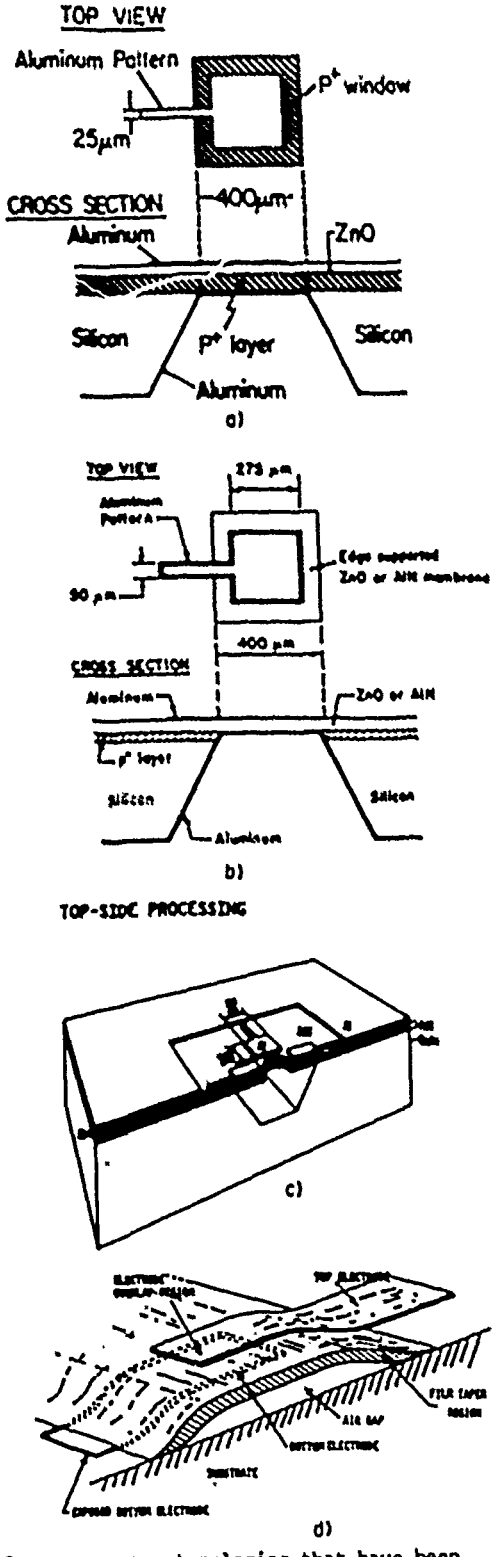


Fig. 2. Resonator topologies that have been implemented using thin film piezoelectrics and microelectronics processing techniques. a) the p^+ membrane platform, b) the piezoelectric membrane with the p^+ removed by plasma etching, c) through-the-hole under-cut chemical etch and d) a bridge under-cut etch process.

Analysis of composite resonator structures by analytical or two-port network cascading of Mason models and transmission lines reveals interesting features not found in the conventional low frequency single plate resonator.

For a simple composite structure composed of a piezoelectric film and single sided delay line, the electrical port impedance is given by,

$$Z = (1/j\omega C_T)\left[1 - \frac{K^2}{\theta}\frac{\sin\theta\,(2\tan\theta + z'\tan\theta')}{\sin\theta + \cos\theta\, z'\tan\theta'}\right] \quad (3)$$

where C_T = low frequency capacitance of film

θ = kd, phase of piezoelectric film

K^2 = stiffened coupling coefficient

$\theta = \theta/2$

z' = mechanical impedance ratio; film/delay-section

θ' = phase in delay section.

The delay section, membrane or substrate, is assumed to be attached to one side of the piezoelectric film. Using (1b) to obtain a general definition of the effective coupling coefficient, the result is

$$K_e^2 = \theta_s/\tan\theta_s \quad (4)$$

where $\theta_s = \frac{\pi}{2} f_s/f_p$.

Here fs and fp are the resonant frequencies obtained by theoretical or experimental means. When applied to (3) or experimental results the effective coupling coefficient gives a useful comparison to conventional resonators. Likewise (2) is taken as the definition of Q and respective resonant frequencies as previously discussed.

The result of applying (3) and (4) to the case of resonator composites is shown in Fig. 3. Note that, unlike a conventional resonator (zero thickness ratio for Fig. 3a) the composite exhibits both even and odd order resonances. Peaks in the effective coupling coefficient occur when the phase of the sound wave derived electric potential more closely matches the applied voltage. The slight peak in the first mode near zero thickness ratio is due to the same effect but is more subtle because the finite thickness of Si "stretches" the phase of the sound wave so that the more linearly varying portion of the wave comes closer to matching the applied voltage which is of course linearly varying across the piezoelectric.

The phase matching effect on K^2 is of technical importance because it allows useful coupling at high mode numbers. For example, the third mode composite has nearly 4 times the coupling coefficient as the n=3 conventional resonator. The reason is that the three half cycles of wave in the conventional resonator largely averages to zero when multiplied by the applied voltage's linear variation. In the composite case the K^2 reduction is instead due to the more rapid shift in phase with frequency due to the longer propagation path. Resonators having mode numbers in excess of 100 still exhibit a useful inductive impedance response around the K^2 peaks. Fig. 3b shows the similar properties for AlN/Si composites.

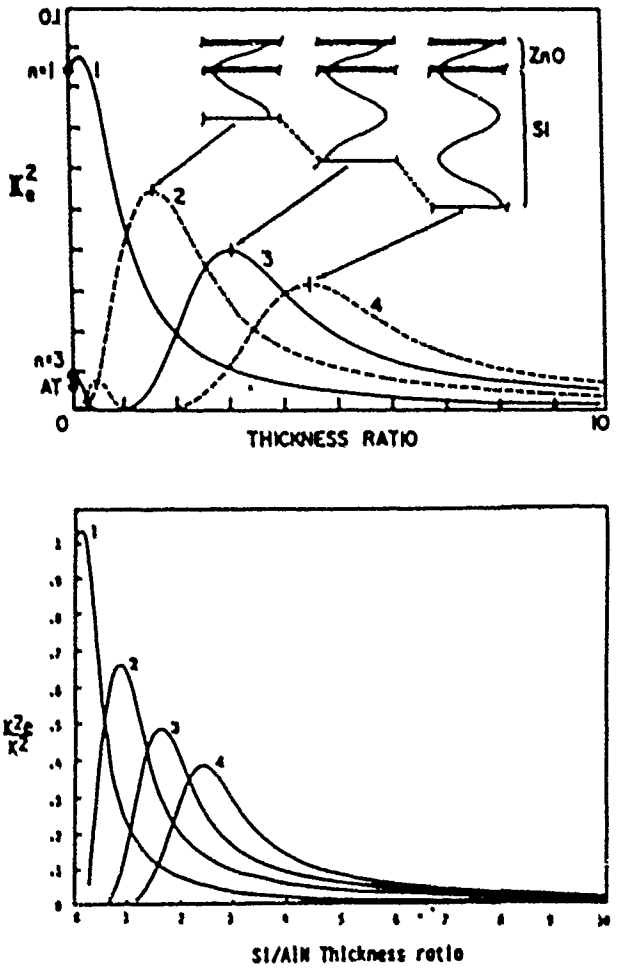


Fig. 3. Effective coupling coefficient of piezoelectric film on Si composite resonators showing even and odd order peaks in piezoelectric coupling a) for ZnO films, b) for AlN films.

Impedance responses from composite resonators using AlN as the piezoelectric are shown in Figs. 4 and 5. In these cases the resonators were energy trapped as evidenced by the response in Fig. 4. In all practical resonators the electrode area is finite and therefore the wave propagation is two dimensional. For resonators, as employed here, the electrode area is approximately 100 times the thickness and evidence of weak transverse wave generation is shown. The transverse waves are generated at the electrode edges where the discontinuity produces an uncompensated transverse motion via the Poisson effect. The transverse wave energy thus generated may propagate away from the resonator (un-trapped and lower Q) or if external propagation is cut-off or limited in the plate then the energy is confined to the electrode area and is trapped (high Q resonance).

The trapped transverse wave components generally exhibit two features. First, the resonances are weak for the higher transverse mode numbers because the displacement currents are averaged to small values by the electrodes which are larger than the wavelength of the transverse wave. (In contrast the electrodes of SAW transducer are arranged to efficiently excite transverse waves.) The lowest order transverse wave

shows the largest or primary resonance. The subsidiary resonances are unequally spaced in frequency (an-harmonic) because the velocity of the transverse wave is dispersive. A number of theoretical studies of energy trapping in simple TFR structures have been reported.[8, 9]

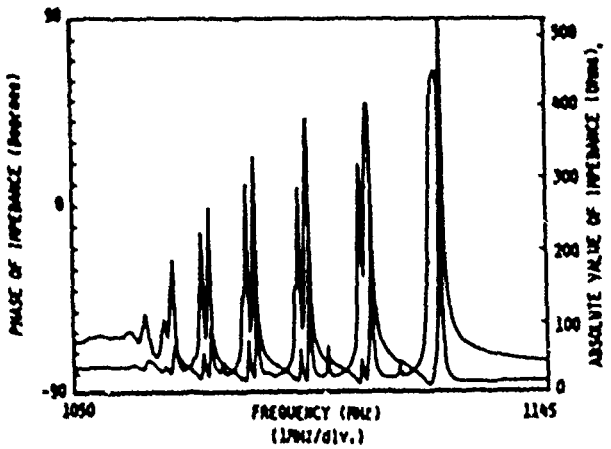


Fig. 4. Complex impedance response of an AlN composite trapped energy resonator.

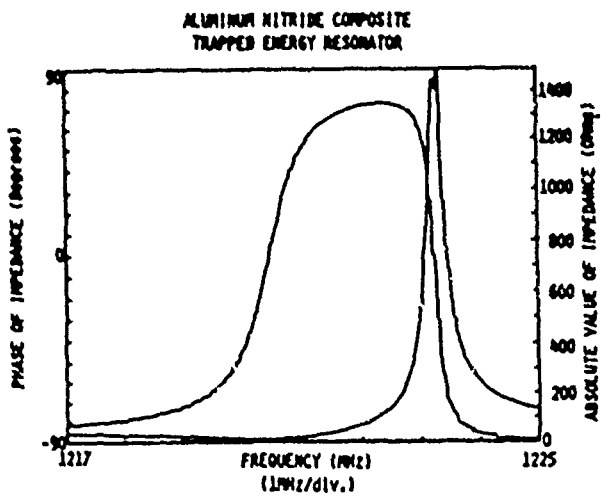


Fig. 5. Complex impedance response of an AlN composite trapped energy resonator, main resonance.

When energy is trapped within the resonator structure the Q is high as shown in Fig. 5. Here, and expanded portion of the primary resonance of a trapped energy resonator (not the one of Fig. 4) is shown. The lower Q of the series resonance (Qs=1500, Qp=5000) is due to the series resistance of the metal trace connecting the top metal electrode. The ripple on the curves is due to frequency quantization in the 8505A Network Analyzer phase locked by the 8662A frequency synthesizer.

In the composite resonator structure temperature compensation may be obtained by balancing the positive TC of p^+ silicon (8ppm/°C) against the normally negative TC of the piezoelectrics for shear waves (-25ppm/°C for AlN, -36ppm/°C for ZnO).[10, 11] The TC response for one AlN/Si composite resonator is shown in Fig. 6 along with the curve for AT quartz. The third order curve will be different for different thickness ratios of AlN to Si. At this time the individual material TC's are not known to sufficient precision to allow the calculation of the third order composite TC curve for various thickness ratios.

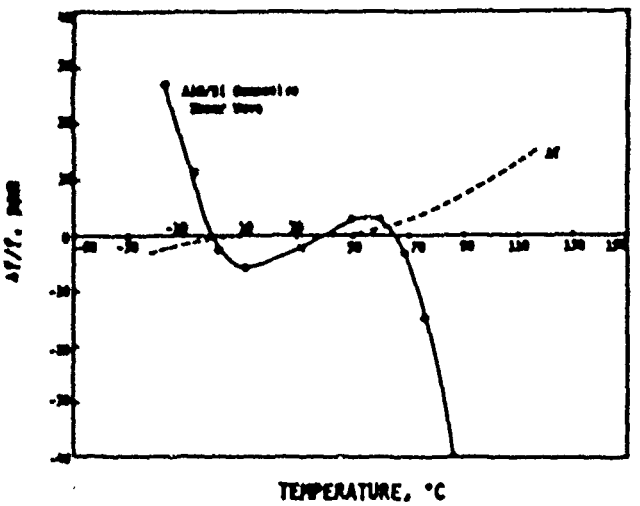


Fig. 6. Temperature variation of Δf/f for a composite resonator.

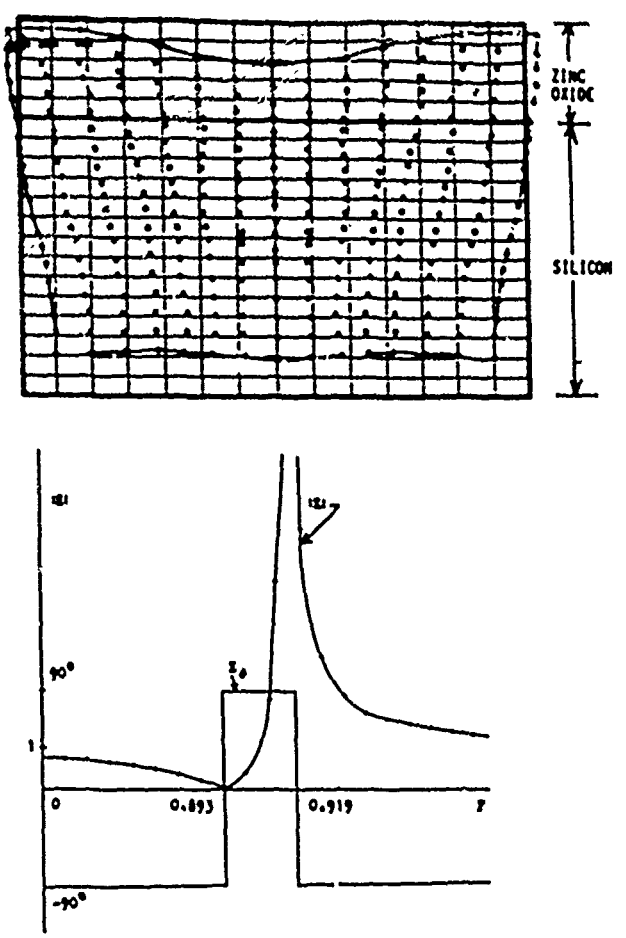


Fig. 7. Finite difference numerical analysis result for the case of a two dimensional free bar composite resonator of ZnO/Si. a) the greatly magnified mesh displacements, b) the electrical impedance normalized by the low frequency capacitative reactance.

Modeling of resonator performance takes two forms. First, in cases of large diameter-to-thickness ratios, one dimensional thickness mode analysis using Mason models and transmission line sections is adequate if it can account for sharp material boundaries. In cases where the resonator is long and narrow, such as in the monolithic crystal filter, a two dimensional analysis is adequate. Preliminary results for a two dimensional numerical analysis based upon a finite difference formulation of the arbitrarily anisotropic inhomogeneous piezoelectric boundary value problem[12] is shown in Fig. 7. This formulation is in the process of being applied to more complex structures as the computer program is moved to a mini-supercomputer (SCS-40).

Summary data on thin film resonator materials used to date are given in Table I. These results are from measured data on piezoelectric films in membrane configurations. There is an obvious need to pursue in the future higher quality piezoelectric materials such as $LiNbO_3$ in film form for increased device performance once optimal device configurations have been determined.

Table I SUMMARY DATA OF PIEZOELECTRIC MATERIALS

MATERIAL	K^2	Q	T/C
AlN	.09	5000 (1GHz)	-24 ppm/°C
AlN (c-axis inclined)	.03	4000 (400MHz)	-25 ppm/°C
AlN/Si (1/3.1)		4000 (300 MHz)	(0)
ZnO	.08	7000 (500MHz)	-59 ppm/°C
ZnO (c-axis inclined)	.03	2500 (900MHz)	-36.2 ppm/°C
ZnO/Si (1/9.4)		4000 (300 MHz)	(0)

Applications of the TFR

Primary applications of the TFR are as feedback control elements in oscillators, as resonator and acoustic coupling elements in filters, and as transducers in a wide variety of sensors.

Filters

Three general classifications of bandpass filters will be discussed here; the classical ladder filter, the monolithic crystal filter, (MCF), and the stacked crystal filter (SCF) as implemented by TFR's.

The ladder filter, shown in Fig. 8, uses resonators as loop or branch coupling elements to control the frequency response. In this mode the crystal resonator is treated as a "black box" having an equivalent circuit often given by the BVD model. Whereas the classic ladder has only pure series or parallel resonator branches, the crystal resonator exhibits both branches. Nevertheless, synthesis techniques are well established.

Implementation of the ladder using the TFR[13] has some potential in that each resonator may be optimized and because there is a possibility for integration of resonators onto a common membrane. Drawbacks might be large chip area requirements and parasitic effects.

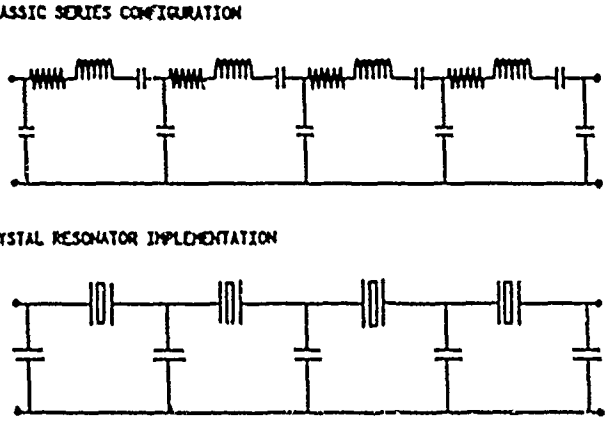


Fig. 8. Schematic of ladder filter for potential TFR implementation.

The monolithic crystal filter, Fig. 9, is well established as a low frequency (up to 200 MHz) device in quartz, $LiTaO_3$ and to a lesser extent in other materials. Design rules for shear wave plate wave devices are well established even for multipole devices.

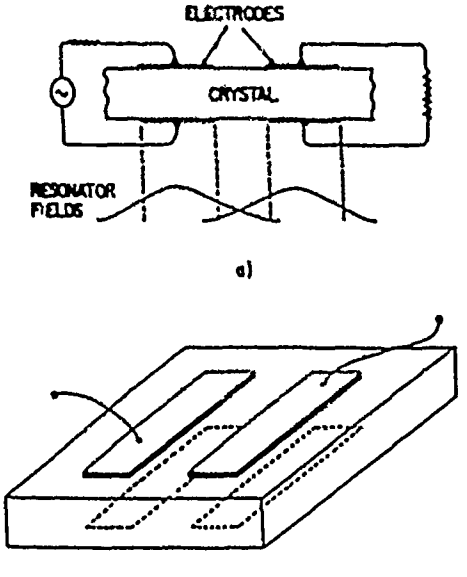


Fig. 9. Classical monolithic crystal filter employing transverse wave coupling of adjacent closely spaced resonators.

In implementing the MCF using the TFR technology,[14] the device shows great promise for the following reasons; only a single crystal film is required, the electrodes required are well within the lithography requirements of current microwave active device fabrication, and high device performance has been exhibited as reported here. However, design rules have not been established because the Lamb/

flexure plate wave nature of the composite is more difficult to predict than the low frequency shear waves, the long narrow electrodes may present some problems with resistance, and finite feature effects.

Experimental results are shown in Figs. 10-12 for a variety of MCF thin film devices. The ZnO composite MCF in Fig. 10 exhibits a modest insertion loss of 2.9dB, bandwidth of 6MHz and center frequency near 900MHz in a 50 ohm measuring system. In a 500 ohm environment the filter response shows a double peak due to the unloading of the two resonators by the increased source and load resistances. (Calibration errors account for the slight gain.) The proper resonator acoustic coupling has not been obtained in this device for the 50 ohm environment but gives some idea of device potential.

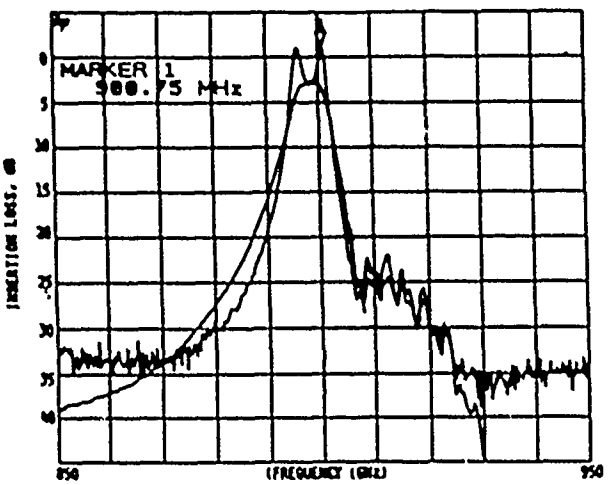


Fig. 10. Response of a ZnO composite monolithic filter composed of two coupled resonators. The insertion loss is less than 2.9 dB and the bandwidth is 6 MHz in the 50 ohm measuring system. In a 500 ohm measuring system the resonators become under-loaded and two peaked responses near zero dB are observed.

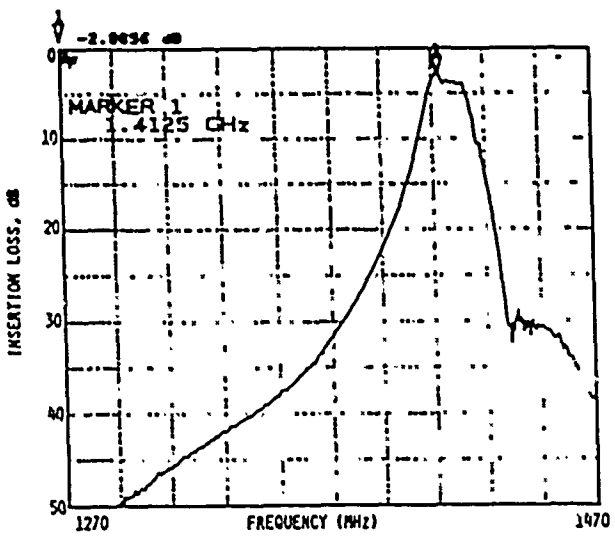


Fig. 11. Response of an AlN composite monolithic filter. The insertion loss is less than 2.9 dB and the bandwidth is 15 MHz.

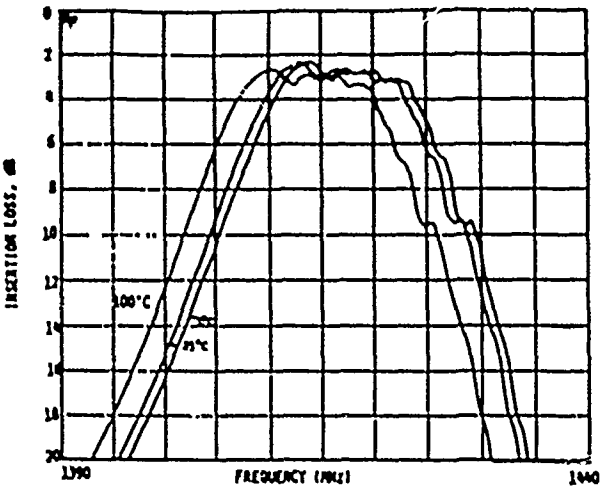


Fig. 12. Response of the filter in Fig. 11 on an expanded scale showing effect of temperature. The temperature variation is -28 ppm/°C.

In Fig. 11 is shown an AlN composite resonator of somewhat higher frequency near one of the GPS channels. Insertion loss is less than 2.9dB and the bandwidth is approximately 15MHz. The temperature response of this filter is shown in Fig. 12 indicating -28ppm/°C and approximately 15MHz of usable bandwidth.

STACKED CRYSTAL FILTER

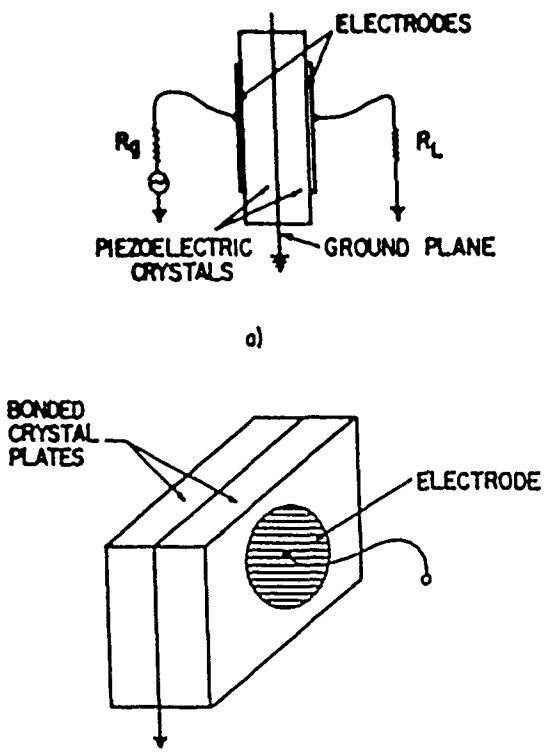


Fig. 13. Conceptual drawing of a stacked crystal filter composed of three metal electrodes (one common ground plane) and two piezoelectric films.

The thin film stacked crystal filter, shown in Fig. 13, is an out-growth of the low frequency bulk wave work on the SCF by Ballato and Lukaszek.[15-20] The SCF consists of two tightly coupled piezoelectric resonators separated by a common ground plane with one radiating directly into the other. From one dimensional Mason models an equivalent circuit can be derived that gives accurate results near center frequency.[21]

From the standpoint of implementing the SCF by the TFR process some features should be noted. The SCF requires two high quality piezoelectric films, one grown on the other with a separating ground plane. At microwave frequencies, this must be viewed as a 5 layer system; 3 metal electrodes and 2 piezoelectric films. From a fabrication standpoint, the SCF is the most complex of the three filter configurations described. However, the SCF takes less chip area because of its vertical construction and is readily adaptable to 50 ohm system requirements.

The equivalent circuit for the SCF near the passband frequencies is shown in Fig. 14 along with a suggested circuit symbol for the device. Series connection of the SCF results in an ideal form of the ladder filter that has only series resonance branches.

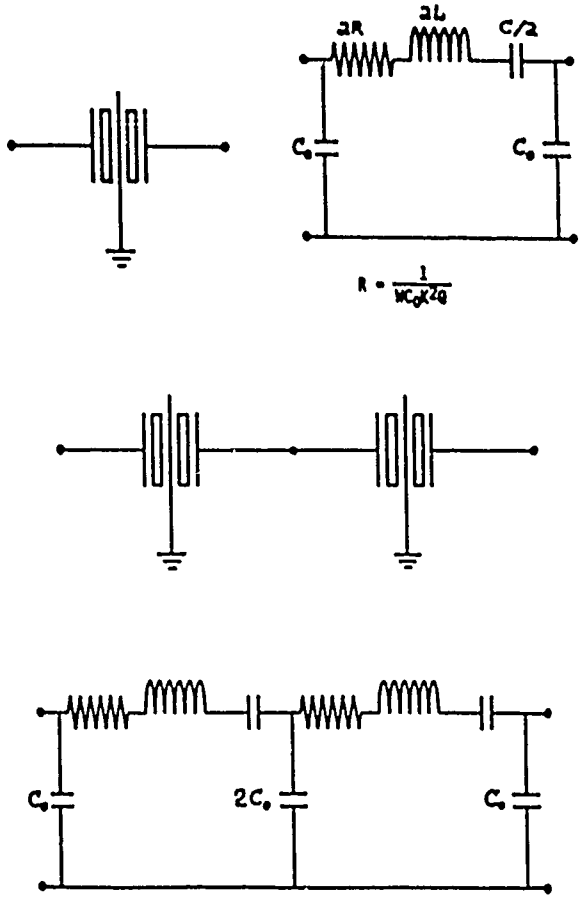


Fig. 14. Symbolic representations and near-in equivalent circuit for a stacked crystal filter. Series connected filters implement ladder filter sections.

Calculated SCF responses using Mason modeling, are shown in Fig. 15 for 1, 2, and 3 sections connected in series electrically. The steep near-in skirt selectively is due to the ground plane effect of the center metalization which electrically isolates input and output ports for frequencies off resonance.

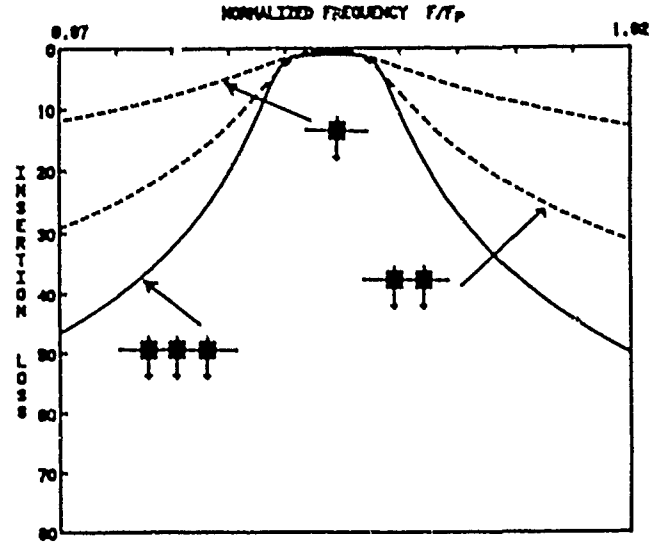


Fig. 15. Theoretical response of stacked crystal filters for a single filter and series connected filters.

Experimental results for an AlN composite SCF are shown in Fig. 16 having an insertion loss of less than 1.9dB in a 50 ohm system and less than 1dB in an optimized source and load impedance. A bandwidth of 20 MHz was obtained in this case.

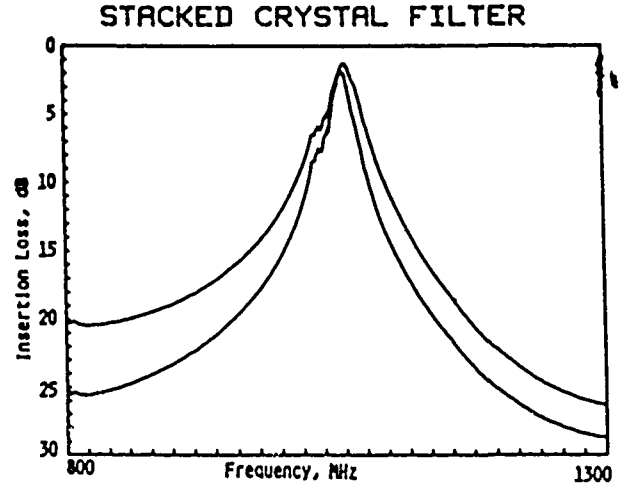


Fig. 16. Experimental results for an AlN composite stacked crystal filter. The insertion loss is less than 1.9 dB in a 50 ohm system, less than 1 dB in an optimized source and load, and with bandwidth of 20 MHz.

A summary of TFR filters implemented to date is shown in Table II. In the future improvements in films and device processing should yield filters with wider bandwidths and insertion losses less than 1dB for both the MCF and SCF configurations.

Table II SUMMARY DATA ON FILTERS

TYPE	INSERTION LOSS	CENTER FREQUENCY	BANDWITH
Monolithic			
ZnO	2.2 dB (50 ohm)	0.9 GHz	0.6 MHz
	0 dB (500 ohm)	0.9 GHz	high ripple
AlN	2.9 dB (50 ohm)	1.4 GHz	15 MHz
Stacked Crystal			
AlN	1.9 dB (50 ohm)	1.1 GHz	21 MHz
AlN	13 dB	1 GHz	50 MHz

Oscillators

Resonators are used in oscillator circuits to control both the average frequency and line width. Resonator control of the average frequency depends on the resonator controlled feedback circuit phase characteristic and upon temperature and aging effects. The resonator controls oscillator line width through its steep phase response as determined by device Q. As the instantaneous phase of the amplifier fluctuates in time the operating point for the Barkhausen phase condition moves along the phase line for the feedback element H. Recent high performance oscillator studies[22, 23] have shown evidence that the acoustic elements may contribute noise as well.

Use of the TFR in oscillator control has been demonstrated in a number of configurations.[24-26]

By combining TFR and microelectronics technology, we have demonstrated an integrated oscillator composed of a 2GHz bipolar transistor and TFR fabricated at the MRC,[27] Fig. 17. Preliminary noise measurements indicate better than -95 dBc/Hz measured at a 1KHz offset.

Materials Processing

Materials processing plays a key role in the development of the TFR to an extent as large or larger than that of device physics and modeling. In many cases the devices to be realized are relatively simple in concept but require techniques used for IC manufacturing. In addition, some specialized processes have been developed and made compatible with IC processing.

The choice of device topology or configuration must be a tradeoff between design concept and what can actually be realized in fabrication. Many fundamental and practical problems remain in TFR materials processing in areas of film growth, patterning, and IC process compatibility.

Film Growth

High quality piezoelectric films are the backbone of the TFR technology. This quality is generally macroscopically measured by the piezoelectric coupling coefficient, K^2, and mechanical loss as determined by Q. For filter applications the K^2Q product is a composite measure of film quality since the resonator series loss is given by $R = X/(K^2Q)$ where X is the reactance of the geometrical capacitance.

Material quality must be considered at two levels. First, what is the quality of the single crystal material and secondly, how well is that realized in film form. Clearly, epitaxial films would be the most desirable in approaching bulk film quality. In this regard it is important to note that films produced for bulk-wave transduction or for SAW delay lines may not be of sufficient quality. A low Q film as satisfactorily employed in a transducer, having a low device loaded Q, may be worthless in a resonator where the device Q is that of the basic material. Similarly, SAW transversal filters may allow insertion-loss/bandwidth tradeoffs that could mask poor material performance unacceptable in low insertion loss TFR bandpass filters.

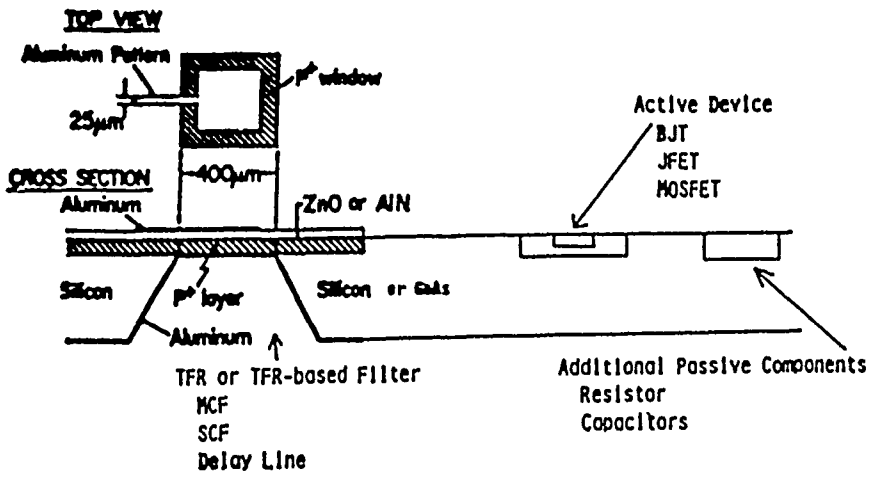


Fig. 17. Thin film resonator integrated oscillator (TFRIO) concept. Noise response of -95 dBc/Hz at 1 KHz was obtained in a preliminary oscillator running near 300 MHz.

In the work reported here, piezoelectric films of ZnO and AlN were obtained by sputter deposition using planar magnetron systems with RF or dc excitation.[28-30] The most sensitive material parameter is the coupling coefficient K^2. Since x-rays cannot distinguish between centrosymmetric and non-centrosymmetric crystal systems the net piezoelectric effect must be measured, usually in a resonator configuration. Excessively twinned crystals (c-axes oppositely directed among the various crystallites or domains) average the piezoelectricity to zero yet Q may still be substantial if the film is dense.

Although well controlled sputtering systems can produce usable films, it would certainly be desirable to synthesize epitaxial films by other techniques such as chemical vapor deposition, molecular beam epitaxy, and new techniques like chemical beam epitaxy.

Patterning

Resonator based devices must be fabricated such that electrodes and reflecting surfaces are provided during the processing. Patterning of the metal electrodes is done by straight forward photolithography as with SAW devices and IC's. The piezoelectric films of ZnO may be etched by wet chemicals (HCl for example) or by dry plasma reactive ion etching (RIE) using hydrogen iodide, HI.[31] Aluminum nitride is not easily chemically etched and is a highly refractory material. However, in thin film form it has been etched by wet chemicals to allow resonator top-side processing in GaAs.[6] A satisfactory RIE process that is selective to AlN has not been found.

IC Process Compatibility

Although there are many applications for TFR based devices outside of the IC active device chip it is still useful to seek those device fabrication processes that are compatible with low cost techniques developed for the IC industry.

Step-height metalization coverage is a problem for thick (low frequency) films on IC substrates that must connect to the substrate. Also, the thick piezoelectric must not cover the entire wafer and preclude electrical contact to other devices. Using film etches that leave a tapered edge is one solution that is employed.

Using IC processing techniques we have been able to fabricate 2GHz Si bipolar transistors and TFR's onto common substrates to demonstrate an integrated oscillator running near 300MHz.[27] More complex circuits and processes are now under development with the goal of obtaining integrated amplifiers, filters, oscillators and other system components on a common die.

Systems Integration

A difficulty in the insertion of TFR's into systems is the fact that comparable devices are not now employed and therefore, one-for-one replacement possibilities are rare. Crystal resonators are generally low frequency and systems are accordingly so designed even though operation at higher frequencies may be desirable. As the feasibility of TFR devices become apparent and the technology more established, applications specific IC's (ASIC), employing more extensive integration, will find increased demand and a new design methodology will evolve around the TFR just as it did in part for SAW devices. The long range goal will be to obtain ASIC's of the form shown in Fig. 18 implying a mix of the new TFR technology and the more established high frequency IC technology. The integrated oscillator and one of the filters discussed in this paper represent less than 1/500 the area of the die shown in Fig. 18.

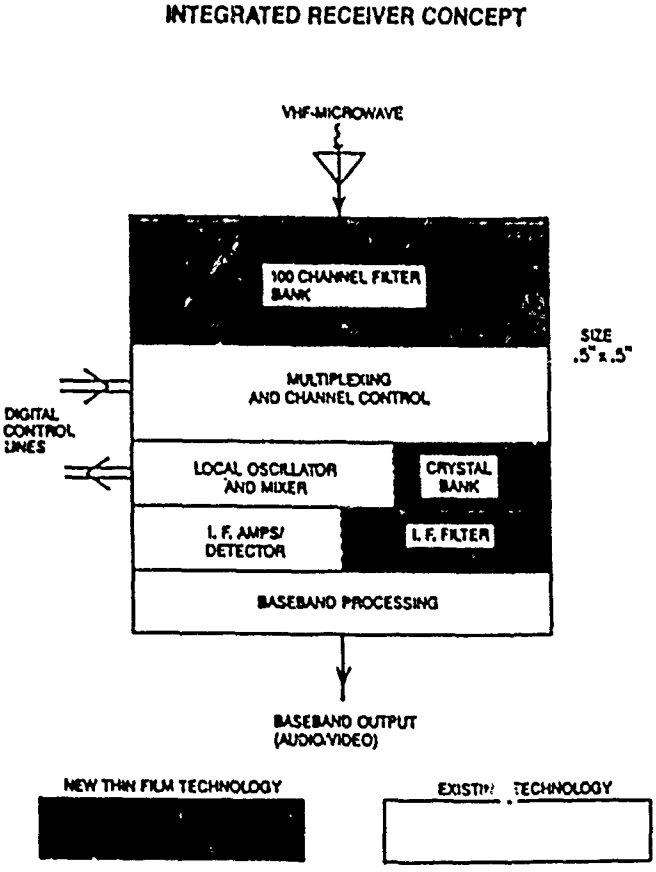


Fig. 18. Systems integration of TFR and active devices to implement an integrated receiver as a schematic example of an applications specific integrated circuit. An integrated oscillator and TFR filter, as reported in this paper, would occupy less than 1/500 the chip area shown.

Prior to eventual systems integration it will be necessary to perfect materials processing and to conduct further benchmark feasibility studies of TFR devices such as the integrated oscillator and the preliminary bandpass filters described in this paper.

Summary

This paper has presented an overview and progress report of the development of the thin film resonator based technology. High Q trapped energy resonators, low insertion loss filters, and an integrated oscillator have been demonstrated and reported in this paper. The future of the TFR technology appears greatest in the area of applications specific IC's as applied to communications, telemetry, sensors, and other high frequency systems. The technology will continue to develop with emphasis on materials processing, device physics and modeling, and systems integration benchmark studies.

References

1. K.M. Lakin and J.S. Wang, "Acoustic Bulk Wave Composite Resonators", Applied Physics Letters, Vol. 39 (3), Feb. 1, 1981.

2. T.W. Grudkowski, J.F. Black, T.M. Reeder, D.E. Cullen and R.A. Wagner, "Fundamental Mode UHF/VHF Miniature Acoustic Resonators and Filters", Applied Physics Letters, Vol. 37 (11), 1980, pp. 993-995.

3. K.M. Lakin, J.S. Wang, G.R. Kline, A.R. Landin, Y.Y. Chen, and J.D. Hunt, "Thin Film Resonators and Filters:, IEEE Ultrasonics Symposium Proceedings, 1982, pp. 466-475.

4. T. Shiosaki, T. Fukuichi, M. Tokuda, and A. Kawabata, "Temperature Compensated High Coupling and High Quality Factor ZnO/SiO_2 Bulk Wave Resonators on High Resistance Substrates", IEEE Ultrasonics Symposium Proceedings, 1984, pp. 405-410.

5. K. Nakamura, Y. Ohashi, and H. Shimizu, "UHF Bulk-Acoustic-Wave Filters Utilizing Thin ZnO/SiO_2 Diaphragms on Silicon", Japanese Journal of Applied Physics, Vol. 25 (3), Mar. 1985, pp. 371-375.

6. G.R. Kline and K.M. Lakin, "1.0 GHz Thin Film Bulk Acoustic Wave Resonator on GaAs", Applied Physics Letters, Vol. 43 (8), Oct. 15, 1983, pp. 750-751.

7. H. Satoh, Y. Ebata, H. Suzuki, and C. Narahara, "An Air Gap Type Piezoelectric Composite Thin Film Resonator", Proceedings of the 39th Annual Symposium on Frequency Control, 1985, pp. 361-366.

8. H.F. Tiersten and D.S. Stevens, "An Analysis of Thickness-Extensional Trapped Energy Resonators With Rectangular Electrodes in the Zinc-Oxide Thin Film on Silicon Configuration", Proceedings of the 37th Annual Symposium on Frequency Control, 1983, pp. 325-336.

9. R.F. Milsom, "Two Dimensional Theory of Thin Film ZnO Resonators on Silicon", 1982 Ultrasonics Symposium Proceedings, pp. 484-488.

10. J.S. Wang and K.M. Lakin, "Low-Temperature Coefficient Bulk Acoustic Wave Composite Resonators", Applied Physics Letters, Vol. 40 (4), Feb. 15, 1982, pp. 308-310.

11. J.S. Wang, A.R. Landin, and K.M. Lakin, "Low Temperature Coefficient Shear Wave Thin Films for Composite Resonators and Filters", 1983 Ultrasonics Symposium Proceedings, pp. 491-494.

12. K.M. Lakin, "Analysis of Composite Resonator Geometries", Proceedings of the 37th Annual Symposium of Frequency Control, 1983, pp. 320-324.

13. M.M. Driscoll, R.A. Moore, J.F. Rosenbaum, S.W. Krischnaswamy, J.R. Szedon, "Recent Advances in Monolithic Film Resonator Technology", 1986 IEEE Ultrasonics Symposium Proceedings, pp. 365-369.

14. K.M. Lakin, G.R. Kline, R.S. Ketcham and S.G. Burns, "Thin Film Resonator Based Low Insertion Loss Filters", 1986 IEEE Ultrasonics Symposium Proceedings, pp. 371-374.

15. A. Balloto and T. Lukaszek, "A Novel Frequency Selective Device: The Stacked Crystal Filter", Proceedings of the 27th Annual Frequency Control Symposium, June 1973, pp. 262-269.

16. A. Ballato "Transmission-Line Analogs for Stacked Piezoelectric Crystal Devices", Proceedings of the 26th Annual Frequency Control Symposium, June 1972, pp. 86-91.

17. A. Ballato, H.L. Bertoni, and T. Tamir, "Systematic Design of Stacked-Crystal Filters by Microwave Network Methods", IEEE Transactions Vol. MTT-22, Jan. 1974, pp. 14-25.

18. A. Ballato and T. Lukaszek, "Stacked Crystal Filters", Proceedings IEEE, Vol. 51, Oct. 1973, pp. 1495-1496.

19. A. Ballato, "The Stacked Crystal Filter", IEEE International Symposium on Circuits and Systems, Apr. 1975, pp. 301-304.

20. A. Ballato and T. Lukaszek, "Distributed Network Modeling of Bulk Waves in Crystal Plates and Stacks", Technical Report ECOM-4311, May 1975, U.S. Army Electronics Command, Fort Monmouth, N.J. 07703.

21. K.M. Lakin, "Equivalent Circuit Modeling of Stacked Crystal Filters", Proceedings of the 35th Annual Frequency Control Symposium, 1981, pp. 257-262.

22. H.L. Salvo, M. Gottlieb, and B.R. McAvoy, "Shear Mode Transducers for High Q Bulk Microwave Resonators", Presented at the 41st Annual Frequency Control Symposium., 1987.

23. T.E. Parker, "Characteristics and Sources of Phase Noise in Stable Oscillators", Presented at the 41st Annual Frequency Control Symposium, 1987.

24. S.G. Burns and R.S. Ketcham, "Performance of Fundamental-Mode UHF Oscillators Using Bulk-Acoustic-Wave Resonators", Proceedings of the 38th Annual Symposium on Frequency Control 1984, pp. 266-270.

25. S.G. Burns and R.S. Ketcham, "Fundamental-Mode Pierce Oscillators Utilizing Bulk-Acoustic-Wave Resonators in the 250-300 MHz Range", IEEE Transactions on MTT, Vol. MTT-32, No. 12, Dec. 1984, pp. 1688-1671.

26. M.M. Driscoll, S.V. Krishnaswamy, R.A. Moore, and J.R. Szedon, "UHF Film Resonator Evaluation and Resonator-Controlled Oscillator and Filter Design Using Computer-Aided Design Techniques", 1985 IEEE MTT-S International Symposium.

27. W.A. Burkland, A.R. Landin, G.R. Kline, and R.S. Ketcham, "A Thin Film Bulk Acoustic Wave Resonator Controlled Integrated VHF Oscillator on Silicon", to be submitted for publication.

28. J.S. Wang and K.M. Lakin, "C-axis Inclined ZnO Piezoelectric Shear Wave Films", Applied Physics Letters, Vol. 42(4), Feb. 15, 1983, pp. 352-354.

29. J.S. Wang and K.M. Lakin, "Sputtered AlN Films for Bulk-Acoustic-Wave Devices", 1981 Ultrasonics Symposium Proceedings, pp. 502-505

30. K.M. Lakin, J.S. Wang, and A.R. Landin, "Aluminum Nitride Thin Film and Composite Bulk Wave Resonators", Proceedings of the 36th Annual Symposium on Frequency Control, 1982, pp. 517-524.

31. J.S. Wang, Y.Y. Chen, and K.M. Lakin, "Reactive Ion Etching of Piezoelectric Films for Acoustic Wave Devices", 1982 Ultrasonics Symposium Proceedings, pp. 346-349

UHF OSCILLATOR PERFORMANCE USING THIN FILM RESONATOR BASED TOPOLOGIES

S.G. Burns, G.J. Kline and K.M. Lakin
Microelectronics Research Center
Iowa State University
1925 Scholl Road
Ames, Iowa 50011

Abstract

Thin-film bulk-acoustic-wave resonators (TFR) synthesized using either ZnO or AlN for piezoelectric transduction have been used as the basis for low insertion loss UHF and L-band bandpass filters. In this paper, we report the design, implementation, and characterization of UHF and L-band generalized Pierce oscillators using two-pole TFR based filter configurations and silicon delay lines in the feedback circuit as a precursor study to direct integration. Circuit analysis and the design procedure use discrete element models implemented with SPICE 2G. Onset of oscillation was computed using the Barkhausen criteria and measured data for both commercial amplifier modules and the TFR filters and delay lines. Experimental measurements show phase noise performance from -90 dBc/Hz to -100 dBc/Hz at 1 kHz with carrier frequencies of 1200 MHz.

Introduction

Thin-film bulk-acoustic wave resonator (TFR) technology[1,2,3] has been used to implement high performance fractional bandwidth filters at L-band.[4,5] These filters have been fabricated using both sputtered AlN and ZnO piezoelectric films with either Si or semi-insulating GaAs[6] substrates. Devices are defined by selective crystallographic-plane chemical etching.

The TFR has also been used as the frequency-control element in a UHF Pierce hybrid oscillator in very much the same fashion that BAW quartz and SAW devices are used.[7,8] This sputtered piezoelectric film technology, used to fabricate the TFR, is directly integrable with active devices, such as the BJT, on semiconductor substrates yielding in this case, an integrated oscillator, but also applicable to the synthesis of a variety of UHF and microwave circuits.

In this paper, we describe the use of monolithic TFR technology-based filters and delay lines as the frequency-control element in generalized Pierce oscillators operating to L-band. The gain block consists of commercially available thin-film hybrid BJT amplifiers. Design and analysis of oscillator performance uses linear equivalent circuits for both the amplifier and filter feedback element. The linear filter phase characteristic can be exploited to provide VCO operation by adjusting total feedback network phase. Noise performance and analysis is also presented.

In addition, we present preliminary data for a monolithic oscillator consisting of a composite ZnO TFR integrated on the same silicon substrate with a BJT in a Pierce configuration.

Oscillator Modeling

The general representation of an oscillator and its operating conditions are shown in Fig. 1. The amplifier or gain element is represented by G and the feedback element by H.

Oscillation occurs under the well known Barkhausen criteria

$$|G|\,|H| = 1 \tag{1a}$$

$$\phi_G + \phi_H = 0 \tag{1b}$$

which must be satisfied in the steady state. In Fig. 1(b) the phase of H, ϕ_H, is represented by a straight line near the operating point and the phase of G, ϕ_G, by a band of uncertainty caused by noise fluctuations. Since ϕ_H is a function of frequency and ϕ_G is a function of time ϕ_H must also be a function of time as suggested by,

$$\phi_G(t) + \phi_H(f(t)) = 0 \tag{2}$$

The implication is that frequency is a function of time, and gives rise to the well known finite oscillator line width. Low noise is achieved by a steep ϕ_H versus frequency characteristic such that operating point on ϕ_H is more narrowly confined in frequency.

The single-sideband phase-noise in dBc/Hz can be described by[9]

$$\frac{P_{sb}}{P_c} = 10 \log \left[N^2 \left(\frac{\alpha}{\omega_m^3 \tau^2} + \frac{GFkT}{P_c \omega_m^2 \tau^2} + \frac{\alpha}{\omega_m} + \frac{GFkT}{P_c} \right) \right] \tag{3}$$

where P_{sb} = single-sideband noise power in a 1-Hz bandwidth

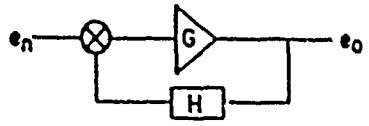
- P_c = amplifier carrier power output level
- N = frequency multiplication factor which for these oscillators is 1
- G = magnitude of the amplifier gain, approximately equal to the feedback network insertion loss
- F = amplifier noise figure
- ω_m = $2\pi f_m$ = Fourier modulation frequency, frequency offset from carrier frequency
- kT = $(1.38\times10^{-23}\ J/°K)(300°K) = 4.14\times10^{-21}$ Joules = 4.14×10^{-21} watts/Hz
- α = flicker noise constant in sec^{-1}
- τ = group delay phase slope = $\frac{1}{2\pi}\frac{d\phi_H}{df} = \frac{Q}{\pi f}$ where

  ϕ_H is the open loop phase and Q is the loaded Q of the feedback element.

Operation of an oscillator as a VCO for TCXO or other applications requires that the feedback path be altered so that an additional phase shift element can be identified or that ϕ_H (or less likely ϕ_G) can be changed. This corresponds to sliding ϕ_H in frequency in Fig. 1(b). For low noise operation ϕ_H must change to a different operating point without changing the slope, $d\phi_H/df$. In simple resonator feedback circuits the resonator can be pulled but unfortunately the resonator phase slope, (Q), is a rapid function of frequency and consequently noise performance can degrade.

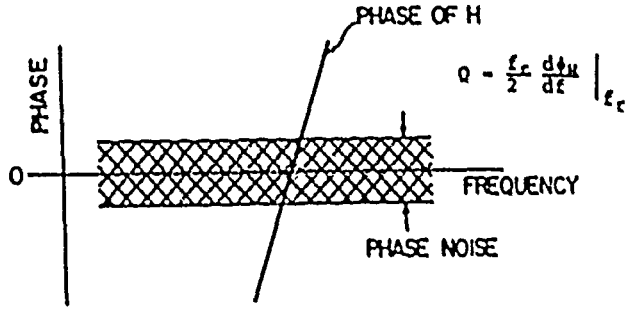
CH2427-3/87/0000-382 $1.00 C 1987 IEEE

Bandpass filters and delay lines with external phase shifters can be used to provide VCO action while maintaining low phase noise over limited frequency ranges. The purpose of the work reported here was to examine those delay-line and filter configurations that could be implemented in a total integration with active devices.



$$\phi_G + \phi_H = 0$$

(a) Feedback Oscillator Where G is the Gain Element and H is the TFR Based Filter Feedback Element



$$\phi_G(t) + \phi_H(t) = 0$$

(b) Phase Relationships

Fig. 1 General Oscillator Noise Concept

MCF Oscillator Experimental Results

A two-pole ZnO monolithic filter with the amplitude characteristic given in Fig. 2, was used as the feedback element across a Hewlett Packard HAMP-1002 BJT thin-film hybrid amplifier, Fig. 3. A variable phase shifter was utilized in the feedback loop to obtain frequency shift versus phase delay data. A buffer amplifier was used to isolate the oscillator from the instrumentation. A battery was used for the oscillator dc supply to reduce n x 60 Hz noise.

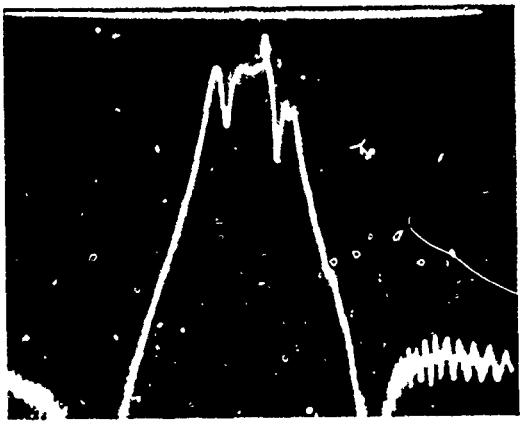


Fig. 2 Amplitude Insertion Loss Characteristics for the Candidate 1185 MHz MCF
Marker -4.9 dB IL. Total Sweep 100 MHz
Total Vertical Scale of 45 dB

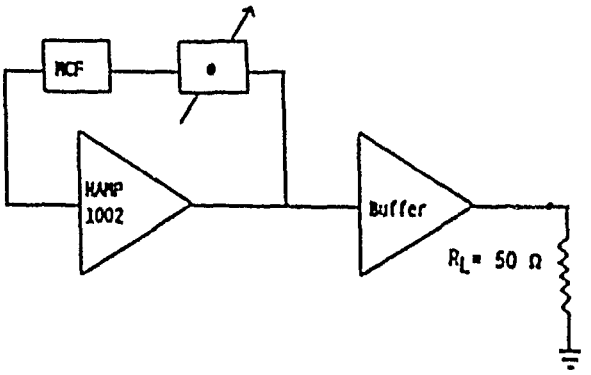
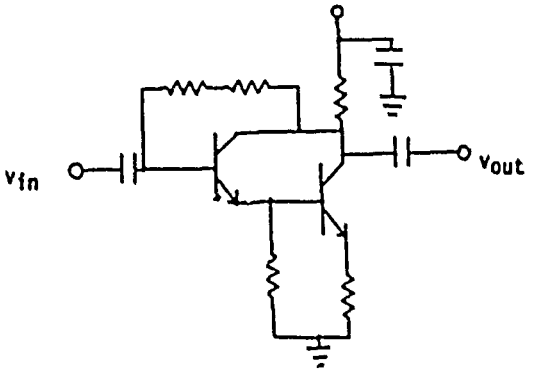
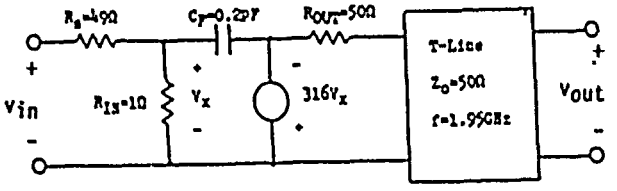


Fig. 3 Basic Oscillator Test Configuration

The HAMP-1002, Fig. 4(a), uses internal thin-film resistors for a negative feedback circuit around a Darlington pair fabricated from HXTR-7011 family BJTs. The two-port linear model, Fig. 4(b) was used to synthesize the S_{21} phase and magnitude characteristics as shown in Fig. 5. Essentially the HAMP-1002 was treated as a uniform 10 dB gain block with a linear phase characteristic for design considerations. It is also possible to model these terminal characteristics by incorporating each of the passive and active circuit elements in a SPICE simulation.



(a) HAMP-1002 BJT Amplifier



(b) Equivalent Circuit Used for SPICE Model

Fig. 4 HAMP-1002 Amplifier and Two-Port Linear Model Used for SPICE 2G Simulation

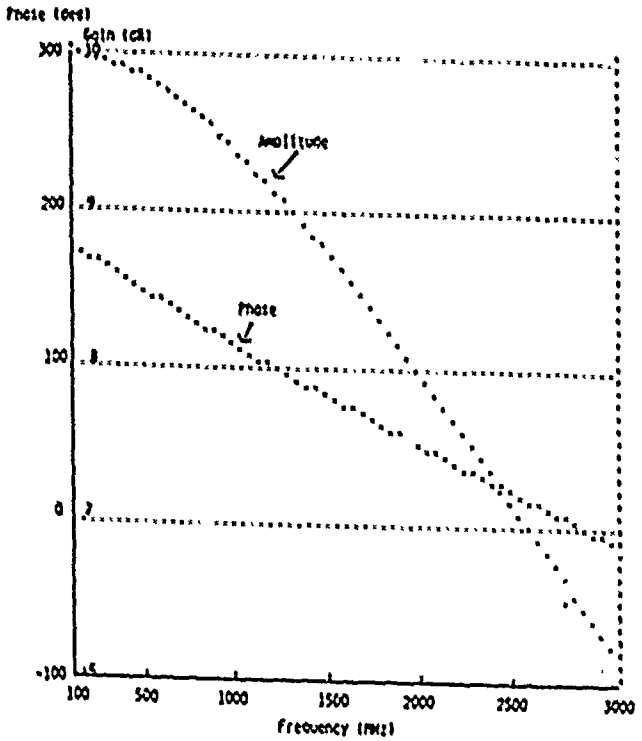
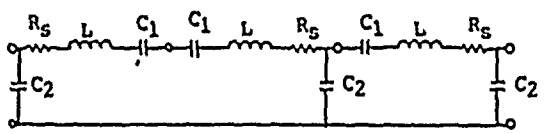


Fig. 5 HAMP-1002 Amplifier SPICE Model Characterization Using Circuit Parameters from Fig. 4(b)

There are a variety of ways to model a two-pole monolithic filter.[10,11] The best match to the bandpass magnitude and phase characteristics of the candidate filter was provided by the two series-resonant arms with three shunt coupling capacitors which is equivalent to a π-connection of three identical half-sections, Fig. 6(a). The model element values for this candidate filter were determined using standard formulae[12] along with measured two-port data. Fig. 6(b) illustrates the resultant MCF model insertion loss and phase response.

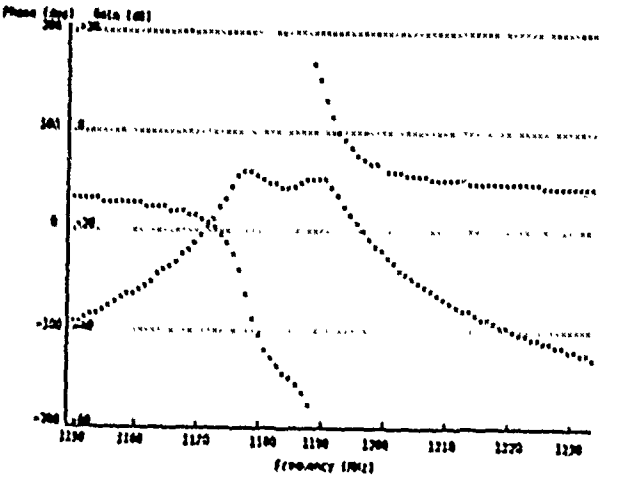
Using the HAMP 1002 model, Fig. 4 and the MCF model, Fig. 6, a closed-loop phase and gain simulation was obtained, Fig. 7, by including a variable-frequency voltage generator in series with the loop. We observe that at the frequency predicted for the onset of oscillation the loop phase shift is 0° and the loop gain exceeds unity.



C_1 = 18.26832 fF L = 1μH

C_2 = 1.2652890 pF R_S = 5Ω

(a) 2-pole MCF Model Used for SPICE



(b) MCF Model Simulation

Fig. 6 Model and SPICE Simulation for the MCF Characteristics Measured as Shown in Fig. 2. Used for the Oscillator Feedback Element.

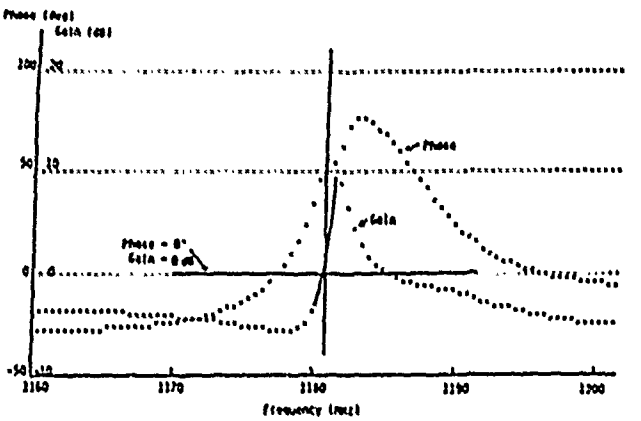
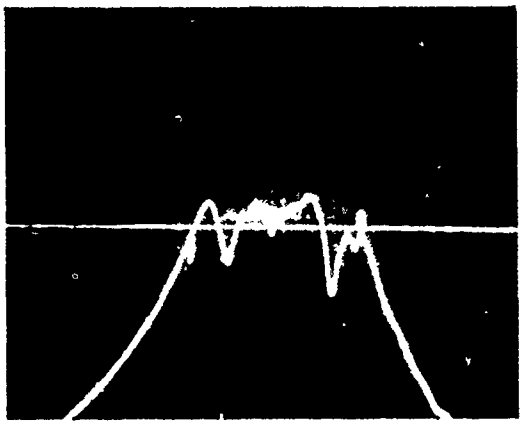


Fig. 7 Closed Loop Phase and Gain Simulation of the HAMP-1002 and the MCF in the Feedback Loop Oscillatory Behavior Predicted Where Eqns. (1a) and (1b) are Satisfied

The measured open-loop gain and phase characteristic for an MCF controlled oscillator is shown in Fig. 8. Using the center set of markers we see the frequency of oscillation occurs at 1185 MHz where there is a gain margin of 1.6 dB. The Barkhausen criteria of $|G|\,|H| \geq 0$ and $\phi_G + \phi_H = 0°$ are satisfied. We can use the measured phase slope of 23°/MHz corresponding to τ = 64 nsec in estimating the phase noise performance from Eqn. (3). Using both the measured and published specifications for the HAMP 1002 where N = 1, F = 5.1 dB = 3.23, $|G|$ = 9 db = 7.94, and P_C = 10.5dBm = 11.22 mW; we estimate the noise performance using Eqn. (3) at about -90 dBc/Hz at f_m = 1 kHz. This phase noise performance is shown in Fig. 9. The $1/f^3$ slope is to be expected in this frequency regime by considering the first term in Eqn. (3) to be dominant with a curve fit value for the flicker noise constant of $\alpha = 10^{-12}$ sec^{-1} which is a reasonable value for the system.

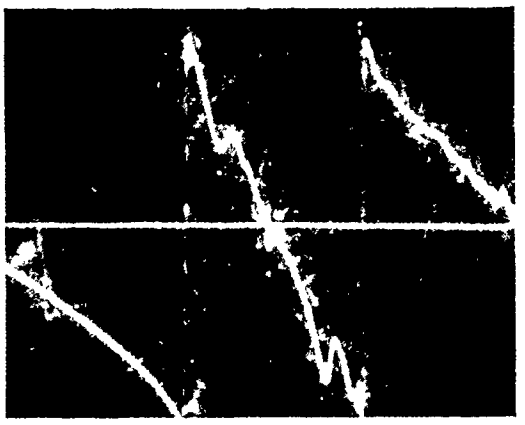
The -130 dBc/Hz pedestal to 500 kHz is an artifact of using the HP 8662A frequency synthesizer in the DCFM mode. The measured noise floor for our HP 8662A in DCFM at 1 kHz is -110 dBc/Hz with a $1/f^3$ characteristic to 20 kHz. Additional close-in phase noise measurements will require the use of the EFC phase-lock function or the use of a delay line functioning as a discriminator.

Because of the linear MCF phase characteristic, we obtained stable oscillator output with -90 dBc/Hz at 1 kHz noise performance at any frequency along the passband, Fig. 8, where $|G|\,|H| \geq 0$ dB by adjusting the variable phase shifter to compensate for MCF phase-shift and thus satisfy the closed-loop Barkhausen phase condition. For this example, we operated anywhere from about 1180 to 1190 MHz in agreement with the predicted linear circuit SPICE simulations.



(a) Amplitude Characteristic

Marker	Freq (MHz)	Gain (dB)
1	1178	-1.1
2	1185	+1.6
3	1193	-1.1



(b) Phase Characteristic

Marker	Freq (MHz)	Phase (deg)
1	1178	+171
2	1185	0
3	1193	-166

Fig. 8 HAMP-1002 in Cascade with an MCF-Measured Amplitude and Phase Characteristics

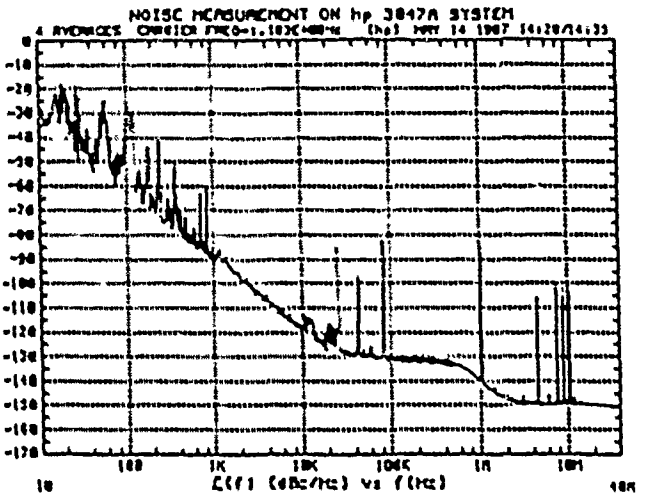


Fig. 9 Noise Spectrum of an 1185 MHz MCF-Controlled Oscillator. Noise Performance Agrees with Theoretical Predictions Suggested by Eqn. (3)

Delay Line Oscillator Experimental Results

Delay line feedback oscillators usually have sufficiently large delay that the phase condition Eqn. (1b) is modified to its equivalent,

$$\phi_H + \phi_G = n2\pi \quad , n = 0,1,2,\ldots \tag{4}$$

$$\phi_H = \phi_S + \phi_D \text{ where} \tag{5}$$
ϕ_S = phase shifter phase delay
ϕ_D = delay line phase = $2\pi f\tau_d$
τ_d = time delay.

Now the phase condition becomes

$$2\pi f_n\tau_d + \phi_S + \phi_G = n2\pi \tag{6}$$

where f_n denotes an average frequency satisfying the oscillation condition. The separation between allowed frequencies is given by

$$\Delta f_n = \Delta f \left[1 - \frac{\Delta\phi_{Sn} + \Delta\phi_{Gn}}{2\pi}\right] \tag{7}$$

where $\Delta f = 1/\tau_d$

$$\Delta\phi_{Sn} = \frac{\partial\phi_S}{\partial f}\Delta f_n$$

$$\Delta\phi_{Gn} = \frac{\partial\phi_G}{\partial f}\Delta f_n.$$

If the phase shifter and amplifier are dispersionless then

$$\Delta f_n = \Delta f = 1/\tau_d$$

and the allowed frequencies are equally spaced and given by,

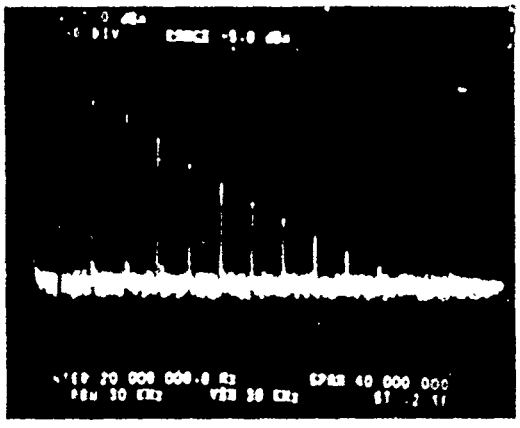
$$f_n = (n - \frac{\phi_n}{2\pi})\,\Delta f \qquad \text{and}$$

$$\phi_n = \phi_{Sn} + \phi_{Gn}$$

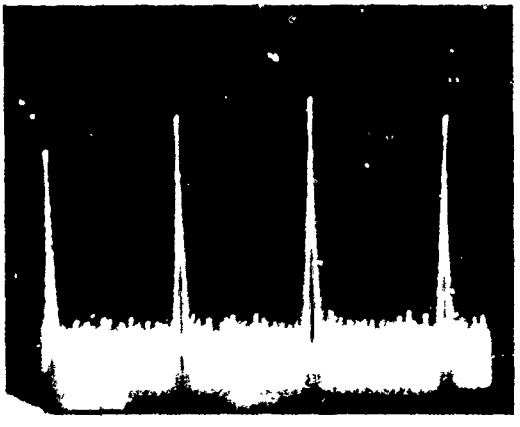
When the phase external to the delay line, ϕ_n is a multiple of 2π then f_n is an integral multiple of Δf. Ideally ϕ_n should be independent of n (or frequency).

Most delay lines have an operating bandwidth much larger than Δf and sufficiently flat so that more than one frequency f_n will meet the time average amplitude condition Eqn. (1a).

Delay lines in silicon substrates using both ZnO and AlN for piezoelectric transduction were utilized as the feedback element in an oscillator. The delay line IL required using an Aventek Model AVP-8000T rather than the HAMP 1002 to provide higher forward gain. Delays of $\tau = 0.33$ μsec yielded both single line and comb oscillator outputs depending upon the amount of additional phase shift incorporated in the feedback loop. For $\tau = 0.33$ μsec, a $\Delta f = 3$ MHz comb at frequencies between 750 and 850 MHz was obtained. Typical spectral output for a silicon delay line with ZnO piezoelectric transduction centered at 785 MHz is shown in Fig. 10. In addition as predicted by Eqn. (6), the entire comb spectrum could be shifted in frequency by using the variable phase shifter in cascade with the delay line. Noise performance of individual spectral lines was measured to be about -95 dBc/Hz at 1 kHz, however, theoretical predictions suggest that the HP 8662A DCFM noise floor was obscuring the actual results. More work is required in this area.



(a) $f_c = 785$ MHz with $\Delta f = 3$ MHz. Decreasing Amplitude Reflects the Buffer Amplifier Frequency Response



(b) $\Delta f = 3$ MHz Expanded Spectrum Plot, Full Scale 50 dB

Fig. 10 Silicon Delay Line Oscillator Comb Generator Spectrum Output ZnO Piezoelectric Transduction

Monolithic Oscillator Circuits

A ZnO composite TFR has been monolithically fabricated with a double-diffused BJT on a silicon substrate in a Pierce oscillation configuration operating at 257 MHz.[13] This concept is illustrated in Fig. 11. This work addressed the critical issues of selective-area high-concentration p^+ diffusions for membrane formation and the patterning of the resonator structure in a manner which allows connection of the TFR electrodes to the transistor circuit. In addition, the processing for the TFR must preserve the material quality of the silicon in areas to be used for active device structures. The successful fabrication of this monolithic oscillator circuit suggests the feasibility of integrating the circuits discussed in this paper.

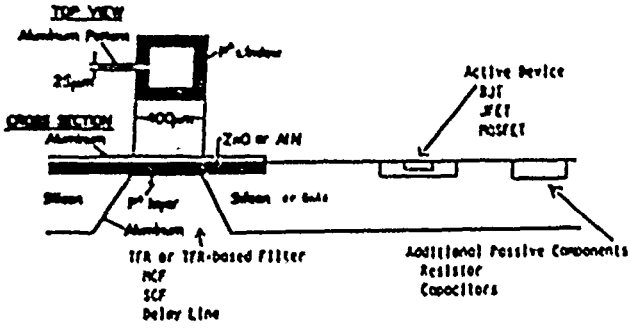


Fig. 11 Integrated Oscillator Concept Using a TFR and Active Device in a Semiconductor Substrate

Conclusions and Directions of Future Work

A series of L-band oscillators using the MCF and delay line in the feedback loop were demonstrated. The MCF and delay lines are based upon TFR and piezoelectric thin-film technology with sputtered ZnO and AlN. Onset of oscillation and phase performance were modeled using a linear SPICE 2G simulation.

Work continues in the development of integrating both the TFR-based filter and signal-processing elements along with the active devices and other passive elements for oscillators and other applications. The development and demonstration of a monolithic low-noise Pierce oscillator using a complete ZnO TFR demonstrates the potential of this technology. Development of these filters in hybrid circuit configurations as well as stand alone circuit elements will be continued in parallel with the monolithic work. We envision exploiting this RFLSI technology in communications systems and signal processing. Concommitant with this is the continued improvement in thin-film quality and compatible processing technology. Considerable progress needs to be made in modeling of the filter and oscillator with particular emphasis on non-linear behavior and anharmonic mode analysis.

References

1. K.M. Lakin and J.S. Wang, "Acoustic Bulk Wave Composite Resonators", Applied Physics Letters, Vol. 38 (3), Feb. 1, 1981.

2. T.W. Grudkowski, J.F. Black, D.E. Cullen, G.W. Drake, J.T. Orr, and R.A. Wagner, "Bulk Acoustic Wave Devices", United Technologies Research Center, Final Technical Report, Sept. 1982.

3. M.M. Driscoll, R.A. Moore, J.F. Rosenbaum, S.V. Krishaswamy, J.R. Szedon, "Recent Advances in Monolithic Film Resonator Technology", IEEE Ultrasonics Symposium Proceedings, Nov. 1986, pp. 365-369.

4. K.M. Lakin, J.S. Wang, G.R. Kline, A.R. Landin, Y.Y. Chen, and J.D. Hunt, "Thin Film Resonators and Filters", IEEE Ultrasonics Symposium Proceedings, 1983, pp. 466-475.

5. K.M. Lakin, G.R. Kline, R.S. Ketcham, and S.G. Burns, "Thin Film Resonator Based Low Insertion Loss Filters", IEEE Ultrasonics Symposium Proceedings, Nov. 1986, pp. 371-375.

6. G.R. Kline and K.M. Lakin, "1.0 GHz Thin-Film Bulk Acoustic Wave Resonators on GaAs", Applied Physics Letters, Vol. 43(8), Oct. 15, 1983.

7. S.G. Burns and R.S. Ketcham, "Fundamental-Mode Pierce Oscillators Utilizing Bulk-Acoustic-Wave Resonators in the 250-300 MHz Range", IEEE Transactions on Microwave Theory and Techniques, Vol. MTT-32, No. 12, Dec. 1984, pp. 1668-1671.

8. M.M. Driscoll, S.V. Krishnaswamy, R.A. Moore, and J.R. Szedon, "UHF Film Resonator Evaluation and Resonator-Controlled Oscillator and Filter Design Using Computer-Aided Design Techniques", 1985 IEEE MTT-S International Symposium.

9. E.A. Gerber and A. Ballato, Editors, Precision Frequency Control Oscillators and Standards, Vol. 2, Academic Press, 1985.

10. R.A. Sykes, W.L. Smith, and J.W. Spencer, "Monolithic Crystal Filters", IEEE International Convention Record, Part II, Mar. 1967.

11. W.D. Beaver, "Analysis and Elastically Coupled Piezoelectric Resonators", Journal Acoustic Society of America, Vol. 43, pp. 972-981, May 1968.

12. Reference Data for Radio Engineers, Fifth Edition, pp. 7-5 through 7-7, Howard W. Sams, 1973

13. W.A. Burkland, A.R. Landin, G.R. Kline, R.S. Ketcham and S.G. Burns, "A Thin-Film Bulk-Acoustic-Wave Resonator Controlled Integrated VHF Oscillator on Silicon", to be submitted for publication.

41st Annual Frequency Control Symposium - 1987

SHEAR MODE TRANSDUCERS FOR HIGH Q BULK MICROWAVE RESONATORS*

H. L. Salvo, Jr.
Westinghouse Defense and Electronics Center
Baltimore, MD 21203

M. Gottlieb and B. R. McAvoy
Westinghouse R & D Center
Pittsburgh, PA 15235

ABSTRACT

The influence of the transducer structure on the loaded Q value and noise performance of high overtone bulk acoustic resonators (HBARs) is being investigated. This investigation has turned, in part, to the fabrication of piezoelectric shear mode transducers which are used to excite nonpiezoelectric substrates such as spinel ($MgAl_2O_4$) to overtone resonance. Spinel and Lithium Tantalate ($LiTaO_3$) are among the promising substrate candidates exhibiting very low propagation loss (High Q) and temperature stable resonance respectively. Three approaches with a number of innovations for each approach are being examined for optimum shear mode transducer performance. The first approach uses vacuum compression bonded lithium niobate on spinel substrates. This provides a transducer structure which is fundamentally different than that provided by RF sputtering of zinc oxide. It is because of this that different noise generating mechanisms are expected from the two transducer types.

Two different approaches to the sputtering of shear mode ZnO transducers, to provide different film growth conditions, have been undertaken. The effect on resonator noise performance, if any, between these two ZnO resonator types are to be evaluated. The first RF sputtering technique employs a diode system with an independent (incoherent) RF biasing of the substrate. The effect of this biasing on film growth will be discussed. The second technique uses a magnetron sputtering system with D.C. substrate biasing. Both systems have yielded ZnO shear mode transducers suitable for HBAR use. The RF diode technique is the most useful so far in that it has provided films which transduce shear mode waves with no longitudinal component which can be measured.

INTRODUCTION

The sources of resonator noise are becoming an important topic as the effort to reduce oscillator noise increases. Figure 1 shows the phase noise performance of an HBAR (High-overtone Bulk Acoustic Resonator) fundamental oscillator operating at 640 MHz. Shown for comparison is the performance of an 80 MHz SC-CUT quartz crystal oscillator with X8 multiplication and with a SAW filter. The heavy dashed line assumes a noiseless HBAR and is the goal of on-going efforts such as the present one. For frequencies near the carrier the principal determinant of the noise is the loaded Q value of the resonator.

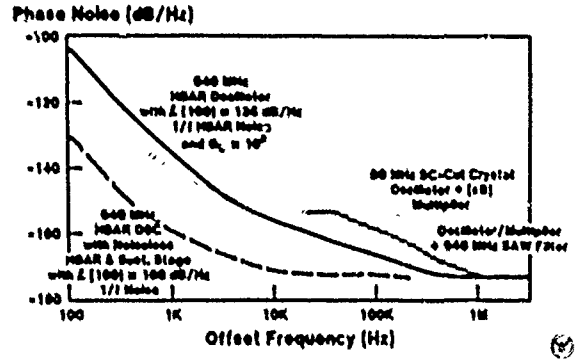


Figure 1. Calculated Noise Performance of Various Oscillators. Measured 80 MHz SC Cut Resonator Performance Shown for Comparison.

Hence one objective of this effort is to find the means of producing the highest Q. Another important objective is to identify the noise sources in the resonator itself. The prime suspect in this search is the transducer. This assertion is substantiated in part by a widely varying noise performance for HBARs fabricated from the same substrate with presumed identical processing. To reach the first objective, i.e. high Q, we are using shear mode transducers on low loss substrates. To explore the influence of the transducer on noise we are examining the effects of shear mode transducers made in different ways. Figure 2 highlights one method of providing shear transducers in an HBAR configuration shown at the upper left. This method uses tilted fiber growth in the ZnO film. At 40° tilt with respect to a normal electric field produced by parallel plate electrodes provides a bulk shear wave. At the lower left in this figure is a short list of HBAR substrate materials. This list is accompanied by ratios showing the shear mode to compressional mode propagation loss. For spinel the anticipated increase in Q is 4 times that over a compressional mode resonator. The increase in Q for YIG (Yttrium Iron Garnet) is also substantial.

YIG has been shown to provide magnetically tunable HBARs for filter applications.[3] There are a number of different methods of making ZnO transducer films which produce shear waves. These are discussed in a subsequent section. Another approach employs vacuum compression bonded lithium niobate plates on substrates such as spinel.

* This effort was sponsored in part by the Electro-magnetic Directorate, Rome Air Development Center, Air Force Systems Command, Hanscom Air Force Base, MA under Contract F-19628-84-C-0099.

CH2427-3/87/0000-388 $1.00 C 1987 IEEE

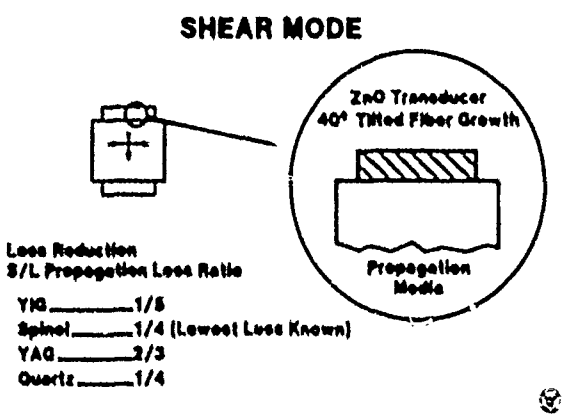


Figure 2. Tilted ZnO film fiber growth produces a shear bulk mode transducer. Increase in Q is the inverse of loss ratios for the resonator materials listed.

PLATE TRANSDUCERS

A complete description of single crystal lithium niobate plate transducers is given in Reference 4. In this section, we describe current results which are an extension of this technique. Figure 3 shows the polishing jig used to reduce the thickness of a bonded transducer to substrates such as spinel. A flat even plate surface with respect to the substrate is obtained by using lithium tantalate stand-offs as shown around the fixture. Figure 4 is a close-up of the bonded plate after polishing to about 0.5 micron thickness. Figure 5 shows the final thinning step with the average transducer thickness of about 3000 angstroms. The plate is then segmented into transducer areas.

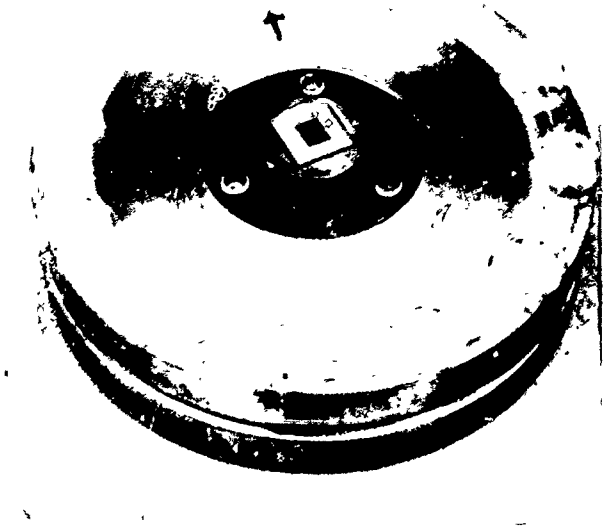


Figure 3. Polishing Fixture for Thinning $LiNbO_3$ Plate Transducers.

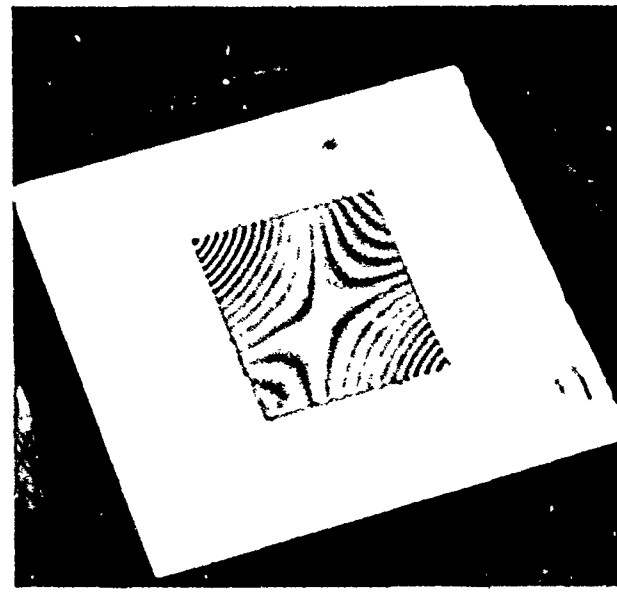


Figure 4. Thinned, Bonded Lithium Niobate Transducer Ready for Plasma Etching Step.

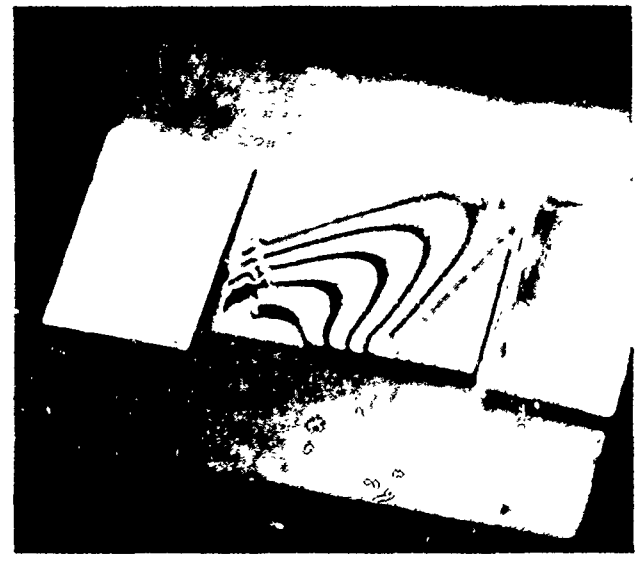


Figure 5. Plasma Etched Lithium Niobate Transducer Thickness is about 3,000 Angstroms.

SHEAR MODE FILM TRANSDUCERS

Modification of existing techniques and equipment is necessary for the fabrication of shear mode producing ZnO films[5]. Two sputtering techniques are being employed. modified RF diode and D.C. biased magnetron sputtering. The basic RF diode system is shown in Figure 6 and the magnetron system is a CVC 601 system with vacuum load lock. To the right in Figure 6 is the ZnO target surrounded by a liquid nitrogen cooled coil (Meisner trap). We use a sputter up configuration. To the left in the figure is the substrate platform which, in the vacuum system, is held over the ZnO target. The substrate is positioned off center near the edge of the ZnO target so that a 45° angle is subtended to the center of the target. Independent Rf power supplies are used to power the target and to bias

the substrate. To date, encouraging results have been obtained using the RF diode system. Low insertion loss transducers (15 dB at resonance) have been fabricated on YIG and YAG (yttrium iron and yttrium aluminum garnet), lithium tantalate and spinel ($MgAl_2O_3$). Transduction tests on all these substrates show good shear mode conversion with no longitudinal mode component. Resonators for noise tests and Q measurements are being fabricated on <111> and<100> spinel. The magnetron system is currently undergoing further modifications.

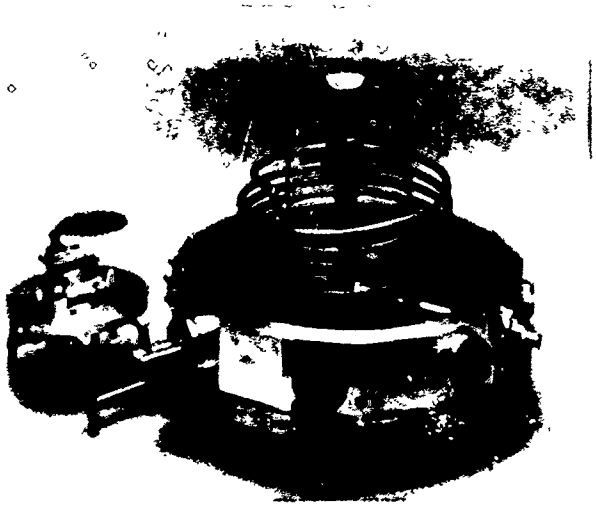


Figure 6. RF diode sputtering system with substrate fixture to the left. This fixture is suspended above the ZnO target off center in the vacuum system.

NOISE TESTS

Noise generation mechanisms such as strain relaxation, generation - recombination and ion migration are expected to be different for films fabricated in different ways. Trends in the noise data due to different processing techniques are being evaluated. Figure 7 is an example of a set of noise data for longitudinal mode MBAR. The number on each bar indicates the population associated with the noise interval indicated. All the resonators were fabricated with identical batch fabrication techniques. The phase noise indicated was measured in a one cycle bandwidth 100 Hz away from the carrier. The nearly 20 dB spread remains unexplained.

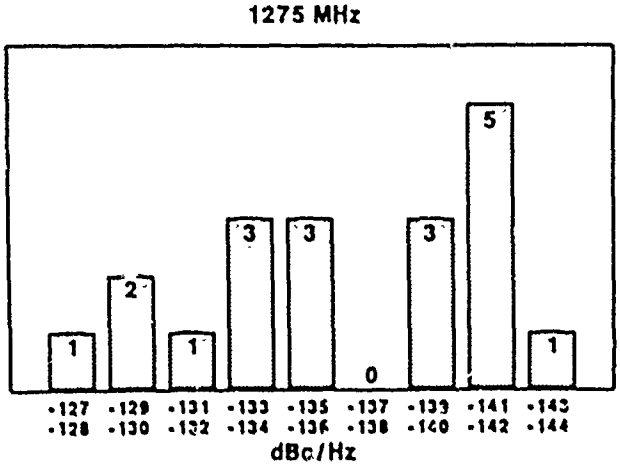


Figure 7 Phase noise distribution for batch processed MBARs. Frequency range of noise measurements is 1270 - 1279 MHz.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the fabrication skills of Jack Conroy, Don Watt and Stan Piesecki. They also value the guidance of M. Driscoll and R. Jelen of the Westinghouse Defense and Electronics Center.

REFERENCES

1. F. A. Gerber, et. al., "Advances in Microwave Acoustic Frequency Sources," IEEE - Trans. on Microwave Theory and Techniques, Vol. MTT-34, No. 10 p. 1002, OCT 1986.

2. S. V. Krishnaswamy, et. al., "Oriented ZnO films for Microwave Shear Mode Transducers," 1982 Ultrasonics Symposium Proceedings, IEEE Cat. No. 82CH1823-4, p. 476.

3. J. D. Adam, et. al., "Magnetically Tunable High Overtone Microwave Resonators," 40th Annual Frequency Control Symposium, May 1986.

4. Z. Turek, et. al, "Properties of an X-cut Single Crystal Lithium Niobate Transducer, IEEE Trans. on Sonics and Ultrasonics, Vol. SU-22, No. 6, p. 402, Nov. 1975.

41st Annual Frequency Control Symposium - 1987

ANALYSIS AND DESIGN OF THE PIEZOELECTRIC CERAMIC RESONATOR OSCILLATORS

Satoru Fujishima, Kazuya Togawa, and Satoshi Ohta

Murata Mfg. Co., Ltd.

2-26-10, Tenjin, Nagaokakyo-shi, Kyoto, Japan

Summary

We have analyzed oscillation stability using gain-phase characteristics in the open-loop circuit and verified this corresponds fairly to the behavior in actual closed loop circuit.

In order to analyze oscillation stability theoretically, we have simulated the gain-phase characteristics in the open-loop circuit on computer.
By this simulation, we can evaluate oscillation stability for the parameters of ceramic resonators and inverters, for which we could not evaluate experimentally.

Furthermore, we have analyzed rise time of ceramic resonator oscillator by the transient theory comparing with crystal resonator oscillator.

Introduction

Piezoelectric ceramic resonators have been widely used for microprocessor IC in clock oscillator in place of quartz crystal resonators.

Small configuration and fast rise time are main merits of piezoelectric ceramic resonators comparing to quartz crystal resonators. However, there have been many unknown matters for designing the piezoelectric ceramic resonator oscillators.

This report serves a dual purpose that is to design piezoelectric ceramic resonators for microprocessor IC in clock oscillator circuit with theoretical evaluation to ensure its performace and stability, and to support clock oscillator circuit design of microprocessor IC.

Oscillation Stability

Oscillator satbility is analyzed by examining the loop gain and phase shift.

Figure 1 shows a generic feedback oscillator. The circuit will oscilate when

(I) Loop Gain : $\alpha \times \beta \geq 1$

(II) Phase Shift : $\theta_1 + \theta_2 = 360° \times n$

$(n = 1, 2, \ldots)$

Figure 2 is a typical CMOS-inverter oscillator with a ceramic resonator. In order to analyze the oscillator stability, we opend the circuit at a connecting terminal ⓐ in Figure 2.

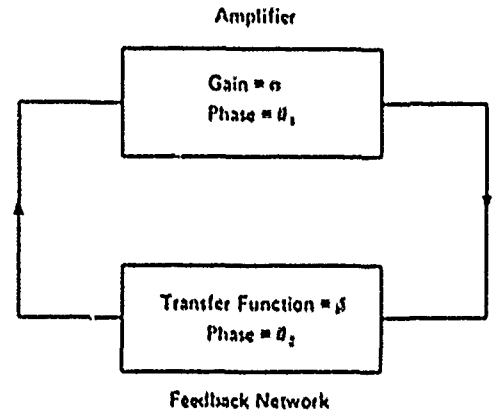
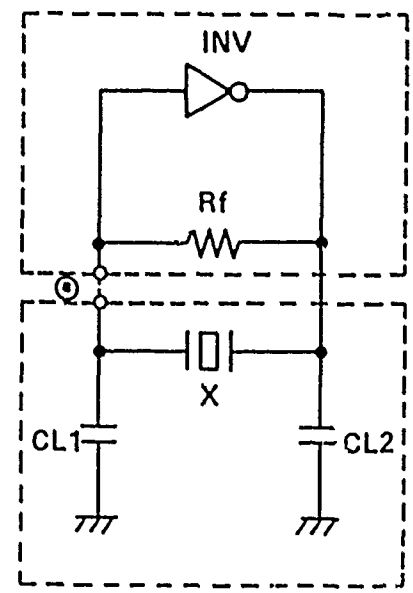


Fig.1 Generic feedback oscillator circuit



INV : Inverter
Rf : Feedback Resistor
X : Resonator
CL1, CL2 : Load Capacitor

Fig.2 Typical oscillator circuit w'th CMOS inverter

CH2427-3/87/0000-391 $1.00 C 1987 IEEE

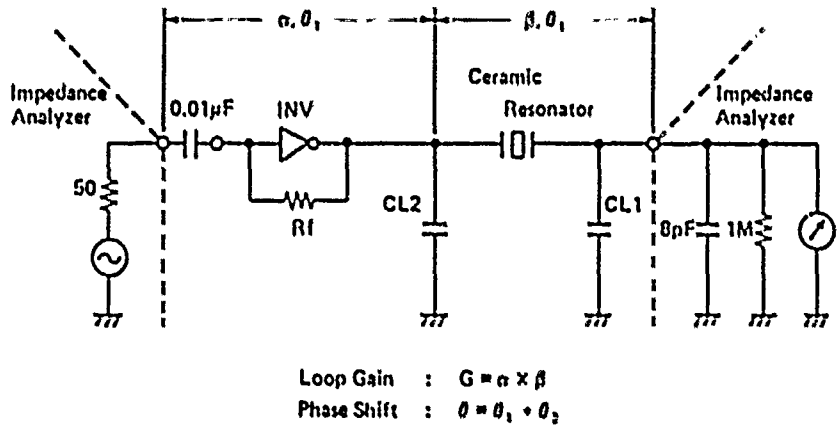


Fig.3 Open loop oscillator circuit to obtain gain-phase characteristic

Figure 3 shows the total network including input and output impedance of measurement equipment to measure loop gain and phase shift
We can evaluate the stability of oscillation using this loop gain at the oscillating frequency.

Then, we consider simulation analysis for the loop gain.

Theoretical Analysis for Gain-Phase Characteristics

To compute the gain-phase characteristics, we need equivalent circuit of the ceramic resonator and the inverter.

The ceramic resonator's equivalent circuit is shown in Figure 4 which is the same configuration as the quartz crystal resonator. (1)

The typical schematic of CMOS inverter is shown in Figure 5. As P-channel and N-channel FET in CMOS inverter are nearly symmetrical to each other, this schematic can be transfered to one source model. And since MOS FETs are transconductance device, the simplest model is a voltage-controlled-current source.
This is shown in Figure 6. (2)

We can calculate the gain-phase characteristics of the open loop circuit using transmission parameters.

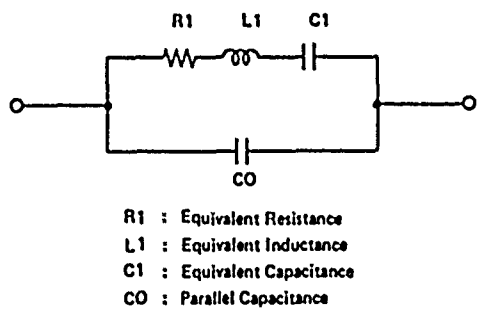


Fig.4 Equivalent circuit of ceramic resonator and crystal resonator

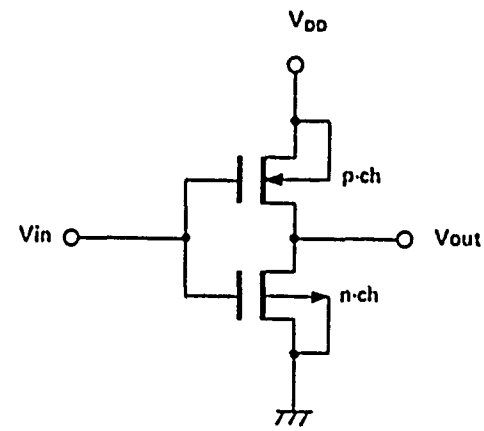
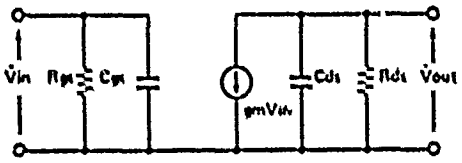


Fig.5 Typical schematic of CMOS inverter



gm : transconductance
Rgs: gate-to-source resistance
Cgs: gate-to-source capacitance
Rgs: drain-to-source resistance
Cds: drain-to-source capacitance

Fig.6 Four-terminal equivalent circuit of CMOS inverter

Consequently, Figure 3 can be expressed using equivalent circuit as Figure 7. When we consider this network to be defined by A-parameters, A-parameters are defined as in equation (1). The loop gain and phase shift are rewrited as equation (2) and equation (3).

$$\begin{pmatrix} \dot{V}in \\ \dot{I}in \end{pmatrix} = \begin{pmatrix} \dot{A}1 \end{pmatrix}\begin{pmatrix} \dot{A}2 \end{pmatrix}\begin{pmatrix} \dot{A}3 \end{pmatrix}\begin{pmatrix} \dot{A}4 \end{pmatrix}\begin{pmatrix} \dot{V}out \\ \dot{I}out \end{pmatrix}$$

$$= \begin{pmatrix} A & B \\ C & D \end{pmatrix}\begin{pmatrix} \dot{V}out \\ \dot{I}out \end{pmatrix} \qquad (1)$$

$$\text{Loop Gain} = \left|\frac{\dot{V}out}{\dot{V}in}\right| \qquad (2)$$

$$\text{Phase Shift} = \angle\dot{V}out \quad \angle\dot{V}in \qquad (3)$$

From equation(1), the relation between input and output is shown as following.

$$Vin = A\,Vout + B\,Iout \qquad (4)$$

The voltage transfer ratio (Vout/Vin) is for an open output, so Iout = 0. Therefore, the second term of equation (4) drops out.

$$Vin = A\,Vout$$

$$Vout/Vin = 1/A \qquad (5)$$

Equation (5) shows that the voltage transfer ratio depends only on parameter A.

With above model, we have computerized the gain-phase simulation. And we have learned that if the model uses measured values of network parameters, the model's predictions and the behavior of the circuit correspond closely. (3)

Figure 8 shows the gain-phase characteristics of actual measurement and computed simulation.
In Figure 8, simulation result corresponds closely with actual measurement result.

As mentioned before, the circuit must satisfy the condition of (I) and (II) to oscillate. Therefore, in Figure 8 oscillation occurs only at point Ⓕ and the gain (ΔG) is the gain margin of oscillation circuit.

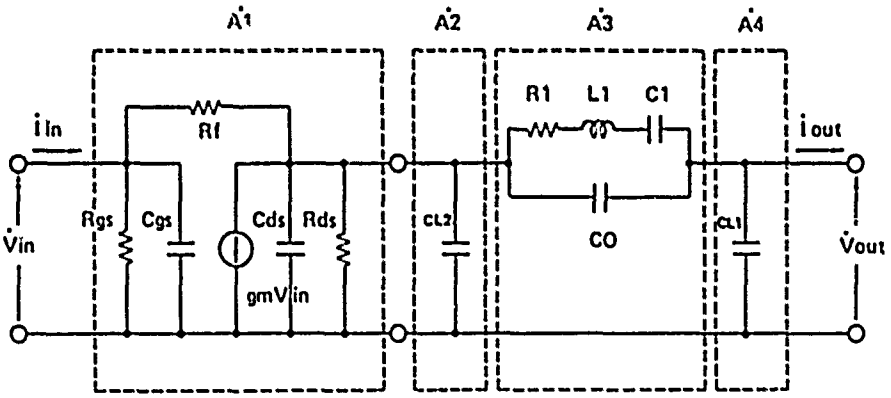


Fig.7 Gain-Phase simulation circuit using A-parameter

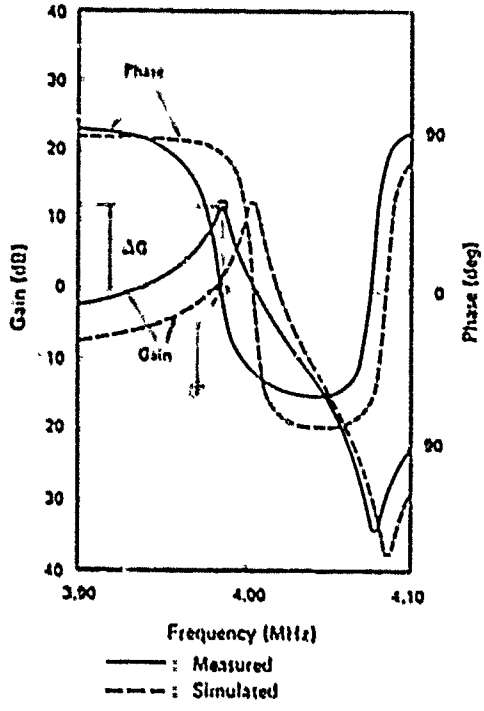


Fig.8 Gain-Phase characteristics by actual measuring and computed simulation

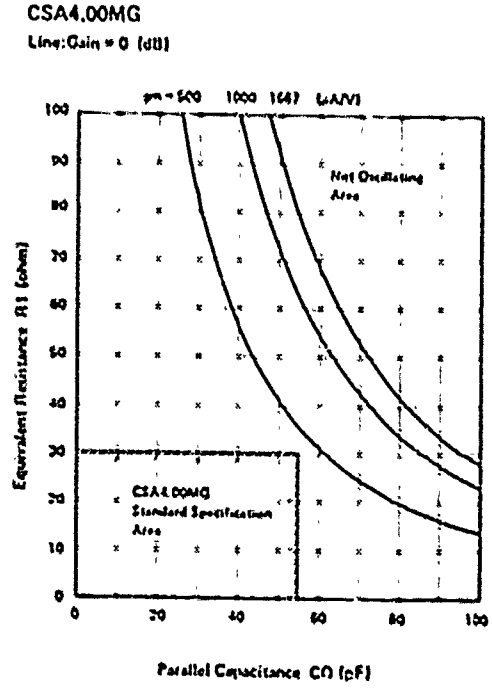


Fig.9 Oscillating region in R1-C0 plane by inverter's transconductances (gm)

Application of the Gain-Phase Simulation

Until now, there has not been effective way to evaluate the stability of oscillation for the ceramic resonator/inverter circuit. This simulation makes it possible do that. At any given parameters of resonator and inverter, we can determine whether or not the network will oscillate.

Figure 9 shows the result of this analysis. The ceramic resonator's equivalent resistance (R1) and parallel capacitance (C0) are plotted against each other, with the inverter's transconductance (gm) as the parameter.

The curves are the bordeline condition of oscillation, where the gain is 1 (=0dB). The region over these curves represents lower gain, and the network won't oscillate.

This data was obtained using 4MHz ceramic resonator (Murata CSA4.00MG). The shaded region shows the specified range of R1 and C0 parameters for this particular resonator. If the inverter's gm is no lower than about 500 μA/V, we can guarantee that a resonator whose parameters fall in this area will oscillate with any circuit.

Likewise, the higher the transconductance, the greater the gain margin, and the broader the resonator's specifications can be. Note, however, that a transconductance of 500 μA/V brings the line of oscillation very close to the acceptable range of resonator parameters. A slight drop in gain, or a minor change in the resonator might cause the oscillator to stop. Therefore, such a design is not very stable.

Of course, we can design (or selec) the resonator to have tighter tolerances. But this reduces the yield and raises the cost. If the inverter would be designd to enter resonator's specifications into the stable region, we can supply most economical and stable oscillator circuit to customers.

Figure 10 shows stability with respect to the changes in the external components. CL2 is plotted against CL1, with the transconductance as the running parameter.

Figure 11 shows how the gain varies with CL1 and CL2, for an assumed transconductance of 1000 μA/V.

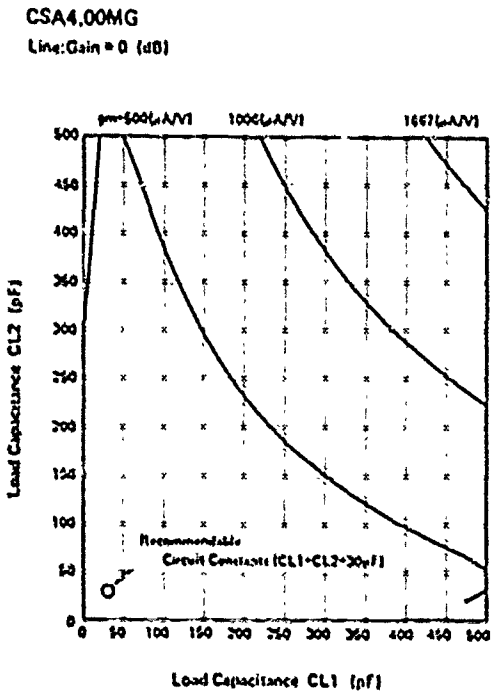


Fig.10 Oscillating region in CL1-CL2 plane by inverter's transconductance (gm)

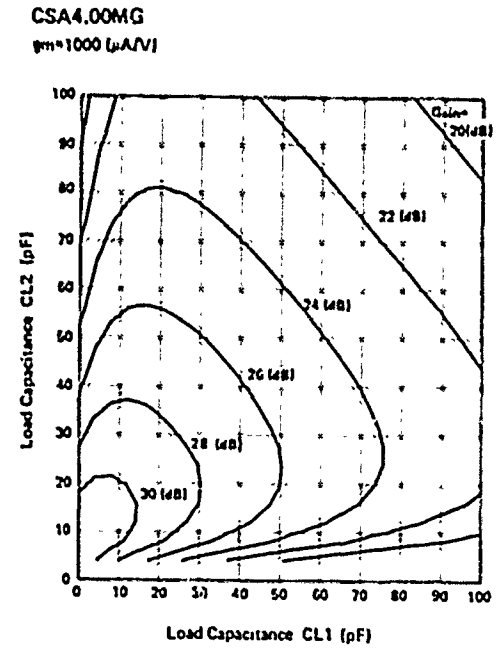


Fig.11 Variation of the gain by CL1 and CL2

Analysis for Rise Time

We will be able to analyze rise time by the transient analysis.

If the gain of inverter would be sufficently high, we could consider the circuit as LCR resonant circuit supplied alternating current voltage. (4)

When the circuit is supplied the voltage of $e(t) = Em \sin\omega t$, the transient response of loop current is calculated from following equation.

$$L\cdot\frac{di}{dt} + R\cdot i + \frac{1}{C}\int i\,dt = Em\cdot\sin\omega t \qquad (6)$$

In the resonant case, the general solusion of loop current $i = i(t)$ is shown as following.

$$i(t) = \frac{Em}{Z}\cdot\frac{1}{\sin\phi}\cdot\varepsilon^{-\alpha t}\cdot\left\{\sin\theta\cdot\sin(\phi-\beta t) - \frac{\beta}{\omega}\cos\theta\cdot\frac{\sin\beta t}{\sin\phi}\right\} + \frac{Em}{Z}\cdot\sin(\omega t-\theta) \qquad (7)$$

Where,

$$Z = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2}$$

$$\theta = \tan^{-1}\left(\frac{\omega L - 1/\omega C}{R}\right)$$

$$\phi = \tan^{-1}\left(\frac{\beta}{\alpha}\right)$$

$$\alpha = R/2L$$

$$\beta = \sqrt{1/LC - (R/2L)^2} = \sqrt{\omega o^2 - \alpha^2}$$

When ω corresponds with the natural frequency of the circuit and ωo is much greater than α,

$$\omega = \beta = \sqrt{\omega o^2 - \alpha^2} \fallingdotseq \omega o$$

$$\theta = 0$$

$$\phi = \frac{\pi}{2}$$

then equation (7) is represented as following.

$$i(t) = \frac{Em}{R}\left(1-\varepsilon^{-\alpha t}\right)\sin\omega t \qquad (8)$$

From equation (8), the amplitude of i(t) is varied by $\varepsilon^{-\alpha t}$.

Therefore, rise time is proportional to $1/\alpha$ as follows:

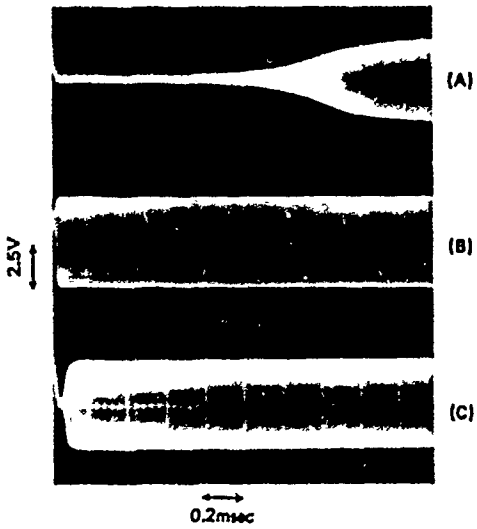
$$Tr = 1/\alpha = 2L/R = 2Q/\omega \qquad (9)$$

Equation (9) shows that rise time is directly proportional to Q and inversely proportional to ω (frequency).
We verified this results by actual experiments.

Figure 12 is a photograph of rise time at 4 MHz, comparing with quartz crystal, ceramic and LiTaO3 resonator.

Figure 13 shows the variation of rise time against resonator's mechaniacl quality factor (Qm). Rise time is directly proportional to Qm.

Figure 14 shows the variation of rise time against oscillation frequency, for ceramic resonators and quartz crystal resonators obtained by the same experiments of Figure 12.
For both cases, rise time is inversely proportional to oscillation frequency.



Rise time of clock oscillator using X'tal, ceramic resonator and $LiTaO_3$ resonator
(A) : 4,000MHz X'tal
(B) : 4,000MHz Ceramic Resonator
(C) : 4,000MHz $LiTaO_3$

Fig.12 Photograph of rise time

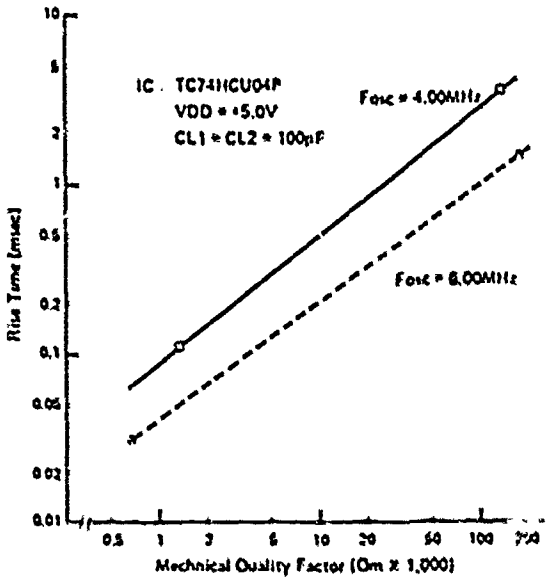


Fig.13 Rise time vs mechanical quality factor(Qm)

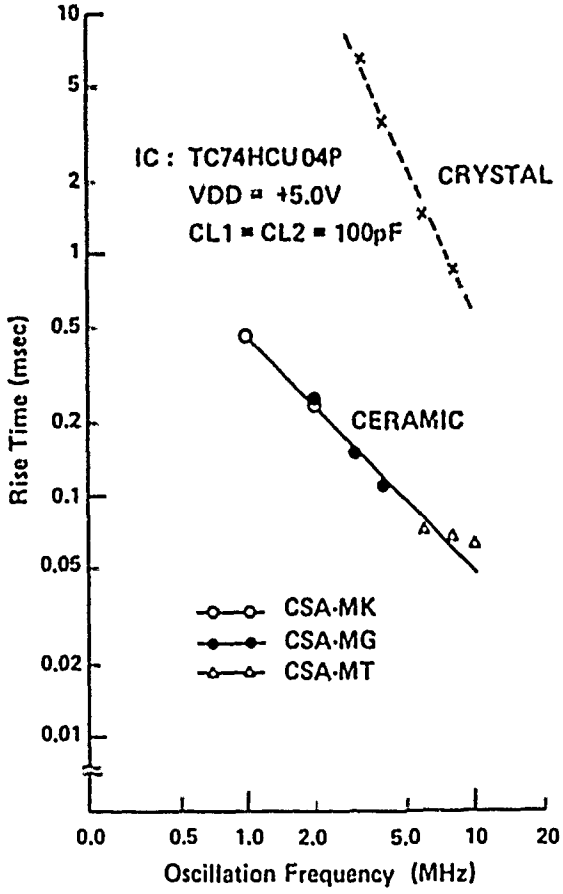


Fig.14 Rise time vs oscillating frequency

Conclusion

Our gain-phase simulation makes it possible to design an inverter oscillator circuit that is matched to the characteristics of resonator. We no longer need to guess whether the circuit will work, or what ranges of inverter and resonator parameters are acceptable.

As for oscillation rise time, we could verify its trend for resonator's Q and frequency theoretically and experimentally.

References

1) S. Fujishima, J. Merlina and J. Miyazaki
"PIEZOELECTRIC CERAMIC RESONATOR AND FILTER"
Frequency Control Symposium
Proc. May, 1984

2) J. E. Meyer
"MOS MODELS AND CIRCUIT SIMULATION"
RCA Rev.32 p.p.42-63 (1971)

3) T. Adachi, M. Hirose and Y. Tsuzuki
"COMPUTER ANALYSIS OF CRYSTAL OSCILLATOR"
IECE Vol.J68-CNo.7 July, 1985

4) M. Toki and Y. Tsuzuki
"MEASURING METHOD OF EQUIVALENT SERIES CAPACITANCE AND NEGATIVE REGISTANCE OF CRYSTAL OSCILLATOR CIRCUITS"
IEEE Electronics Letters,19,23,p.p.956-960 (Nov.,1987)

THE ACCELERATION SENSITIVITY OF QUARTZ CRYSTAL OSCILLATORS: A REVIEW

Raymond L. Filler
US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

SUMMARY

The acceleration-sensitivity of quartz crystal oscillators is assuming increasing significance in modern communications, navigation, radar, and identification systems. This paper is a tutorial and a review of the various aspects of acceleration-sensitivity. The topics to be discussed are 1. the consequences of acceleration-sensitivity in crystal oscillators, including the effects of sinusoidal and random vibration, on the Allan Variance, phase noise and integrated phase jitter, 2. the vector nature of quartz resonator sensitivity, 3. the theoretical description of the cause of the acceleration-sensitivity of quartz resonators, 4. techniques for the measurement of acceleration-sensitivity, and 5. the effect of frequency multiplication on acceleration effects.

In addition, various techniques currently being used or developed for reducing the effective acceleration sensitivity will be discussed. The techniques fall into three general categories: reduction of the acceleration-sensitivity of the resonator: passive techniques that use compensating elements in the oscillator feedback loop, e.g. a second resonator or an acceleration sensitive capacitor: and active acceleration compensation schemes that sense the acceleration and feed back a compensating signal to a tuning network.

INTRODUCTION

There has been an awareness of acceleration effects in frequency sources at least since the advent of missile and satellite applications,[1-6] Doppler radars,[7,8] and other systems requiring extremely low noise.[9,10] There has not, however, been a general appreciation of the magnitude of the problem. As a case in point, there is little or no mention of acceleration sensitivity in the general texts on quartz crystal oscillators and resonators or of vibration-induced phase noise in radar textbooks.

Time dependent acceleration, i.e., vibration, can cause a large increase in the noise level of a quartz crystal oscillator. In fact, in frequency sources operating on mobile platforms, the vibration-induced phase noise is usually greater than all other noise sources combined.

High-stability frequency sources, including atomic standards, contain quartz crystal resonators. One result of the evolution of electronics, i.e., the transition from tubes to transistors, and from point-to-point wiring to printed circuits, is the establishment of the quartz crystal resonator as the most acceleration-sensitive component in frequency sources. This paper will review the causes and effects of acceleration sensitivity of bulkwave quartz crystal resonators, and the methods that reduce or compensate for that sensitivity. Most of what is discussed is equally relevant to surface acoustic wave (SAW) and shallow bulk acoustic wave (SBAW) devices.

THE EFFECT OF ACCELERATION ON A CRYSTAL RESONATOR

A quartz crystal resonator subject to a steady acceleration has a slightly different series resonant frequency than the same resonator experiencing zero acceleration.[1] Furthermore, it has been observed that the magnitude of the frequency shift is proportional to the magnitude of the acceleration, and is also dependent upon the direction of the acceleration relative to a coordinate system fixed to the resonator.[11] It has been shown, empirically, that the acceleration sensitivity of a quartz crystal resonator is a vector quantity.[12] Therefore, the frequency during acceleration can be written as a function of the scalar product of two vectors

$$f(\vec{a}) = f_0(1 + \vec{\Gamma}\cdot\vec{a}), \qquad (1)$$

where $f(\vec{a})$ is the resonant frequency of the resonator experiencing acceleration $\vec{a}$, f_0 is the frequency with no acceleration (referred to as the carrier frequency), and $\vec{\Gamma}$ is the acceleration-sensitivity vector. It can be seen from Eq. (1) that the frequency of an accelerating resonator is a maximum when the acceleration is parallel to the acceleration-sensitivity vector; it is a minimum when the acceleration is antiparallel to the acceleration-sensitivity vector. An important result of Eq. (1) is that the frequency shift, $f(\vec{a})-f_0$, is zero for any acceleration in the plane normal to the acceleration-sensitivity vector.

The frequency shift described in Eq. (1) is also induced by the acceleration due to gravity (even without motion). This is commonly referred to as "2g-tipover." During "2g-tipover," the magnitude of the acceleration is 1g in the direction towards the center of the earth. (The magnitude of acceleration given in this paper will be in units of g, i.e., the magnitude of the earth's gravitational acceleration at sea level, 980 cm/sec^2.) When a resonator is rotated 180° about a horizontal axis, the

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

scalar product of the acceleration and the unit vector normal to the initial "top" of the resonator goes from -1g to +1g, which is a difference of 2g. Figure 1 shows actual data of the fractional frequency shift of a resonator (operating in an oscillator) when the oscillator is rotated about three mutually perpendicular axes in the earth's gravitational field. For each curve, the axis of rotation is held horizontal. The sinusoidal shape of each curve is a consequence of the scalar product being proportional to the cosine of the angle between the acceleration-sensitivity vector and the acceleration due to gravity.

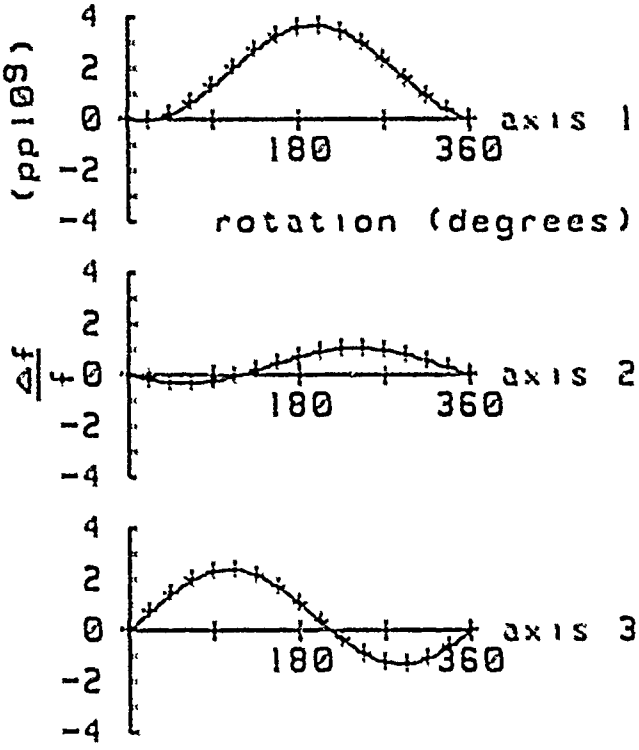


Figure 1. "2g Tipover," Frequency change vs. rotation in the Earth's gravitational field for three mutually perpendicular axes.

THE EFFECT OF VIBRATION ON A CRYSTAL RESONATOR

In most applications, the magnitude of the acceleration is time-dependent. The magnitude of the acceleration-sensitivity vector, for acceleration amplitudes commonly encountered, is independent of acceleration amplitude.[11] The time-dependent frequency shift due to a complex vibration can, therefore, be determined from the sum of the individual sinusoidal components, i.e., the system is linear and superposition holds. Simple sinusoidal vibration will be discussed first; the extension will then be made to random vibration.

Simple harmonic motion will be assumed, with an acceleration given by

$$\vec{a} = \vec{A}\cos(2\pi f_v t), \qquad (2)$$

where $\vec{A}$ is the peak acceleration vector in units of g, f_v is the frequency of vibration in Hertz, and t is time in seconds. The variation of the frequency with time can be determined by combining Eq. (2) with Eq. (1), resulting in

$$f(\vec{a}) = f_o\,(1 + (\vec{\Gamma}\cdot\vec{A})\cos(2\pi f_v t)). \qquad (3)$$

This can be rewritten as

$$f(\vec{a}) = f_o + \Delta f\cos(2\pi f_v t), \qquad (4)$$

where

$$\Delta f = f_o\,(\vec{\Gamma}\cdot\vec{A}) \qquad (5)$$

is the peak frequency shift due to the acceleration $\vec{A}$.

The behavior of the device can be described by defining a rectangular coordinate system fixed to the resonator. The acceleration-sensitivity vector and the acceleration vector can then be described in terms of the three unit vectors defined by that coordinate system. Therefore Eq. (3) can be transformed into a scalar equation containing the three (i, j, and k) components of $\vec{A}$ and $\vec{\Gamma}$, i.e.

$$f(\vec{a}) = f_o(1+\Gamma_i A_i+\Gamma_j A_j+\Gamma_k A_k) \qquad (6)$$

There are three quantities with units of frequency to keep in mind: f_o, f_v, and Δf. It can be seen from Eq. (5) that the output frequency deviates from the center frequency, f_o, by the amount $\pm\Delta f$, at a rate of f_v. This is shown schematically in Figure 2. Each plot is the instantaneous output frequency of a crystal oscillator while undergoing a vibration at frequency f_v. Figure 3b is the voltage vs. time at the output of the same crystal oscillator showing, in a much exaggerated way, the variation in the frequency with acceleration amplitude. Figure 3a is the acceleration waveform.

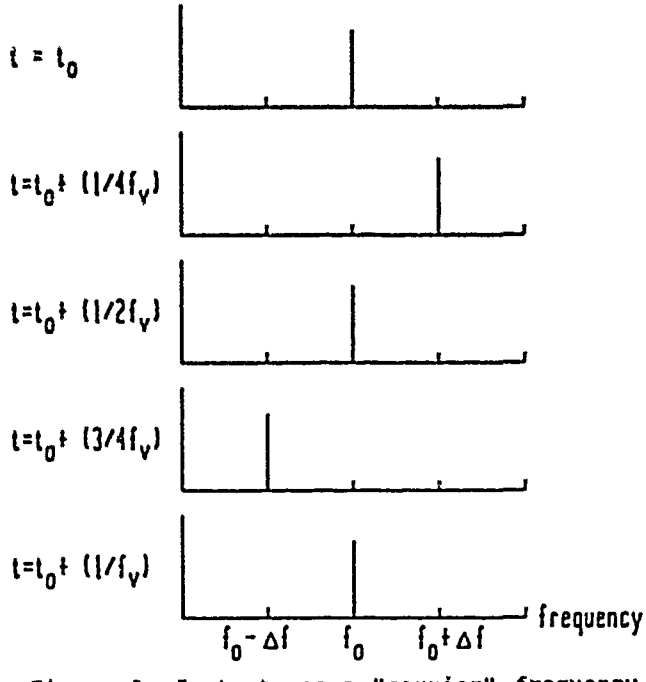


Figure 2. Instantaneous "carrier" frequency for several instants during one cycle of vibration.

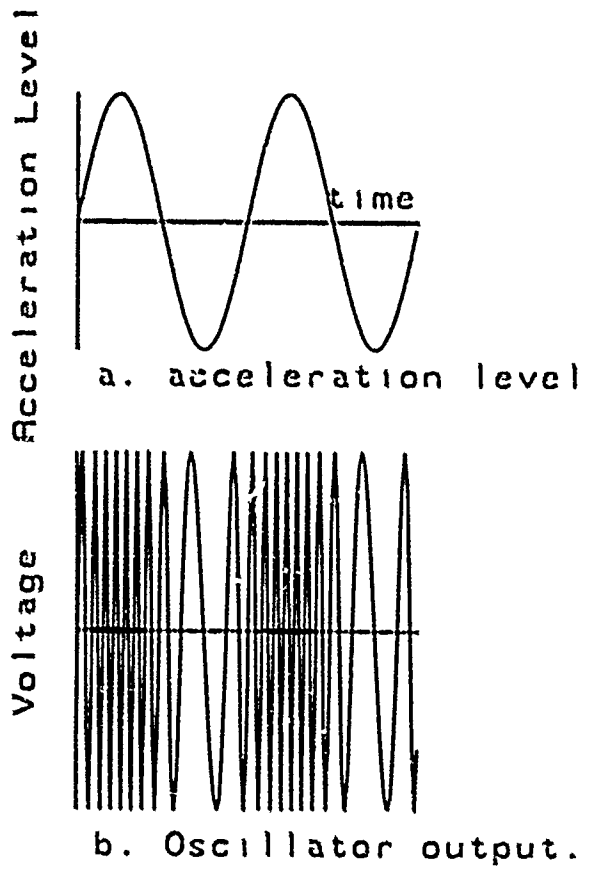


Figure 3. a. Acceleration level vs. time.
b. Resulting oscillator output.

FREQUENCY DOMAIN

It is very useful to transform the effect of vibration into the frequency domain. This will allow the formulation of a convenient measurement scheme and allow comparison of vibration effects to more familiar forms of phase noise.

The voltage appearing at the output of an oscillator is given by

$$V(t) = V_o \cos(\phi(t)), \qquad (7)$$

where the phase $\phi(t)$ is derived from the frequency by

$$\phi(t) = 2\pi \int_0^t f(t')\, dt'. \qquad (8)$$

When the oscillator frequency is varying due to simple harmonic acceleration modulating the resonant frequency of the resonator, the phase in Eq. (7) becomes, using Eqs. (5) and (8),

$$\phi(t) = 2\pi f_o t + (\Delta f/f_v)\sin(2\pi f_v t). \qquad (9)$$

When Eq. (9) is inserted into Eq (7), the result is

$$V(t) = V_o \cos[2\pi * \frac{\Delta f}{f_v} \sin(2\pi f_v)] \qquad (10)$$

Eq. (10) is the expression for a frequency-modulated signal. It can be expanded in an infinite series of Bessel functions[13] resulting in[14]

$$\begin{aligned} V(t) = V_o\, [&J_o(\beta)\cos(2\pi f_o t) \\ &+ J_1(\beta)\cos(2\pi(f_o+f_v)t) \\ &+ J_1(\beta)\cos(2\pi(f_o-f_v)t) \\ &+ J_2(\beta)\cos(2\pi(f_o+2f_v)t) \\ &+ J_2(\beta)\cos(2\pi(f_o-2f_v)t) + \ldots] \end{aligned} \qquad (11)$$

where $\beta = \Delta f/f_v = (\vec{\Gamma}\cdot\vec{\Delta}) f_o/f_v$ is the modulation index (from standard FM theory.)

The first term in Eq. (11) is a sine wave at the carrier frequency with an amplitude, relative to V_o, of $J_o(\beta)$. The other terms are vibration-induced sidebands at frequencies $f_o + f_v$, $f_o - f_v$, $f_o + 2f_v$, $f_o - 2f_v$, etc. In analogy with common phase-noise notation,[16] the ratio of the power in the nth vibration-induced sideband to the power in the carrier, denoted by $\mathcal{L}_v^n$, is given by

$$\mathcal{L}_v^n = (J_n(\beta)/J_o(\beta))^2 \qquad (12)$$

or, more commonly expressed in decibels as

$$\mathcal{L}_v^n(\text{dBc}) = 20 \log(J_n(\beta)/J_o(\beta)), \qquad (13)$$

where dBc refers to dB relative to the carrier and the subscript v refers to the vibration.

SMALL MODULATION INDEX

Several approximations can be made if the modulation index is less than 0.1. This is the case for most frequency standards in the HF band experiencing "normal" accelerations of 10 g or less at acceleration frequencies above a few hertz. The approximations are

$$\begin{aligned} J_o(\beta) &= 1 \quad : \ \beta<0.1 \\ J_1(\beta) &= \beta/2 \quad : \ \beta<0.1 \\ J_n(\beta) &= 0 \quad : \ \beta<0.1,\ n>2. \end{aligned} \qquad (14)$$

Therefore, after combining Eqs. (13) and (14):

$$\begin{aligned} \mathcal{L}_v^1(\beta) &= 20 * \log((\vec{\Gamma}\cdot\vec{\Delta}) f_o/(2f_v)) \\ \mathcal{L}_v^{n>1}(\beta) &= 0 \end{aligned} \qquad (15)$$

For small modulation index, therefore, most of the power is in the carrier and a small amount is in the first sideband pair (upper and lower). The amplitudes of the higher sidebands are negligible. The condition of small modulation index permits $\mathcal{L}(f)$ to be equated to $S_\phi(f)/2$.[16]

As an example, consider a 5 MHz oscillator with a vibration sensitivity of 2×10^{-9}/g. If the oscillator experiences an acceleration of 5 g along the direction of the acceleration-sensitivity vector, the

value of Δf is $(5 \times 10^6$ Hz) * (5 g) * $(2\times10^{-9}/g)$ = 0.05 Hz. The modulation index and $\mathcal{L}_v^1$ for several vibration frequencies are shown in Table I.

TABLE I

First sideband level, in dBc, for a 5 MHz oscillator with an acceleration sensitivity of $2\times10^{-9}/g$ experiencing an acceleration of 5g peak.

f_v(Hz)	β	$\mathcal{L}_v^1$(dBc)
5	.01	-46
25	.002	-60
50	.001	-66
500	.0001	-86

It can be seen that $\mathcal{L}_v^1$ falls off at 6 dB per octave (20 dB per decade.) A consequence of linearity at low vibration levels is that these sinusoidal accelerations are independent. The amplitude at any frequency is the same if the bright lines are excited separately or in combination. The first three of these vibration-induced sidebands and the carrier are shown in a typical spectrum analyzer output in Figure 4.

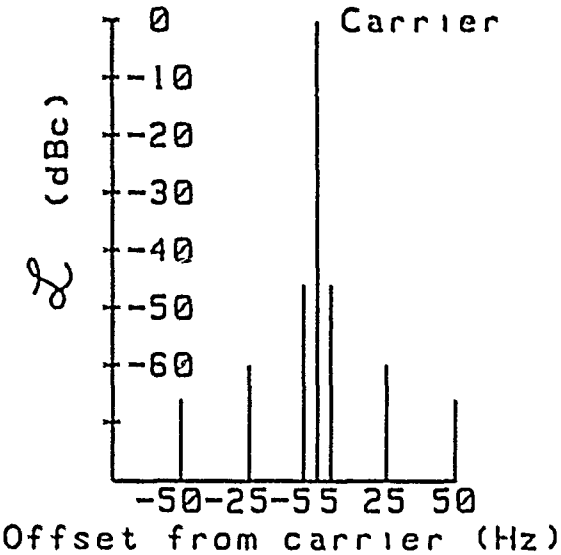


Figure 4. Carrier and vibration-induced sidebands resulting from sinusoidal acceleration at 5, 25, and 50 Hz.

FREQUENCY MULTIPLICATION

Phase noise during vibration is of great concern in systems such as radar, navigation, and satellite communications, where the frequency of the crystal oscillator is multiplied up to the microwave region. Upon frequency multiplication by a factor of N, the vibration frequency f_v is unaffected since it is an external influence. The peak frequency change due to vibration, Δf, however, becomes

$$\Delta f = (\vec{\Gamma}\cdot\vec{A})Nf_o. \qquad (16)$$

The modulation index β is therefore increased by a factor of N. Expressed in decibels, frequency multiplication by a factor N increases the phase noise by 20 log N.

The relationship between the vibration-induced phase noise of two oscillators with the same vibration sensitivity and different carrier frequencies is

$$\mathcal{L}_2 = \mathcal{L}_1 + 20\log(f_2/f_1). \qquad (17)$$

Where $\mathcal{L}_1$ is the sideband level, in dBc, of the oscillator at frequency f_1, and $\mathcal{L}_2$ is the sideband level of the oscillator at frequency f_2. This is identical to the "20 log N" term associated with the increase in phase noise due to frequency multiplication. Again, this relationship holds only if $\beta < 0.1$.

LARGE MODULATION INDEX

If the modulation index b is larger than about 0.1, the approximations made in Eq. (14) are not valid. This often occurs in UHF and higher frequency systems.[9] Frequency multiplication to a higher frequency is indistinguishable from direct frequency generation at the higher frequency. For example, when a $2 \times 10^{-9}/g$ sensitivity 5 MHz oscillator's frequency is multiplied by a factor of 308 to generate a frequency of 1575 MHz, its output will have identical vibration-induced sidebands as a 1575 MHz SAW oscillator with a sensitivity of $2\times10^{-9}/g$. At 5 g acceleration, for example, the vibration-induced sidebands produced by the 1575 MHz oscillator and the 5 MHz oscillator multiplied by 308 are shown in Table II.

TABLE II

First sideband level, in dBc, for a 1575 MHz oscillator with an acceleration sensitivity of $2\times10^{-9}/g$ experiencing a peak acceleration of 5 g.

f_v(Hz)	β	$\mathcal{L}_v^1$ (dBc)
5	3.15	+2.3
25	0.63	-9.6
50	0.315	-16.
500	0.0315	-36.

It can be seen that the values of $\mathcal{L}_v^1$ in Table II are much greater than those in Table I for the same vibration frequency. This is a consequence of the ratio of output frequency to vibration frequency being much larger. It is possible, as can be seen in this example, for the sidebands to be larger than the carrier. Also, there are even conditions where the carrier disappears and the value of $\mathcal{L}_v^1$ goes to infinity, e.g., when β equals 2.4, that is, all of the power is in the sidebands and none is in the carrier.

ALLAN VARIANCE

The effect of sinusoidal phase modulation on the Allan variance of a frequency standard has been shown[15] to be given by

$$\sigma_y = \frac{\Phi}{\pi f_o \tau}\sin^2(\pi f_v \tau) \qquad (18)$$

where ϕ is the peak phase deviation, f_v is the frequency of the phase modulation, and the measurement time is τ. It can be seen from Eq. (9) that the magnitude of the maximum phase deviation for a single vibration-induced sideband is $\Delta f/f_v$. Therefore, the vibration-induced Allan variance of a frequency standard with an acceleration sensitivity of Γ, subjected to acceleration $\bar{A}$ at a frequency of f_v, is

$$\sigma_y = \frac{\Gamma \cdot \bar{A}}{\pi f_v \tau} \sin^2(\pi f_v \tau). \qquad (19)$$

This effect is shown in Figure 5. The frequency standard is assumed to have an Allan variance of $(1\times10^{-12}/\tau)^2$ when not being accelerated.

A plausibility argument for the occurrence of the peaks and valleys in Figure 5 is that the average frequency, as given in Eq. (5), is f_o for an averaging time equal to an integer multiple of the period of the vibration. Since this is a constant, the variance is zero. The average frequency departure from f_o is a maximum when the averaging time is an integer multiple of one-half the period of the vibration. Therefore, the peaks occur when $\tau = (2n+1)/(2f_v)$ and the valleys occur when $\tau = (n+1)/f_v$, where $n = 0,1,2,3,\ldots$.

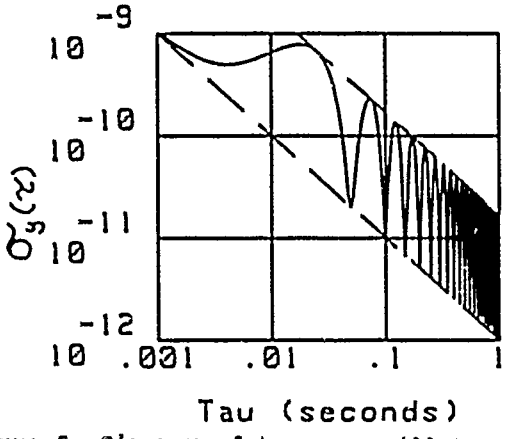


Figure 5. Sigma y of tau an oscillator with an acceleration-sensitivity of 1ppb/g during a 1 g vibration at 20 Hz.

RANDOM VIBRATIONS

In most situations the acceleration experienced by a frequency standard is not simple harmonic motion; it is random, i.e., the vibratory power is randomly distributed over a range of frequencies, phases, and amplitudes. The acceleration can be described by its power spectral density G(f). The power spectral density of frequency fluctuations, $S_y(f)$, can be obtained by multiplying the power spectral density of acceleration by the square of the acceleration sensitivity in $(Hz/g)^2$. σ_y^2 is related to S_y, if the modulation index is small, by[16]

$$\sigma_y^2 = S_y/(2f_v^2). \qquad (20)$$

Therefore, σ_y^2 for random vibration is given by

$$\sigma_y^2 = (\Gamma f_o)^2\, G(f)/(2f_v^2). \qquad (21)$$

As an example, consider the random acceleration spectrum given in the upper right of Figure 6. G(f), in units of g^2/Hz, is given by

$$\begin{aligned}
G(f) &= 0.04 && 5 < f_v < 220 \text{ Hz}\\
G(f) &= 0.07 \times (f_v/300)^2 && 220 < f_v < 300 \text{ Hz}\\
G(f) &= 0.07 && 300 < f_v < 1000 \text{ Hz}\\
G(f) &= 0.07 \times (f_v/1000)^{-2} && 1 < f_v < 2 \text{ KHz}.
\end{aligned} \qquad (22)$$

If the vibration sensitivity is $1 \times 10^{-9}/g$ and the oscillator is operating at 10 MHz, the peak frequency deviation is 0.01 Hz. Therefore, σ_y^2 is

$$\begin{aligned}
\sigma_y^2 &= \frac{(0.04)(0.01)^2}{(2f_v^2)} && 5 < f_v < 220 \text{ Hz}\\
\sigma_y^2 &= \frac{(0.07)(0.01/100)^2}{2} && 220 < f_v < 300 \text{ Hz}\\
\sigma_y^2 &= \frac{(0.07)((0.01)^2}{(2f_v^2)} && 300 < f_v < 1000 \text{ KHz}\\
\sigma_y^2 &= \frac{(0.07)(0.01*1000)^2}{(2f_v^4)} && 1 < f_v < 2 \text{ KHz}
\end{aligned} \qquad (23)$$

which is shown in Figure 6. Note that outside of the vibration frequency range defined by the given acceleration spectrum, the phase noise is identical to that of a non-accelerated device.

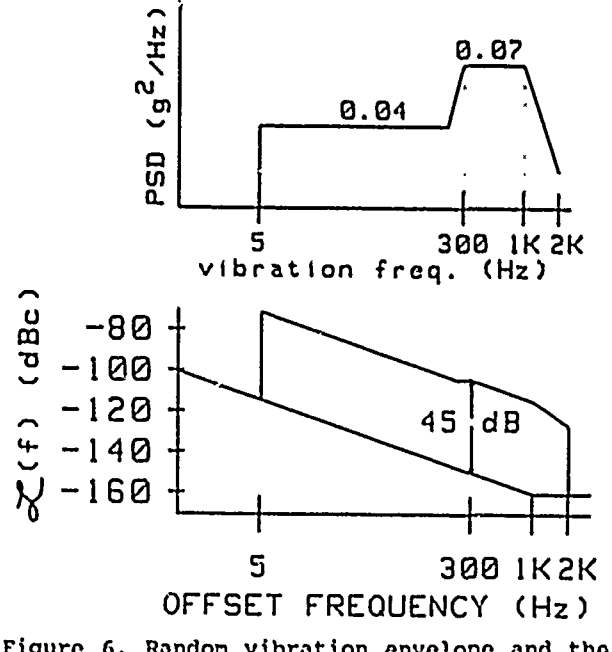


Figure 6. Random vibration envelope and the random vibration-induced phase noise.

INTEGRATED PHASE NOISE AND PEAK PHASE EXCURSION

Specialists in crystal resonators and

oscillators generally characterize phase noise by $S_\phi(f)$ or $\mathscr{L}(f)$. Many users of crystal oscillators, however, characterize phase noise in terms of "phase jitter." Phase jitter is the phase noise integrated over the system bandwidth. Similarly, in phase-lock-loops, it is the magnitude of the phase excursions that determines whether or not the loop will break lock under vibration.

One can use the previous example to investigate the effect of vibration on integrated phase noise. Integrated phase noise is defined, for the band f_1 to f_2, as

$$\phi_i^2 = \int_{f_1}^{f_2} S_\phi(f)\, df, \tag{24}$$

where S_ϕ is the spectral density of phase, equal to $2\mathscr{L}_v$. In the frequency band of 1 Hz to 2 KHz, the phase noise of the nonvibrating oscillator from Figure 6 is given by

$$\begin{aligned} \mathscr{L}_v &= 1\times10^{-10}/f^2 \qquad f < 1 \text{ KHz} \\ \mathscr{L}_v &= 1\times10^{-16} \qquad f > 1 \text{ KHz} \end{aligned} \tag{25}$$

and the integrated phase noise in the same band is

$$\phi_i^2 = 2 \times 10^{-10} \text{ radians}^2. \tag{26}$$

Therefore,

$$\phi_i = 1.4 \times 10^{-5} \text{ radians}. \tag{27}$$

While the oscillator is vibrating, the phase noise is given by Eq. (25) in the band from 1 Hz to 5 Hz, and by Eq. (23) in the band from 5 Hz to 2000 Hz. The integrated phase noise is

$$\begin{aligned} \phi_i^2 = 2 &\int_{f_1}^{f_2} (1\times10^{-10}/f^2)\, df \\ &+ \int_{f_2}^{f_3} (0.04)(.01/f)^2\, df \\ &+ \int_{f_3}^{f_4} (0.07)(0.01/300)^2\, df \\ &+ \int_{f_4}^{f_5} (0.07)(.01/300f)^2\, df \\ &+ \int_{f_5}^{f_6} (0.07)(0.01*1000)^2/f^4\, df \\ &= 8 \times 10^{-7} \text{ radians}^2. \end{aligned} \tag{28}$$

Therefore,

$$\phi_i = 9\times10^{-4} \text{ radians}. \tag{29}$$

While the oscillator is vibrating, it can be seen that the integrated phase noise is 4000 times that of the noise when it is not vibrating and the rms phase deviation, ϕ_i, is about 60 times larger during vibration.

When the oscillator is subjected to a simple sinusoidal vibration, the peak phase excursion follows from Eq. (9), i.e.,

$$\phi_{peak} = \Delta f/f_v. \tag{30}$$

For example, if our 10 MHz, 1 x 10^{-9}/g oscillator is subjected to a 10 Hz sinusoidal vibration of amplitude 1 g, the peak vibration-induced phase excursion is 1 x 10^{-3} radians. If this oscillator is used as the reference oscillator in a 10 GHz radar system, the peak phase excursion at 10 GHz will be 1 radian. Similarly, when an oscillator's frequency is multiplied, the integrated phase noise is increased by the multiplication factor. At 10 GHz, for example, the integrated phase noise in the above example increases from 9 x 10^{-4} radians to 0.9 radians! Such large phase excursions are detrimental to the performance of many systems, such as those which employ phase lock loops or phase shift keying.

MEASUREMENT

The sidebands generated by vibration can be used to measure the acceleration sensitivity. Eq. (15) can be rearranged to get

$$\Gamma_i = (2f_v/A_i f_o)\ 10^{\mathscr{L}_v/20}, \tag{31}$$

where Γ_i is the component of the acceleration sensitivity vector in the i direction. Three measurements, along mutually perpendicular axes, are required to characterize the vector, $\vec{\Gamma}$, which becomes

$$\vec{\Gamma} = \Gamma_i \hat{i} + \Gamma_j \hat{j} + \Gamma_k \hat{k} \tag{32}$$

with a magnitude of

$$|\vec{\Gamma}| = (\Gamma_i^2 + \Gamma_j^2 + \Gamma_k^2)^{1/2} \tag{33}$$

One scheme for measuring $\vec{\Gamma}$ is shown in Figure 7. The local oscillator is used to mix the carrier frequency down to the range of the spectrum analyzer. If the local oscillator is not modulated, the relative sideband levels are unchanged by mixing. The frequency multiplier is used to overcome dynamic range limitations of the spectrum analyzer, using the "20 log N" enhancement discussed previously. The measured sideband levels must be adjusted for the multiplication factor prior to insertion into Eq. (31). It must be stressed that Eq. (17) is valid only if $\beta < 0.1$. A sample measurement output and calculation is given in Figure 8.

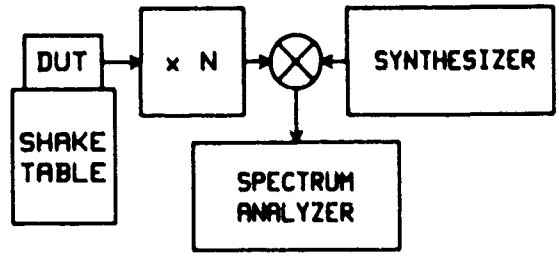


Figure 7. Typical configuration for the measurement of acceleration-sensitivity using sinusoidal excitation.

Other measurement schemes include passive excitation[17] of the resonator, the use of random vibration,[18] and the 2-g tipover test. Many oven-controlled oscillators are not suitable for characterization by the 2-g tipover test, however, because rotation of the oscillator results in temperature changes inside the oven that can mask the effects due to the acceleration changes. Vibration tests are also subject

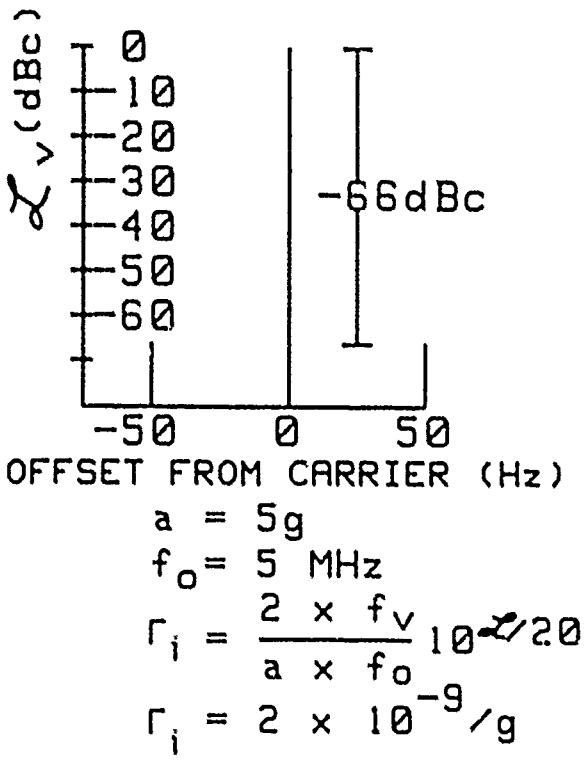


Figure 8. Sample acceleration-sensitivity measurement using sinusoidal acceleration.

to pitfalls since resonances in the oscillator or shake table assembly can produce false results. It is, therefore, important to perform the test at more than one vibration frequency.

THEORY

The theoretical understanding of acceleration sensitivity does not yet enable one to predict the acceleration sensitivity of "real" resonator designs. Theoretical activity started with the study of in-plane forces on simplified resonator models. The first studies were concerned with the closely related force-frequency effect. This effect, first reported by Bottom[19] in 1947, is the change in frequency induced by a pair of opposed forces in the plane of a resonator plate applied at the rim. Those forces distort the quartz plate and, because of the nonlinear elastic behavior, change the acoustic velocity.[11] Since the frequency of a resonator is a function of the acoustic velocity and the dimensions of the quartz plate, the forces change the frequency.

The first attempt at an analytical solution to the force-frequency problem was made by Mingins, Barcus, and Perry in 1962.[20] They assumed an AT-cut crystal plate of infinite lateral dimension and used a perturbation technique with linear elastic coefficients. In 1963, they discussed nonlinear theory and the need to include the third-order elastic coefficients.[21] Those coefficients were measured in 1966 by Thurston, McSkimin, and Andreatch.[22] The first calculation to use the nonlinear theory was performed by Keyes and Blair in 1967.[23] In 1973, Lee, Wang, and Markenscoff[24] published the first of a continuing series of papers (by Lee and his students,) in which they calculated the force-frequency coefficient (as defined by Ratajski[25]) as a function of azimuth, using the general theory of incremental elastic deformations superimposed on finite initial deformations. The next year, Lee et al. investigated the frequency change due to cantilever bending of the plate.[26] In 1977, a variational analysis of the force-frequency effect, including the effect of material anisotropy on initial stress, was given by DeNiese, Ballato, and Lukaszek[27] for doubly rotated cuts. In 1978, Janiaud, Nissim, and Gagnepain[28] obtained analytic solutions for the biasing stress in singly and doubly rotated plates subject to diametric forces. Those results allowed them to calculate Ratajski's force-frequency coefficient, and also the in-plane acceleration sensitivity. Lee and Wu[29] extended the work of reference 24 to treat plates of any cut. In that paper, the solution for an arbitrary number of ribbon supports was obtained.

Under acceleration, body forces in the quartz plate are balanced by reaction forces from the mounting structure. As in the force-frequency effect, distortion of the crystal lattice causes the resonant frequency to change. For bulk-wave resonators, only the in-plane case has been treated. The first analysis was by Valdois, Besson, and Gagnepain[11] in 1974, who showed that the effect is linear. Several papers by Lee and Wu[29-31] considered resonators with three- and four-point mounts. Two-point mounts were considered by Janiaud, Nissim, and Gagnepain.[28] For SAW resonators, recent work by Shick and Tiersten[32] and by Sinha and Locke[33] has treated acceleration both in the plane of the plate and normal to it.

There seems to be a significant dependence of the acceleration sensitivity on small variations in the mounting. Analyses of bulkwave resonators, to-date, have considered only point supports. Real resonators have supports that are distributed over a finite area. In addition, initial stress conditions are difficult to determine and may not be well-reproduced. There is an ongoing effort by Lee and Tang[34] to use finite element analysis to more accurately model the mounting structure. It is hoped that this will allow computer simulations that will permit the determination of the optimum design parameters for minimizing acceleration sensitivity.

EXPERIMENTAL RESULTS

Several papers have reported experimental results on the force-frequency effect[20,25,35] and the effects of bending moments.[36] All used point mounts in special fixtures, and results agree fairly well with the theoretical analyses. The reported experimental observations of the actual acceleration sensitivity of real resonators, on the other hand, is remarkable in that it defies simple explanation.

The experimental effort on acceleration-sensitivity started in the 1960's with Smith, Spencer, and Warner.[1-5] Valdois, Besson, and Gagnepain[11] made measurements on

resonators and oscillators to demonstrate that the resonator was the acceleration-sensitive element. Results from 2-g tipover experiments were reported for SC-cut resonators by Kusters, Adams, Yoshida, and Leach[37] in 1977. In 1979, Warner, Goldfrank, Meirs, and Rosenfeld[38] reported results on SC- and FC-cut devices. In 1981, the first of a series of papers intended to explore a wide range of resonator fabrication parameters was published by Filler and Vig.[39] In that same year, Nakazawa, Lukaszek, and Ballato[40] reported constant-acceleration results that suggested an acceleration sensitivity that is not linear with acceleration level. The only significant parameter for the reduction of acceleration-sensitivity was reported by Filler, Kosinski, and Vig[41] in 1982. As one makes plano-convex and biconvex AT-cut resonators flatter, the acceleration sensitivity decreases. It must be noted, however, that other effects, as yet undetermined, cause large scatter in the data. A survey of a large number of resonator parameters, such as SC-cut contour, thickness, drive level, temperature, and angle of cut, was published in 1983 by Filler, Kosinski, and Vig.[42] No significant correlations were found.

The most recent paper, by Weglein[43] in 1984, reported the acceleration sensitivity of VHF AT- and SC-cut resonators. All of the resonators in that study were disassembled after the acceleration sensitivity measurements to examine fabrication details. That investigation, which seems to summarize the efforts to date, showed "little dependence on any recognizable parameter," except that the lowest average acceleration sensitivity was found in the group of AT units. The large spread in the data, typical of all experimental observations of acceleration-sensitivity, seems to be caused by a combination of sometimes offsetting subtle effects, which are difficult to control during fabrication.

REDUCTION OF THE ACCELERATION SENSITIVITY OF RESONATORS

The introduction of the stress-compensated cut of quartz, i.e., the SC-cut, was accompanied by the wide-spread expectation that the SC-cut would have significantly lower acceleration sensitivity than other cuts. Unfortunately, that expectation has not been realized. The lowest acceleration sensitivity achieved with AT-cut resonators[41-43] equals that achieved with SC-cut resonators; both cuts can have low parts in 10^{10}/g sensitivity.

Efforts to reduce the sensitivity of individual resonators to the effects of acceleration have stressed the support structure. Lukaszek and Ballato[44] proposed a plate geometry that would assure the proper support configuration to reduce the force-frequency effect. Besson, Gagnepain, Janiaud, and Valdois[45] proposed a support structure that insured symmetry with the median plane of the resonator plate. Debaisieux, Aubry,and Groslambert[46] and Aubry and Debaisieux[47] have reported results using QAS resonators, a variation of the BVA,[48] which insures symmetry of mount as well as accurate mount locations for reducing the force-frequency effect. Their results show a marked reduction in the scatter of the measured acceleration sensitivity.

ACCELERATION COMPENSATION OF OSCILLATORS

The lack of progress in reducing the acceleration sensitivity of the resonator below the low parts in 10^{10}/g level has spawned several techniques for compensation of the effect. There are two general classes of compensation, passive and active. The first compensation results were published by Gagnepain and Walls in 1977.[49] They used the passive method of mechanically arranging two resonators such that the components of the acceleration sensitivities normal to the plates were antiparallel. The resonators were electrically connected in series in a single oscillator. Przyjemski[12] and Emmons[50] used an active technique. They sensed the acceleration magnitude with an accelerometer aligned with the direction of the acceleration sensitivity vector of the resonator. The accelerometer signal was fed into a tuning circuit in the oscillator in order to counter the acceleration-induced frequency changes. A limitation of this technique is the requirement on the linearity of the tuning network at all operating points. Emmons also employed the dual-resonator technique and suggested using an acceleration-sensitive capacitor in the tuning network of the oscillator. The latter technique is available as an option on a commercial cesium-beam frequency standard.

Ballato[51] suggested a method for compensation in all directions using a resonator pair made of enantiomorphs. He argued that opposite handedness is the only way to have all three crystallographic axes line up antiparallel. This was extended by a series of patents by Ballato and Vig.[52-54] A simplification was patented by Filler,[55] who showed that, since the acceleration-sensitivity has vector properties, all that is required is that the vectors be aligned antiparallel, independent of the handedness of the quartz. Vig and Walls[56] extended this work by suggesting a method to accommodate resonators with different acceleration sensitivity magnitudes.

Rosati patented[57] an active technique of compensation which was further developed by Rosati and Filler.[58] This method makes use of the polarization effect in doubly rotated resonators, i.e., that the resonant frequency of a doubly rotated resonator is a function of the voltage applied to the electrodes.[59] If one senses the acceleration using an accelerometer and feeds that signal with appropriate amplification and phase reversal directly to the resonator electrodes, compensation can be achieved. The advantage of feeding the correction signal to the resonator electrodes rather than a varactor is that the polarization effect has superior linearity. When the correction signal is applied to a varactor, the nonlinearity of the varactor and of the frequency-capacitance function causes sidebands at harmonics of the vibration

frequency. One implementation of the polarization effect technique was used to compensate a rubidium oscillator.[60]

CONCLUSION

Vibration effects are a significant problem in modern communication, navigation, and radar systems. Progress has been made in understanding the causes of acceleration sensitivity but a full explanation for real devices has been elusive. A great deal of effort has been expended on compensating for acceleration-sensitivity, but more work is needed to improve the level and bandwidth of the compensation.

ACKNOWLEDGMENTS

For many helpful discussions, the author thanks Mr. Robert Smythe of Piezo Technology, Inc.,and Dr. John Vig and Mr. Vincent Rosati of US Army LABCOM. In addition, the author thanks Mr. Martin Bloch for generating the initial interest in this problem.

REFERENCES

1. A. W. Warner and W. L. Smith, "Quartz Crystal Units and Precision Oscillators for Operation in Severe Mechanical Environments,"14th ASFC*, pp 200-216, 1960.

2. W. L. Smith, "An Ultra-Precise Standard of Frequency," Final Report, Contract DA 36-039 sc-73076, US Army, 1960.

3. W. J. Spencer and W. L. Smith, "Precision Crystal Frequency Standards," 15th ASFC*, pp 139-155, 1961.

4. W. J. Spencer and W. L. Smith, "Precision Quartz Crystal Controlled Oscillator for Severe Environmental Conditions," 16th ASFC*, pp406-421, 1962.

5. W. L. Smith and W. J. Spencer, "Quartz Crystal Controlled Oscillators," Final Report, Contract DA 36-039 sc-85373, US Army, 1963.

6. D. B. Leeson, "Aerospace Crystal Environmental Requirements," 19th ASFC*, pp49-57, 1965.

7. D. B. Leeson and G. F. Johnson, "Short-Term Stability for a Doppler Radar: Requirements, Measurements, and Techniques," Proc. IEEE, 54, pp 244-248, 1966.

8. G. F. Johnson, "Vibration Characteristics of Crystal Oscillators," 21st ASFC*, pp 287-293, 1967.

9. J. Moses, "NAVSTAR Global Positioning System Oscillator Requirements for the GPS Manpack," 30th ASFC*, pp 390-400, 1967.

10. J. M. Przyjemski and P. L. Konop, "Limitations on GPS Receiver Performance Imposed by Crystal-Oscillator g-Sensitivity," NAECON '77, 1977.

11. M. Valdois, J. Besson, and J. J. Gagnepain, "Influence of Environment Conditions on a Quartz Resonator," 28th ASFC*, pp 19-32, 1974.

12. J. M. Przyjemski, "Improvement in Syst Performance using a Crystal Oscillat Compensated for Acceleration Sensitivity, 32nd ASFC*, pp 426-431, 1978.

13. M. Abramowitz and I. A. Stegu "National Bureau of Standards, Appli Mathematics Series, No. 55," Governme Printing Office, Washington, DC, 196 Revised by Dover, New York, 1965.

14. R. L. Filler, "The Effect of Vibrati on Frequency Standards and Clocks," 35 ASFC*, pp 31-39, 1981.

15. S. Stein, private communication, 198 using eq. 23 from Barnes et. al. "Characterization of Frequency Stability, IEEE Trans. IM, vol. IM-20, 1971.

16. D. Halford, J. S. Shoaf, and A. Risley, "Spectral Density Analysis: Frequency Domain Specification and Measurement Signal Stability," 27th ASFC*, pp 421-43 1973.

17. F. L. Walls and A. E. Wainwright "Measurement on the Short-term Stability Quartz Crystal Resonators and the Implica tion for Crystal Oscillator Design," IEE Trans. on I&M, IM-24,pp 15-20, 1975.

18. D. J. Healy, III, H. Hahn, and Powell, "A Measurement Technique for Deter mination of Frequency vs. Acceleratio Characteristics of Quartz Crystal Units, 37th ASFC*, pp 284-289, 1983.

19. V. Bottom, "Note on the Anomalous Ther mal Effect in Quartz Oscillator Plates, American Mineralologist,Vol. 32 , pp 590 591, Sept - Oct 1947.

20. C. R. Mingins, L. C. Darcus, and R. Perry, "Effects of External Forces on th Frequency of Vibrating Crystal Plates, 16th ASFC*, pp 47-76, 1962.

21. C. R. Mingins, L. C. Darcus, and R. Perry, "Reactions of a Vibrating Piezoelec tric Plate to Externally Applied Forces, 17th ASFC*, pp 51-87, 1963.

22. R. N. Thurston, H. J. McSkimin, and Andreatch, Jr., "Third -Order Elastic Coef ficients of Quartz,": J. Applied Phys. vo 37, pp 267-275, January 1966.

23. R. W. Keyes and F. W. Blair, "Stres Dependence of the Frequency of Quart Plates," Proc. IEEE (letters), pp565-566 1967.

24. P. C. Y. Lee, Y. S. Wang, and J Markenscoff, "Elastic Waves and Vibration in Deformed Crystal Plates," 27th ASFC*, p 1-6, 1973.

25. J. M. Ratajski, "The Force Sensitivit of AT-Cut Quartz Crystals," 20th ASFC* pp33-48, 1966.

26. P. C. Y. Lee, Y. S. Wang, "Effects o Initial Bending on the Resonance Frequencie of Crystal Plates," 28th ASFC*, pp 14-18 1974.

27. A. Ballato, E. P. EerNisse, and T. Lukaszek, "The Force-Frequency Effect in Doubly Rotated Quartz Plates," 31st ASFC*, pp 8-16, 1977and E. P. EerNisse, T. Lukaszek, and A. Ballato, "Variational Calculation of Force-Frequency Constants of Doubly Rotated Quartz Resonators," IEEE Trans. Sonics Ultrason., Vol. SU-25, No. 3, pp 132-138, May 1978.

28. D. Janiaud, L. Nissim, and J.-J. Gagnepain, "Analytical Calculation of Initial Stress Effects on Anisotropic Crystals Application to Quartz Resonators," 32nd ASFC*, pp 169-179, 1978.

29. P. C. Y. Lee and K. M. Wu, "Nonlinear Effect of Initial Stresses in Doubly - Rotated Resonator Plates," 34th ASFC*, pp 403-411, 1980.

30. P. C. Y. Lee and K. M. Wu, "Effects of Acceleration on the Resonance Frequencies of Crystal Plates," 30th ASFC*, pp 1-7, 1976.

31. P. C. Y. Lee and K. M. Wu, "The Influence of Support Configuration on the Acceleration Sensitivity of Quartz Resonator Plates," 31st ASFC*, pp 29-34, 1977.

32. D. V. Shick and H. F. Tiersten, "An Analysis of the Acceleration Sensitivity of ST-Cut Quartz Surface Wave Resonators Supported along the Edges," 40th ASFC*, pp 262-268, 1986.

33. B. K. Sinha and S. Locke, "Acceleration and Vibration Sensitivity of SAW Devices," IEEE Trans. on UFFC, vol UFFC-34, 1, pp 29-38, January 1987.

34. P. C. Y. Lee and M. S. H. Tang, "Initial Stress Field and Resonance Frequencies of Incremental Vibrations in Crystal Resonators by Finite Element Method," 40th ASFC*, pp 152-160, 1986.

35. A. D. Ballato, "Effects of Initial Stress on Quartz Plates ' Vibrating in Thickness Modes," 14th ASFC*, pp 89-114, 1960.

36. 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," 33rd ASFC*, pp 346-350, 1979.

37. J. A. Kusters, C.A. Adams, H. Yoshida, and J. G. Leach,"TTC's - Further Developmental Results," 31st ASFC*, pp 3-7, 1977

38. A. Warner, B. Goldfrank, M. Meirs, and M. Rosenfeld, "Low "g" Sensitivity Crystal Units and Their Testing," 33rd ASFC*, pp 306-310A, 1979.

39. R. L. Filler and J. R. Vig, "The Acceleration and Warmup Characteristics of Four - Point - Mounted SC and AT-Cut Resonators," 35th ASFC*, pp 110-116, 1981.

40. M. Nakazawa, T Lukaszek, and A Ballato, "Force - and Acceleration- Frequency Effects in Grooved and Ring - Supported Resonators," 35th ASFC*, pp 71-91, 1981.

41. R. L. Filler, J. A. Kosinski, and J. R. Vig, "The Effect of Blank Geometry on the Acceleration Sensitivity of AT & SC-Cut Quartz Resonators," 36th ASFC*, pp 215-219, 1982.

42. R. L. Filler, J. A. Kosinski, and J. R. Vig, "Further Studies on the Acceleration Sensitivity of Quartz Resonators," 37th ASFC*, pp 265-271,1983.

43. R. D. Weglein, "The Vibration Sensitivity of VHF Quartz Crystals for Missile Applications," 38th ASFC*, pp 73-79, 1984.

44. T. J. Lukaszek, and A. Ballato, "Resonators for Severe Environments," 33rd ASFC*, pp 311-321, 1979.

45. R. Besson, J.- J. Gagnepain, D. Janiaud, and M. Valdois, "Design of a Bulk Wave Quartz Resonator Insensitive to Acceleration," 33rd ASFC*, pp37-345, 1979.

46. A. Debaisieux, J. P. Aubry, and J. Groslambert, "Design of SC Cut 10 MHz H.Q. Crystals with g Sensitivity Better than 2.10^{-10}/g," Proc. 15th Proc. Time & Time Int. Conf. (PTTI), pp 635-650, 1983.

47. J. P. Aubry and A. Debaisieux, "Further Results on 5 MHz and 10 MHz Resonators with BVA and QAS Designs," 38th ASFC*, pp 190-200, 1984.

48. R. J. Besson, "A New Piezoelectric Resonator Design," 30th ASFC*, pp 78-83, 1976,

49. J.- J. Gagnepain and F. L. Walls, "Quartz Crystal Oscillators with Low Acceleration Sensitivity," NBSIR 77-855, National Bureau of Standards, 1977.

50. D. A. Emmons, "Acceleration Sensitivity Compensation in High Performance Crystal Oscillators," 10th Proc. Time & Time Int. Conf. (PTTI), 1978.

51. A. Ballato, "Crystal Resonators with Increased Immunity to Acceleration Fields, "IEEE Trans S&US, vol SU-27,No. 4, pp 195-201, July 1980; or "Resonators Compensated for Acceleration Fields," 33rd ASFC*, pp 322-336, 1979.

52. A. Ballato and J. R. Vig, "Acceleration Resistant Combination of Opposite - Handed Piezoelectric Crystals," US Patent No. 4,344,010, 1982.

53. A. Ballato and J. R. Vig, "Method of Fabricating Acceleration Resistant Resonators , Resonators so Formed," US Patent No. 4,365,182, 1982.

54. A. Ballato and J. R. Vig, "Method of Fabricating Acceleration Resistant Crystal Resonators," US Patent No. 4,409,711, 1983.

55. R. L. Filler, "Acceleration Resistant Crystal Resonators," US Patent No. 4,410,822, 1983.

56. J. R. Vig and F. L. Walls, "Acceleration Insensitive Oscillator," US Patent No. 4,575,690, 1986.

57. V. J. Rosati, "Suppression of Vibration Effects on Piezoelectric Crystal Resonators," US Patent No. 4,453,141, 1984.

58. V. J. Rosati and R. L. Filler, "Reduction of the Effects of Vibration on SC-Cut Quartz Crystal Oscillators," 35th ASFC*, pp 117-121, 1981.

59. J. Kusters, "The Effect of Static Electric Fields on the Elastic Constants of Alpha Quartz," 24th ASFC*, pp 46-54,1970.

60. C. Colson, "Vibration Compensation of the Seaktalk Rubidium Oscillator," 36th ASFC*, pp 197-199, 1982.

*ASFC - Annual Symposium on Frequency Control, copies available from

1956-1977, 1982	National Technical Information Service Sills Building 5285 Port Royal Road Springfield, VA 22161
1979 - 1891	Electric Industries Association 2001 Eye Street Washington, DC 20006
1983 - 1986	Institute of Electrical & Electronic Engineers 445 Hoes Lane Piscataway, NJ 08854

SUPPRESSION OF VIBRATION-INDUCED PHASE NOISE IN CRYSTAL OSCILLATORS: AN UPDATE

Vincent J. Rosati
US Army Electronics Technology and Devices Laboratory
Fort Monmouth, NJ 07703-5000

Summary

Many military systems, such as those based on coherent Doppler radar, require very low phase noise frequency sources. However, quartz crystal resonators are sensitive to acceleration, so on vibrating platforms the the oscillator is frequency modulated, severely degrading the oscillator phase noise.

Several methods have been proposed for reducing vibration (acceleration) effects without recourse to bulky and massive vibration isolators. Among these are: improved resonators; resonator pairs[1,2,3]; varactor tuning[4]; capacitive-accelerometer tuning[5]; and polarization-effect tuning[6]. Some progress has been made in all of these areas during the past several years. However, the ultimate goal of building an oscillator that has an effective g-sensitivity of about 2 x 10^{-12}/g remains unattained.

This paper describes recent results obtained using the "polarization-effect tuning" method. Using this method, a commercial oscillator's acceleration sensitivity was reduced from $8X10^{-10}$/g to $8X10^{-12}$/g, over a vibration-frequency band of 5 Hz to 350 Hz. An analysis of the amplitude and phase accuracy for deep suppression of vibration-induced sidebands is also presented.

Introduction

The phase noise performance of quartz crystal oscillators under vibration has become recognized as one of the serious limiting factors in the performance of certain radar and navigation systems. For example, assuming a 10 MHz oscillator that uses a resonator with an acceleration sensitivity of $5X10^{-10}$ per g, a 1g vibration at 100 Hz produces a bright line that is only 92 dB below the carrier, a level that is about 60 dB worse than required.[7,8]

Figure 1 shows the single-sideband phase noise of a good commercial oscillator at rest, and the phase noise of the same oscillator while undergoing only 0.01 g^2/Hz random vibration from 10 Hz to 1 kHz . The degradation of phase noise below 100 Hz is more than 50 dB. Clearly, methods for reducing vibration effects must be developed further, with repeatability, producibility, and cost in mind.

Polarization-Effect Tuning

The polarization-effect tuning method was described previously by Rosati and Filler[6]. Simply stated, the method consists of sensing the acceleration with an accelerometer, amplifying and phase inverting the sensing signal, and feeding it directly to the electrodes of the SC-cut resonator in the oscillator, through isolation resistors. The SC-cut resonator, via the polarization effect, is tuned to oppose the frequency change induced by the acceleration.

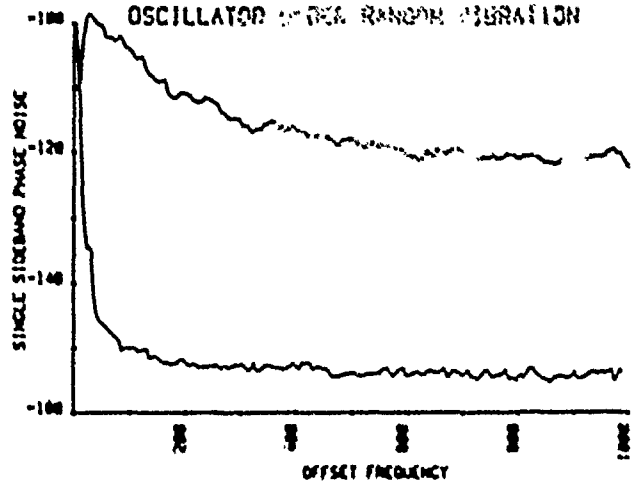


Figure 1. Single-sideband phase noise of a quality oscillator at rest (lower trace), versus random vibration (upper trace).

Early attempts at suppression of vibration-induced phase noise were limited by the accelerometer and the circuitry used for amplification. Figure 2 is representative of the early work, showing only 15 dB of suppression at low vibration

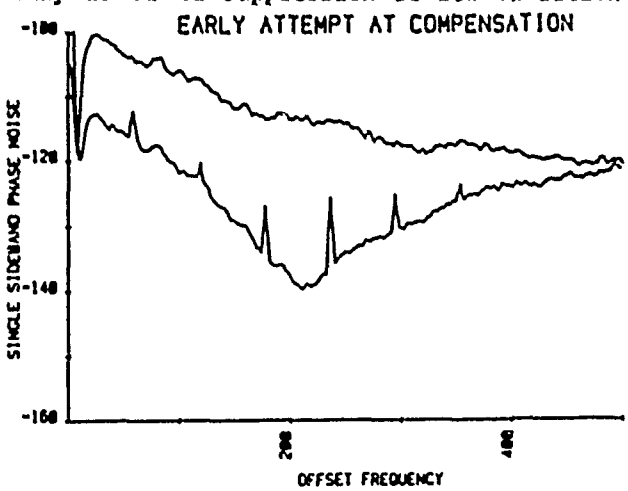


Figure 2. Results of an early attempt at reducing vibration-induced phase noise using an accelerometer-derived feedback signal directly to the resonator electrodes. The lower trace is the phase noise after the suppression signal is applied.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

Figure 7 is a simplified schematic diagram of the amplifiers used in the present work. Input to the charge amplifiers is from a three-axis piezoelectric accelerometer, having charge sensitivity of 12 picocoulombs/g. The 20 pF capacitor is supplied by the accelerometer cable. The 300 megohm resistor tailors the accelerometer low-frequency response and provides a path for bias current of the input FET. The gain amplifier may be switched from inverting to noninverting to provide the correct phase for the feedback signal for its axis. A gain of about 40 was sufficient for the resonator used. Not shown on the schematic is that the feedback resistor of the gain amplifier consists of two variable resistors, one for coarse- and one for fine gain control. The summing amplifier output is fed to each crystal electrode through 100 kilohm resistors.

Figure 8 shows the close-to-the-carrier result, obtained by retrofitting an existing oscillator to accept the suppression signal. The upper trace is the vibration-induced phase noise due to 0.01 g^2/Hz random vibration between 10 Hz and 250 Hz. The lower trace is the phase noise when suppression is applied, showing 35 to 40 dB of suppression. The *'s at the lower left corner of the graph represent suppressed sinusoidal vibration at 5, 10, and 15 Hz. Sinusoidal excitation was used because of the low-frequency cutoff of the random-noise generator driving the shake table. The figure shows, then, that 35 to 40 dB of suppression is obtained over more than 5 octaves, from 5 to 250 Hz

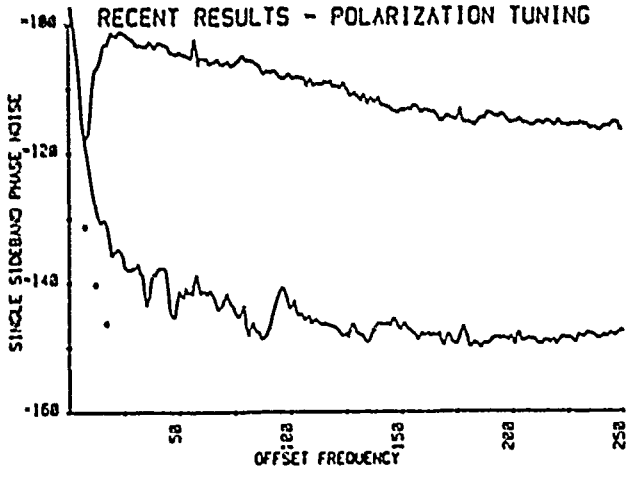


Figure 8. Reduction of vibration-induced phase noise using the circuit of Fig. 7.

Figure 9 is from the same data as Figure 8, except that noise to 1 kHz is shown. The reason for the rolloff of suppression above 300 Hz is not yet fully identified. However, it is thought at this time to be due to the suppression signal (on the order of 1/2 volt) leaking through to the oscillator ALC circuitry. The oscillator is then mistuned via the amplitude-frequency effect. To test this hypothesis a constant-amplitude, swept audio-frequency signal was applied to two non-vibrating oscillators of different circuit design. The FM sidebands thus generated should fall as $1/f^2$. In one

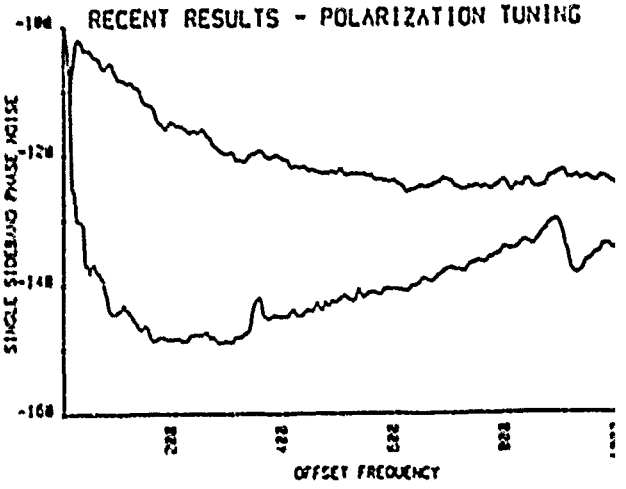


Figure 9. The effect of the suppression circuit to 1 kHz is shown. The rolloff of suppression at higher vibration frequencies is not fully identified.

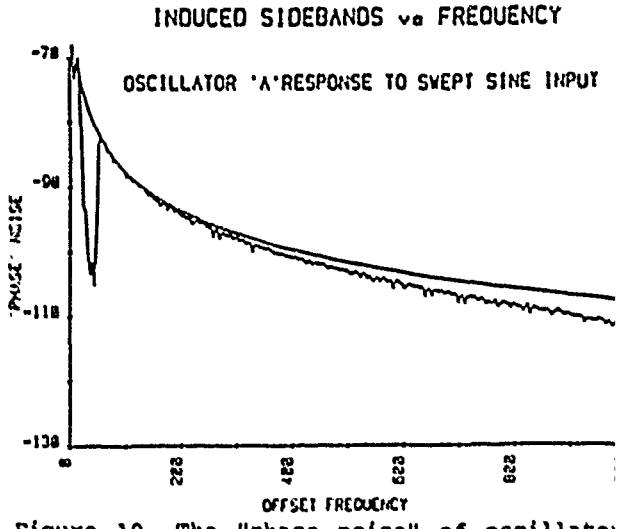


Figure 10. The "phase noise" of oscillator "A" due to applying a constant-amplitude, swept-frequency signal to the resonator. The solid curve is the computed $1/f^2$ rolloff.

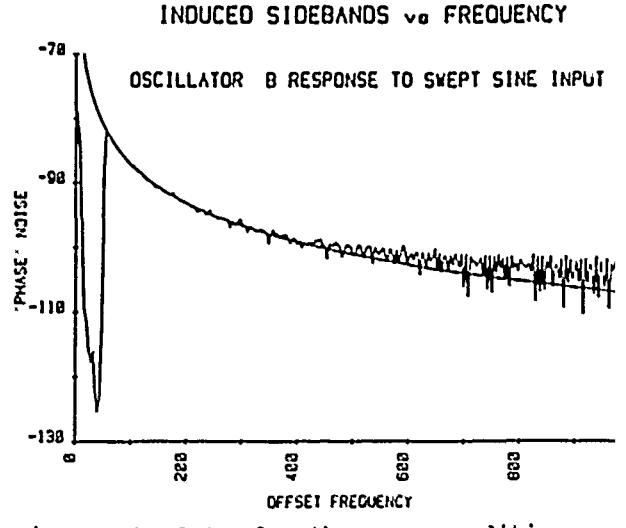
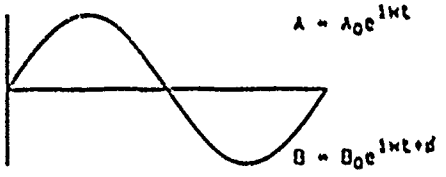


Figure 11. Data for the same conditions as Fig. 10, but for oscillator "B".

frequencies, increasing to 25 dB at 200 Hz, then decreasing toward 0 dB at 500 Hz. Indeed, beyond 500 Hz, the "suppression" circuit actually made the noise worse!

Subsequent to the early work, which demonstrated that the method is valid, an analysis was performed to determine the sensitivity of the method to phase and/or amplitude changes in the accelerometer feedback path. The analysis model, which should apply to all methods of suppression (except vibration isolation), is the cancellation of one sine wave with another, as shown in Figure 3.

PHASE AND AMPLITUDE REQUIREMENTS FOR SINE WAVES TO CANCEL



'INTENSITY' OF RESULTANT IS: $S^2 = A_0^2 + B_0^2 + 2A_0B_0\cos(\delta)$

RATIO OF INTENSITIES IS: $R = S^2/A_0^2$

LETTING $A_0 = 1$: $10 \log R = 10 \log(1 + B_0^2 + 2B_0\cos(\delta))$

Figure 3. The constraints on the amplitude and phase accuracy for cancellation of sine waves are derived from this model.

Figures 4 and 5 show the depth of suppression obtainable as the amplitude or phase is varied. Note that to achieve 50 dB of suppression, the amplitude must be held constant to better than 0.3 percent, while phase must be held to within better than 0.2 degrees. Figure 6 is a qualitative depiction of the two-parameter sensitivity "surface".

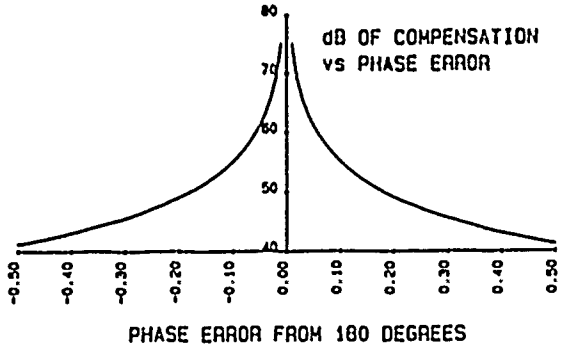


Figure 4. The variation of compensation vs. phase error is shown.

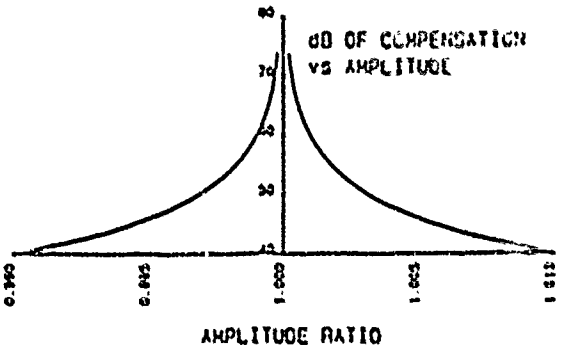


Figure 5. The variation of compensation vs. amplitude ratio is shown.

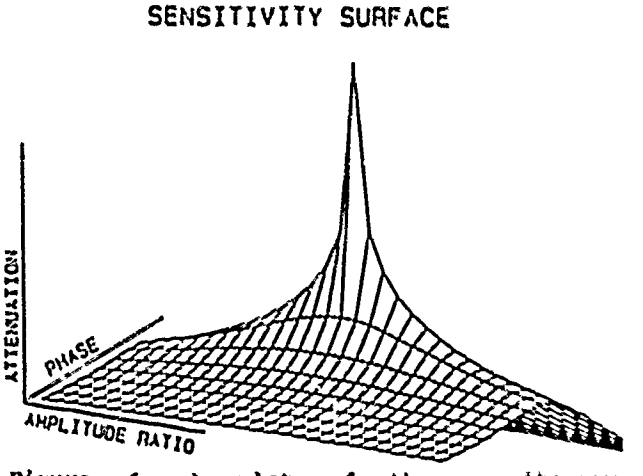


Figure 6. A plot of the sensitivity "surface" of suppression vs. amplitude and phase accuracy.

The analysis makes clear the need for high gain-bandwidth-product amplifiers, rigid oscillator construction to avoid mechanical resonances, and close proximity of the resonator and accelerometer to avoid phase shifts due to the finite velocity of propagation of the vibration wave between the resonator and accelerometer. Close proximity is also necessary in order to minimize acceleration level differences due to rotations.

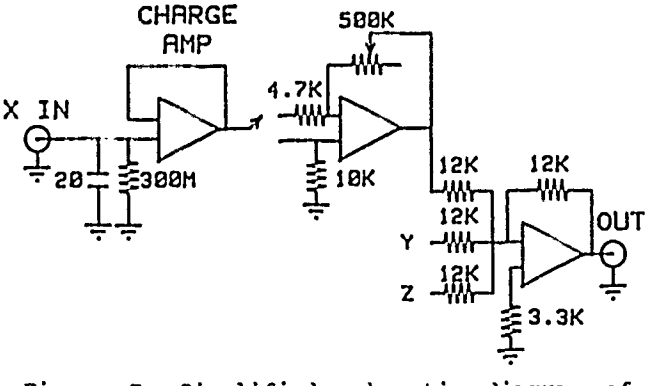


Figure 7. Simplified schematic diagram of the suppression circuit. The X-channel is identical to the Y and Z channels. The X, Y, and Z channels are summed in the last stage.

oscillator the rolloff was slower than $1/f^2$, while in the other it was faster than $1/f^2$, as shown in Figures 10 and 11. This suggests that the rolloff is oscillator circuit dependent, and that careful design is necessary to avoid this interaction.

Concern has been expressed that the mere presence of the suppression circuitry would degrade the phase noise of the non-vibrating oscillator because no amplifier is noise free. However, the mechanism of adding noise to the oscillator is the transfer function of the polarization effect, which is about $5X10^{-9}$/volt. Therefore, 1 microvolt of noise from the amplifier contributes to oscillator frequency instability by $5X10^{-15}$, which is well below the level of concern. Figure 12 shows the phase noise of a test oscillator with the suppression circuitry connected, and then short-circuited. Note that the phase noise is slightly lower when the circuit is active. This is possibly due to suppression of low-level building vibrations. Similar curves have been observed out to 50 kHz, the limit of our baseband spectrum analyzer.

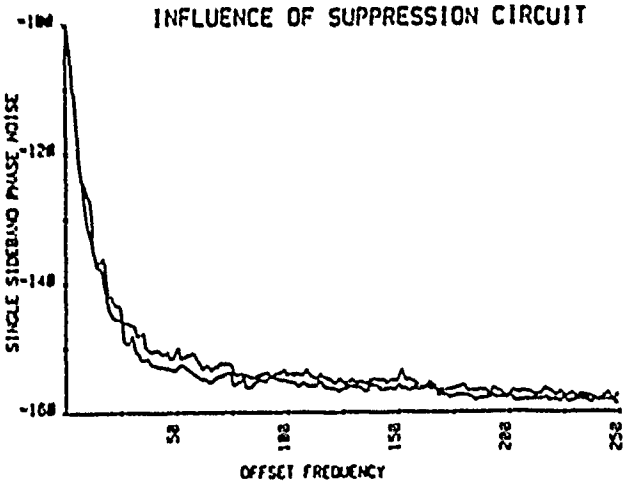


Figure 12. Behavior of a (non-vibrating) quality oscillator with and without the suppression circuit active. The lower trace is obtained when the circuit is active.

Conclusion

The data that have been presented shows that the acceleration sensitivity of a retrofitted oscillator can be improved by two orders of magnitude, at least between 5 and 250 Hz, and that this approach is viable. Future work will be concerned with designing an oscillator with this suppression technique in mind, with the goal of extending suppression to at least 2 kHz. To achieve better than 40 dB of suppression over a broad band, it is possible that this method will need to be combined with some degree of mechanical isolation.

Acknowledgements

The author wishes to thank Mr. Dennis Bowman for the layout and construction of the amplifier circuitry, Mr. Peter Thompson for conducting the surgery on the retrofitted oscillators, and Mr. Stanley Schodowski for suggesting that interaction with the ALC circuit is responsible for the rolloff of suppression at higher frequencies.

REFERENCES

1. A. Ballato, "Resonators Compensated for Acceleration Fields", Proc. 33rd ASFC, pp 322-336, 1979.

2. J.R. Vig and F.L. Walls, "Acceleration Insensitive Oscillators", U.S. Patent No. 4,575,690, 1986.

3. R.L. Filler, "Acceleration Resistant Crystal Resonators", U.S. Patent No. 4,410,822, 1983.

4. J.M. Przyjenski, "Improvement in System Performance Using a Crystal Oscillator Compensated for Acceleration Sensitivity", Proc. 32nd ASFC, pp 426-431, 1978.

5. D.A. Emmons, R.M. Garvey, and B. Milliren, "Acceleration Sensitivity Compensation and Gamma Symmetry In Quartz Oscillators", Proc. 38th ASFC, pp 80-91, 1984.

6. V.J. Rosati and R.L. Filler, " Reduction of the Effects of Vibration on SC-Cut Quartz crystal Oscillators", Proc. 36th ASFC, pp 117-121, 1981.

7. R.L. Filler, "The Effects of Vibration on Frequency Standards and Clocks", Proc. 35th ASFC, pp 31-39, 1981.

8. R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators:A Review", these proceedings.

41st Annual Frequency Control Symposium - 1987

CRYSTAL OSCILLATORS FOR AIRBORNE APPLICATIONS

by Guy CARET, André DEBAISIEUX, Evelyne GERARD and Gilles ROBICHON
CEPE - 44, Avenue de la Glacière 95100 ARGENTEUIL Cedex FRANCE

SUMMARY

Requirements of state of the Art radar systems for military applications, involve the use of high stability and high spectral purity sources. These performances are achieved by using a high Q resonator.
The purpose of this paper is to review the state of the Art as regard to the selection of these resonators and the definition of oscillators.
CEPE proposes several new configurations which are designed in order to work in the frequency range 45 to 140 MHz.
A modular conception, using different hybrid functions (thermal regulator, oscillators, amplifiers, switch) allows to have the oscillators performances in accordance to customers requirements. For each design, a finite elements analysis allows to garantee an acoustical behaviour up to 20 KHz and a mechanical transfer function up to 3 KHz.
These new oscillators are compatible with the following parameters :

- small size (< 55 cm³ without absorber)
- vibration and acoustical environment
- purity or spectral density (floor <-170 dBc/Hz Fo = 125 MHz)
- switching time (< 1 µs)
- low g sensitivity (1.10^{-10} < s / g < 5.10^{-10})

I - INTRODUCTION

Requirements of the state of the Art radar systems for military applications (aircraft, helicopters tanks, and other vehicles) involve the use of high stability and high spectral purity sources. The use of bulk acoustic wave oscillators has solved the system'sneeds during years 80's.

As regard 90's exigences, the increment of levels and frequency range of random vibrations imply new manufacturing concepts, in order to reduce sources of frequency unstability due to mechanical stress.

Our investigations have been concentrated on the different causes of degradations. Three axis of analysis were developped :

- the dynamic analysis of resonator
- mechanical structure
- printed circuit,

during the random vibrations test.

The purpose of this paper is to present the work performed at CEPE labs on a new family of oscillators, the originality of which is to obtain a prediction of phase noise performance whatever the mechanical environment is.

II - RESONATOR SELECTION

The obtaining of phase noise performance of an oscillator under mechanical environment reveals the great importance of the resonator's electrical and mechanical characteristics.

II.1 Influence of the electrical parameters

(static measurement)

Without acceleration and vibrations, the spectral purity of an oscillator is characterized by $\mathcal{L}$(fm), defined as the ratio between the effective phase noise power (in a sideband) in a 1 Hz window at a distance fm from the carrier and a total power of the signal. In this case, $\mathcal{L}$ (fm) is expressed in dBc / Hz.

However, the phase noise performance close to the carrier depends on the value of loaded resonator Q. The experimental phase noise measurement (figures 1 and 2) shows that the effective resonator Q is typically 50 per cent of the unloaded resonator Q. In these conditions, the crystal cut selection is quite important, so at CEPE, we choose AT, BT or SC cuts.

CH2427-3/87/0000-413 $1.00 C 1987 IEEE

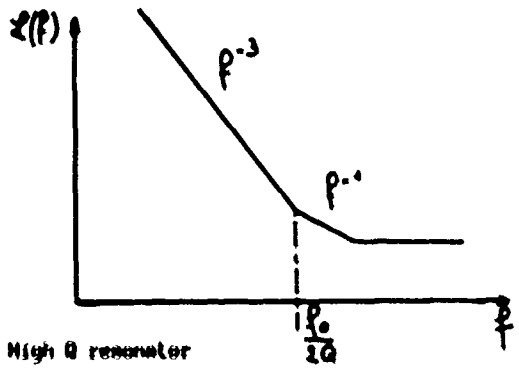


Figure 1

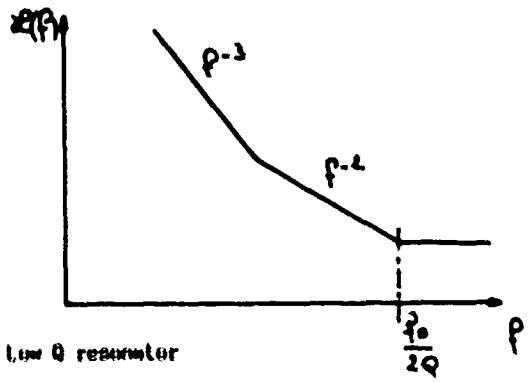


Figure 2

II.2 Influence of the mechanical parameters
(dynamic measurement)

The static phase noise performances observed on oscillators are spoiled when they are exposed to mechanical and acoustic environment tests.

II.2.1 Degradation due to random vibration
(carrier than 3 KHz)

A phase noise built up from the carrier to 3KHz depends on the g sensitivity of the resonator (if not limited by the natural noise of the oscillator)

$$\mathcal{L}(f_m) = 20 \log \underbrace{\frac{\sqrt{A_{max} \cdot \gamma \cdot F_o}}{2 f_v}}_{D} \qquad \text{For } D<0{,}01 \qquad [1]$$

γ : accelerometric sensitivity of the resonator

fo : nominal frequency

$A_{max} = \sqrt{2}\ \sqrt{A^2_{rms}}$: peak acceleration value

fv : vibration frequency

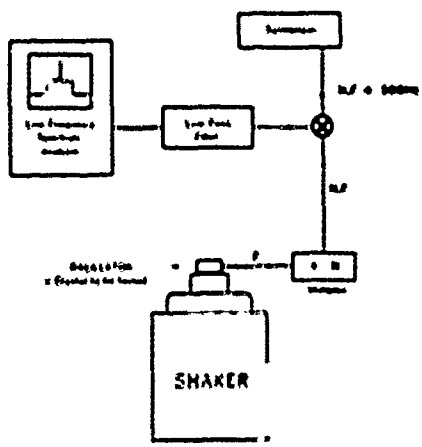
Numerous studies have been devoted to force frequency effect and acceleration sensitivity of quartz resonators. It's not necessary to remind here the theoritical results. But experimental analysis on a lot of resonator's batch enable us to conclude that the g sensitivity is depending on the mounting support (2,3 or 4 points) and the know-how. Though the orientation of the support is optimized, manufacturing limits involve a large disperion concerning the g sensitivity (1.10^{-10} / g to 8.10^{-10} / g).

g sensitivity measurement method

Under an environmental disturbance, the resonator can be used as an acceleration transducer. With this approach a new technic of g sensitivity measurement has been improved (synoptic figure 3).

SYNOPTIC

Figure 3



In order to obtain the g sensitivity evaluation, the crystal oscillator is submitted to a random spectrum (figure 4).

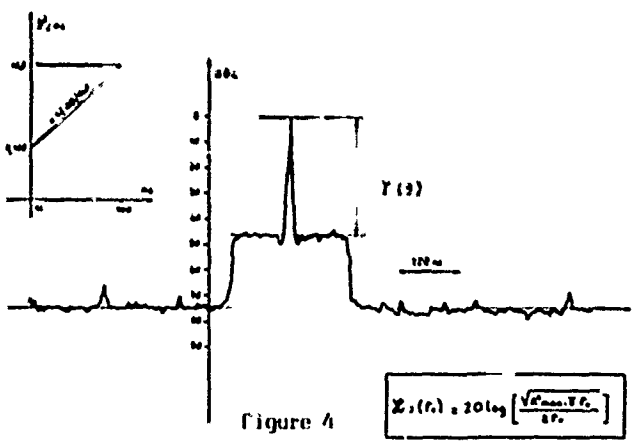


Figure 4

the power spectral density slope is fixed to 12 dB/oct. In this case, the theory shows that the phase noise obtained with the spectrum presents an increase of the noise floor (fig 4) which depends on the g sensitivity of the resonator. With this method, the limitation is about 3.10^{-11}/g and the accuracy is evaluated to ± 15%.

II.2.2 Degradation due to the microphonics sensitivity (2 KHz up to 20 KHz)

One of our R & D matters is to reduce the volume of oscillators. the consequence is a reduction of the acoustic insulation that the oscillator case provides, at the level of the resonator.

Also, in order to minimize the acoustic perturbation, it follows from the above mentioned remark that the future resonator structure must present one first acoustic resonance beyond 15 KHz. Finite elements modelization and experimental results (table 1) show that the best solution, with a classical package (HC 40), is a plate quartz on a 4-post support. [2] and [3]

mechanical resonance mode	4-post support	
	theoritical results	experimental results
1	16 KHz	16 KHz < f_1 < 21 KHz
2	25 KHz	21 KHz < f_2 < 25 KHz
3	26 KHz	28 KHz < f_3 < 29 KHz
4	34 KHz	30 KHz < f_4 < 32 KHz
5	36 KHz	33 KHz < f_5 < 35 KHz

Acoustic sensitivity measurement bench (synoptic figure 5)

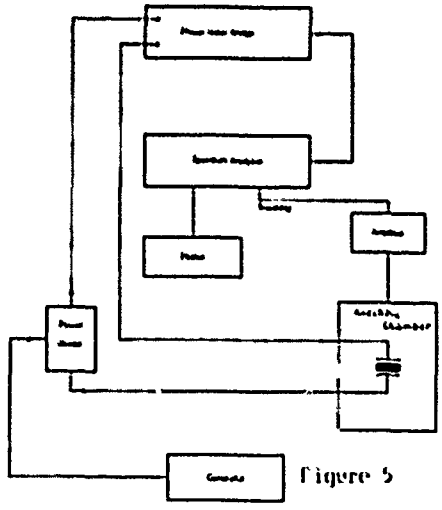


figure 5

The experimental acoustic tests are performed in a anechoic chamber (figure 6).



figure 6

Unfortunately, decrease of the resonator acoustic sensitivity, implies, today, a degradation of the accelerometric sensitivity. In fact, we see that the resonator choice for one oscillator is depending on phase noise performances requirements but also on the vibration and acoustic environment requirements.

III - OSCILLATOR CONCEPTION

The remarkable progress in theoritical modelization, test bench and measurement systems allow to gain an understanding of the behaviour of piezoelectric oscillators. New scientific technics are based on the separate analysis of respectively electronic and mechanical parts in order to optimize the electrical performances.

III.1 Electronic concept

All crystal oscillator circuits may be considered as two essential parts :

- a feedback network that provides positive feedback at the frequency of oscillator,
- an amplifier.

For some kind of oscillators, an additional part must be considered : the switch circuit.

III.1.1 Feedback network

The oscillator is a Clapp type, combined with a selecting circuit in order to suppress the fundamental mode and undesirable overtone mode.

On figure 7 is representated the detailed schematic diagram.

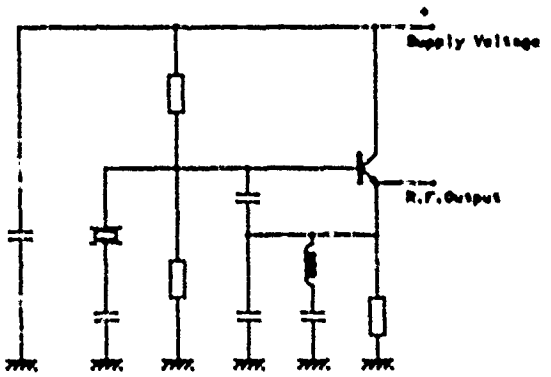


Figure 7

III.1.2 Amplifier circuit

The phase noise performance involves that we make a signifiant progress in the noise level of different electronic parts (DC supply, amplifier ...). We have selected for the most part of CEPE oscillators one common base type amplifier with a selective circuit adaptation at the input and output.
The figure 8 shows a shematic diagram.

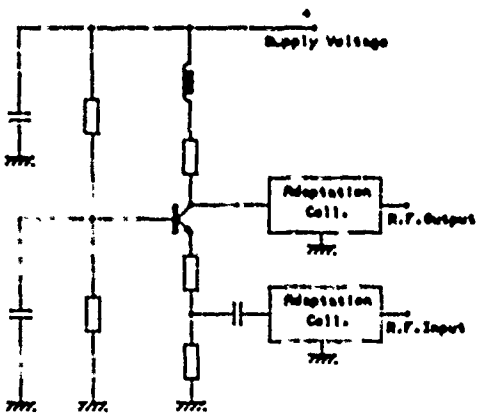


Figure 8

Typical characteristics are as following :

- gain / 50 ohms : 13 dB
- band-pass (at 3 dB) = 3 MHz
- insulation : - 25 dB

III.1.3 Optional switch circuit

For some applications, it is interesting to include after the output amplifier a switch circuit in order to commutate rapidly the RF output signal. In this case, CEPE has developped a switch circuit in DIL 16 package.

The main characteristics are :

- attenuation : > 70 dB
- switch time : < 1 µs
- RF signal leakage on switch control : - 60 dBm

III.2 Phase noise results without vibration

The oscillator circuit configurations described above permit to meet successfully the phase noise specification. One exemple of phase noise performance is given on figure 9, up to 100 KHz and on figure 10 up to 5 MHz.

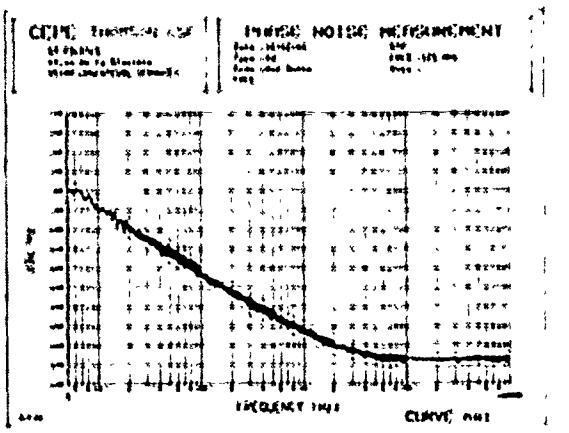


Figure 9

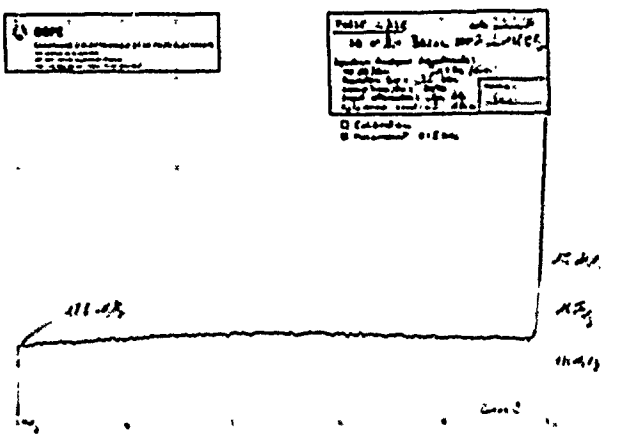


Figure 10

The phase noise test bench is shown on figure 11.

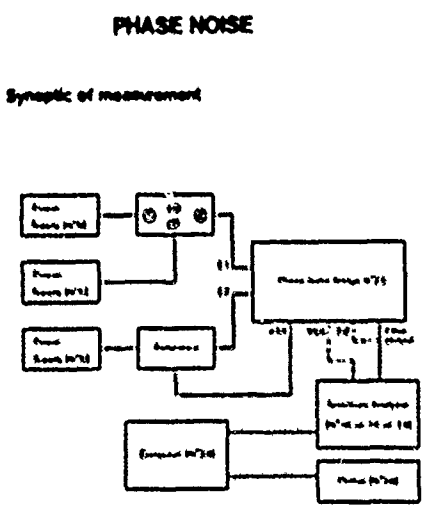


Figure 11

III.3 Mechanical concept

The main preoccupation is to limit the phase noise degradation during the acoustical and random vibration tests. As we have seen above, after the resonator selection, the phase noise performance is directly linked with an acceleration density applied to the resonator. In this case, it is clear that it is very important to identify the mechanical behaviour of the package and of the printed circuit during the test.

Finding the best mechanical configuration involves one finite elements modelization (figures 12 and 13).

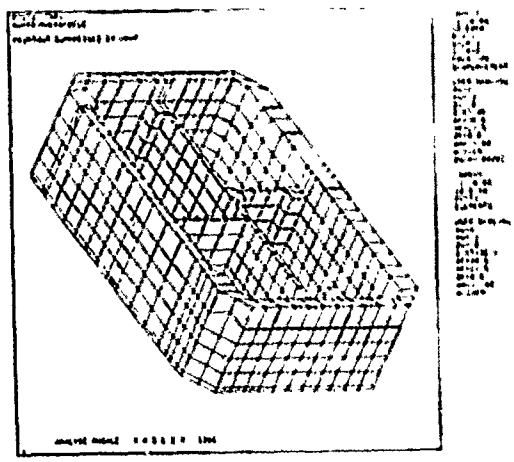


Figure 12

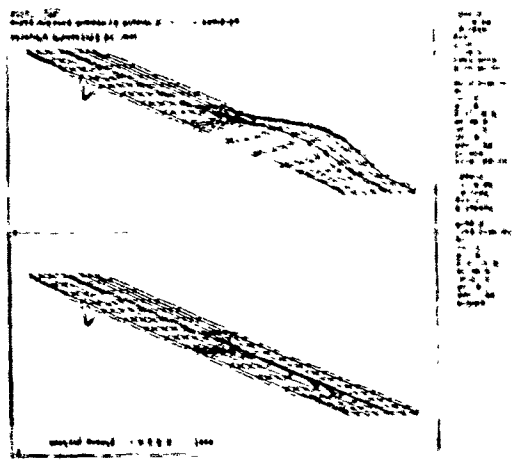


Figure 13

Studies underway at the oscillators R & D department now confirm that we are able to evaluate, in first step the frequencies of our structures and in second step, we can establish at circuit level, a cartography of the acceleration isovalues for a fixed direction (figure 14).

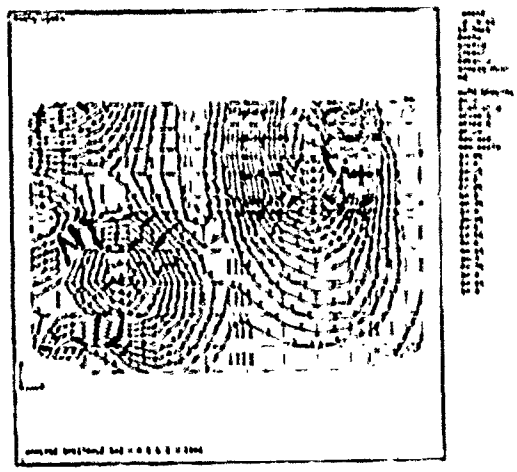


Figure 14

This information can be used in theory to calculate the resultant phase noise degradation.

Figures 15,16, 17 and 18 meet the typical phase noise performance for one XO oscillator during the acoustic and random vibration tests.

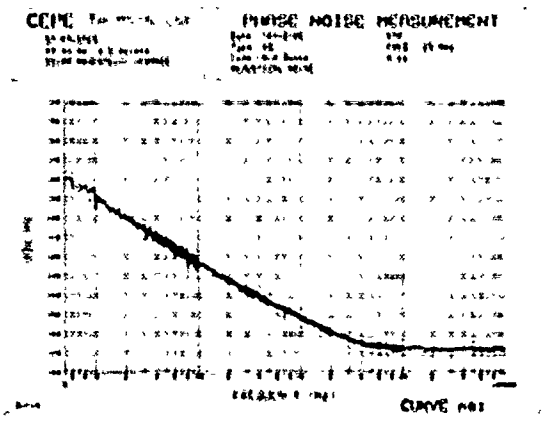


Figure 15

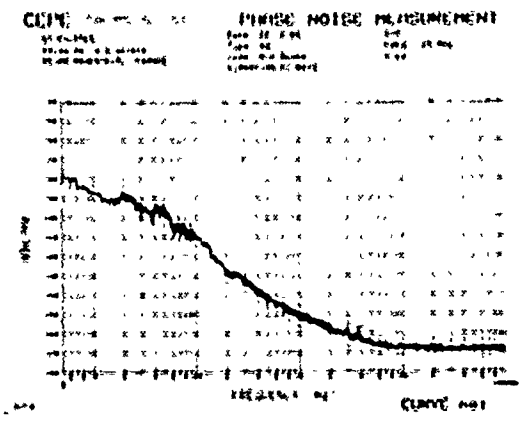


Figure 18

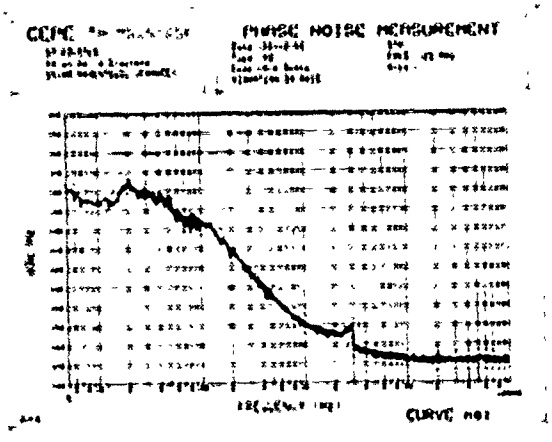


Figure 16

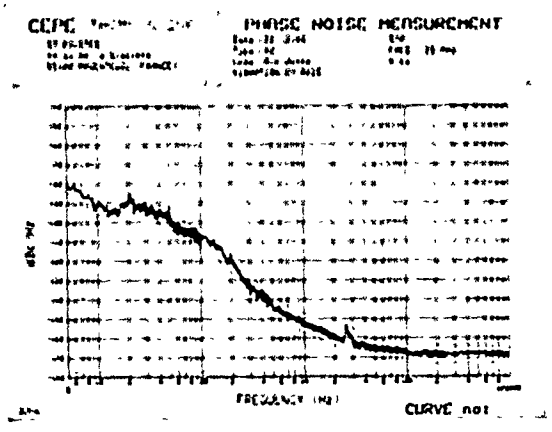


Figure 17

IV - TYPICAL SPECIFIATION (XO)

IV.1 Electrical characteristics

1) Frequency range : 100 to 130 MHz
2) Frequency stability : $\pm$ 5.10^{-6}
3) DC supply : 12 V $\pm$ 5 %
4) Output power : 10 dBm $\pm$ 1,5 dB
5) RF signal leakage through terminal : - 65dBm
6) Electromagnetic box attenuation : 85 dB
7) Phase noise without vibration

10 Hz	- 85 dBc / Hz
100 Hz	- 120 dBc / Hz
1 KHz	- 140 dBc / Hz
10KHz up to 5MHz	- 160 dBc / Hz

8) Phase noise with random vibration , spectrum of IV.3.1

10 Hz	- 85 dBc / Hz
100 Hz	- 95 dBc / Hz
1 KHz	- 150 dBc / Hz
3 KHz	- 150 dBc / Hz

Points 1 - 6 are specified under environmental conditions

IV.2 Dimensions (figures 19 and 20)

Size : 20 x 40 x 60 mm
Weight : 60 g



Figure 19



Figure 20

IV.3 Environmental conditions

1) Random vibration spectrum : 1,5 g rms
   0,05 g^2 / Hz near 50 Hz
   10^{-6} g^2 / Hz from 1 KHz to 3 KHz
2) Acoustic spectrum : 90 dB max
3) Ripple on DC supply : 1 mV rms
4) Functional temperature range : - 40 to + 85°C
5) Storage temperature range : - 55 to + 105°C

V - CONCLUSION

Before the wide variety of military applications for crystal oscillators, CEPE has used theorical and experimental means in order to take into account the new technical environmental specifications and to concept new oscillators with the choice of the best technical compromises.

These means are the more as important in order to reduce the size of the futur oscillators, it will be necessary to use new materials (ceramics, electronic subtrats, new aluminium alloys ...).

REFERENCES

[1] R.L. Filler, "The effect of vibration on quartz crystal resonators" Electronic Technology & Devices Lab.

[2] R.D. Weglein - W.C. Hu and A.F. Margagliano "Microphonics in UHF Quartz resonators and Oscillators" 1984 Ultrasonics Symposium.

[3] J.P. AUBRY - G. CARCI - G. MAROIEL, BNM Temps - Fréquence (1986)
Oscillateurs à quartz de faible sensibilité à l'environnement pour les applications Temps - Fréquence

[4] M. Driscoll "Low noise, VHF Crystal-Controlled Oscillator Utilizing Dual SC-cut resonators" IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control. vol UFF-C 33/6/11/86/

[5] J.P. VALENTIN - Thèse de Doctorat es Sciences Physiques Besançon 1983.
"Analyse de quelques aspects du comportement thermique des résonateurs à quartz à ondes de volume.

41st Annual Frequency Control Symposium - 1987

THE RELATIONSHIP BETWEEN RESONATOR AND OSCILLATOR NOISE, AND RESONATOR NOISE MEASUREMENT TECHNIQUES

G. Stephen Curtis

Hewlett-Packard Company, Spokane Division
1620 Signal Drive, Spokane, WA 99220

Abstract

Flicker, or 1/f noise in just one component, the acoustic resonator, is the dominant source of frequency fluctuations in a well designed oscillator using a bulk acoustic wave or surface acoustic wave device. Measurement of resonator noise allows prediction of oscillator noise and performance improvements through selection of devices.

Leesons model for oscillator noise is revieved, then modified to incorporate the effect of resonator noise. Typical oscillator spectra are examined using the model. Two models for resonator noise are discussed, and data presented demonstrating that the cause of resonator instability is center frequency fluctuations, not phase fluctuations. The effect of loaded Q on oscillator and resonator noise is discussed.

Three means to measure resonator noise are revieved, these being, installing the resonator in an oscillator, a dual resonator bridge, and a single resonator bridge. Calibration of each method is discussed and a simple calibration method for the single resonator bridge, which lends itself to automated testing, is proposed. Noise floor limitations of the three methods are analyzed, and typical levels to be expected are given. A correlation technique to further improve the noise floor is suggested. An improved single resonator bridge which automatically sets and holds quadrature for fast, temperature stable measurements is presented. Use of a commercial phase noise measurement system to improve accuracy and ease calibration is discussed.

Introduction

A problem that has plagued oscillator designers for years is that of accurately predicting phase noise. Part of the difficulty is in modeling the oscillator, an inherently nonlinear circuit. One must also know the noise levels generated by each component in the oscillator. Shot noise sources are very predictable but 1/f (flicker) noise sources can vary orders of magnitude between seemingly identical components. If one component dominates the oscillator phase noise, it may be possible to select devices to improve performance.

In oscillators utilizing acoustic resonators, three primary noise contributors have been observed. The noise of the oscillator at large offsets from the carrier is limited by the shot and thermal noise generated by the amplifier circuit in the oscillator and any buffer amplifiers following the oscillator. A region of 1/f power spectral density may be seen, as in Figure 1, also being generated in the oscillator circuitry or the buffer amp. In most oscillators, $1/f^3$ or $1/f^2$ spectral densities can be observed, which may be caused by the conversion of 1/f and white phase fluctuations in the oscillator circuitry to frequency fluctuations at offset frequencies less than the resonator half bandwidth. Generally, the oscillators being discussed here do not exhibit a $1/f^2$ region, but do exhibit a region of $1/f^3$ spectral density, because the oscillator circuitry is dominated by 1/f noise at offsets inside the narrow bandwith of the resonant circuit. Short term 1/f frequency fluctuations of the resonator also result in a $1/f^3$ spectrum, and are usually the dominant cause of this behavior in the oscillator, if care has been taken in the design to reduce the contribution of other noise sources[1,2]. Because the level of noise in the $1/f^3$ region is critical to the performance of many systems, the resonator performance must be specified.

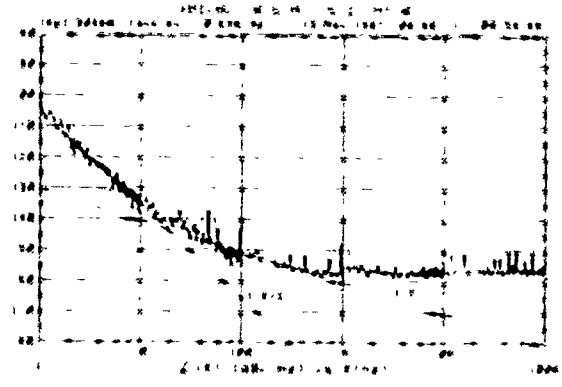


Figure 1. Typical crystal oscillator phase noise spectrum.

When discussing phase noise, a variety of measures are used. In the frequency domain, the most common are single sideband power spectral density to carrier power ratio , $\mathcal{L}(f)$, power spectral density of phase fluctuations, $S_\phi(f)$, power spectral density of frequency fluctuations, $S_\nu(f)$, and the power spectral density of fractional frequency fluctuations, $S_y(f)$. Since manufacturers of oscillators and signal generators tend to use $\mathcal{L}(f)$, that will be used here to in order to simplify comparison of various sources. Other noise measures may be preferrable to $\mathcal{L}(f)$ for a given application, but mainly $\mathcal{L}(f)$ is used for clarity. Equations to convert between $\mathcal{L}(f)$ and the other measures mentioned are supplied below, with the caution that the conversions are only valid for small angle modulation (total carrier phase modulation much less than one radian).

$$\mathcal{L}(f) = 1/2 * S_\phi(f) \quad (1)$$

$$\mathcal{L}(f) = 1/2 * S_\nu(f)/f^2 \quad (2)$$

$$\mathcal{L}(f)=1/2*S_y(f)(\nu_0/f)^2 \quad (3)$$

The carrier frequency is denoted by ν_0, and f is the frequency offset from the carrier. Graphs of $\mathcal{L}(f)$ in this paper will be presented logarithmically in dBc/Hz, but equations will be in the linear form to simplify the presentation.

The Relationship Between Resonator Noise And Oscillator Noise

Leesons Model

Leesons model for oscillator noise considers the oscillator as a phase servo loop, where the frequency of oscillation, ν_0, is adjusted so that the total phase shift through resonator and amplifier is an integer multiple of 2π radians[3]. If a phase disturbance occurs, then the frequency must shift accordingly. In using the model, it is often assumed that the amplifying stage is the dominant source of phase fluctuations, and consists of a combination of white and 1/f noise. The single sideband phase noise due to the amplifier is usually described by an equation of the form

$$\mathcal{L}_a(f)=\mathcal{L}_a(1)/f + FkT/P_s \quad (4)$$

where FkT/P_s represents the white noise level and $\mathcal{L}_a(1)$ the 1/f noise level at 1 Hz. Referring to Figure 2, the noise at oscillator output 1 is

$$\mathcal{L}_o(f)=\mathcal{L}_a(f)(1+(\nu_0/2Q_1f)^2) \quad (5)$$

where Q_1 the loaded Q of the resonator. If the output is taken at 2, the resonator will filter the noise outside the half bandwith, so that

$$\mathcal{L}_o(f)=\mathcal{L}_a(f)(1+(\nu_0/2Q_1f)^2)/(1+(2Q_1f/\nu_0)^2) \quad (6)$$

simplifying to

$$\mathcal{L}_o(f)=\mathcal{L}_a(f)(\nu_0/2Q_1f)^2 \quad (7)$$

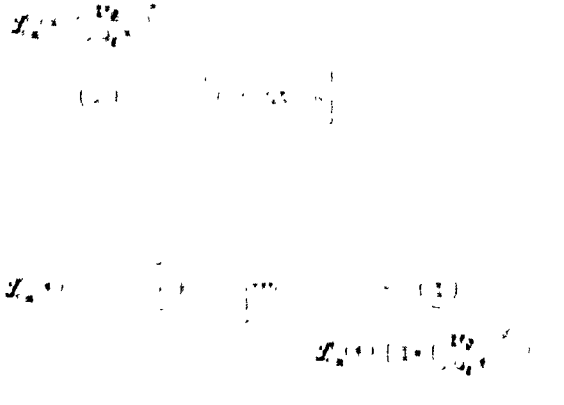


Figure 2. Leesons model.

This behavior cannot be observed at large offset frequencies because other noise sources, such as buffer amplifiers are dominant. It also is important to note that loaded Q can not be inferred from the noise spectrum observed at the output of the resonator because of the cancellation of the $1+(\nu_0/2Q_1f)^2$ term.

In Figure 3, the noise described by equations 4-7 is plotted.

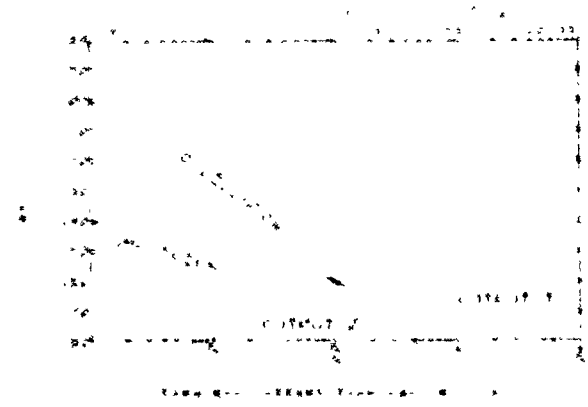


Figure 3. Oscillator noise calculation.

Problems in using equations 4-7 to predict oscillator noise arise because $\mathcal{L}_a(f)$ has not been properly defined. The quantity $\mathcal{L}_a(f)$ is the single sideband phase noise of the amplifier when driven at the same level and with the same load impedance as in the oscillator circuit. In general, Equation 4 is not valid, since rapidly changing resonator impedance outside the passband causes the amplifier noise to be a function of offset frequency, making it wrong to assume that $\mathcal{L}_a(f)$ consists of just 1/f and white noise spectral densities. Equations 5 and 7 are still valid, since no assumptions were made as to the nature of $\mathcal{L}_a(f)$.

Resonator 1/f noise can dominate the oscillator noise, if larger than that of the amplifier. Figure 4 superimposes resonator phase noise on the oscillator noise predicted with and without a noisy resonator. Inside the half bandwidth, the resonator noise is increased by the factor $(\nu_0/2Q_1f)^2$, outside the half bandwidth the resonator noise is decreased by the same factor, as it filters its own noise. This is similar to Equation 7, and a continuous 30 dB/decade slope will be observed around the half bandwidth. The oscillator noise will be the sum of resonator caused noise and amplifier caused noise, which can be written as

$$\mathcal{L}_o(f)=\mathcal{L}_r(f)(\nu_0/2Q_1f)^2 + \mathcal{L}_a(f)(1+(\nu_0/2Q_1f)^2) \quad (8)$$

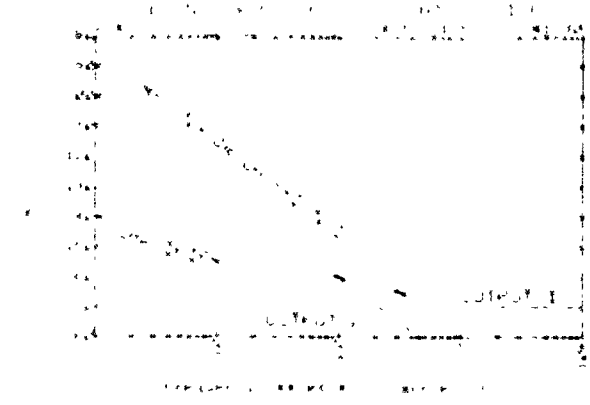


Figure 4. Resonator caused oscillator noise.

if amplifier noise is not filtered by the resonator, or

$$\mathcal{L}_0(f)=\mathcal{L}_r(f)(\nu_0/2Q_l f)^2 + \mathcal{L}_a(f)(\nu_0/2Q_l f)^2 \quad (9)$$

if it is. $\mathcal{L}_r(f)$ is the resonator noise referred to its input.

Modeling Resonator Noise

Resonator noise can be modeled as either fluctuations in resonant frequency, or phase fluctuations, which will also cause oscillator frequency fluctuations. Leesons model assumes a phase disturbance in the oscillator, so this is a possible resonator noise model to evaluate. Figure 5 illustrates the phase vs. frequency response of a simple resonator that has been phase modulated by $\Delta\theta$. In an oscillator, the frequency will change such that phase shift around the loop returns to an integer multiple of 2π radians. The amount of frequency shift is inversely proportional to $d\theta/d\omega$, the phase slope, which equals $2Q_l/\omega_0$. As seen on right side of figure 5, a lower Q oscillator will have larger frequency fluctuations. The phase fluctuations of the resonator will not change with loading of Q, if it behaves as described here.

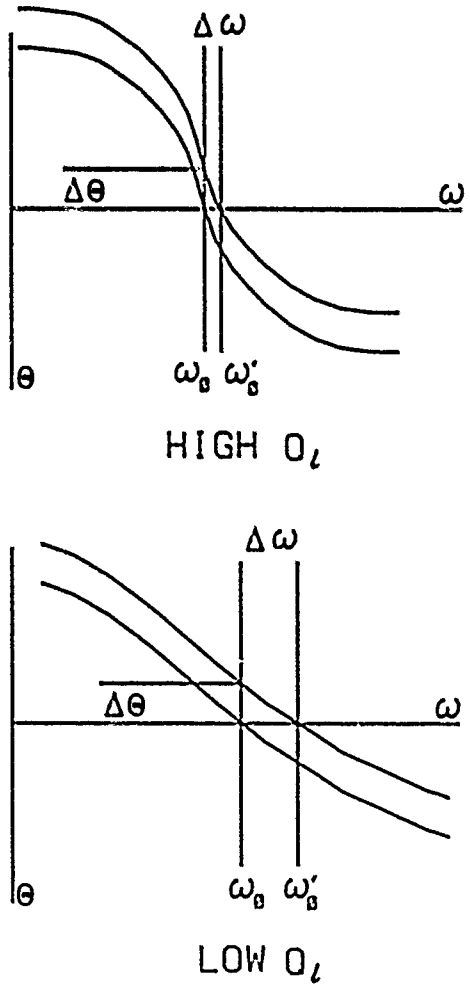


Figure 5. Resonator phase fluctuations.

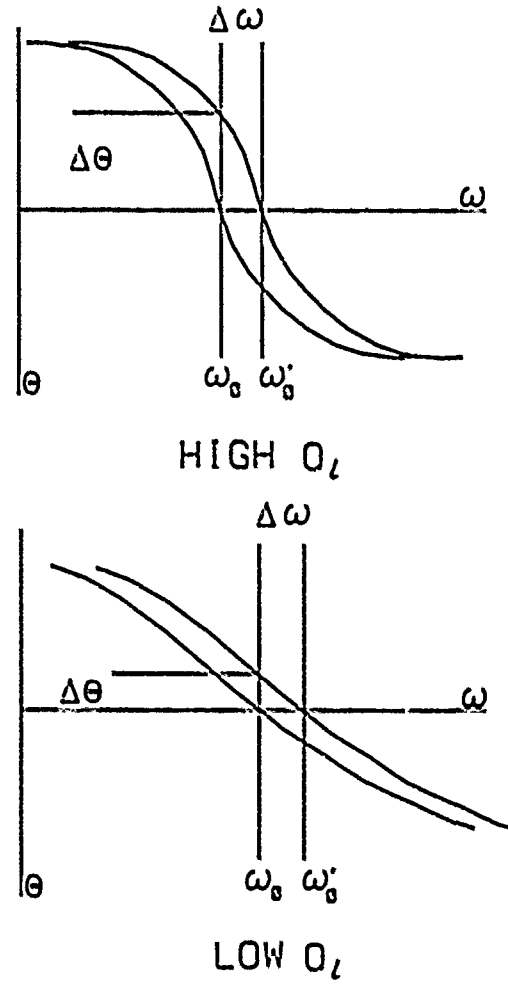


Figure 6. Resonator frequency fluctuations.

Figure 6 illustrates a resonator exhibiting resonant frequency fluctuations. The oscillation frequency will track that of the resonator, in order to maintain a constant phase shift around the loop. Loading the resonator does not affect the amount of oscillator frequency shift, as on the right side of Figure 6. It follows that oscillator noise will be independent of loaded Q, just the opposite of the previous model. It is important to note that in this case resonator phase fluctuations decrease with loading, being proportional to $d\theta/d\omega$. For oscillators where resonator noise dominates, behavior indicating that resonator frequency fluctuations are the mechanism has been observed[1,4]. An example of this behavior is Figure 7, where the phase fluctuations of a VHF crystal resonator has been reduced 6 dB by lowering the loaded Q a factor of two.

This result has some implications for prediction of oscillator noise. If oscillator fluctuations are to be predicted from measured resonator phase fluctuations, the measured resonator noise, $\mathcal{L}_{rm}(f)$ must be scaled by the factor $(Q_l/Q_{lm})^2$, because phase fluctuations depend on loading, and the oscillator Q_l will probably be different than Q_{lm}, the loaded Q in the measurement circuit. Equation 10 defines $\mathcal{L}_r(f)$ in terms of $\mathcal{L}_{rm}(f)$, including the effect of a different Q_l.

$$\mathcal{L}_r(f) = \mathcal{L}_{rm}(f)(Q_l/Q_{lm})^2(1+(2Q_{lm}f/\nu_0)^2) \quad (10)$$

Since $\mathcal{L}_r(f)$ has been defined here as the noise referred to the input of the resonator, the measured resonator noise has the filtering effect of the resonator removed by multiplying by the reciprocal of the resonator transfer function, H(f), where

$$H(f) = 1/(1+(2Q_{lm}f/\nu_0)^2) \quad (11)$$

which accounts for the additional factor in Equation 10. Oscillator noise can be predicted using Equation 8 or 9, if $\mathcal{L}_r(f)$ is defined as in (10). If the right hand side of equation 10 is substituted into equations 8 and 9, they become

$$\mathcal{L}_o(f) = \mathcal{L}_{rm}(f)(\nu_0/2Q_{lm}f)^2/H(f) + \mathcal{L}_a(f)(1+(\nu_0/2Q_lf)^2) \quad (12)$$

at output 1 and

$$\mathcal{L}_o(f) = \mathcal{L}_{rm}(f)(\nu_0/2Q_{lm}f)^2/H(f) + \mathcal{L}_a(f)(\nu_0/2Q_lf)^2 \quad (13)$$

at output 2. Note that the resonator noise terms are independent of the oscillator loaded Q, as expected, since the resonator exhibits frequency fluctuations.

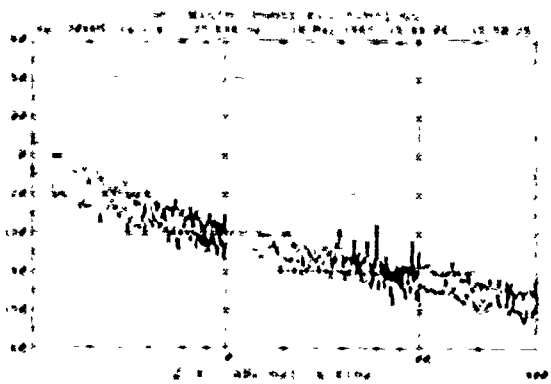


Figure 7. Effect of loaded Q on resonator phase fluctuations

Equations 8,9,12, and 13 can be simplified somewhat if resonator noise is expressed as frequency fluctuations, $S_{\nu r}(f)$. The first term on the right hand side of all the equations can be replaced with $S_{\nu r}(f)/2f^2$ which equals the single sideband noise caused by the resonator. Note that there is no Q_l present because the frequency fluctuations are indpendent of Q. A modified version of Leesons model to account for resonator frequency fluctuations is shown in Figure 8.

Summarizing this section -

Resonator noise should be modeled as center frequency fluctuations of the device.

Oscillator noise is independent of loaded Q, when resonator noise dominates.

Since resonator phase fluctuations are proportional to Q_l, any measurement of resonator phase fluctuations must include a measurement of the loaded Q in the test fixture. To avoid having to specify Q_l, the

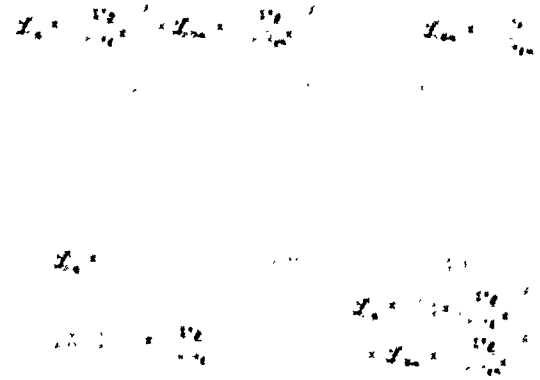


Figure 8. Modified Leesons Model.

oscillator noise caused by the resonator can be calculated from a resonator phase fluctuation measurement, or measured directly in an oscillator.

Leesons model can be modified to include resonator phase fluctuations, and to account for differences in loaded Q between the oscillator and the resonator phase fluctuation measurement.

Resonator Test Methods

Oscillators

The most common method of specifying the noise level of a resonator is to install the device under test (DUT) into the final oscillator circuit, and measuring the latters phase noise. This is the safest method to guarantee producing oscillators with a given level of phase noise, but is not general purpose, being specific to a given oscillator design, and may not be the best method for a resonator vendor to use. One solution to this problem is to build the oscillator with a broadband, 50 Ohm input and output impedance, low noise amplifier, a limiter, an overtone selector (band pass filter), and a matching network for the resonator. Driscoll has reported on this type of oscillator, which simplifies characterization of the various oscillator components[4]. At low offset frequencies, resonator noise should dominate, yielding the same result as in the original oscillator circuit. One advantage of the oscillator method is that it allows direct measurement of resonator frequency fluctuations. It is not necessary to measure loaded Q to determine oscillator noise, although it can help verify the test system performance.

This type of test requires a reference source which must have either lower noise than the DUT or be a known quantity, in order to determine the DUT noise. A phase detector is used to generate a voltage proportional to the phase fluctuations of the sources, which is amplified, as in figure 10. The power spectral density of the output voltage $S_v(f)$,which can be observed using a spectrum analyzer, is related to the oscillator single sideband noise to carrier ratio, $\mathcal{L}_o(f)$, by

$$\mathcal{L}_o(f) = S_v(f)/2(K_\phi G)^2 \quad (14)$$

To maintain quadrature at the mixer, it is necessary to form a phase locked loop (PLL) by feeding the output voltage back to an electronic tuning input on the reference oscillator. The PLL effectively

highpass filters the output voltage, so either the loop response must be corrected for, or the loop bandwidth must be made much narrower than the frequency offset of interest.

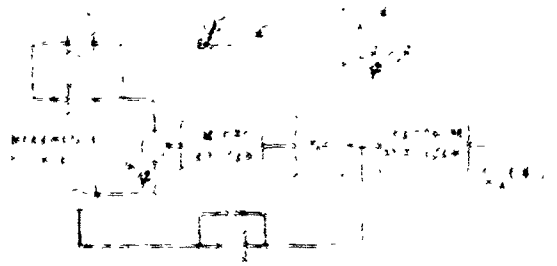


Figure 10. Oscillator method.

Dual Resonator Bridge

Having a need to measure the inherent short-term stability of crystal resonators, Walls and Wainwright designed a method to measure this without the influence of oscillator electronics, using a pair of crystal resonators in a passive circuit[6]. Figure 11 illustrates the system used. A low noise signal source drives two closely matched crystals at their series resonant frequency, ω_0. The resonant frequency and Q_l of one arm can be adjusted to minimize any difference in phase or group delay between the two arms. The output of the mixer when its inputs have been set to quadrature will be relatively insensitive to AM and FM noise of the source. Any resonant frequency fluctuation in either crystal will cause a phase change at the input to the phase detector. The phase change for a given frequency shift is proportional to the phase slope (or group delay), $2Q_l/\omega_0$, of the devices. The power spectral density of the output voltage is related to the phase fluctuations of each resonator by

$$\mathcal{L}_{rm}(f) = S_v(f)/2(K_\phi G)^2 \qquad (15)$$

As mentioned earlier a resonator will filter its own noise, and $\mathcal{L}_{rm}(f)$ will have an additional $1/f^2$ slope outside the resonator bandwidth. The spectrum observed should consist of a $1/f$ portion and a $1/f^3$ portion as in fig. 12, a measurement of a pair of 125 MHz AT cut crystals. The oscillator noise caused by each resonator is (as in Equations 12 and 13)

$$\mathcal{L}_o(f) = \mathcal{L}_{rm}(f)(\nu_0/2Q_{lm}f)^2/H(f) \qquad (16)$$

or

$$\mathcal{L}_o(f) = (S_v(f)/2(K_\phi G)^2)(\nu_0/2Q_{lm}f)^2/H(f) \qquad (17)$$

for offset frequencies much less than the resonator half bandwidth, Equation 8 simplifies to

$$\mathcal{L}_o(f) = (S_v(f)/2(K_\phi G)^2)(\nu_0/2Q_{lm}f)^2 \qquad f<\nu_0/2Q_{lm} \qquad (18)$$

In this measurement, the reference is a resonator, which must have very low or known noise, since the total noise of both units is measured. Measurements of three devices in three combinations can be used to determine the noise of the reference resonator.

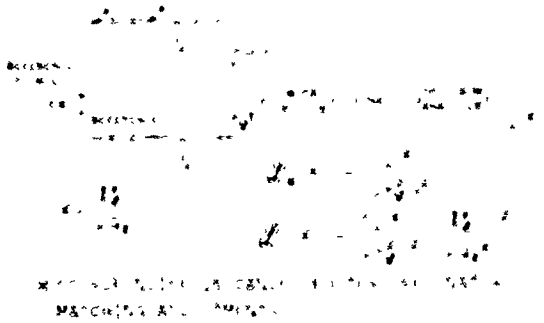


Figure 11. Dual resonator bridge.

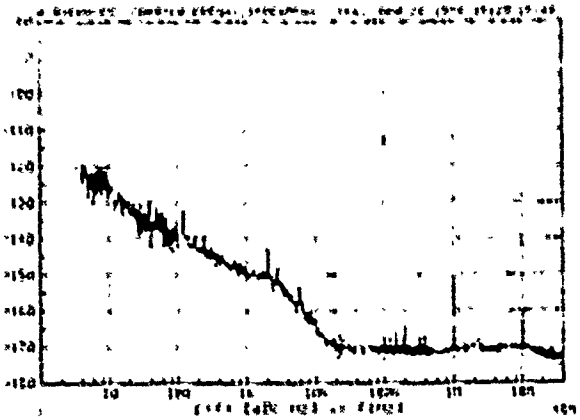


Figure 12. Dual resonator bridge measurement.

Single Resonator Bridge

To avoid the ambiguity of measuring two resonators against each other, Elliot and Bray developed a method to measure phase noise requiring only one resonator[7]. The task of tuning a second resonator to match the DUT is also avoided. Illustrated in Figure 13, this system consists of a very low noise signal source, a splitter, a variable phase shifter, one resonator (the DUT), and a phase detector. With its two inputs set to quadrature, any phase fluctuations caused by the resonator will produce a proportional voltage at the output of the phase detector. This relationship is described by equation 15, the same as with the dual resonator bridge. Oscillator noise caused by the resonator can be calculated using equations 16-18.

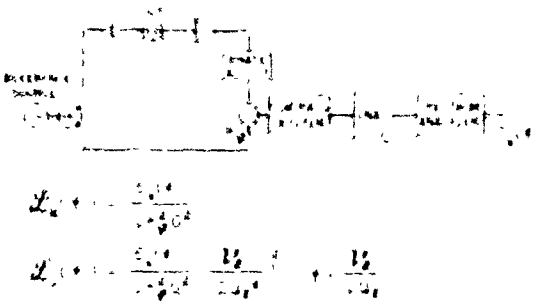


Figure 13. Single resonator bridge.

The system will also act as a frequency discriminator, and will produce an output proportional to any source frequency fluctuations. In fact, the discriminator sensitivity to source frequency fluctuations is exactly the same as the sensitivity to resonator frequency fluctuations, being

$$S_\nu(f)=S_\nu(f)(K_\phi G)^2(2Q_{1m}/\nu_0)^2|H(f)| \qquad (19)$$

where $S_\nu(f)$ can be either resonator frequency fluctuations or source frequency fluctuations. If Equation 19 is solved for $S_\nu(f)$, then converted to $\mathcal{L}_0(f)$ using Equation 2, the result is identical to Equation 17. A 250 MHz SAW resonator is measured using the single resonator bridge in figure 14.

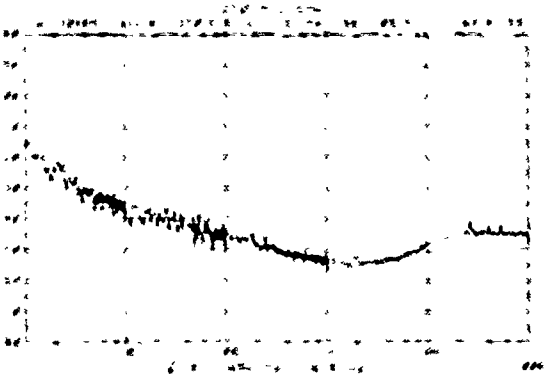


Figure 14. Single resonator bridge measurement.

Calibration

Calibration for any of the three methods described requires measuring the phase detector sensitivity, K_ϕ and amplifier gain G. The loaded Q of the resonator in the test circuit must be measured for either bridge circuit, in order to calculate resonator caused oscillator noise. In addition, resonator Q and center frequency must be matched to reject source noise for the dual resonator method.

Phase Detector Calibration

The easiest way to calibrate K_ϕ for the oscillator method is to offset the frequency of the reference oscillator relative to the DUT oscillator and measure the beatnote out of the phase detector. For a sinusoidal beatnote, the peak signal voltage is equal to K_ϕ in V/radian. For other shapes, K_ϕ can be determined from the slope of the detector output at the zero crossing. The positive and negative slopes may not be the same, so whichever is used in the measurement is the one to be measured. As the beatnote signal can saturate the LNA, it may be necessary to switch in an attenuator before the amplifier during calibration, or measuring the detector constant and amplifier gain separately. The beatnote method can also be used in the bridge measurements, by substitution of an equal level signal source for the signal coming from the splitter.

Another way that a second RF source can be used for calibration of K_ϕ is to add a spur onto the carrier using a coupler at the detector input. The level relative to the carrier at the detector input must be measured. It can be shown that the spur consists of equal level AM and FM components, each containing half of the power. The phase detector usually rejects AM by 20 dB or more, so just the FM is detected. For example, a -40 dBc spur results in .01 radian peak phase deviation at the detector input, and a peak voltage at the output of $(.01)K_\phi$ Volts. This method should generally give better results than the substitution of a second source, because the system is set to quadrature and configured exactly as it would be in the noise measurement, except a small perturbation has been added (the spur). The spur can be made small enough that the LNA will not saturate, allowing simultaneous measurement of K_ϕ and G.

Any means to introduce a calibrated phase shift can also be used to measure K_ϕ in the bridge circuits, if another RF source is not available.

Loaded Q Measurement

The loaded Q of the resonator should be measured under conditions identical to the test circuit. This may be done by measuring the amplitude after the resonator and varying the source frequency to determine the bandwith. Another possibility is to use a vector voltmeter or network analyzer, and directly measure the phase slope at resonance, where $d\theta/df=2Q_{1m}/\nu_0$.

A different approach uses the bridge to measure the resonator bandwidth. If second RF source is available, it can be used to inject a spur before the resonator. The resultant phase modulation is filtered by the resonator then detected. By sweeping the spur frequency, the resonator bandwidth can be observed at the bridge output using a spectrum analyzer.

Additional Calibration

Additional calibration steps are required by the oscillator and dual resonator bridge techniques. Because an oscillator measurement requires a PLL to hold quadrature at the phase detector, frequency response effects of the loop must be accounted for, if measurements are to be made at offset frequencies close to the loop bandwidth. This involves injecting an error signal into the loop and measuring the response. Measured noise data must then be divided by the loop gain. This task is tedious, but is handled readily by an automated phase noise measurement system, as was used to make the measurements in this paper.

To achieve maximum performance from the dual resonator bridge, the two resonators must be matched closely in frequency and Q. To match the resonator frequencies, an iterative procedure can be used. First, the source frequency should be adjusted to maximize the signal through the resonator not being tuned. An impedance equivalent to the crystal at resonance can be substituted for the DUT. This allows an approximate setting of quadrature using a phase shifter in one side of the bridge. The DUT can then be placed in the bridge, and its frequency adjusted to restore quadrature at the phase detector. A signal generator with FM and AM capabilities is needed for the next steps. An FM modulated carrier will be detected by the bridge if the group delay of the bridge arms are not equal. Adjusting the Q_1 of the DUT for minimum

detected modulation signal at the bridge output will minimize the difference between the devices. It may be necessary to adjust the resonant frequency and Q_L of the DUT again if more cancellation is needed. AM noise rejection can be maximized by adjusting the phase shifter to null out any modulation detected by the phase detector with an amplitude modulated source applied to the bridge. At the AM null there may be a significant amount of DC at the output of the phase detector. This must be checked to avoid overloading the LNA.

Simplified Calibration Procedure

The calibration procedure can be greatly simplified by using the methods described next. It was previously noted that the single resonator type of measurement will discriminate source frequency fluctuations with the same sensitivity as resonator frequency fluctuations. By setting quadrature, then either changing the source frequency, or modulating the source frequency, a voltage change out of the mixer can be measured, allowing the discriminator sensitivity, K_d, to frequency fluctuations to be determined as,

$$K_d = \Delta V_{out}/\Delta\nu \qquad (20)$$

K_d is constant for offset frequencies less than half the resonator bandwidth, but decreases at larger offsets due to resonator filtering. Resonator (and source) frequency fluctuations are related to the output power spectral density, as follows,

$$S_\nu(f) = S_v(f)/(K_d)^2. \qquad (21)$$

Comparing this to equation 19, it can be seen that

$$(K_d)^2 = (K_\phi G)^2(2Q_{1m}/\nu_0)^2 H(f). \qquad (22)$$

Equation 21 can also be written in terms of oscillator noise,

$$\mathcal{L}_o(f) = S_v(f)/2(fK_d)^2. \qquad (23)$$

The important features of this calibration technique is that only one measurement is necessary to calibrate the bridge, and that a programmable source, such as a low noise synthesizer with FM capability can be used to calibrate automatically. If the modulation source can be swept, this can be used to determine K_d as a function of offset frequency. In Figure 14, the simplified calibration method was used to directly determine oscillator noise caused by a resonator. The resonator was also placed in an oscillator circuit and measured. Both results are plotted, and show excellent agreement. At low offset frequencies, the results diverge somewhat because of differences in the measurement setups. The discriminator had the crystal in open air, which caused increased thermal fluctuations relative to the oscillator measurement.

The measurement system used has a frequency discriminator measurement option in the software which allows the user to calibrate using an FM tone. The system measures the demodulated voltage to determine K_d. It then assumes the noise it measures to be frequency fluctuations proportional to $1/K_d$. It can then plot the output as $\mathcal{L}(f)$, or other format, if desired.

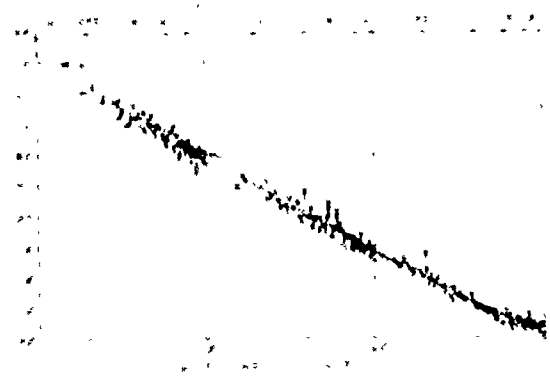


Figure 14. Resonator In Bridge and In Oscillator Noise Measurement.

Noise Floor Limitations

In order to examine the limitations of the test systems discussed, it will be necessary to define the "state of the art". In Figure 15, $\mathcal{L}(1)$ data for a variety of acoustic oscillators and resonators is graphed vs. carrier frequency. The data was culled from the literature, manufacturers data sheets and measurements made by the author. Only the lowest noise devices were included, more commonplace devices may be up to 40 dB worse. The line drawn through the data was derived by making the following assumptions about resonator noise. First, $\mathcal{L}_r(1)$ is constant for all carrier frequencies, and that $\mathcal{L}_r(1) = 5*10^{-13} = -123$ dBc, under the condition that $Q = Q_u$. Employing equation 16,

$$\mathcal{L}_o(1) = 5*10^{-13}(\nu_0/2Q_u)^2 \qquad (24)$$

The frequency dependence of Qu in quartz can be approximated as

$$Q_u = 10^{13}/\nu_0 \qquad (25)$$

Substituting into Equation 24,

$$\mathcal{L}_o(1) = 2.5*10^{-39}\nu_0^4 \qquad (26)$$

The line described by Equation 26 appears to be a reasonable estimate of the best performance available, although it might be possible to find a device 10 db better than this, since there appears to be such a large variation in noise, even among identical devices fabricated with the same process[17].

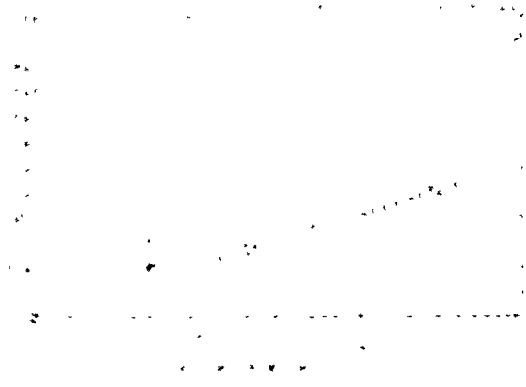


Figure 15. $\mathcal{L}(1)$ vs. carrier frequency.

The three methods to test resonator noise have similar noise floor limitations. These are the reference, be it an oscillator or another resonator, mixers and amplifying stages. In oscillator and single resonator bridge measurements, the reference is a signal source. One way to generate a signal lower in noise than the DUT is to lock a low noise synthesizer to a low frequency reference. The low frequency reference noise increases by the square of the output frequency, instead of the fourth power of frequency observed in fundamental oscillators, as in Equation 26. In Figure 15, the noise of a high quality, multiplied 5 MHz reference is plotted, illustrating that the multiplied signal should be adequate for measuring any source above 10 MHz. Equation 26 can be re-written to account for frequency multiplication by n, as follows

$$\mathcal{L}_\phi(1) = 2.5 * 10^{-38} \nu_0^4 n^2 \qquad (27)$$

The two resonator method was developed to allow use of a source noisier than the DUT, cancelling out its frequency fluctuations by adding a resonator in the other side of the bridge. The second resonator then becomes the reference against which the DUT is measured. The noise detected will be the sum of the noise of the two resonators, assuming the resonator Q factors have been matched closely enough to neglect source noise. Here the second resonator should be selected for low noise.

The other primary contributor of noise in resonator measurements is the phase detector and amplifier noise. In oscillators, the amplifier phase fluctuations must be much less than resonator phase fluctuations. Detector phase fluctuations are not critical here because oscillator noise is much higher at the offset frequencies of interest. In bridges, buffer amplifiers and the phase detector must exhibit much lower phase fluctuations than the resonators to be measured. Earlier, it was discussed how resonator phase fluctuations are proportional to Q_l. The more the loaded Q is decreased, the closer resonator phase noise will be to amplifier and detector noise. In this section, the assumption was made that the best resonators have $\mathcal{L}_r(1) = -123$ dBc under the condition that the Q_l equals the unloaded Q. A more likely condition might be that $Q_l = .25\ Q_u$, which will reduce the resonator noise by 12 dB (noise power is proportional to $(Q_l)^2$), to -135 dBc. A low level mixer or an amplifier may only be as good as -135 dBc at 1 Hz. This illustrates how the noise floor can quickly become limited because of inadequate loaded Q. Feedback amplifiers and high level mixers can be used to get better noise performance, as good as $\mathcal{L}(1) = -145$ dBc, up to UHF frequencies. Performance tends to degrade at higher frequencies, for various reasons.

In single resonator bridges and oscillator measurements it is useful to think of the noise floor in terms of oscillator noise. This allows source noise, resonator caused oscillator noise, and amplifier or phase detector noise converted to oscillator noise (either by calculation or by an oscillator) to be plotted on one graph, allowing direct comparison. When thinking in terms of oscillator noise, resonator caused oscillator noise and source noise will remain constant, but amplifier or detector phase fluctuations are converted to frequency fluctuations inside the resonator bandwidth. The lower the Q, the wider the bandwidth, the higher the noise. This is illustrated in Figure 16. The application of this idea is the same for the bridges and oscillators. This is just a straightforward application of Equation 8 or 9, where detector or amplifier noise is substituted for $\mathcal{L}_a(f)$.

The system noise due to amplifiers and the phase detector for bridges can be measured fairly directly by substituting an impedance equal to the resonator impedance at resonance, and then measuring the system noise using the calibration constants, K_ϕ and Q_l, or K_d, as were used with the resonators. In the case of the oscillator, the loop must be broken, then open loop phase fluctuations and Q_l determined, in order to calculate what the amplifier contribution would be.



Figure 16. Noise floor degradation due to low Q.

The low noise amplifier following the phase detector can also add to the noise floor, unless it has been designed for low 1/f noise. Signal levels out of the phase detector should be large enough so that the LNA noise does not contribute significantly to the measurement, while being careful to keep the resonator drive level low enough to avoid nonlinear resonator operation.

Determining the reference noise is simply a matter of measuring the source phase noise for the one resonator bridge and the oscillator methods. The dual resonator method requires a comparison of three resonators, two at a time, to determine the noise of the reference resonator, unless one can be found that is much better than any DUT to be measured.



Figure 17. Correlation Method To Improve Noise Floor.

Reduction Of Noise By Correlation

One possible method to improve the noise floor limitations on resonator measurements is to employ a correlation technique. The system shown in Figure 17 measures the noise of the DUT against two other resonators. Source noise is cancelled out, as with the dual resonator bridge. Each detector output is fed into a correlator/spectrum analyzer. If the cross power spectral density is averaged a sufficient number of times, the only component left will be DUT noise, since it is the only noise source common in both paths to the correlator. All other noise components should average to zero.

Improved Single Resonator Bridge

One of the problem in using a bridge circuit is that setting quadrature manually is time consuming. The single resonator bridge also suffers from temperature fluctuations which can cause the operation point to shift away from quadrature during a measurement, requiring re-adjustment. These problems can be solved by replacing the manual phase shifter with an electronic one. Feedback from the phase detector can be used to force the detector error voltage to zero. An integrator with a pole approximately at zero can be used to keep the error very small. The modified block diagram is shown in figure 18.

The feedback loop reduces the effective phase detector sensitivity. With an integrator in the loop, the bridge will have a high-pass response to phase fluctuations. The corner frequency can be made very low in frequency (however, this increases the time it takes the loop to lock initially), or the frequency response can be characterized. This circuit has a similar effect on the discriminator as the PLL has on an oscillator measurement.

The phase shifter can be made from a quadrature hybrid and two varactor diodes. The diodes are placed on the hybrid outputs. The signal is output from what is normally the isolated port of the hybrid. The phase shift is proportional to the varactor capacitance, which is tuned by coupling in a voltage through RF chokes. The phase shifter is a low noise device as long as the source has low AM noise and the tune line is well filtered. It does convert these noise sources to phase modulation, so care must be taken.



Figure 17. Improved Single Resonator Bridge.

Conclusions

Resonator 1/f frequency fluctuations are the dominant source of noise in a well designed oscillator. Three methods to measure resonator noise were discussed, each having advantages. The single resonator bridge may prove to be the best for production testing of resonators, since it may be quickly calibrated using the method presented here, and yields unambiguous results.

Analysis of resonator noise under different loading conditions has yielded results that were not commonly known. Resonator phase fluctuations were found to be proportional to loaded Q, and oscillator phase noise caused by the resonator was found to be independent of loaded Q. It is still important to maintain high loaded Q because the effect of amplifier and detector noise is increased relative to resonator noise when Q_L is reduced. This applies to oscillators and for measurements made in bridge circuits.

It is probably best to specify resonator noise in terms of resonator caused oscillator phase noise, if an oscillator is the intended application. If resonator phase fluctuations are specified, the loaded Q in the test must also be measured to get an accurate prediction of noise, since the target circuit may have a different loaded Q. This can be accommodated for by scaling the measured noise data.

1. T.E. Parker, "1/f Frequency Fluctuations in Acoustic and Other Stable Oscillators", Proceedings 39th Annual Frequency Control Symposium, pp. 97-106, May 1985.

2. M.M. Driscoll and Bruce W. Kramer, "Spectral Degradation in VHF Crystal Controlled Oscillators Due to Short-Term Instability in the Quartz Resonator", 1985 Ultrasonics Symposium Proceedings, pp. 340-345.

3. D.B. Leeson, "A Simple Model of Feedback Oscillator Noise Spectrum", Proceedings of the IEEE, Vol. 54, No. 2, pp. 329-330, Feb. 1966.

4. C. Moulton, "Analysis and Prediction of Phase Noise in Resonator and Oscillators", Hewlett-Packard RF and Microwave Symposium, 1985. Available from Hewlett-Packard, Signal Analysis Division, 1424 Fountain Grove Parkway, Santa Rosa, CA 95401.

5. M.M. Driscoll, "Low Noise Crystal Oscillators Using 50-Ohm Modular Amplifier Sustaining Stages", Proceedings 40th Annual Frequency Control Symposium, pp. 329-335, May 1986.

6. 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.

7. S.S. Elliot and R.C. Bray, "Direct Phase Noise Measurements of SAW Resonators", 1984 Ultrasonics Symposium Proceedings, pp. 180-185, Oct. 1984.

Fabrication and Methods for Evaluation and Circuit Utilization of Prototype Lateral Field Resonators

M.M. Driscoll
Westinghouse Defense and Electronics Center
and
W.P. Hanson
Piezo Crystal Company

SUMMARY

Lateral field SC cut resonators may have improved aging due to the reduction of electrode material in the active region of the plate.[1] Several electrode configurations exist for the lateral field. Single-side electrode lateral field resonators (SLFR) have the electrodes plated on the contoured side of the blank. This configuration allows the resonators to be adjusted to frequency with standard plating equipment.

Precision clean room processes were used in manufacturing the resonators to minimizing aging. Phase noise data taken on an HP-3047 system, Allan Variance, and Long Term aging are being compiled on these resonators.

The higher impedance level exhibited by these resonators would normally make device evaluation and use difficult. A straightforward method has been used to transform (reduce) the resonant impedance of these devices. The technique involves use of lumped element approximation, quarter wavelength transmission lines. Use of maximum permissible line impedances results in desirably large transformation circuit bandwidth, and the resonator static capacitance can be absorbed into the circuitry in order to obtain a symmetric resonant impedance characteristic.

Use of this technique has allowed accurate determination of prototype resonator motional parameters and measurement of device self noise (short term stability) using conventional, 50 ohm test equipment and has also been applied to the design or 5 MHz oscillators incorporating the prototype resonators. The measured oscillator signal spectrum is characterized by a noise floor level of -168 dB/Hz. The flicker of frequency level at fm = 1 Hz is -105 dB/Hz and is due to measured [Sy (f $-$ 1 Hz) = 2.5×10^{-9}] short term instability in the resonators themselves.

INTRODUCTION

The lateral field resonator (also known as the Parallel Field resonator) has been around since the 1940's. Early workers in lateral field resonators include R. Bechmann,[2] V. Ianouchevsky,[3] A. Warner,[4] A. Ballato,[5] and others. Recent work by B. Goldfrank and A. Warner[6] at Frequency Electronics resulted in lateral field SC cut resonators with respectable Q's and electrical parameters. A clear advantage of the lateral field design with SC cut resonators is the suppression of the B-mode in the SC cut. The control of the B-mode is accomplished by the orientation of the electrodes to minimize or maximize the lateral field coupling. Since the C mode is desired the coupling of the B-mode to the laterally applied electric field must be minimized. In the SC cut, using a four point mount 90 degrees apart, the coupling of the B-mode is reduced significantly by rotating five degrees clockwise from the X-axis when the $-$X side is down.

Two basic designs were tried. The single sided lateral field and the composite lateral field. The single sided lateral field simply has electrodes on one side while the composite lateral field has identical electrodes on both sides. Frequency adjustment was accomplished in two ways. On the single sided LFR's a central spot was plated on the unelectroded side. This may reduce the effectiveness of the lateral field to age well and is no longer being used. The other method is to evaporate two spots, one on each lateral field electrode. This method is capable of adjusting the frequency of the LFR's up to 150 Hz without affecting the performance of the resonators.

The measured Q's of the single sided and double sided LFR's were similar; the Rs were very different.

Single Sided
Mean Q : 1.9 million
3 sigma : .600 Million (1.3 M to 2.5 M)
Mean Rs : 3,200 ohms

Double Sided
Mean Q : 1.6 million
3 sigma : 1.0 million (.6 M to 2.6 M)
Mean Rs : 666 ohms

LFE RESONATOR EVALUATION/OSCILLATOR DESIGN

CIRCUIT UTILIZATION DIFFICULTIES

Analyses have shown that the degree of acoustic coupling obtainable with LFE resonators is comparable to that of conventionally configured devices. In simple terms, this means that, in the equivalent resonator electrical circuit, similar capacitance ratio is obtainable. The capacitance ratio (Co/Cm in figure 1) relates the impedance level of the resonant (motional arm) portion of the equivalent circuit to that of the inter-electrode capacitance, Co.

The inter-electrode capacitance for the LFR resonator is extremely small as a consequence of electrode placement on the same surface. HF LFE resonators have been fabricated by Piezo Crystal Company (and others) that typically exhibit impedances 10 to 50 times higher than conventionally configured (thickness excitation) devices.

The ultra high LFE resonator electrical impedance prevents accurate resonator evaluation (equivalent circuit determination) using either conventional equipment such as the crystal impedance (CI) meter, or newer methods utilizing 50 ohm network analyzers. Not only

CH2427-3/87/0000-429 $1.00 C 1987 IEEE

is device evaluation difficult, but actual device use in practical oscillator and filter circuits is not easily implemented as a consequence of high device impedance level.

Simple magnetically coupled transformers do not provide an effective means of accurately transforming (reducing) the entire resonant impedance characteristic of LFE resonators. The (useful) electrical resonant impedance of a typical LFE HF resonator that must be transformed varies from off-resonant reactance values approaching ± j 20,000 ohms to at-resonance resistive values of 500 to 2000 ohms. The circuit described in the following section constitutes a means for accurately transforming the entire useful LFE device resonant impedance characteristic to a practical level.

IMPEDANCE MATCHING USING QUARTER WAVELENGTH LINES

Figure 2 shows a well known method used in microwave circuits for the purposes of impedance transformation and matching. As shown in the figure, a cascaded connection of two quarter wavelength lines can be used to transform a given circuit impedance by the square of the ratio of each of the quarter wavelength line characteristic impedances.

At HF and VHF frequencies where the physical length of a quarter wave transmission line is large, the quarter wavelength lines of figure 2 can be configured as lumped element approximations. The use of lumped element approximations also allows necessary (as will be shown) selection of large value characteristic impedances not easily implemented in distributed form.

Prototype LFE resonators were designed and fabricated by Piezo Crystal Company for evaluation. The estimated crystal series resonant (5 MHz) resistance was on order of 2000 ohms. An impedance step-down ratio of 36 to 1 was considered in order to obtain a transformed resonator series resistance of 55 ohms.

Figure 3 shows the circuit that was used to implement the impedance transformation. Super Compact network analysis software was utilized to compute, for the quarter wavelength line circuits, the effects of: (1) number of LC sections, (2) absolute characteristic impedance level (holding ratio Z01/Z02) constant, (3) single ended versus balanced circuits, and (4) circuit losses and parasitics (distributed capacitance and lead inductance). The results may be summarized as follows.

1. In order to not only effectively transform the crystal in-band resonant impedance characteristic, but also to maintain (the transformed) high out-of-band impedance of the resonator over the widest possible bandwidth, the characteristic impedances of the (lumped element) quarter wave lines should be as high as possible. Maintenance of high out-of-band impedance over widest possible bandwidth is desirable in connection with avoidance of spurious oscillation at quarter waveline LC resonant frequencies when the resonator with quarter wave networks is used in oscillator circuitry.
2. A two section approximation for each quarter waveline suffices. Use of additional sections does not dramatically improve transformation or bandwidth performance. For high impedance lines, a balanced approximation works well, allowing the high impedance line total required large inductance value to be distributed between four inductors.
3. The use of two (lumped element approximation) quarter wave lines provides widest bandwidth compared to both maximum bandwidth bandpass structures and structures consisting of more than two quarter wave lines.
4. The use of the disclosed impedance transformation circuitry has been tried experimentally and has allowed successful, accurate, and simple determination of resonator equivalent circuit parameters, measurement of resonator self noise level, and fabrication and test of a resonator stabilized oscillator.
5. The terminating capacitance used in the high impedance quarter wave line circuit allows absorption of the resonator static capacitance, thus providing a desired symmetrical resonant response at the input to the quarter wave line input terminals.

Figures 4 and 5 show, respectively, the computed and measured LFE resonator (Smith Chart) impedance transformed (reduced) by 36:1 using the figure 3 circuit. As shown in figure 4, out-of-band non-crystal controlled resonances occur due to the transformation circuitry at 3.8 and 5 MHz. In an oscillator circuit, then, additional selectivity is required to prevent oscillation at these frequencies, as well as those occurring at the unwanted crystal resonant modes.

DISCRETE TRANSISTOR OSCILLATOR DESIGN

Figure 6 shows a simplified AC schematic diagram for the oscillator circuit that was designed for use with the prototype LFE resonators and including the impedance transformation circuit of figure 3.

As shown in the figure, a self-limiting, grounded collector, (crystal) series resonant oscillator design was selected, with the exception that output signal current is extracted via the sustaining stage transistor collector via a common base buffer stage. The buffer stage is biased at higher DC current than sustaining stage in order to achieve linear (class A) buffer operation under conditions of self (current) limiting in the sustaining stage transistor. In the figure 6 configuration the sustaining stage may be regarded as a means for generating a sufficient negative resistive impedance across the crystal (plus impedance transformation circuit) connection terminals. Viewed this way, the negative resistance magnitude is determined by proper selection of tank circuit element values. In addition, tank circuit loaded Q should be as low as possible (for minimum oscillator frequency sensitivity to tank circuit component value variation with time and temperature), yet sufficiently high to eliminate the possibility of oscillation at crystal (plus transformation circuit) out-of-band resonant frequencies. For the circuit of figure 6, the tank circuit Q was approximately 9 (including inductor self loss), and crystal dissipation was approximately 2 mw.

Figure 7 shows the measured phase noise sideband spectra for two, phaselocked, figure 6 type oscillators. The spectrum is characterized by a flicker of frequency region with £(1 Hz) = −100dB/Hz, a flicker of phase region with £(100 Hz) = −142 dB/Hz, and a white phase noise (floor) level of −165 dB/Hz. The flicker of phase

spectral level was higher than anticipated and was not improved via linear sustaining stage transistor operation (using an auxiliary, schottky diode limiter for oscillator ALC). The flicker of frequency noise (as will be shown) portion of the spectrum is due to short term instability in the resonators themselves and is similar to that measured for similar conventional, thickness excitation devices. [4]

LFE OSCILLATOR USING MODULAR AMPLIFIER

There are notable advantages associated with the use of a 50-ohm modular amplifier as the oscillator sustaining stage. Among these are easily implemented oscillator circuit open loop characterization and evaluation using 50-ohm test equipment, and the fact that modular amplifiers exhibiting very low levels of flicker of phase noise are readily available.[5]

Lumped element approximation, quarter wave impedance transformation circuitry can also be used in the design of the modular amplifier oscillator. In the design approach utilized, no additional frequency selective tuned circuitry is required to prevent oscillation at unwanted crystal resonant modes. Oscillator frequency sensitivity to (L-C) component value change is therefore greatly reduced. One reason this is possible is that for the LFE SC-cut resonator, suppression of nearby B mode resonance is achieved via proper electrode orientation.[6,7]

Figure 8 shows a simplified AC schematic diagram for the oscillator circuit. As shown in the figure, the resonator is included in the oscillator feedback circuit and is operated at series resonance such that its series resistance constitutes the series arm of a π type attenuator whose attenuation is equal to the net feedforward gain of the amplifier/power splitter.The amplifier/power splitter signal phase shift is approximately 180 degrees. Two lumped element approximation, quarter wavelength lines are used to: (1) match the attenuator characteristic impedance, Z_A, to 50 ohms, and (2) furnish the requisite additional 180 degree (90 degrees each) circuit signal phase shift. For an amplifier/power splitter gain of 16 dB and a nominal crystal series resistance of 500 ohms (exhibited by second iteration, prototype resonators), the 16 dB attenuator shunt resistance is:

$$R_P = .45\ R_S \approx 220 \text{ ohms} \quad [1]$$

and the attenuator characteristic impedance is:

$$Z_A = .325\ R_S \approx 162 \text{ ohms} \quad [2]$$

The quarter wavelinc characteristic impedance is given by:

$$Z_O = \sqrt{(Z_A)(50 \text{ ohms})} = 90 \text{ ohms} \quad [3]$$

In order to operate the amplifier linearly, a schottky diode limiter is used at the attenuator input such that, at a prescribed signal level, the ALC circuit reduces the attenuator net input shunt resistance to the prescribed (approx. 220 ohm) steady state value. A prototype oscillator was fabricated using a Q-BIT QBH-124 amplifier and a 500 ohm, 3rd overtone LFE SC-cut resonator. Predicted circuit operation was easily and accurately achieved. Oscillation at crystal unwanted resonant modes is prevented due to the fact that the quarter wave line circuitry does not provide required circuit phase shift and/or impedance match at the crystal fundamentals and unwanted higher overtone resonant mode frequencies.

RESONATOR FLICKER NOISE MEASUREMENT

Independent measurement of prototype LFE resonator short-term frequency stability was also made by driving pairs of device from a low noise synthesizer in a balanced transmission circuit.[8,9] The same quarter wave, impedance transformation networks in the oscillator circuits of figures 6 and 8 were also used in the measurement apparatus, as shown in figure 9. No attempt at fine frequency set-on was made in the fabrication of the initially utilized LFE resonators. External (varactor diode) reactive tuning sensitivity was limited to several ppm. It was necessary to operate the resonators at different temperatures in order to achieve the degree of frequency match required for measurement of resonator (figure 9) and phase locked oscillator (figure 7) short-term frequency stability.

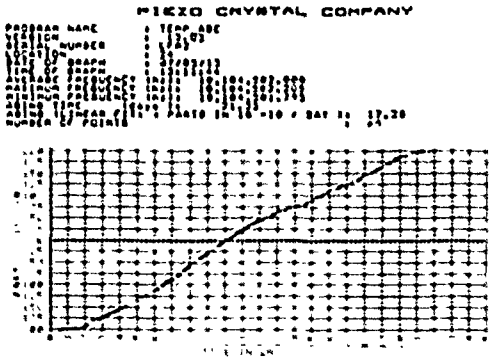
Figures 10 through 12 show typical resonator short-term stability measurement results. At this time it has not been determined to what extent laboratory environmental effects (such as low level ambient temperature fluctuations and vibration) may have affected the results. Figure 13 shows the measured signal phase noise sideband spectra using the (figure 12) 500 ohm crystals in two figure 8 type oscillators.

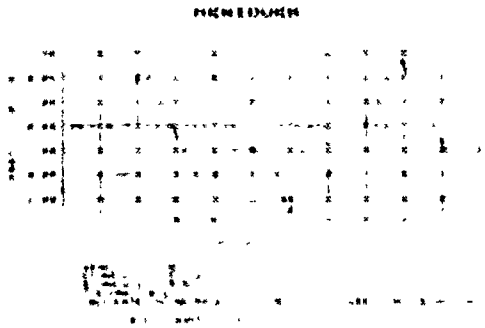
As shown in the figures, the initial data indicates a 18 dB variation in resonator short-term stability with good correlation between individual resonator and resultant oscillator signal Sy(f = 1 Hz) frequency stabilities in the flicker of frequency portion of the spectra.

MODERATE AND LONG TERM STABILITY

AGING

Two aging experiments were conducted with the double sided lateral field resonators. One resonator, finished early, was placed in an oscillator and aged for several weeks. The resonator was left open during some initial testing and was not expected to age well. The aging after 19 days is shown in the next figure.





The aging rate was 1.7 parts in 10^{9} per day. This is disappointing, but not surprising since the resonator was not completely finished in a clean room. The second aging experiment was an attempt to compare the aging of thickness mode resonators to LFR's. Twenty-four crystals, 13 thickness mode and 11 LFR's were manufactured at the same time with precision resonator manufacturing processes. All resonators were tested over six days at 160°C. The resonators were plated, sealed and tested at the same time. The results are summarized below:

Thickness Mode Resonator

Mean Aging	: 1.35 × 10^{-9}/day (@ 160°C)
Range	: 2.00 × 10^{-9}/day to
	4.40 × 10^{-9}/day
3 sigma	: 4.03 × 10^{-9}/day

Lateral Field Resonators

Mean Aging	: 7.45 × 10^{-9}/day (@ 160°C)
Range	: 0.00 × 10^{-9}/day to
	1.80 × 10^{-9}/day
3 sigma	: 1.77 × 10^{-9}/day
Allan Variance	: 1 × 10^{-11} @ 1 second
G-sensitivity	: 1 × 10^{-9}/g

From this data the LFR's aged 1.8 times better than the thickness mode resonators. The difficulty with this data is evaluating the aging rate at lower temperatures. This can not be done easily. Piezo Crystal Co.'s standard 5 MHz 3rd overtone SC cut has a mean aging rate of 6 × 10^{-11}/day at 85°C. If the ratio of aging rates is maintained at lower temperatures than the LFR's should age at 3 × 10^{-11}/day. This has not been confirmed yet by other methods.

CONCLUSIONS

Long term aging of precision lateral field resonators were measured to be 1.8 times better than thickness mode resonators at 160°C. Long term aging at lower temperatures may not correlate with high temperature aging.

The use of lumped element approximation, quarter wave transmission line networks has been shown to provide a practical means for reducing LFE resonator resonant impedance levels, allowing device use in conventional, discrete transistor and 50 ohm modular amplifier oscillator circuitry. In the case of the modular amplifier circuit, the inherent suppression of the SC-cut B mode resonance obtainable using the LFE configuration, coupled with the broadband nature of the oscillator and transformation circuitry, results in greatly reduced frequency sensitivity to drift in the non-resonator portion of the circuitry.

Transmission line resonator impedance transformation circuitry has also provided an effective means for accurate determination of equivalent electrical circuit parameters and measurement of device self-noise. This approach should prove especially useful in the design of oscillator and resonator measurement apparatus incorporating higher frequency LFE resonators.

REFERENCES

1. Frequency Electronics Inc., "Lateral Field, SC-Cut, Resonator Development", Contract DAAK20-83-C-0418 with ET&D Laboratory, ERADCOM, Ft. Monmouth, NJ 07703
2. R. Bechmann, "Über Dickenschwingungen piezoelektrischer Kristallplatten," Archiv der Elektrischen Übertragung, Vol. 6, September 1952, pp 361-368.
3. V. Ianouchevsky, "Parallel Field Excitation," Proc. IRE, Vol. 48, June 1960, p. 1165.
4. A.W. Warner, "Use of Parallel-Field Excitation in the Design of Quartz Crystal Units," Proc. 17th AFCS, May 1963, pp 248-266.
5. A.D. Ballato and R. Bechmann, "Piezoelectric Crystal Element," U.S. Patent No. 3,202,846, issued 24 August 1965; U.K. Patent No. 1,028,102, issued 4 May 1966.
6. B. Goldfrank and A. Warner, "Development of SC-cut Lateral Field Resonators," Research and Development Technical Report, ERADCOM, U.S. Army Electronics and Research Command, Fort Monmouth, NJ, May 1984 and June 1985.
7. F.L. Walls and A.E. Wainwright, "Measurement of the Short Term Stability of Quartz Crystal Resonators", IEEE Trans. on I. and M., Vol. IM-24, No. 1, March, 1975, pp 15-20.
8. J.J. Gagnepain, "Fundamental Noise Studies of Quartz Crystal Resonators", Proc. 30th Frequency Control Symp., June, 1976, pp 84-87.
9. M.M. Driscoll, "Low Noise Crystal Oscillators Using 50 Ohm Modular Amplifier Sustaining Stages", Proc. 40th Frequency Control Symp., May, 1986, pp 329-335.
10. A. Ballato, et. al., "Simple Thickness Plate Modes Driven by Lateral Fields", Proc. 39th Frequency Control Symp., May, 1985, pp 162-172.
11. A. Ballato, et. al., "Lateral Field Equivalent Networks and Piezo Coupling Factors of Quartz Plates Driven in Simple Thickness Modes", IEEE Trans. on UFFC, Vol. UFFC-33, No. 4, July, 1986, pp 385-393.
12. M.M. Driscoll and B.W. Kramer, "Spectral Degradation in VHF Crystal Controlled Oscillators Due to Short Term Instability in the Quartz Resonator", Proc. 1985 IEEE Ultrasonics Symp., Oct., 1985, pp 340-344.

Figure 1. Crystal Resonator Equivalent Electrical Circuit

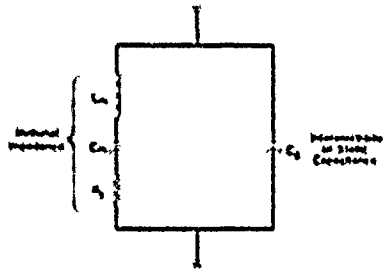


Figure 2. Impedance Transformation Using Quarter Wave Transmission Lines

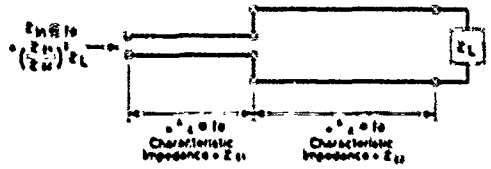


Figure 3. Lumped Element Realization for Quarter Wave Line Impedance Transformer

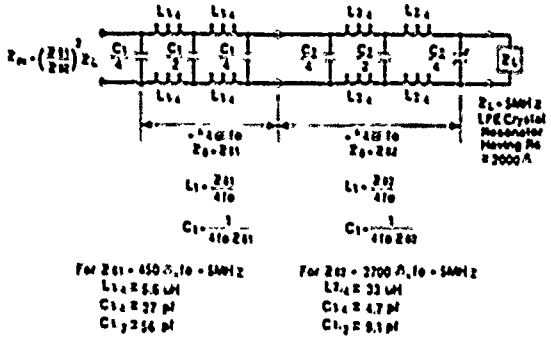


Figure 4. Calculated LFE Resonator Impedance Transformed (reduced) by 36:1 (Including Parasitics and Losses in Lumped Element Quarterwave Lines)

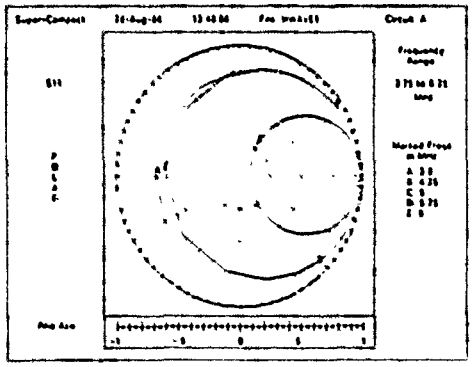


Figure 5. Measured LFE Resonator Impedance Reduced by 36:1 Using Lumped Element Quarterwave Lines

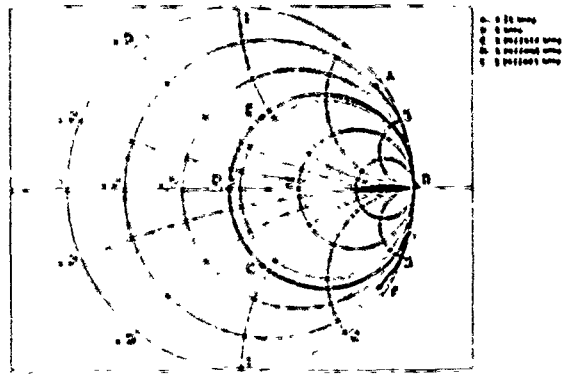


Figure 6. Simplified Functional Schematic Diagram for Discrete Transistor LFE Crystal Oscillator

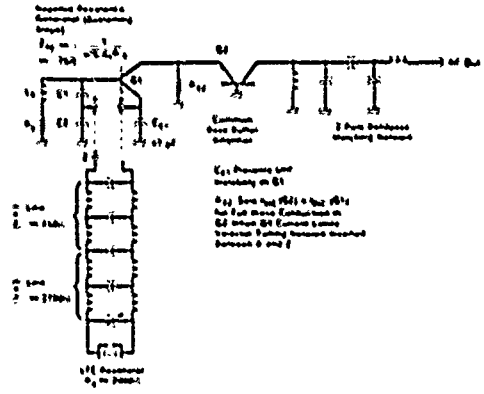


Figure 7. LFE Oscillator Measured Output Signal Noise Spectrum (Compare to Predicted Level in Figure 9 Based on Crystal Self Noise)

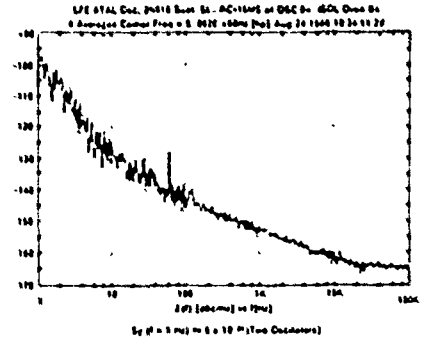


Figure 8. Simplified Functional Schematic Diagram for LFE Crystal Oscillator Using 50 Ohm Modular Amplifier

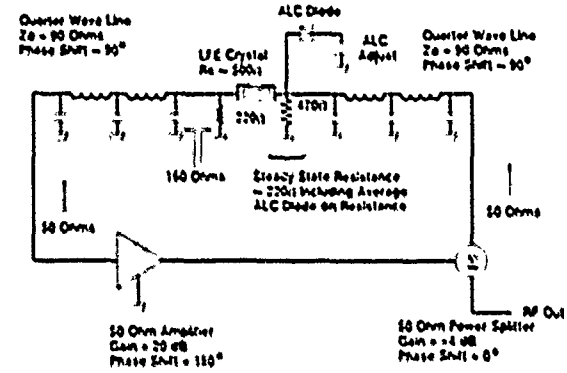


Figure 9. Quarterwave Transformers Used in LFE Resonator Noise Measurement

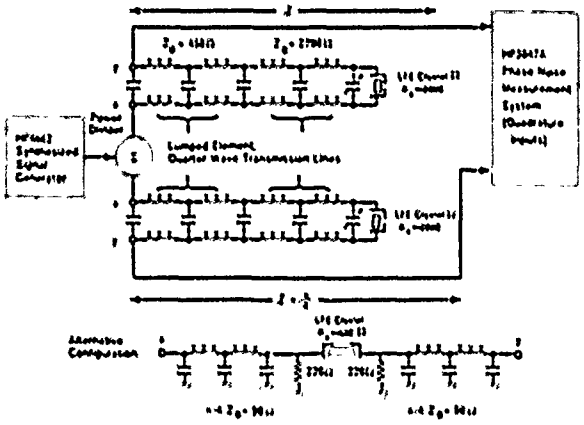


Figure 10. Noise Measurement Results for 2000 Ohm Prototype LFE Resonators

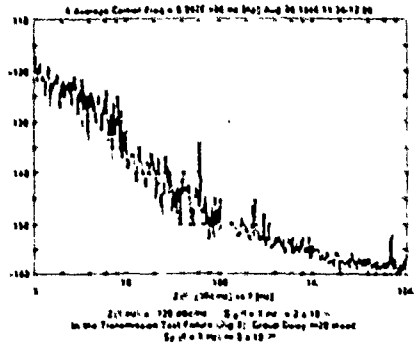


Figure 11. Noise Measurement Results for 500 Ohm Prototyp Resonators

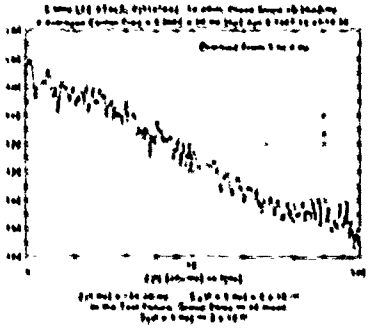


Figure 12. Noise Measurement Results for Second Set of 500 Ohm LFE Resonators

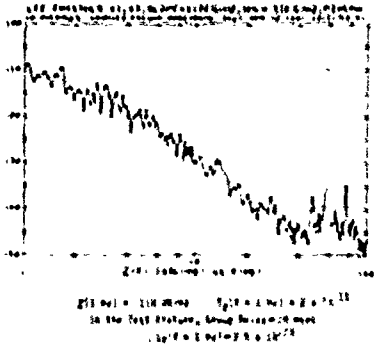
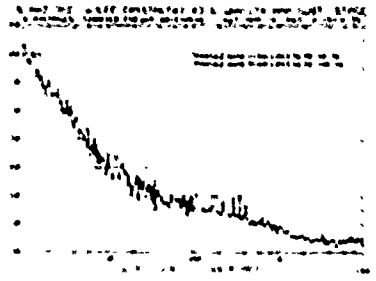


Figure 13. Measured Signal Phase Noise Sideband Spectra fo 500 Ohm LFE Crystals Installed in (Figure 8) Oscillators



41st Annual Frequency Control Symposium - 1987

A DIGITALLY COMPENSATED TCXO USING A SINGLE CHIP LSI

T.HARA, T.KUDO, S.URIYA, H.SAITA, S.OGOU and Y.KATSUTA

NEC Corporation

Kawasaki , 211 Japan

Summary

This paper describes a digitally compensated TCXO using a single chip LSI. Recently, with progress in developing communication systems, such as a mobile telephone system, the oscillator has been required to be smaller in size, have greater frequency stability and lower power consumption.

Temperature compensated crystal oscillators (TCXOs), which include crystal units, have usually been used in these systems.

The authors have developed a single chip LSI for a TCXO, which has digitally compensated frequency temperature characteristics for an oscillator using LSI technology, and have achieved a small digitally compensated TCXO(DTCXO).

The single chip LSI has functional blocks consisting of digital and analog circuits. It includes a temperature sensor, a 9 bit A/D converter, a 512 word x 8bit EPROM, an 8 bit D/A converter, a reference voltage supply circuit, an amplifier circuit for the oscillator and a divider.
The temperature sensor has been made from bipolar transistors. The desired temperature compensating accuracy is 0.5 °C. This LSI chip size is 3.5mm x 5.3mm. About 16,000 transistors have been integrated in it, using CMOS 2.0um design rules.

Important features for the DTCXO with this LSI and a 12.8MHz fundamental AT cut crystal unit are small frequency variations with changing temperature (±0.3ppm from -30 °C to +85 °C), low power consumption (25mW), good long term frequency stability (±0.4ppm/year), insensitivity to vibration and shock (±0.05ppm) and small size(10mm x 12mm x 25mm).

Introduction

TCXOs have been widely used in communication systems. Most TCXOs use thermistors as temperature sensors and a variable capacitance diode (varactor) as a variable reactance element.

At least three thermistors are necessary to compensate for the change in the AT cut crystal unit, whose frequency temperature characteristics are approximated by a third order function. The varactor should have good linearity. The calculation of values for each component is rather complicated and usually requires a computer.

These requirements for an expensive varactor with good linearity, many thermistors, a selected crystal unit with good frequency temperature characteristics and the complicated calculation to determine the individual component values make an oscillator considerably complicated and hence expensive to manufacture.

This paper presents a digitally compensated TCXO using a single chip LSI. It can be easily applied to compensate for changes in frequency temperature characteristics for an AT cut crystal unit.

The DTCXO design features are shown in Table 1.

Principle

The DTCXO using the single chip LSI consists of three basic sections, as shown in Fig.1 , a temperature sensor section, a digital frequency temperature compensating section and a voltage controlled crystal oscillator (VCXO) section. A functional diagram of the digital frequency temperature compensation circuit is shown in Fig.2. It includes the 9 bit A/D converter, the 512 word x 8bit EPROM, the 8 bit D/A converter, paralell to serial converters and serial to paralell converters.

There are two methods for including a temperature sensor in the LSI.

(a) Resistance and capacitance temperature coefficients are used.

(b) A bipolar transistor P-N junction is used.

The authors have selected method (b), considering the element stability due to aging, to generate about -8mV/°C temperature sensitivity. Initial output voltage values for the temperature sensor is adjustable. In order to obtain the desired dynamic range for the A/D converter, the adjustment is achieved by blowing polysilicon fuses.

CH2427-3/87/0000-435 $1.00 C 1987 IEEE

The temperature sensor is integrated in the LSI in order to accurately detect temperature changes. This temperature data is digitized and characterized into 2^{9} break points, where 9th is the number of bits used for the A/D converter. The EPROM is programed with a customized look-up table, which provides corrections directly to each digitized temperature word. The word, from 0 to 511, is used to as the address for the 512word x 8bit EPROM. The EPROM then determines the correction voltage needed at that temperature and outputs the digital correction to the 8 bit D/A converter. The D/A converter converts this value to an analog voltage for application to the varactor in the VCXO.

It is not difficult to show that the maximum frequency error, due to finite word length, is given by[3]:

$$\Delta F = \frac{S}{2}\left(\frac{T}{2^{bt}}\right) + \frac{F}{2}\left(\frac{1}{2^{bv}}\right) \qquad ---(1)$$

Where:

- bt = number of bits in the temperature word.
- bv = number of bits in the crrection voltage word.
- S = maximum frequency-temperature slope for the VCXO, in ppm per °C.
- T = total temperature range, over which the oscillator must be compensated.
- F = maximum peak to peak frequency excursion of the crystal in ppm.
- ΔF = maximum frequency error allowed, in ppm.

Thus, the 9 bit A/D converter and 8 bit D/A converter would be considered, if these were the only errors involved.

Experimental Results

The authors have developed the single chip LSI for the DTCXO. Characteristics of the LSI are shown in Table 2. A photograph of the LSI is shown in Fig.3.

Compensation experiments, using an AT cut 12.8MHz crystal unit, are presented. Fig.4 and 5 indicate overall compensated result. The less than ±0.3ppm frequency variation is achieved over the from -30 °C to 85 °C temperature range.

Fig. 6 shows the frequency shift due to voltage variation.

Fig. 7 shows the long term frequency stability.

Fig. 8 shows the short term frequency stability.

A internal view of the DTCXO, using an AT cut crystal unit, is shown Fig 9. It is about 10mm x 12mm x 25mm.

Conclusions

The DTCXO using the single chip LSI has been developed. The DTCXO has the following excellent characteristics:

1) No thermistor is needed.

2) It has an easily adjustable frequency temperature compensation circuit.

3) Wide temperature range compensation is possible.

4) It is suitable for use in a small size TCXO.

Acknowledgment

The authors express their thanks to Dr.K.Sugibuchi ,Dr.N.Yoshida and Mr.S.Ito, for their encouragement.

References

1. T.Kudo, S.Fujii and S.Nanamatsu, "A New Frequency Temperature Compensation Method for Oscillators." Proc. 37th Annu. Symp. on Frequency Control, pp. 448-453,1983.

2. Buroker,G.E.,Frerking,M.E.,"A Digitally compensated TCXO" Proc. 27th Annu. Symp. on Frequency Control, pp. 191-198,1973.

3. Frerking,M.E.,"The Application of Microprocessors to Communications Equipment Design" Proc. 33rd Annu. Symp. on Frequency Control , pp.431-435,1979.

4. J.Lee, J.E.Andrews, K.W.Lee and W.R.Day, "Digital and Analog Frequency-Temperature Compensation of Dielectric Resonator Oscillators ", 1984 IEEE MTT-S Digest,pp, 277-279.

5. Alan B. Mroch and Glenn R. Hykes, "A Miniature High Stability TCXO using Digital Compensation", 30th Anmu. Symp. on Frequency Control, pp, 292-300.

Frequency Range	10~16MHz (12.8MHz)
Frequency Stability	±1.5ppm
Temperature Range	-20~85℃
Supply Voltage	+5V±5%
Power Consumption	125mW
Size	10×12×25mm (3cm³)

Table 1 DTCXO design features

Design Rules	C MOS 2μm
Chip Size	3.5mm×3.5mm
Temp. Sensitivity	-8mV/℃
A/D Converter	8bit
D/A Converter	8bit
EPROM Capacity	512W×8bit (4Kbit)
Reference Voltage	+3.5V
Supply Voltage	+5V±5%
Power Dissipation	120mW

Table 2 Characteristics of LSI

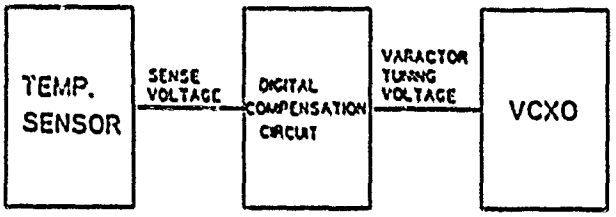


Fig. 1 Block diagram of DTCXO

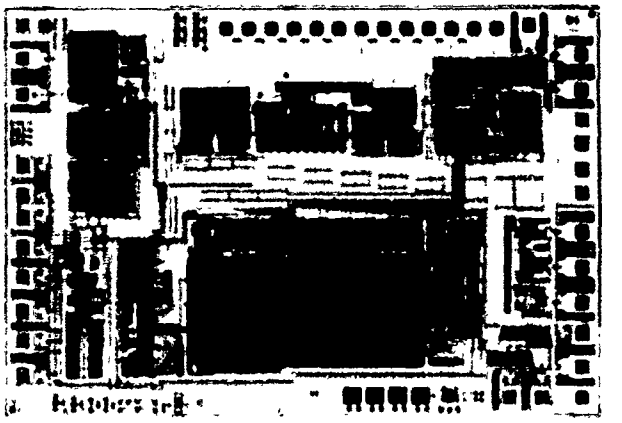


Fig. 3 Photograph of chip

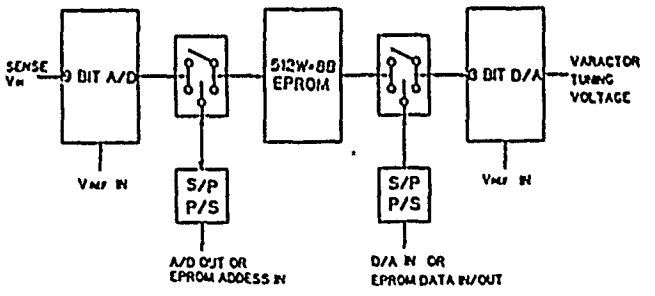


Fig. 2 Block diagram of digtal compensation circuit

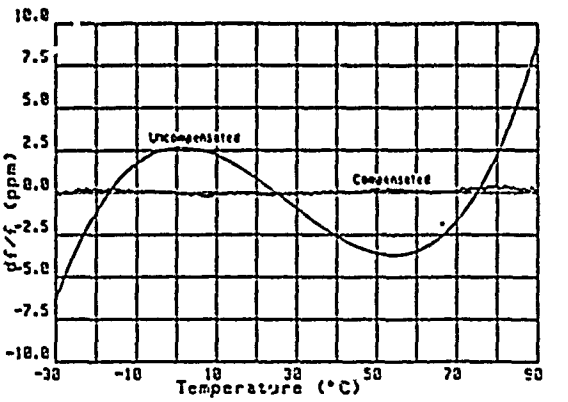


Fig. 4 Frequency vs. temperature characteristics

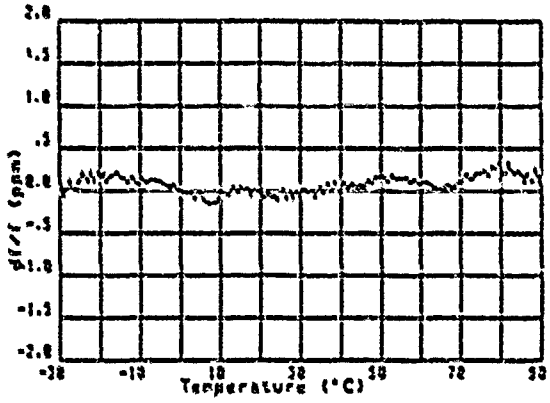


Fig. 5 Compensated frequency vs. temperature characteristics

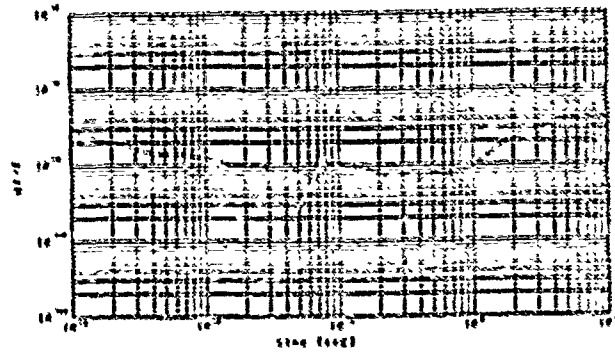


Fig. 8 Short term frequency stability

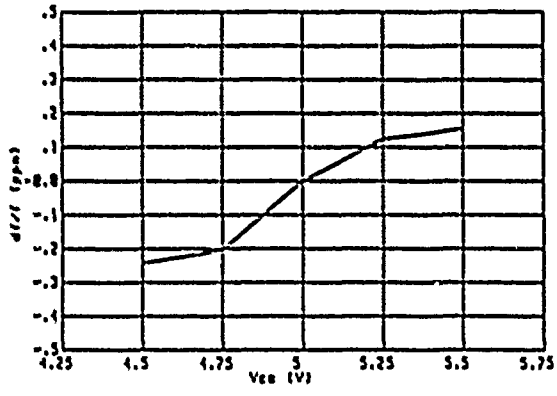


Fig. 6 Frequency shift due to voltage variation

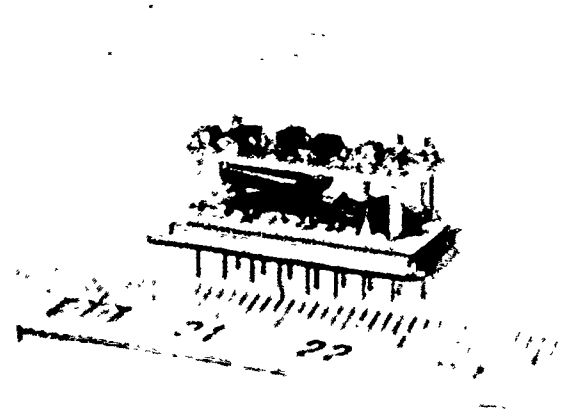


Fig. 9 Internal view of DTCXO

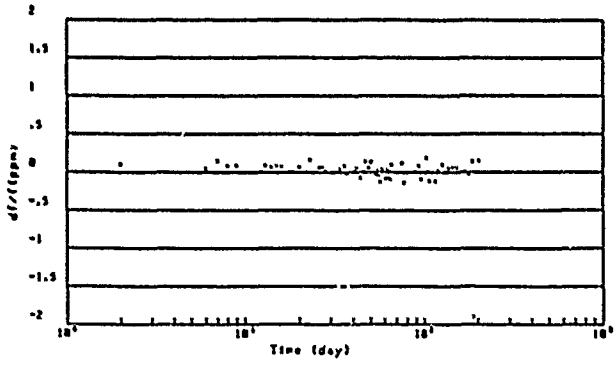


Fig. 7 Long term frequency stability

ANALYSIS OF HIGH PERFORMANCE COMPENSATED THERMAL ENCLOSURES

Fred L. Walls
Time and Frequency Division
Boulder, Colorado 80303

Abstract

Approximate analysis of the conventional thermal enclosures such as ovens and cryostats reveals that the limitation to achievable thermal regulation is in many cases not the gain of the thermal servo loop, but rather the fact that the experiment under observation within the thermal enclosure is still coupled to the outside temperature. So, even if the thermal enclosure is perfectly stable in temperature, the experiment is not. A new configuration is suggested which uses an additional sensor to measure changes in the outside temperature and compensate the temperature set point of the thermal enclosure in order to just correct for the temperature error induced by the coupling to the outside.

Approximate Analysis of Conventional Thermal Enclosures

The following approximate analysis is intended to illustrate how additional thermal sensors can be used to compensate a high performance thermal enclosure such as an oven or cryostat in order to substantially improve the effective thermal regulation of the enclosed experiment. Figure 1 shows the basic elements of the problem for a conventional single layer oven. The temperature of the outer shell T_0, the temperature of the heater/cooler T_H, and the temperature of the experiment T_E are initially assumed to be uniform. R_{HS} represents the thermal impedance between the oven heater/cooler and the sensor; R_{OS} the thermal impedance between the sensor and the outside; R_{HE} the thermal impedance between the heater/cooler and the experiment; R_{OE} the thermal impedance between the outside and the experiment, τ_{HS} the thermal lag between the sensor and the heater/cooler due to R_{HS} and the heat capacity of the sensor; τ_{HE} the thermal time constant of the experiment due to its heat capacity and the effective thermal impedance which is the parallel combination of R_{HE} and R_{OE}. The effects of small time dependent thermal gradients are also included in the model of the compensated thermal enclosure discussed later and schematically illustrated in figure 5 by the term $\delta_T(\tau)$.

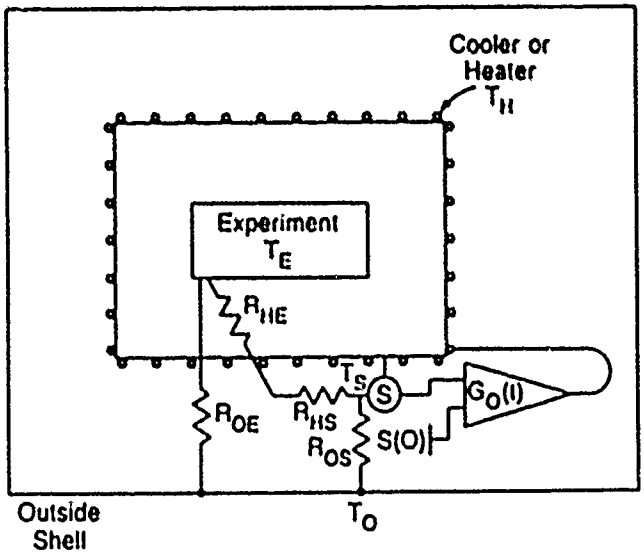


Fig. 1 Schematic representation of an experiment contained within an oven shell. The various temperatures are represented by T's and the thermal impedances by R's. $T_S(0)$ is the set point of the oven, and S is the thermal sensor.

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

The thermal performance of the enclosure is often characterized by thermal gain, $G_0(\tau)$, which is loosely defined by

$$T_s(\tau) - T_s(0) = \frac{T_0(\tau) - T_0(0)}{1 + G_0(\tau)} \qquad (1)$$

where the left side of the equations is the change in the sensor temperature, T_S, at a time delay "τ" from the initial value due to a change in the outside temperature, T_0, from its initial value at time $t=0$. Note that $G_0(\tau)$ is a function of the averaging time and of the general form

$$G_0(\tau) = \frac{\tau + \tau_0}{\tau} G_E(\tau) \qquad (2)$$

where τ_0 is the thermal delay time from the outside shell to the oven and $G_E(\tau)$ is the step response function of the electronic servo gain. From servo theory [1-3] one can show that the heater servo is stable when the response time (unity gain time) of $G_E(\tau)$ is about 4 times slower than the heater sensor delay time, τ_{HS}. $G_E(\tau)$ typically crosses the unity gain point with a slope of 6 dB per octave, although any slope below 12 db per octave is stable. Given this limitation on the gain slope and the need to reduce the effects of high frequency noise in the servo, $G_E(\tau)$ generally has a functional form

$$G_E(\tau) = \left[\frac{(\tau_0/\tau_{HS})\tau}{4\tau + 4\tau_0}\right]\left[\frac{\beta\tau}{\beta\tau + \tau_{HS}}\right] \qquad (3)$$

$$\left[1 + \frac{\tau}{\beta\tau_{HS}}\right]\left[1 + \frac{\tau}{\beta^2\tau_{HS}}\right]$$

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

where the first bracket comes from using linear feedback with the gain set to cross unity at 6 dB per octave and $\tau=4\tau_{HS}$, the second bracket contains the terms yielding additional electronic filtering of the high frequency fluctuations for times shorter than τ_{HS}/β, and the following terms show the effects of adding integrators to increase the gain slope at longer and longer times. β is typically chosen to between 4 and 10. This is illustrated in figure 2. By using a carefully adjusted balance between the proportional part of the gain and that of the integrator one can make the loop optimally fast with imperceptible overshoot [2]. From this representation of $G_E(\tau)$ it is clear that the maximum value of $G_E(\tau)$ that can be obtained at long averaging times or delay times is scaled as $(\tau/\tau_{HS})^{n+1}\,\beta^{-n}$, where n is the number of integrators in the $G_E(\tau)$. The smaller the delay time τ_{HS}, the easier it is to make $G_E(\tau)$ and hence $G_0(\tau)$ large. This analysis explicitly assumes that the sensor temperature error produces a linear correction in the applied heater/cooler temperature.

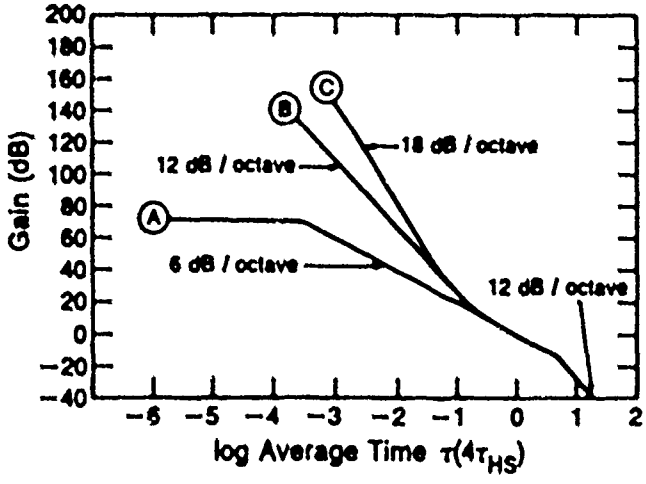


Fig. 2 Oven servo gain $G_0(\tau)$ in dB versus averaging time (measurement time) in units of $4\tau_{HS}$ for proportional gain in curve A, for proportional gain plus one integrator with $\beta=4$ in curve B, and proportional gain plus 2 integrators with $\beta=4$ in curve C.

It should also be noted that the sensor temperature is not exactly that of the heater/cooler, T_H, even for infinite $G_0(\tau)$, because the sensor has finite thermal resistance to the oven heater/cooler characterized here by R_{HS} and coupling to the outside characterized here by R_{OS}. The thermal error of the heaters/coolers in steady state is then of order

$$T_H - T_S = (T_0 - T_H)\,\frac{R_{HS}}{R_{OS}+R_{HS}}\,. \qquad (4)$$

The difference between the temperature T_E of the experiment and that of the sensor versus $G_0(\tau)$ is often characterized by a graph similar to that shown in figure 3. The asymptotic value of T_E as $G_0(\tau)\to\infty$ is approximately given by

$$T_E(\tau)-T_S(0) = \left(T_0-T_H\right)\left(1+\frac{R_{HE}}{R_{OE}+R_{HE}}\right)\left(\frac{R_{HE}}{R_{HE}+R_{OE}}\right)\left(\frac{\tau}{\tau+\tau_{HE}}\right)\,. \qquad (5)$$

Eq. 5 shows that even in the limit of $G_0(\tau)\to\infty$, there can still be a significant thermal error. The first term of the second bracket comes from the thermal error between the heater/cooler and the sensor (eq. 4) and the second, and often more important term, caused by the thermal coupling of the experiment to the outside. The role of τ_{HE} in the 4th bracket is seen to be a filtering of the thermal transients in the thermal control servo at very short times (high frequencies). In steady state with $G_0(\tau)\to\infty$, $R_{HS}/R_{OS}\ll 1$, (i.e., $T_H = T_S(0)$, $R_{HE}\ll R_{OE}$), this reduces to

$$T_E(\tau)-T_S(0) = \underbrace{\left[\left(T_0(\tau)-T_0(0)\right)\right.}_{6a} + \underbrace{\left.\left(T_0(0)-T_S(0)\right)\right]}_{6b}\,\frac{R_{HE}}{R_{OE}} \qquad (6)$$

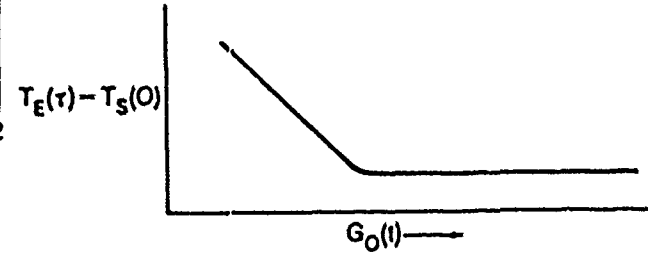


Fig. 3 Temperature error of the experiment versus thermal gain of the oven, $G_0(\tau)$.

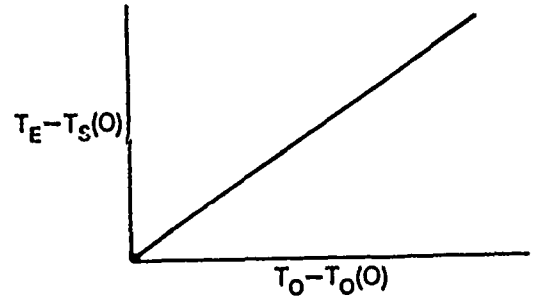


Fig. 4 Steady state temperature error of the experiment, T_E versus changes in the outside temperature, T_0, from its nominal value $T_0(0)$. The slope of this line yields a good estimate of R_{HE}/R_{OE}

The confirmation that this is the correct model of a conventional oven is demonstrated by the fact that, for most ovens, the temperature change of the experiment due to changes in the outside temperature is basically linear.

A Compensated Thermal Enclosure

The formulation of the problem, as given above, for a conventional thermal enclosure illustrates that achieving large thermal gains with an enclosure is not necessarily sufficient to keep the enclosed experiment at the correct or even at a constant temperature. If R_{OE}/R_{HE} is larger than $G_0(\tau)$, then the effects outlined above will dominate the temperature performance of the enclosure that surrounds the experiment. Clearly R_{HE} should be chosen as small as possible consistent with attenuating the high frequency thermal noise in the thermal regulation servo and reducing the thermal gradients in the shell of the thermal enclosure. These thermal gradients scale roughly as R_{OVOV}'/R_{HOV} (R_{EE}'/R_{HE}) where R_{OVOV}' is the nominal thermal impedance from one endcap of the oven to the other, R_{HOV} is the thermal impedance from the heater to the oven shell, and R_{EE}' is the thermal impedance from one end of the experiment to the other. In some cases it is possible to decrease R_{HE} and increase the thermal heat capacity of the experiment in order to preserve the same value of τ_{HE}, and still decrease the relative temperature error which is proportional to R_{HE}/R_{OE}. There is also another solution to the problem which preserves the thermal filtering of R_{HE}; that is to measure $T_0(\tau) - T_H$ and use the result to change the thermal sensor set point by an amount equal to the right hand side of equation 6 [4]. This is shown in figure 5. In practice this is easily accomplished since one need only measure the relative temperature changes of the outside shell versus a nominal setpoint $T_0(0)$ (term 6a). Term 6b only contributes an offset which, for many applications, can be ignored. The effects of a time varying thermal gradient between the position of measurement of the outside temperature and the effective position of the thermal connection between the experiment and the outside are included in Eq. 7 via $\delta_T(\tau)$. The temperature error of the experiment using the scheme shown in Figure 5 is given by equation 7.

$$T_E(\tau)-T_E(0) = [(T_0(\tau)-T_0(0)]\frac{R_{HT}}{R_{OT}} \cdot$$

$$[(T_0(\tau)-T_0(0) + \qquad (7)$$

$$\delta_T(\tau)]G_{FB} + [(T_0(0)-T_E(0)]\frac{R_{HT}}{R_{OT}}$$

The temperature coefficient of the experiment due to changes in the temperature of the outside is approximately given by

$$T_C = \frac{dT_E}{dT_0} = \frac{R_{HT}}{R_{OS}} \cdot \left(1 + \frac{d\delta_T(\tau)}{dT_0}\right)G_{FB} \qquad (8)$$

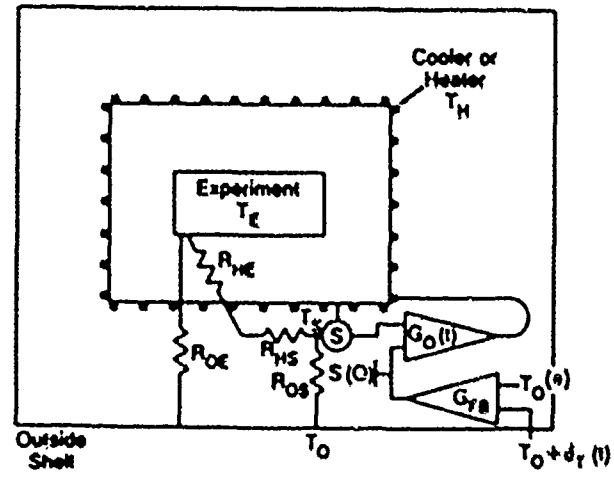


Fig. 5 Schematic representation of an experiment contained within an oven shell using feedback to compensate $T_S(0)$, the set point of the oven. The various temperatures are represented by T's and the thermal impedances by R's. S(0) is the set point of the oven, $G_0(\tau)$ is the oven gain, $T_0(0)$ is the nominal temperature of the outside shell, S2 is the second sensor, G_{FB} is the gain of the second feedback circuit which modifies the temperature of the primary enclosure set point S(0), and $\delta_T(\tau)$ is the time varying thermal gradient along the outside shell between the second sensor and the effective point of thermal coupling from the outside to the experiment.

One notes that T_C can be adjusted from plus to minus or zero by changing the feedback gain, G_{FB}, to within a precision limited only by the time variations of $\delta_T(\tau)$. Additionally, variations of $\delta_T(\tau)$ with T_0 are unimportant. Improvements in T_C are always possible if the open loop performance (i.e., $G_{FB} = 0$) conforms with Figure 4. Adjusting T_C to approximately zero is equivalent to increasing the thermal gain at the experiment to infinity. In practice it has proven possible to achieve thermal gains in excess of 10^5 using a single oven shell fabricated from 1.6 mm thick aluminum cylinder about 25 cm long and 18 cm in diameter. The ends were about 0.63 cm thick and carried the heaters. The experiment under control was an alumina cylinder of ~ 4 kg weight, R_{HE} was of order 1 hour. Particular care was taken to reduce thermal gradients across the experiment and to heat sink the cables connecting the experiment to the oven to increase R_{OE}.

The question which naturally arises when compensating the oven temperature in this manner is: what happens to the thermal transient response of the experiment? The reason that τ_{HE} is finite (hours in some cases) is that a long time constant is perceived to be necessary in order to reduce the thermal transients due to noise in the oven servo and thermal gradients in the experiment. The transient response can easily be understood by modeling it as an equivalent electrical circuit in (figure 6) where $T_E \rightarrow V_E$, $T_H \rightarrow V_H$ etc.

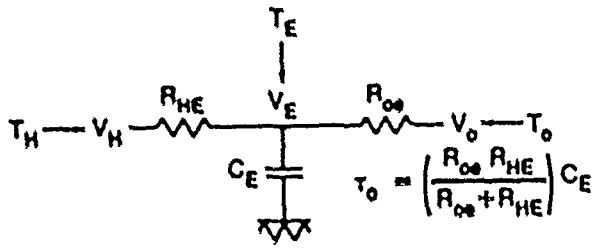


Fig. 6 Electrical analog for the thermal circuit of the heater/cooler, thermal sensor, experiment showing the various couplings.

If V_H or V_O changes, the value of V_E moves towards the new equilibrium value of $V_H + (V_O - V_H)$ $R_{HE}/(R_{HE} + R_{OE})$ with a time constant of $\tau_{HE} = (C_E R_{HE} R_{OE})/(R_{HE}+R_{OE})$. If however V_H were to mirror V_O as $V_S = (V_O - V_S)R_{HE}/(R_{HE}+R_{OE})$, V_E would stay perfectly constant. This can be approximated in the real oven by having $T_H(\tau) - T_S(\tau)$ track the compensated value of $T_S(0) + [T_O(\tau) - T_O(0)]G_{FB}$ with time constant τ_R. The magnitude of the thermal transient at the experiment due to a change in T_O will be of order

$$\Delta T_E = \left(\frac{1}{e}\right)\left(\frac{R_{HE}}{R_{OE}}\right)\left(\frac{\tau_R}{\tau_{HE}}\right)\left(T_O(\tau) - T_O(0)\right) . \quad (9)$$

Clearly this is reduced by making τ_R as small as possible. The lower limit of τ_R is $\sim 4(\tau_{HE} + \tau_{S2})$, where τ_{S2} is the thermal response of the second sensor measuring the relative temperature of the outside.

Therefore, if instead of using the oven compensation outlined here, one uses a perfect thermometer on the experiment to measure the temperature error due to T_O changing, the smallest value that could be obtained for τ_R is about $4\tau_{HE}$. Since τ_{S2} can be orders of magnitude less than the thermal lag of the experiment, τ_{HE}, this compensated oven approach will produce much better thermal transient reduction at the experiment than having a perfect thermometer on the experiment. This is illustrated in Figure 7, and qualitatively verified on the single oven system mentioned above. The sensor on the experiment can also be used in conjunction with the method outlined here to obtain superior long term performance.

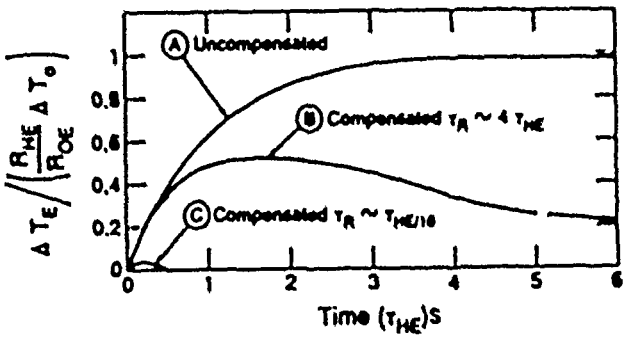


Fig. 7 Curve A shows the calculated thermal transient response of the experiment in an uncompensated thermal enclosure for a step change on the temperature of the outside shell normalized by R_{HE}/R_{OE}. Curve B shows the calculated response using a perfect thermometer on the experiment with the fastest stable feedback response in the compensated network, namely, $\tau_R = 4\tau_{HE}$. Curve C shows the calculated response for the compensated servo described in the text assuming that the second thermal sensor on the outer shell has a time constant of $\tau_{HS}/10$. In general the peak height for $\tau_R \ll \tau_{HE}$ is given in eq. 9.

Discussion

Based on the above analysis one can formulate guidelines for achieving very high thermal gains in single enclosures. These guidelines are:

1) Make τ_{HS}, the time constant between the heater/cooler and the thermal sensor as small as possible and use at least one integrator in order to make $G_O(\tau)$, the electronic gain, large.

2) Make the thermal resistance between the heater (or cooler) and experiment, R_{HE}, as small as possible and still achieve the desired reduction of high frequency noise and thermal gradients in the experiment.

3) Make R_{OE}, the thermal resistance from the outside to the experiment, as large as possible,

4) Make the thermal gradient, $\delta_T(\tau)$, between the measurement point of T_O for the compensation network and the effective point of thermal connection for R_{OE}, as small and as constant in time as possible.

5) Adjust the servo gain G_{FB} (of eq. 8) so as to minimize the thermal coefficient of the "experiment".

Following these guidelines generally, it is possible to achieve thermal regulation (gain) at the experiment which is 1 to 2 orders of magnitude better than that achieved in the same enclosure using the conventional approach. The thermal transient response can be similarly improved. In several experiments using a large single oven approximately 25 cm long and 18 cm in diameter, thermal gains in

excess of 10^5 were obtained. Multiple enclosures can be treated as successive single enclosures.

Acknowledgements

I am indebted to many colleagues for fruitful discussions on this concept, especially Dr. L.L. Lewis who first suggested adding a sensor on the external shell in addition to the one on the experiment in order to compensate for external temperature changes of the primary thermal enclosure.

References

[1] F. M. Gardner, Phase Lock Technique, (John Wiley, New York, New York, 1966).

[2] Private communication with John Hall, Joint Institute for Laboratory Astrophysics, University of Colorado, Boulder, CO 80302.

[3] U.L. Rohde, Digital PLL Frequency Synthesizers, Prentice-Hall, Inc., Englewood Cliffs, NJ, 07632.

[4] It has been brought to our attention that a similar scheme (patent # 3970818) was developed by Jerry Friedricks of Motorola, 2553 N. Edgington, Franklin Park, IL 60131.

41st Annual Frequency Control Symposium - 1987

THE AGING OF RESONATORS AND OSCILLATORS UNDER VARIOUS TEST CONDITIONS

R. FILLER, R. LINDENMUTH, J. MESSINA, V. ROSATI and J. VIG,

US Army Electronics Technology and Devices Laboratory (LABCOM)
Fort Monmouth, New Jersey 07703-5000

Abstract

During measurement of the aging of resonators and oscillators, it has been customary to minimize the effects of all frequency-determining parameters other than time. It is important to perform the experiments at constant temperature, drive level, etc. in order to be sure that the frequency changes measured are due to "aging," and not to some other changing parameter. In many real-world applications, however, the oscillators' aging occurs simultaneously with exposure to changing environments.

Aging experiments on oscillators indicate that power on-off cycling and temperature cycling of the oscillators during aging does not worsen the total frequency change during the aging period, in most cases. Changing the resonator drive level in discrete steps did not affect the aging rates of precision SC-cut oscillators, up to 2.5 ma of drive current. The aging of precision SC-cut resonators did not improve significantly when the aging was performed at temperatures about 100°C lower than the temperatures at which ovenized oscillators are usually operated.

Introduction

Aging is defined as[1] "The relationship between oscillator frequency and time when the oscillator frequency is measured under constant environmental conditions." Historically, during the measurement of aging, it has been necessary to minimize the effects of all frequency-determining parameters other than time. Only by doing so can one be sure that the frequency change measured is due to "aging", and not to some other changing parameter.

In many real-world applications, aging occurs simultaneously with other changing conditions. The major goals of the investigations described below were: 1. to compare the aging under constant conditions with the aging under changing conditions, and 2. to determine if a significant improvement in aging rate can be obtained by lowering the aging temperature to below zero Celsius. The "changing conditions" were on-off cycling, and drive level variations.

Since in some of the experiments, the "aging" conditions were intentionally varied, the term "aging" is used loosely in this paper, i.e., "aging" includes frequency variations with both the time and the condition being varied.

The experimental methods[2], the ceramic flatpack enclosed resonators[3], the special test oscillators[4], and the Tactical Miniature Crystal Oscillators[5] used in these investigations have been described previously.

Aging With On-Off Cycling

In many applications, oscillators are used intermittently. When the equipment is turned off, so is the power to the oscillator. There is virtually no information in the literature on how frequent on-off cycling affects the aging of oscillators.

TMXO

A group of 40 Tactical Miniature Crystal Oscillators (TMXO) were placed on aging. These oscillators contained ceramic flatpack enclosed 10 MHz, 3rd overtone, SC-cut resonators. The resonator drive currents were 0.3 ma to 0.4 ma. The oven temperatures within the TMXO were set to the crystal units' upper turnover temperatures, which ranged from 98°C to 115°C.

The oscillators were placed into a temperature chamber and cooled to -45°C. After an initial continuously-on stabilization period of 11 days, the oscillators were on-off cycled at 24 hr intervals, i.e., on 24 hrs then off 24 hrs, for the next 37 days. After this initial on-off cycling, the oscillators were kept continuously-on for 12 days, then the on-off cycling was continued for another 7 days. The chamber was then heated up to +70°C, the oscillators were kept continuously-on for 6 days, followed by on-off cycling, at 24 hr intervals, for the final 27 days. The frequencies were measured 5 minutes after turn-on, then at 80 minute intervals.

Representative results are shown in Figures 1 to 4. The upper turnover temperatures, i.e., the oven set-points, are indicated on the figures. The results show that the initial aging after turn-on was sometimes worse when the chamber temperature was -45°C (e.g., in Figure 1,) sometimes it was worse when the chamber temperature was +70°C (e.g., Figures 2 and 3,) and sometimes the results were comparable at the -45°C and +70°C temperatures (e.g., Figure 4.) Of the 40 units, 20 were better at -45°C, 4 were better at +70°C, 13 were about the same at the two temperatures, and 3 units failed during the experiment.

If one fits curves to the points at the end of each 24 hr period and compares the aging rate, as represented by those curves, to the aging rates during the continuously-on periods, no degradation in aging rates can be attributed to the on-off cycling.

The aging rates reversed sign for 15 of the 40 TMXO's. This is a significantly higher incidence of reversals than has been observed during the normal, continuously-on aging of other ceramic flatpack enclosed crystal units.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

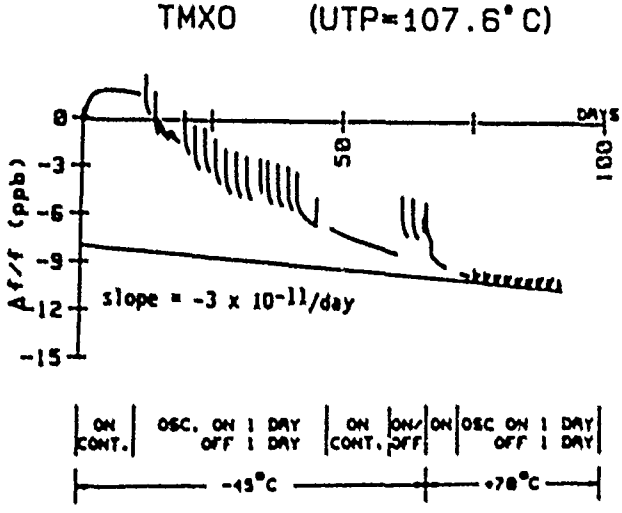


Figure 1 - Aging with on-off cycling. Slope is denoted for reference.

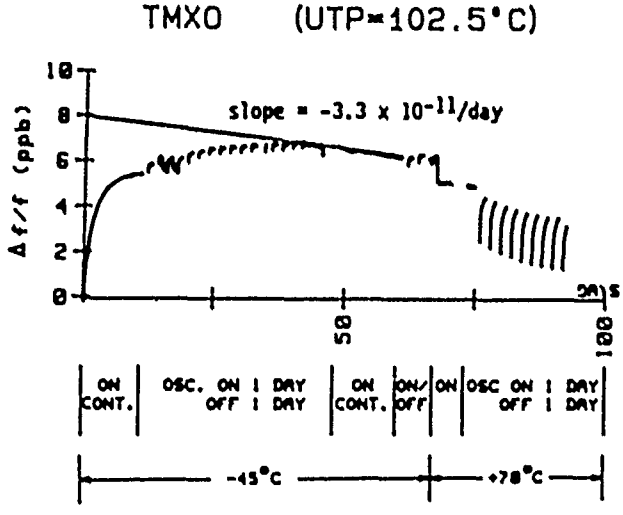


Figure 2 - Aging with on-off cycling. Slope is given for reference.

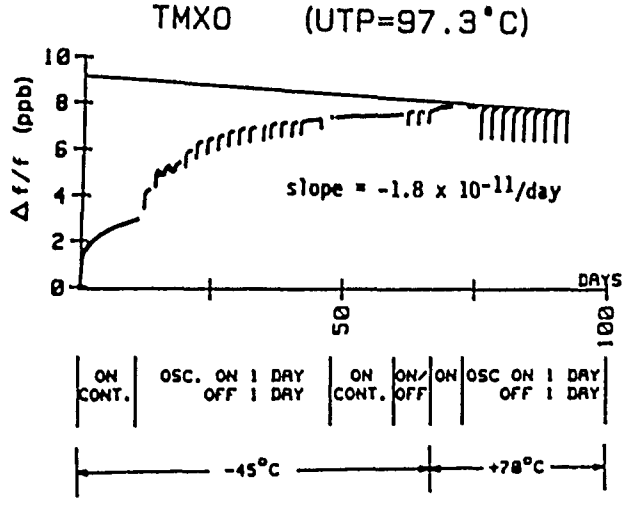


Figure 3 - Aging with on-off cycling. Slope is denoted for reference.

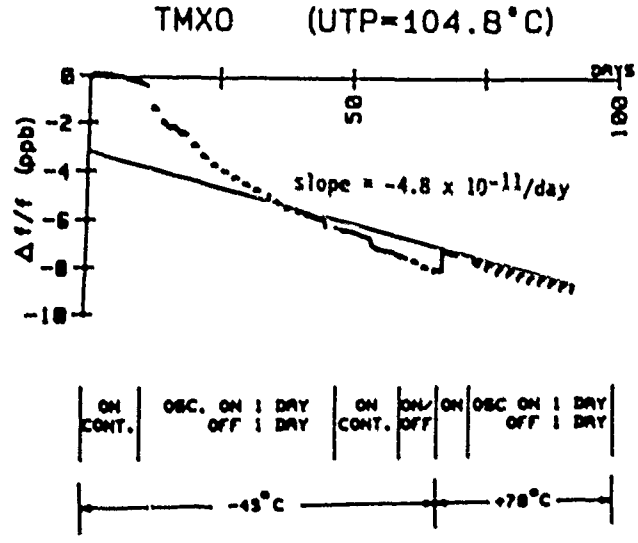


Figure 4 - Aging with on-off cycling. Slope is denoted for reference.

In another experiment, nine TMXO's, in a laboratory ambient environment, were aged continuously-on for 40 days, then were turned off and stored at -40°C for ten days, then were turned on and kept on continuously for 105 days, then were on-off cycled, five days on and two days off, for 4 complete cycles, then were kept on continuously for the next 267 days. Representative results are shown in Figures 5 and 6. Again, the temperature cycling did not worsen the aging rates. The slopes given on the curves are the best fit straight lines to the last third of the data.

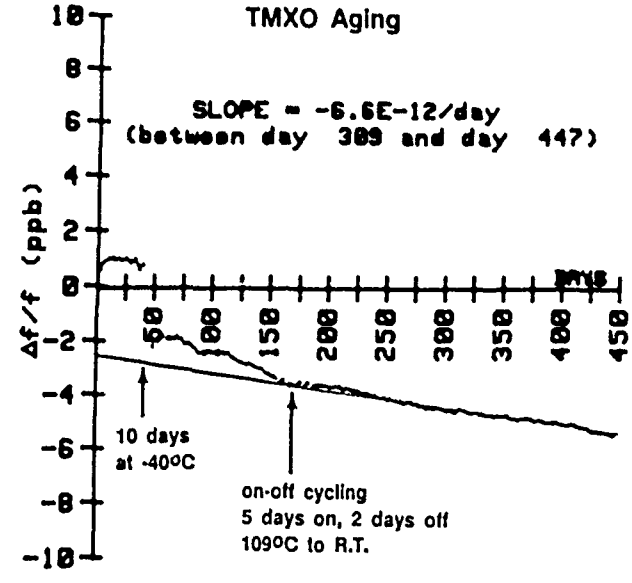


Figure 5 - Aging with on-off cycling. Slope is linear fit between days noted.

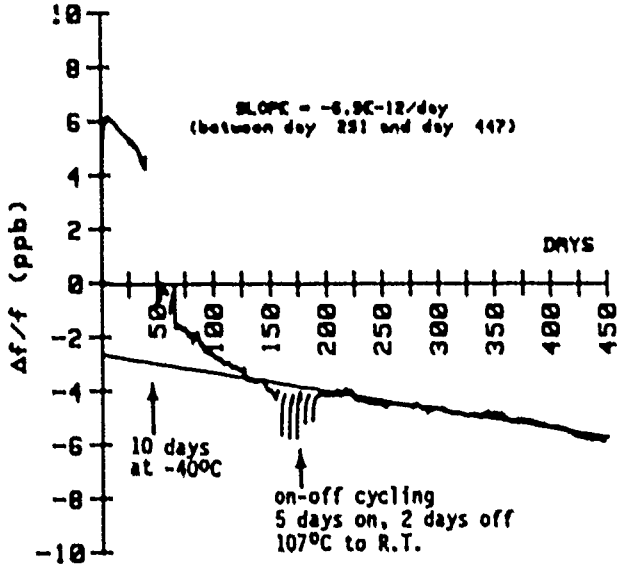


Figure 6 - Aging with on-off cycling. Slope is linear fit between days noted.

Resonators in Test Oscillators

On-off cycling aging experiments were also performed with the special test oscillators on five SC-cut 10 MHz 3rd overtone ceramic flatpack enclosed resonators. In order to minimize the effects of oscillator circuitry on the aging, the resonators were operated near series resonance. The network was a Colpitts design with a b-mode rejection circuit, as shown in Figure 7. The drive currents ranged from 0.4 to 0.6 ma. The oven (i.e., turnover) temperatures ranged from 80°C to 91°C. The oscillators remained at laboratory ambient temperature during the experiment. During the last phase of the experiment, the oscillator circuitry alone was cycled, without cycling the ovens. Representative results are shown in Figures 8 to 11. The on and off periods are noted on the curves.

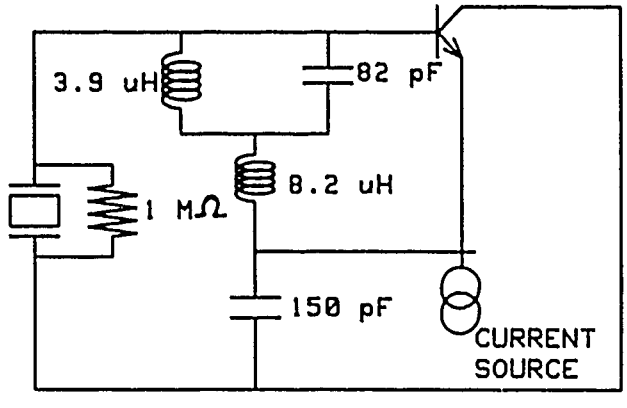


Figure 7 - Crystal network in test oscillators.

Again, the on-off cycling did not degrade the aging rates. The aging rates noted on the curves are comparable to the rates observed for similar resonators that have been aged under continuously-on conditions for comparable periods. In this group of resonators, none exhibited an aging rate reversal.

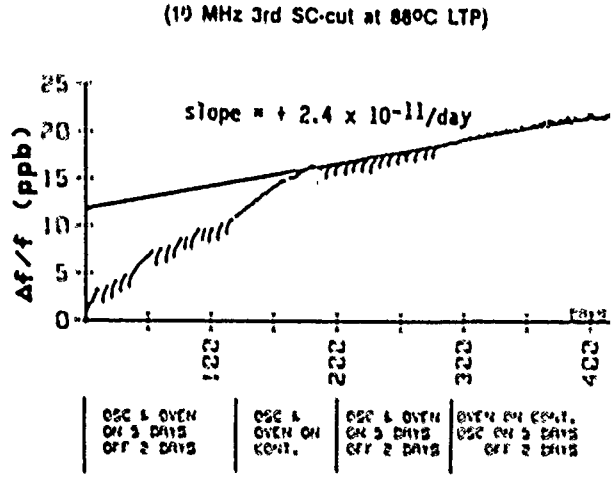


Figure 8 - Aging with on-off cycling. Slo denoted is for reference.

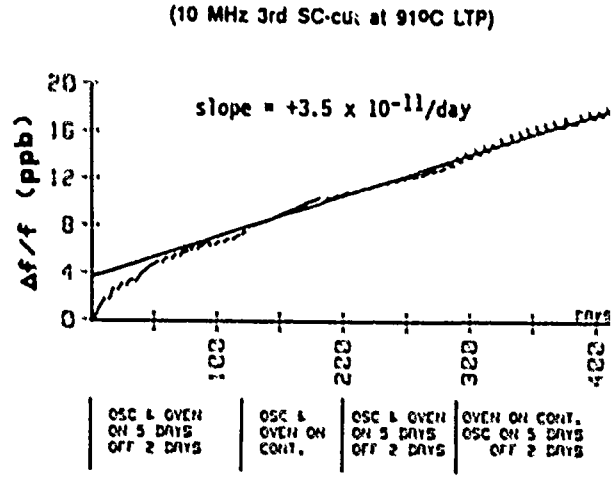


Figure 9 - Aging with on-off cycling. Slo denoted is for reference.

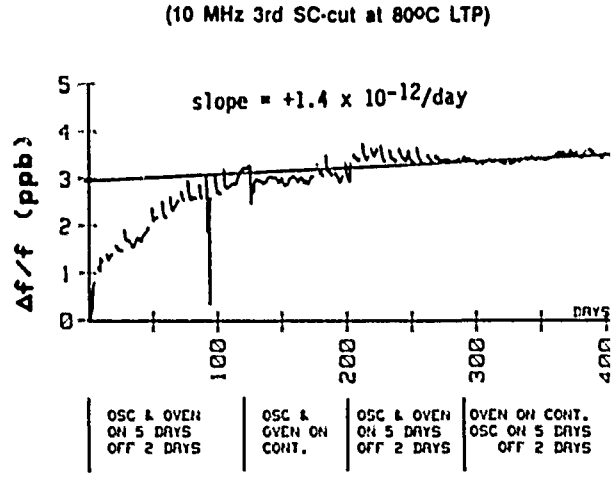


Figure 10 - Aging with on-off cycling. Slo denoted is for reference.

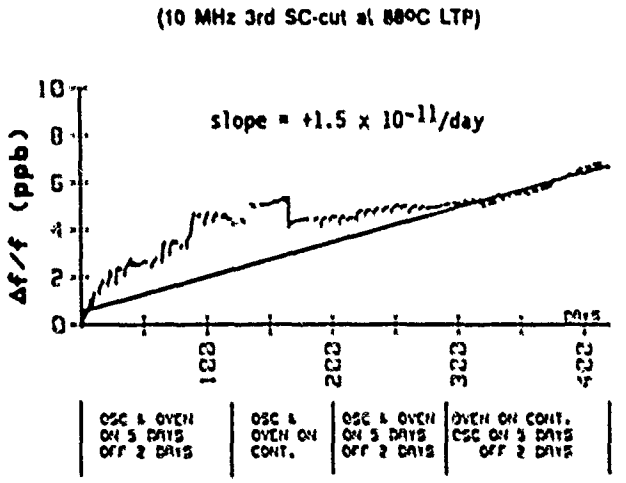


Figure 11 - Aging with on-off cycling. Slope denoted is for reference.

TCXO

In another project, on the evaluation of commercial TCXO's, the long term stability of 3.2 MHz TCXO's (27 total, from four vendors) was evaluated during a period of 400 days. The sequence of events during the evaluation was as follows:

Days	Activity
0 to 30	Aging
31 to 44	2 f vs T runs
45 to 56	Aging
57 to 103	5 f vs T runs
104 to 138	Aging
139 to 150	1 f vs T run
151 to 185	Aging
186 to 200	1 f vs T run
201 to 247	Aging
248 to 267	2 f vs T runs
268 to 315	Aging
316 to 365	4 f vs T runs
366 to 400	Aging

During the aging periods, the TCXO's were in a temperature chamber that was stabilized at 58°C ±1°C. Each frequency vs. temperature (f vs. T) run was conducted according to the following procedure:

1. Cool chamber from room temperature to -55°C in about 20 minutes, without collecting data.

2. Dwell at -55°C for 60 min; measure frequencies at the end of the 60 min.

3. Collect data while heating the chamber to +85°C, then cooling it to -55°C, in 2°C steps. After each 2°C step, stabilize for four minutes and collect data for an additional four minutes.

4. Heat chamber to room temperature in about 20 minutes, without collecting data.

Eleven of the TCXO's failed during the evaluation (seven out of the 11 were units made by one of the four vendors.) Six of these stopped operating during the f-T runs, three stopped operating during one of the aging periods, and two were deemed to have failed because the aging rates increased significantly after one of the f-T runs. The

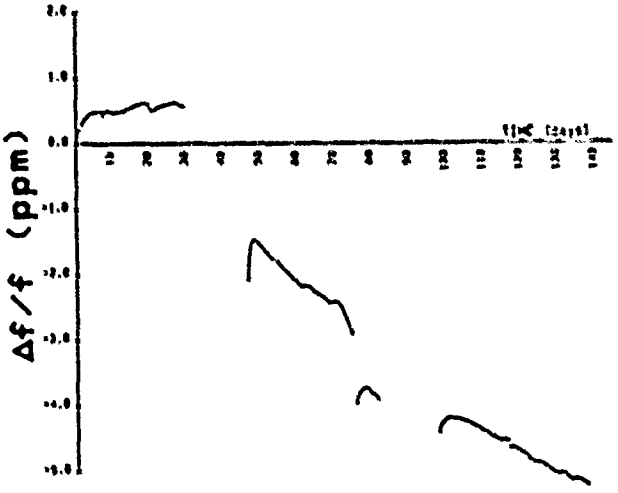


Figure 12 - TCXO aging with temperature cycling.

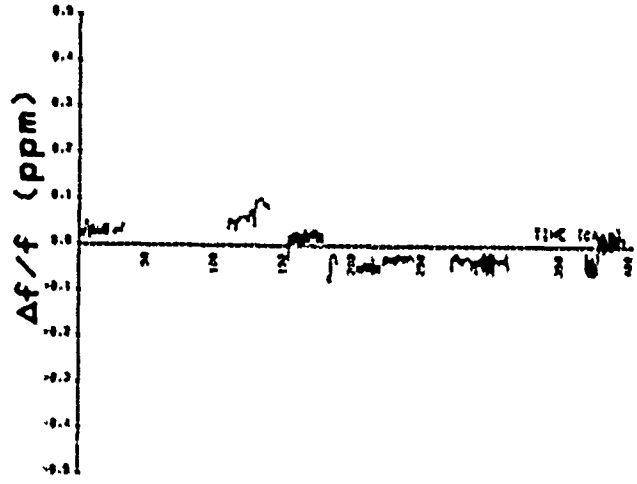


Figure 13 - TCXO aging with temperature cycling.

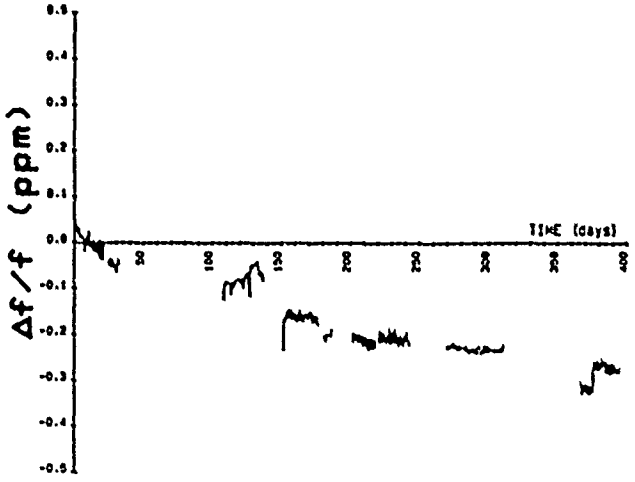


Figure 14 - TCXO aging with temperature cycling.

results for one of the latter two are shown in Figure 12. Representative results for the 16 TCXO's which stayed within a ±0.5 ppm range during the 400 days are shown in Figures 13 and 14. Note that the y-axis scale is in parts per million (ppm) whereas the earlier scales were in parts per billion (ppb).

Aging at Low Temperature (at about -10°C)

In a previous report[2], results for aging rates at about -10°C for thirty-day periods were described, and the rationale for possible aging rate improvements at -10°C was also discussed. The intent of the earlier study was to investigate the temperature dependence of aging rates.

The earlier investigations have now been extended to long term aging. A group of nine 5 MHz 3rd overtone SC-cut crystals, with lower turnover temperatures in the vicinity of -10°C, were placed on long term aging in the special test oscillators. The test oscillators allowed for operation in a freezer, i.e., were built so that the controls for setting the oven temperature and drive level were outside the freezer.

After the initial 200 days of aging, the experiment was interrupted because some of the day to day frequency changes were erratic, as can be seen in Figure 15. It was discovered that the usual high-driving (about 20 ma for about 30 seconds) prior to installation into an oscillator had not been applied to this group of resonators. After high-driving, the resonators were reinstalled into the test oscillators and the aging was restarted. Representative results for the subsequent one year of aging are shown in Figures 16 to 18. The aging rates, as indicated by the best fit to the data during the latter half of the aging period, ranged from low parts in 10^{11} per day, as in Figure 16, to parts in 10^{13} per day, as in Figure 17. While these rates are respectable, comparable rates have also been achieved at the upper turnover temperatures of SC-cut crystals, as can be seen, for example, in Figure 5.

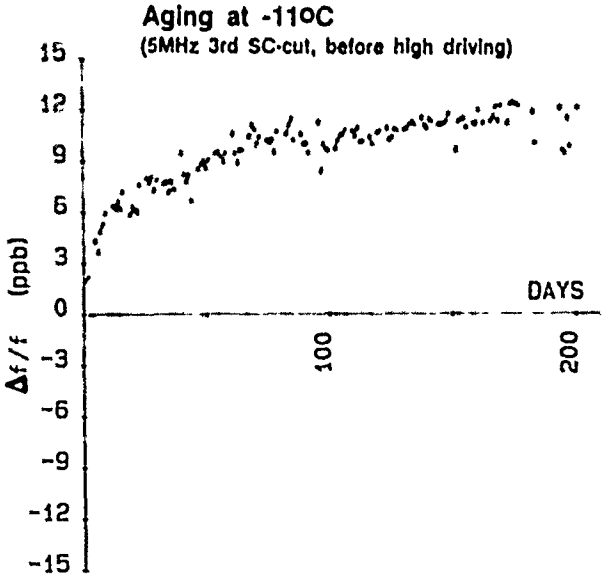


Figure 15 - Erratic aging.

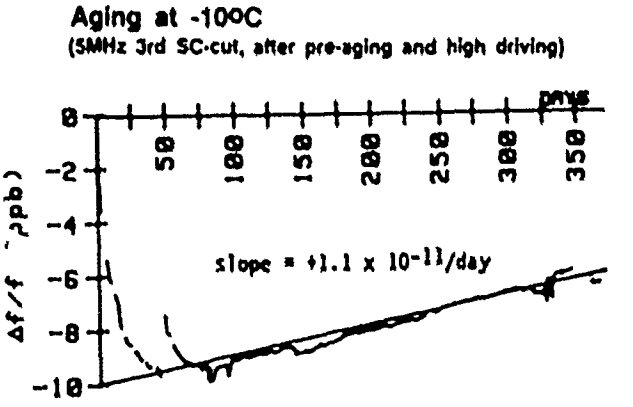


Figure 16 - Low temperature aging of 5 MHz SC-cut 3rd O/T.

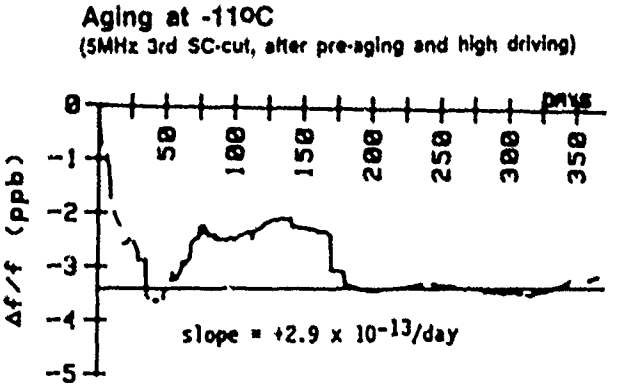


Figure 17 - Low temperature aging of 5 MHz SC-cut 3rd O/T.

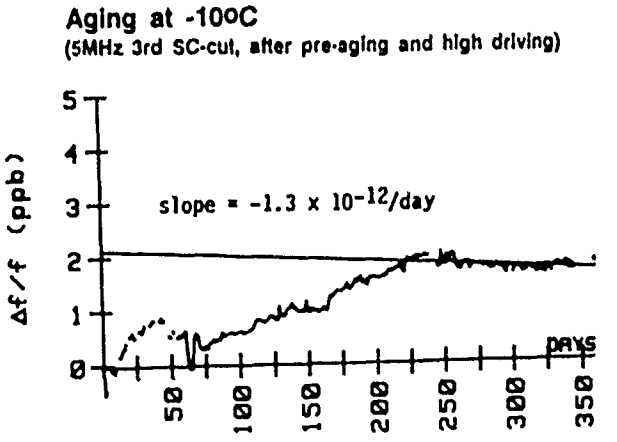


Figure 18 - Low temperature aging of 5 MHz SC-cut 3rd O/T.

Drive Level Dependence of Aging

In precision AT-cut oscillators, it has been standard practice to use a low drive level in order to attain the lowest possible aging rate[6,7]. For example, the recommended drive level[8] for a 5 MHz 5th overtone AT-cut unit that has been used in several manufacturers' high precision ovenized oscillators is 70 μa ±20%.

The low drive level is necessary because at higher drive levels, the AT-cut crystal's nonlinearity results in instabilities. A 1967 review paper[9] on AT-cut quartz frequency standards, states: "Obviously, it is advisable to operate high-precision crystal units at the lowest possible drive level. A change in crystal current (or vibrational amplitude) in 5- and 2.5-MHz fifth-overtone crystal units also changes the aging behavior. An increase in current from 75 μa by one order of magnitude changes the monthly aging from 1 part in 10^{10} to 1.5 parts in 10^{9}."

SC-cut crystals' drive level sensitivity can be significantly lower than the sensitivity of AT-cut units[10]. Moreover, whereas AT-cut units are very susceptible to activity dips at high drive levels (i.e., above about 1.5 ma of drive current) in SC-cut units, activity dips are an extremely rare occurrence, even at the higher drive levels.

The aging of a group of ceramic flatpack enclosed 10 MHz 3rd overtone SC-cut units was measured in the special test oscillators. The group consisted of four resonators that were aged at about 100°C and nine resonators that were aged at about -10°C. The aging of the 13 resonators was measured at three drive levels. Representative results are shown in Figures 19 to 22. The initial period was a stabilization period during which the drive levels, noted on the figures, were in the range of a few tenths of a milliampere. On day 53, the drive levels were decreased to the values indicated. On day 87, the drive levels were increased, and the aging was continued to day 137. The drive levels during the final period ranged from 1.4 ma to 2.5 ma.

The results of this experiment indicate that the aging of these resonators have no significant dependence on drive level, up to the 2.5 ma maximum drive current used during the experiment.

In another aging vs. drive level experiment, a group of five ceramic flatpack enclosed 10 MHz 3rd overtone SC-cut resonators were aged for 223 days in the special test oscillators. The aging temperatures ranged from 71°C to 108°C. On day 43, the drive levels were changed from about 0.03 ma to about 1.2 ma, and on day 123, the drive levels were changed back to the original values.

Representative results are shown in Figures 23 and 24. In all cases, no significant discontinuities in the aging rates were observed. Of course, all drive level changes resulted in a frequency offset at the moment of the change. The offsets are a consequence of the amplitude-frequency effect[10] in the resonators, and possibly, of the changes in the oscillator circuitry due to the effects of the variable-gain amplifier. The drive level

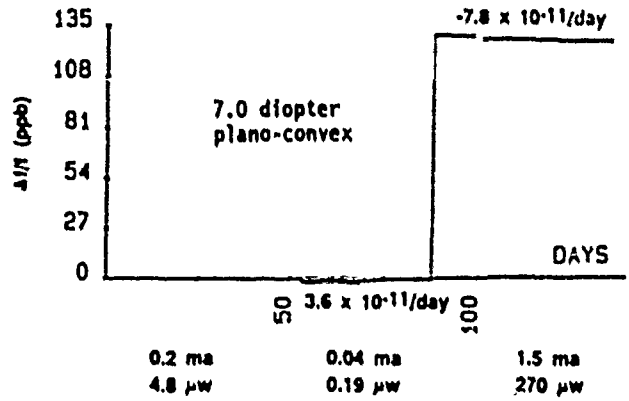


Figure 19 - Aging vs. drive level at 100°C.

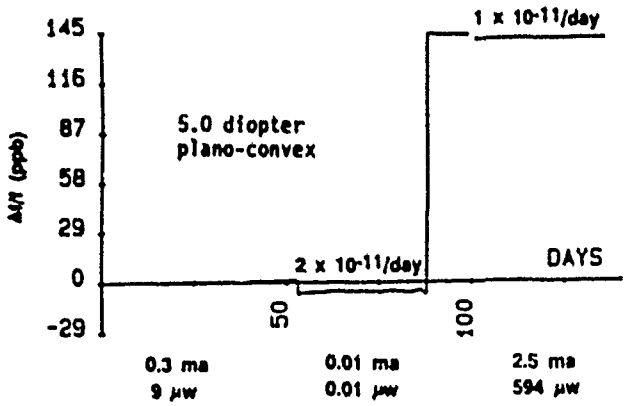


Figure 20 - Aging vs. drive level at 100°C.

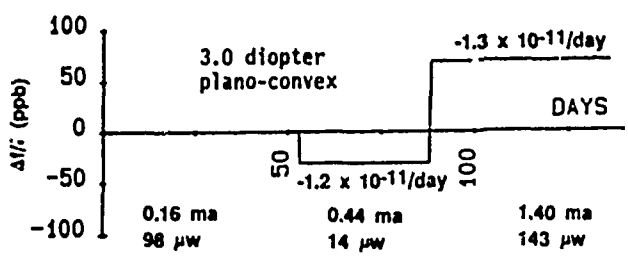


Figure 21 - Aging vs. drive level at -8°C.

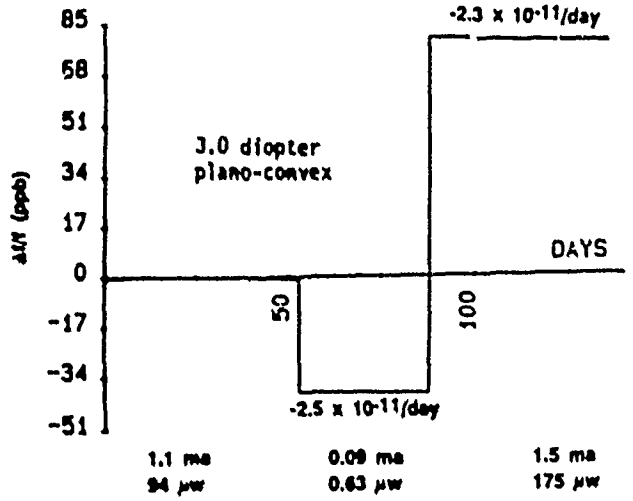


Figure 22 - Aging vs. drive level at -9°C.

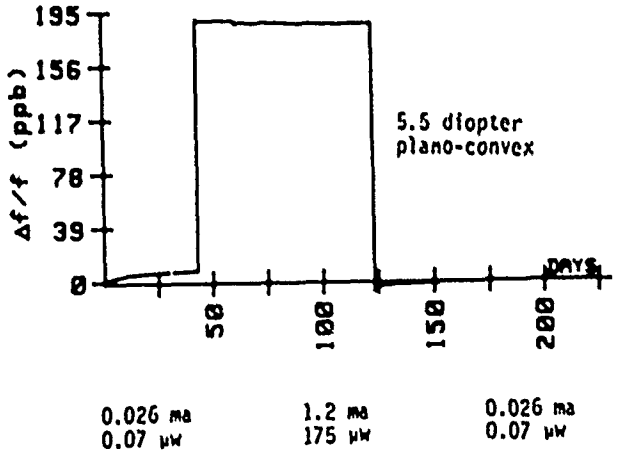


Figure 23 - Effects of drive level changes on aging at 91°C.

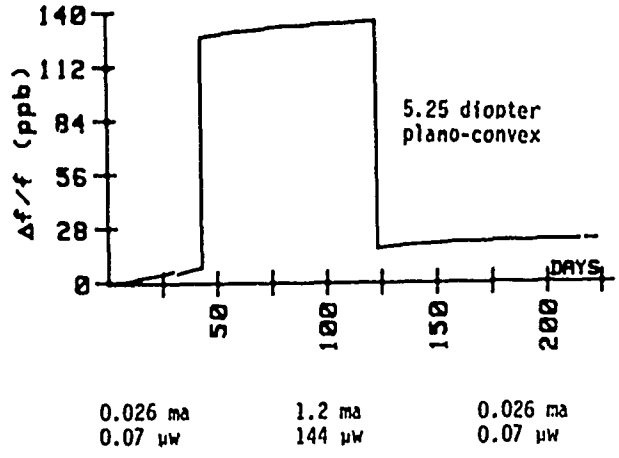


Figure 24 - Effects of drive level changes on aging at 92°C.

changes resulted in no frequency offset upon return to the initial drive levels. The aging rate during the final periods were about what one would obtain by extrapolating the aging during the initial periods. That is, the change in drive levels during the middle periods did not have a significant effect on the aging rates during the final periods.

The crystal plates of the 10 MHz 3rd overtone SC-cut crystal units used in these drive level vs. aging experiments were plano-convex. The plates were 14 mm in diameter and the contours ranged from 1.0 diopters to 7.0 diopters. The drive level sensitivity coefficients[10], of resonators similar to these can be expected to range from 57 ppb per ma^2 to 228 ppb per ma^2. Even the 228 per ma^2 resonators showed no dependence of aging rate on drive level. All of the crystal plates were cut from cultured quartz, both swept and unswept. No correlation was found between quartz material and the drive level dependence of aging.

Conclusions

Aging with on-off cycling

On-off cycling during aging did not worsen the total frequency change during the aging period, except in a few TCXO's. The initial aging after turn-on varied with the temperature during the off-period, but not in a systematic manner. That the initial aging was worse, for some of the oscillators, when the off-period was at +70°C than when it was at -45°C was unexpected.

One possible explanation is that the initial aging is due primarily to stress relief. Although the -45°C temperature produces a larger stress than the +70°C, since stress relief rates are temperature dependent, stress relief at -45°C can be lower when the effect of lower temperature on stress relief rate outweighs the effect of higher stress.

Since aging and retrace are highly dependent on the resonator's processing and design, on-off cycling may or may not affect the aging rates of other types of resonators. Therefore, the above results should not be indiscriminately extrapolated to other types of resonators.

Aging at Low Temperature

That the aging was not significantly lower at the -10°C aging temperature than when the aging took place at temperatures 100°C higher indicates that the aging was not primarily due to a single thermally activated process. Since the resonators had been processed at high temperatures in order to minimize contamination inside the resonator enclosure, the resonators may have been exposed to large mounting stresses at -10°C.

The resonators used in these experiments were in four-point mounts. The mounting directions were the psi = 0 , 90 , 180 , and 270 degree directions, which are not the zero force-frequency-effect directions for SC-cut crystals[11]. Stress relief may, therefore, have been a significant aging mechanism in these resonators. The experiment needs to be

repeated with resonators designed to minimize the effects of thermal stresses, i.e., either with resonators that are mounted along the zero force-frequency-effect directions, or with resonators of the BVA-type[12,13] designs.

Drive Level Dependence

That the aging rates did not vary with drive level is probably due to the SC-cut's low amplitude-frequency effect and to stable oscillator circuitry. That drive levels up to 2.5 ma did not degrade the aging rates has potentially significant implications for the design of low noise-floor oscillators. Experiments are planned on drive level vs. aging and vs. short term stability, at higher drive levels.

Acknowledgements

The authors wish to acknowledge the contributions of John A. Kosinski, who participated in the early parts of some of the experiments.

References

1. Military Specification: MIL-O-55310B, Oscillators, Crystal, General Specification for, para. 6.3.41, 1987; available from Naval Publications and Forms Center, Code 3015, Tabor Ave., Philadelphia, PA 19120.

2. R.L. Filler, J.A. Kosinski, V.J. Rosati and J.R. Vig, "Aging Studies on Crystal Resonators and Oscillators," Proc. 38th Ann. Symp. on Freq. Control, pp. 225-231, 1984.

3. R.L. Filler, L.J. Keres, T.M. Snowden, and J.R. Vig, "Ceramic Flatpack Enclosed AT- and SC-cut Resonators," 1980 Ultrasonics Symposium, pp. 819-824, 1980.

4. H.W. Jackson, "Update of the Tactical Miniature Crystal Oscillator Program," Proc. 36th Ann. Symp. on Freq. Control, pp. 492-498, 1982.

5. D. Brown, E. Laszlo, R. McGill, and P. Stoermer, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillator," Proc. 38th Ann. Symp. on Freq. Control, pp. 380-386, 1984.

6. J.P. Buchanan, Handbook of Piezoelectric Crystals for Radio Equipment Designers, Wright Air Development Technical Report 56-156, October 1956, p. 70, AD 110448.

7. A.W. Warner, "Design and Performance of Ultraprecise 2.5-mc Quartz Crystal Units," Bell System Technical Journal, Vol. 39, pp. 1193-1217, Sept. 1960.

8. Type BG61AH-5S, Bliley Electric Co., P.O. Box 16508, Erie, PA 16508.

9. E.A. Gerber and R.A. Sykes, "Quartz Frequency Standards," Proc. IEEE, Vol. 55, pp. 783-791, 1967.

10. R.L. Filler, "The Amplitude-Frequency Effect in SC-cut Resonators," Proc. 39th Ann. Symp. on Freq. Control, pp. 311-316, 1985.

11. T.J. Lukaszek and A. Ballato, "Resonators for Severe Environments," Proc. 33rd Ann. Symp. on Freq. Control, pp. 311-321, 1979.

12. R.J. Besson, "A New 'Electrodeless' Resonator Design," Proc. 31st Ann. Symp. on Freq. Control, pp. 147-152, 1977.

13. J.P. Aubry and A. Debaisieux, "Further Results on 5 MHz and 10 MHz Resonators with BVA and QAS Designs," Proc. 38th Ann. Symp. on Freq. Control, pp. 190-200, 1984.

41st Annual Frequency Control Symposium - 1987

THE DESIGN AND ANALYSIS OF VHF/UHF CRYSTAL OSCILLATORS

A. Benjaminson

Systematics General Corporation

Crystal City, Arlington, Virginia

Summary

This program is written for use on HP Series 200 Computers with a minimum of BASIC 2.0. The 5.25" flexible disk also contains files for storing input and output data.

The current program produces a design for either of two series resonant crystal oscillators using either Bulk Acoustic Wave Resonators (BAWR), or Surface Acoustic Wave units (SAWR). The latter may be either 1-Port or 2-Port devices of either phase. One design is for Meacham Bridge oscillators (BXO) the other for the conventional SRO's used for VHF/UHF SAWRO's with impedance transforming inductors at collector and base terminals.

These designs are based on the configuration shown in Figure 1. The BXO uses bridge resistor R_1, while the SRO does not. The design programs use some simplifications to achieve a workable set of component values, and then interact with the analysis section, to provide feedback that will modify the design, particularly with respect to the transistor input admittance, Yin.

The analysis section is used to accurately analyze the performance of the selected design that fits the configuration of Figure 1. (such as the Pierce Oscillator) The six network reactances can be either inductive, or capacitive, depending on the circuit involved. Components may be eliminated by entering zero for a closed path, or a large number, i.e. 1E+20, for an open path. It is also possible to enter common-base parameters and analyze grounded-base oscillator circuits.

When resonator data is entered for either design or analysis, the frequency is assumed to be the series-resonant frequency, and a value of Li is calculated to resonate with the entered value of Ci at this frequency.

The value of L6 is also calculated in analysis to resonate with Co, if entered. It may also be modified later, as may all input values, when changes in the circuit are to be tested, via the analysis routine. When a 2-Port SAWR is to be used, the values for CS1 and CS2 are entered instead of Co, along with the phase reversal, if any.

Both design and analysis sections require the transistor's Y-parameters which may be entered directly as common-emitter or common-base factors, or alternatively the S-parameters may be entered. The program will then convert common-emitter S-parameters to common-emitter Y-parameters.

Since these parameters are dependent on Vcc, V_B, and I_C, values for these may also be entered. The program will then compute values for the bias resistors and by-pass capacitors as shown in the schematic diagram, Figure 2.

The analysis routine generates a plot of gain and phase versus frequency and also determines the frequency offset (from the Fs value) required to produce the necessary 360 degree phase shift. The curve (shown in Figures 3 thru 10) also show how the gain and phase vary within a few bandwidths around the Fs value, and indicate whether the networks are peaking at Fs. By modifying the suspected element values, the gain and phase can be adjusted as required.

The sensitivity routine can be used to determine which elements are most sensitive in terms of frequency or gain control, and provides a means for assessing the effect of component tolerance changes.

The analysis routine is a purely linear algorithm, and not sensitive to signal level. Estimates of signal output power, transistor, and crystal dissipation are made in the design section based on the assumption of an open-loop gain of two, combined with base-emitter limiting.

An Approach to Oscillator Design and Analysis

The design of low frequency oscillators such as the Pierce family of positive reactance oscillators, and the Butler and Differential groups of series-resonant oscillators can be accomplished with a minimum of transistor data. The transconductance and input impedance are all that are required and they can easily be derived from the dc emitter current and the nominal HFE. Such designs have developed into the earlier computer programs of this CODA (Crystal Oscillator Design and Analysis) series. As written, they are useful to 50 Mhz or so. The latest CODA-BXO, which requires some Y-parameter inputs, is accurate to over 100 Mhz.

The CODA-SRO however is intended for use to at least 1 Ghz and requires the full set of Y-parameters for the transistor being considered. Since most transistor data is derived from tests on network analyzers, it is usually given in S-parameters, based on a characteristic impedance of 50 ohms. This program will accept transistor S-parameters, convert them to Y-parameters, and store them in that form for use with the selected resonator, whose data can also be stored in the mass storage medium.

An oscillator circuit is basically a narrow-band amplifier that uses the resonator as a filter element to provide gain at the resonator's operating frequency. The gain must be sufficient to supply external load power as well as the input power required by the amplifier. A second requirement is that the phase-shift through the amplifier must be 0 (or 360) degrees, so that the power returned to the input will be at the correct angle to reproduce the output level, or "regenerate" the signal.

CH2427-3/87/0000-452 $1.00 © 1987 IEEE

In practice the small-signal, open-loop gain (from input to the point on the output that will connect to the input) is usually calculated to be two or greater. This will insure reliable starting under all conditions and result in a rapid build-up of oscillation until some limiting action, either transistor saturation, external limiting diodes, or ALC, reduces the gain to unity at a sustainable signal level.

Simple low frequency oscillators, such as the Pierce, use the resonator as an inductance operating in conjunction with two capacitances to form a single tapped resonant circuit. This takes the form of the familiar Pi-section and produces close to 180 degrees of phase-shift at resonance. Since the transistor (when connected as a common-emitter amplifier) produces a phase-reversal of 180 degrees, oscillation occurs very close to the frequency at which the net inductive reactance of the resonator equals the net capacitive reactance of the two capacitors. See Figure 3.

Since the only inductance in the Pierce oscillator is provided by the crystal resonator, the circuit is capable of operating on other than the selected overtone, if sufficient gain is available.

Series-resonant oscillators however, employ a low Q parallel-tuned circuit (connected to the collector of the transistor) to achieve gain at the nominal frequency of operation, while the resonator acts as an additional high Q series L-C circuit that dictates the exact frequency of oscillation. It does this by introducing the appropriate type of reactance (+/-) needed for a precise 360 degrees. See Figures 4 and 5.

As the required frequency of oscillation approaches an appreciable fraction of the transistor's transition frequency, f_t, the phase shift through the transistor increase beyond the 180 degrees of low-frequency reversal. The angle of Y_{fe} at 1GHz for example, may be -50 (+180) degrees for an f_t of 4 GHz.

The design used in this program accounts for this additional lag, by introducing a second L-C network at the base of the transistor. This network serves a dual purpose. It provides a lead angle to reduce the total phase shift to 360 degrees, and also provides a more suitable impedance ratio between the crystal and the input base of the transistor. This can be understood by reference to Figure 1, in which all the elements of the high-frequency SRO are illustrated.

The collector of the transistor connects to both the external load (as represented by R_L), and the Pi-connected tuned circuit, C_1, L_B, C_2. The latter provides two impedance-transforming functions; to reduce the impedance of R_L (in parallel with g_{oe}) to a lower impedance suitable for driving the resonator, and conversely to transform the resonator (in series with its load) to a higher impedance facing the collector. The Pi-network can accomplish this if it operates at a fairly high Q (typically 10 to 20) to hold the phase-shift at nearly 180 degrees.

The second L-C Pi-network (formed by C_4, L_3, C_5) also serves as an impedance transformer to convert the input impedance, Yin, of the transistor to a value suitable for loading the crystal resonator, and in addition provides the phase lead needed to offset the phase lag of Y_{fe}.

The value of Yin, (as shown in Figure 1) depends on the load Y_L, and the y-parameters, but it cannot be calculated directly since Y_L is also dependent on Yin. The program solves this by using y_{ie} as a starting value for Yin. After a first value for Y_L is found, the program analyzes the circuit, recomputes Yin and Y_L, and iterates through this loop until Yin stabilizes. Two or three such "passes" are usually sufficient to achieve a stable value.

The CODA-SRO program provides for either of two designs, the Bridge Oscillator (BXO), or the Series-Resonant Oscillator (SRO). The most apparent difference between the two, is that R_1 is used in the BXO to form an opposing arm of the bridge circuit, but is not used with the SRO. Since the current through R_1 opposes that through the crystal, it provides negative feedback and has a profound effect on the design algorithm.

The BXO should be considered for lower frequencies and higher crystal resistances, where its ability to raise the loaded Q, rather than reduce it, (with respect to the unloaded Q) is most apparent. Both design programs should be exercised however, and a choice made based on application. The SRO will be considered first.

Running the Design Section

The program is loaded from the flexible diskette, after BASIC 2.0 (or higher) is in place; use LOAD "CODA-SRO".

A list of menu choices is displayed, the choice is made via the soft keys in the upper row, left, of the keyboard. It is advisable to select "Draw Schematic" first to study the circuit. The schematic diagram is also shown in Figure 2.

"HELP" menus are also provided, that explain the operation of each section of the program, when additional clarification is required.

If new values for the transistor, crystal, and circuit parameters are to be entered, these may be entered separately, or in one data group. Select for data input either "Y-PARAMETERS" or "S-PARAMETERS" followed by CRYSTAL PARAMETERS" (which also include dc supply voltages). Alternatively, "ALL PARAMETERS" may be selected, and all values including crystal and Y-parameters, etc. entered in one group.

The advantage of separate transistor and crystal parameter files lies in its flexibility to allow later "mix and match" selections of transistor and crystal combinations. The user may now select the "DESIGN" section and then either the "BXO", or "SRO" design.

SRO Design

Running the design algorithm will require more decisions from the user than just providing the parameters of the transistor and resonator. Values for V_{cc}, and V_B, plus a factor, K_L, will be requested (the latter indicates what fraction of the total power generated, should go to the external load, R_L). Suggested values range from 0.2 to 0.8 and will be influenced by the available gain of the circuit, (whose final value is also requested).

When excess gain is available, the design "trades it off" by reducing the load impedance R_L, (thus increasing K_L). If the initial gain is too

low, then the Q-factor, M_Q, is reduced. If it drops below 0.2, the load, R_L, is raised instead (reducing K_L). .If K_L falls below 0.1, a warning message is displayed, and the user will have to reconsider his input decisions.

The loop, after processing these trade-offs, prints the values involved; the gain factor, M_G, the Q-factor, M_Q, and the load resistance, R_L. The user may study the listing and consider changing the transistor data to reduce y_{fe} by reducing the operating current, thereby reducing power consumption as well as power output.

Upon completion of this routine, the "Design Section Outputs" are displayed in tabular form. This includes crystal and transistor labels, followed by a table of component values keyed to the schematic diagram, including the gain and Q-factor. See Figures 6 and 7. The equations used to effect the SKO design are listed in Appendix I.

BXO Design

The bridge oscillator is similar in appearance to the SKO except for addition of resistor R_1, which transforms the Pi-section (C_1, L_s, C_2) into a source of in-phase and out-phase voltages that, combined with R1 and the Rs of the crystal, form a Heachan Bridge. The crystal as a high Q resonator, offers a very high impedance to current flow, except at series resonance, and since the current through R1 differs by 180 degrees the feedback is negative except in the narrow region around resonance. In that region the current through Rs must exceed that through R1 by a sufficient margin to provide the required in-phase gain.

Two ratios have been established to control the relative feedback, Kv and Ke. Kv is the voltage ratio across the collector to crystal network, (C1 L8, C2) while Ke is the ratio of R1 to Rs (See Appendix II). The bridge is balanced when Ke*Kv is equal to -1 (Kv is always negative) and therefore must exceed -1 for oscillation to occur. The exact value depends, of course, upon Y21 and the impedances involved.

The algorithm sets acceptable limits on these factors, and searches for the best product to achieve maximum gain with minimum unbalance. This is reflected in the value for M_Q the Q-multiplication factor, which depends upon the gain and negative feedback ratio. See Figures 8 and 9.

The BXO design algorithm runs the same way as the SRO algorithm. The trade-offs involve adjusting the product Ke*Kv to achieve the gain specified. The load factor, K1, is not adjusted in this procedure. It is left to the designer to select as required.

Running the Analysis Section

The analysis section is used to accurately analyze the performance of the selected design, but may also be used to analyze any design that fits the configuration of Figure 1. (such as the Pierce Oscillator) The six network reactances can be either inductive, or capacitive, depending on the circuit involved. Components may be eliminated by entering zero for a closed path, or a large number, i.e. 1E+20, for an open path. It is also possible to enter common base parameters and analyze grounded-base oscillator circuits.

When resonator data is entered for either design or analysis, the frequency is assumed to be the series-resonant frequency, and a value of Li is calculated to resonate with the value of Ci at this frequency.

The value of Lo is also calculated in analysis to resonate with Co, if entered. It too, may be modified later, as may all input values, when changes in the circuit are to be tested via the analysis routine. When a 2-Port SAW is to be used, the values for CS1 and CS2 and the phase type (0^0 or 180^0 are entered instead of Co.

The analysis procedure is centered around a subroutine that expands the equations shown in Figure 1 to provide data for plotting both gain and phase against frequency.

The center frequency of the plot (as shown in Figure 3) is the series-resonant frequency of the resonator. Its unloaded Q is used to determine the frequency scale, which is eight bandwidths (F_s/Qu) wide. The subroutine is used to first plot open-loop gain, in dB, against frequency, then the overall phase-shift, also against the same scale.

The same algorithm is then used to determine the exact frequency at which the phase shift becomes zero and the difference between this and fop. The phase at the point is then used to calculate the loaded Q. As shown in Figure 3, all this data is tabulated, some on the plot, and some in a following table.

If no phase reversal is located, the algorithm asks permission to widen the search. If granted it will then search two bandwidths around F_s, and repeat the procedure upon request, expanding the search range each time until a phase reversal is located.

After the phase zero is localized, data on the magnitudes (transfer coefficients) of each of the three sections of the circuits, is also tabulated. Ideally, the phase shift through the center section (A_b) will be 180 degrees, with the collector and base sections (A_c and A_c) adding up to -180 degrees at f_{op}.

The user is also given the option of adjusted the balance of the bridge oscillator design. (BXO). A procedure is provided that automatically adjusts R1, the negative feedback arm of the bridge. This directly affects the magnitude A_b, and indirectly both A_c and A_c. If R_1 is used to bring the gain to the design value (usually 2) and requires a large change, it is advisable to rerun the analysis to see if any secondary effects occur.

Running a Sensitivity Analysis

The sensitivity analysis routine determines the effect of a change in value of any component on gain and frequency of oscillation. It is expressed as a change per one percent of the component value involved and applies only to the ac components, excluding the bias resistors and bypass capacitors. The sensitivity to component value change is important in determining the tolerance placed upon the component specifications, and can also be used to determine value changes required in the design. The analysis routine is reran each time a change is made

Appendix I
Equations Used In SKO Design

Let: Ro = Source Resistance driving Crystal

Ri = Lead Resistance facing Crystal

Rl = Load Resistance, representing external load

Kv = Crystal/Collector Voltage Ratio

Rs = Crystal Resistance at Series Resonant Frequency

MQ = Ratio of loaded/unloaded Q of Crystal

Mg = Open-loop Voltage Gain

Kl = Out Power/Total Power

Then: Ro = Rs (1-MQ)/2MQ

Ri = Ro

Rl = ((Rs+Ri)(1/Kl-1)-Ro)/(Ro G22)

Rll = Rl/(Rl G22+1)

Kv = SQR(Ro/Rll)

Rls = $(Rs+Ri)/Kv^2$

Kl = Rls/(Rls+Rl)

Mg = y21 Rlls Kv^2SQR(Rll/Gin)/(Rs+Ri)

Theta = arctan B21/G21

Zoc = SQR(Rll Ro)

XL8 = Zoc Sin160

X1 = Zoc^2Sin160/(Ro Cos160-Zoc)

X2 = Zoc^2Sin160/(Rll Cos160-Zoc)

Zoi = SQR(Ri/Gin)

X3 = Zoi Sin(20+Theta)

X5 = Zoi^2 Sin(20+Theta)/(Ri Cos(20+Theta)-Zoi)

X4 = Zoi^2 Sin(20+Theta)/
(Cos(20+Theta)/Gin-Zoi

Cl = fop Ci/2df

Vo = Ic Rlls

Po = Vo^2/Rl

Pc = Vo^2 G22

Pb = Vo^2Ri/(Rls(Rs+Ri))

Px = Vo^2Rs/(Rls(Rs+Ri))

Appendix II
Equations Used In SKO Design

Let: Kc = Ri/Rs

Kv = Crystal/Collector Voltage Ratio

Rs = Crystal Resistance at Series Resonant Frequency

Rll = Resistance at Collector due to Rl

R22 = Resistance at Crystal due to Rs

Rc = Bridge Resistance at Collector

Kl = Output Power/Crystal plus Output Power

MQ = Ratio of Loaded/unloaded Q of Crystal

Mg = Open-loop Voltage Gain

Then: Zr = Rs Gin

Zb = Zr/Gin

Rl = Kc Rs

Rll = (KcRs^2+RsZb(Kc+1))/Rs+Zb(1-Kv))

R22 = (KcRs^2+RsZb(Kc+1))/(KcRs+Zb(1-1/Kv))

Rc = Rll R22/(R22+Kv^2Rll)

Ri = Rc(1/Ki-1)

Rld = RcRl/(Rc+Ri)

Mg = -Y21RldSQR(Zb/Gin)(1+KcKv)/(Ri+Zb(1+Kc))

MQ = Y21RldSQR(Zb/Gin)/
(Ri+Zb(1+Kc)(1-SQR(Zb/Gin)/(Ri+Zb(1+Kc)

Theta = arctan B21/G21

X_1 = ($R22^2/Kv^2$)Sin170/(R22 Cos170-R_{22}|Kv|)

X_2 = ($R22^2/Kv^2$)Sin170/(R22/Kv^2)Cos170-R22/|Kv|)

XL8 = (R22/|Kv|)Sin170

X5 = (Zr/Gin^2)Sin(10+Th)/
(Cos(10+Th)Zr/Gin-SQR Zr/Gin)

X4 = (Zr/Gin^2)Sin(10+Th)/(Cos(10+Th)/
(Gin-SQR Zr/Gin)

X3 = (SQR (Zr)/Gin)Sin(10+Th)

Cl = fopCiCxRl/(2dfRl cx+fopRsCi)

Cx = fopCiClRs/((fopCi-2dfCl)R_1)

Vo = Ic Rld

Po = Vo^2/Rl

Pc = Vo^2 G22

Vbl = Zb(1+KcKv)/R_1+Zb(1+Kc))

Px = $(Vo(Kv-Vbl)^2/Rs$

Pb = $(VoKvVbl)^2/Zb$

and the gain/phase plot reflects the effects of the change.

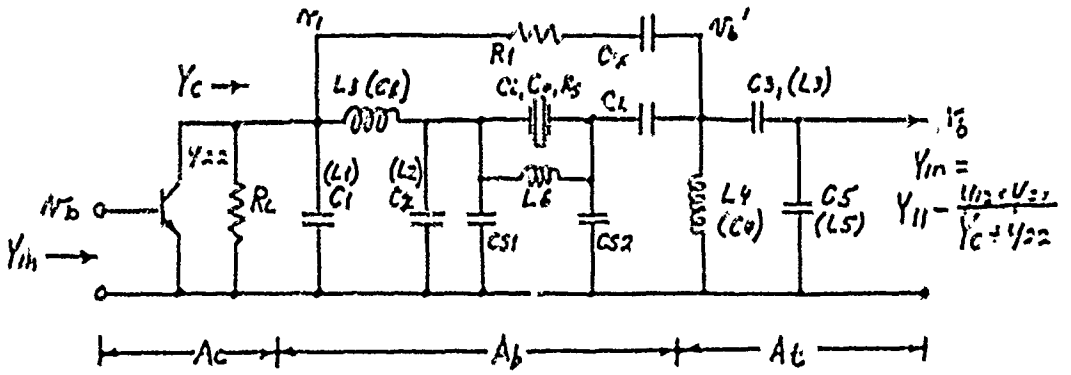
There are two sensitivity parameters: 1) Phase Zero Sensitivity, and 2) Gain Zero Sensitivity. The first refers to the effect of a parameter change on the frequency of oscillation (which always occur at the frequency of zero phase shift). While the second refers to the change in gain (Hg) at zero phase.

Acknowledgement

The author wishes to acknowledge the support of the U.S. Army Laboratory Command under Contract DAAL01-85-C-0445.

References

(1) L.A. Meacham, "Bridge-Stabilized Oscillator" Proceedings of the IRE 26(10) pp. 1278-1294 (1938).



$$Ac = \frac{-Y_{21}}{Y_c}$$

$$Ab = \frac{\pm V_b'}{V_1}$$

$$A_t = \frac{g}{g+h+Yin}$$

$$Yc = \frac{1}{RL} + G_b + G_{22}$$

$$A_b = \frac{\pm D(A(B+E)+B(C+E))}{CD(A+B)+E(A+B)(C+D)+AB(C+D)}$$

$$Y_t = \frac{f+g(h+Yin)}{g+h+Yin}$$

$$+j(B_b+B_1+B_{22})$$

$$Y_b = \frac{A(B+D+E)+E(B+C+D)+C(B+D)}{CD(A+B)+E(A+B)(C+D)+AB(C+D)}$$

$$f = jB_4$$

$$A = jX_8 \qquad B = jX_2$$

$$g = jB_3$$

$$C = R_1 + jX_x \qquad D = 1/Y_t$$

$$h = jB_5$$

$$E = \frac{Rs+j(\omega Li-1/\omega Ci)(\omega^2 LiCo-Co/Ci)-j(\omega CoRs^2)}{(\omega CoRs)^2+(\omega^2 LiCo-Co/Ci-1)^2} - j/Ci$$

$$Hg = Ac * Ab * At$$

Figure 1. CODA _ SRO Basic Circuit

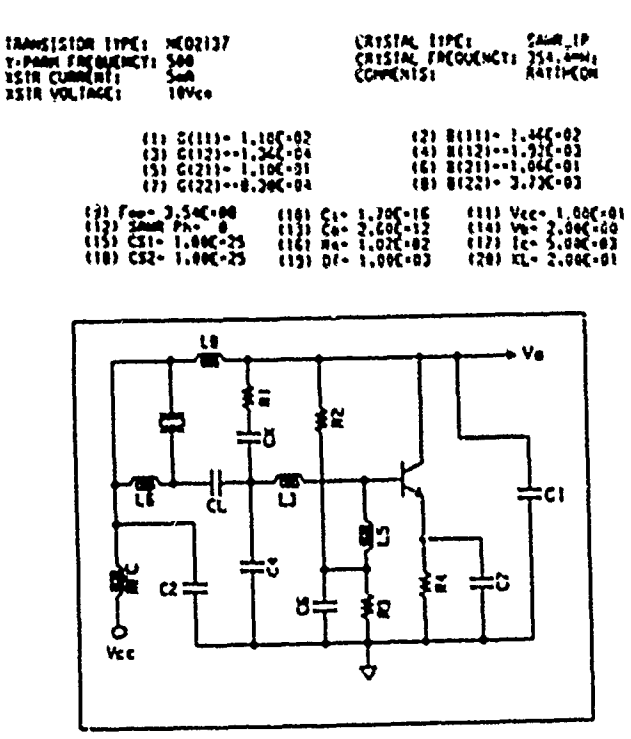
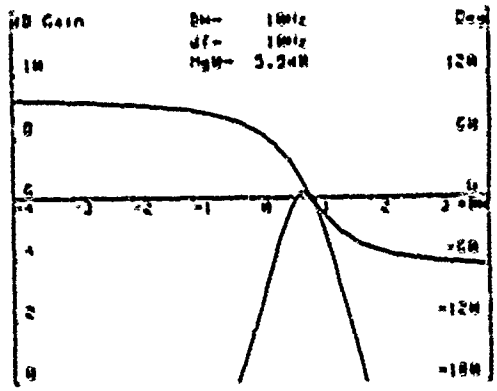


Figure 2. Input Format and Schematic Diagram



MAG Ac= 4.0
MAG Ab= .50
MAG At= .99

PHASE Ac= 127 DEG
PHASE Ab= 232 DEG
PHASE At= 1 DEG

LOADED Q IS 8.0E+05
UNLOADED Q IS 1.0E+06

Mg= .78
PHASE SLOPE=-2.6E-02

ANALYSIS OUTPUTS

R1= 100.E+05	CRYSTAL TYPE:	SC_3	C1= 100.E-14		
R2= 952.E+01	FREQUENCY:	10MHz	C2= 180.E-12		
R3= 286.E+01	TRANSISTOR:	2N918	C3= 100.E-11		
R4= 100.E+00	CURRENT:	1mA	C4= 100.E-12		
RL= 400.E+00	COMMENTS:	DCXO	LS= 100.E-06		
Cs= 144.E-16			LG= 140.E-04		
CL= 100.E-11			CG= 227.E-12		
Mg= 2.0			C7= 118.E-11		
Df= 10Hz			C8= 100.E-11		
QL= 801776			QU= 1026805		
Gin= 7.0E-04			Bin= 5.5E-04		

Figure 3. Analysis of 10 MHz Pierce Oscillator

TRANSISTOR TYPE: 2N5179
Y-PARM FREQUENCY: 100MHz
XSTR CURRENT: 5mA
XSTR VOLTAGE: 10Vcc

CRYSTAL TYPE: AT_5
CRYSTAL FREQUENCY: 100mHz
COMMENTS: CCC

COMMON BASE Y-PAR.

(1) G(11)= 1.20E-01
(2) B(11)=-4.00E-02
(3) G(12)=-1.00E-05
(4) B(12)=-6.00E-04
(5) G(21)=-1.20E-01
(6) B(21)= 6.00E-02
(7) G(22)= 1.00E-05
(8) B(22)= 1.00E-03

(9) Fop= 1.00E+08
(10) C1= 4.00E-16
(11) Vcc= 1.00E+01
(12) SAWR Ph= 0
(13) Co= 2.80E-12
(14) Vb= 2.00E+00
(15) CS1= 1.00E-25
(16) Rs= 4.00E+01
(17) Ic= 5.00E-03
(18) CS2= 1.00E-25
(19) Df= 1.00E+03
(20) KL= 3.00E-01

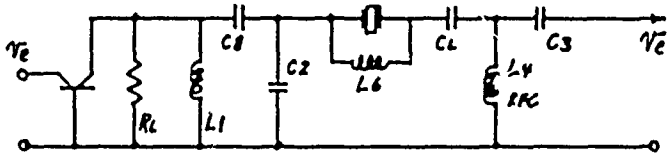
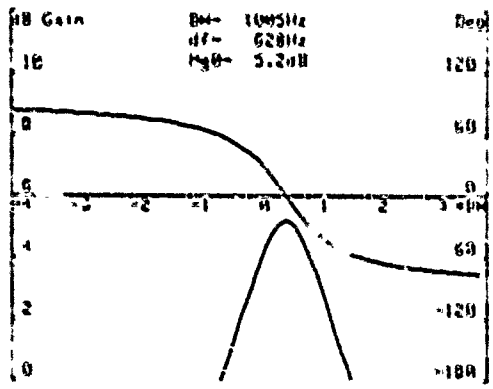


Figure 4. Analysis Input for Grounded-Base Series-Resonant Oscillator at 100 MHz



MAG Ac= 95.4
MAG Ab= .02
MAG At= 1.01

PHASE Ac= 2e DEG
PHASE Ab= 25 DEG
PHASE At= 1 DEG

LOADED Q IS 7.1E+04
UNLOADED Q IS 9.9E+04

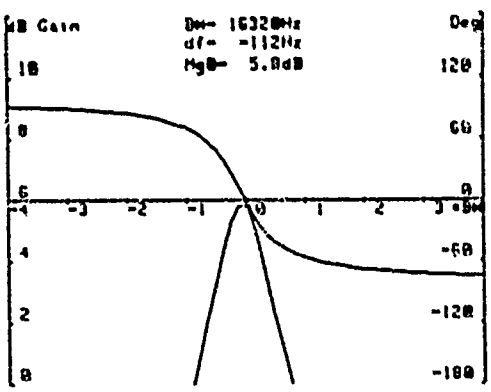
Mg= .71
PHASE SLOPE=-2.3E-04

ANALYSIS OUTPUTS

R1= 100.E+18	CRYSTAL TYPE:	AT 5	L1= 870.E-10
R2= 100.E+02	FREQUENCY:	100MHz	C2= 330.E-12
R3= 300.E+01	TRANSISTOR:	2N5179	C3= 100.E-10
R4= 200.E+00	CURRENT:	5mA	L4= 100.E-06
RL= 800.E+00	COMMENTS:	CCC	C5= 100.E-14

Cx= 100.E-20	L6= 900.E-09
CL= 320.E-13	C6= 100.E-11
Mg= 1.8	C7= 687.E-12
Df= 628Hz	C8= 300.E-13
QL= 70940	QU= 99471
Gin= 7.8E-02	Bin=-6.2E-02

Figure 5.
Analysis Output for Grounded-Base Series-Resonant Oscillator at 100 MHz



MAG Ac= 12.6
MAG Ab= .14
MAG At= 1.10

PHASE Ac= 112 DEG
PHASE Ab= 18 DEG
PHASE At= 49 DEG

LOADED Q IS 2.6E+04
UNLOADED Q IS 2.6E+04

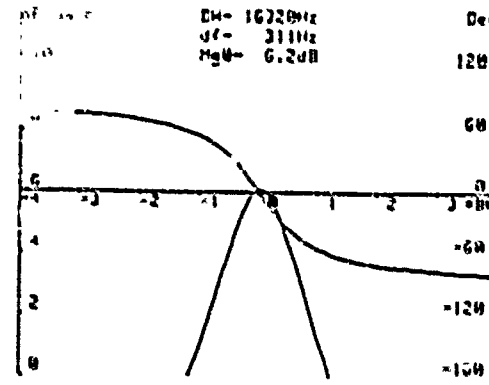
Mg= .98
PHASE SLOPE=-1.9E-05

ANALYSIS OUTPUTS

R1= 100.E+05	CRYSTAL TYPE:	SAW 2P	C1= 100.E-14
R2= 952.E+01	FREQUENCY:	500MHZ	C2= 100.E-14
R3= 286.E+01	TRANSISTOR:	NE856	C3= 150.E-13
R4= 186.E+00	CURRENT:	7MA	L4= 120.E-10
RL= 500.E-01	COMMENTS:	RAYTH 180°	L5= 700.E-11

Cx= 144.E-16	L6= 140.E+04
CL= 158.E-12	C6= 241.E-12
Mg= 2.0	C7= 118.E-11
DF= -112Hz	L8= 100.E-12
QL= 25612	QU= 26024
Gin= 2.7E-02	Bin= 3.7E-02

Figure 7.
2-Port SAWR (180°) Series-Resonant Oscillator at 454 MHz



MAG Ac= 16.6
MAG Ab= .11
MAG At= 1.09

PHASE Ac= 157 DEG
PHASE Ab= 156 DEG
PHASE At= 7 DEG

LOADED Q IS 1.9E+04
UNLOADED Q IS 2.6E+04

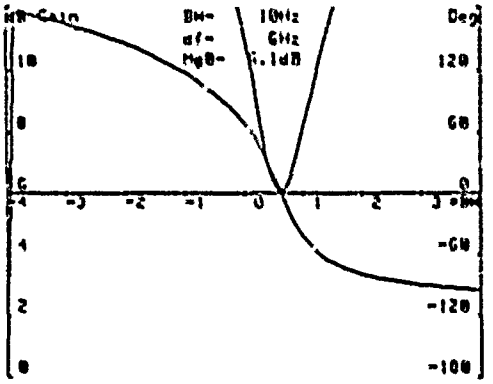
Mg= .73
PHASE SLOPE=-1.4E-05

ANALYSIS OUTPUTS

R1= 100.E+05	CRYSTAL TYPE:	SAW 2P	C1= 302.E-13
R2= 952.E+01	FREQUENCY:	500MHZ	C2= 513.E-13
R3= 286.E+01	TRANSISTOR:	NE856	C3= 857.E-13
R4= 186.E+00	CURRENT:	7MA	C4= 503.E-14
RL= 854.E 01	COMMENTS:	RAYTH 0°	L5= 582.E-11

Cx= 144.E-16	L6= 140.E+04
CL= 158.E-12	C6= 241.E-12
Mg= 2.1	C7= 118.E-11
DF= 311Hz	L8= 610.E-11
QL= 18895	QU= 26024
Gin= 2.6E-02	Bin= 4.3E-02

Figure 6. 2-Port SAWR (0°) Series-Resonant Oscillator at 454 MHz



MAG Ac= 33.6
MAG Ab= .06
MAG At= 1.00

PHASE Ac= 171 DEG
PHASE Ab= 188 DEG
PHASE At= -0 DEG

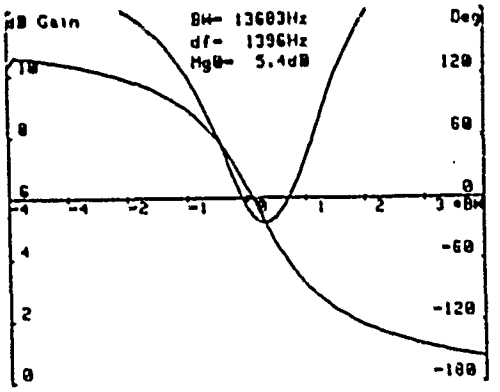
LOADED Q IS 1.3E+06
UNLOADED Q IS 1.0E+06

Mg= 1.24
PHASE SLOPE=-4.0E-02

ANALYSIS OUTPUTS

R1= 119.E+01	CRYSTAL TYPE:	SC 3	C1= 134.E-12
R2= 250.E+02	FREQUENCY:	10MHz	C2= 868.E-12
R3= 200.E+02	TRANSISTOR:	2N918	L3= 159.E-3S
R4= 130.E+01	CURRENT:	1mA	L4= 253.E+07
RL= 807.E+02	COMMENTS:	OCXO	L5= 667.E-08
Cx= 155.E-12			L6= 110.E-06
CL= 141.E-12			C6= 380.E-12
Mg= 2.0			C7= 576.E-11
DF= GHz			L8= 214.E-08
QL= 1268575			QU= 1026805
Gin= 1.4E-03			Bin= 2.8E-03

Figure 8. Bridge Oscillator at 100 MHz



MAG Ac= 28.7
MAG Ab= .14
MAG At= .45

PHASE Ac= 123 DEG
PHASE Ab= 184 DEG
PHASE At= 53 DEG

LOADED Q IS 2.4E+04
UNLOADED Q IS 2.6E+04

Mg= .92
PHASE SLOPE=-2.1E-05

ANALYSIS OUTPUTS

R1= 449.E+00	CRYSTAL TYPE:	SAWR_ ?	C1= 150.E-13
R2= 133.E+02	FREQUENCY:	500	C2= 288.E-13
R3= 400.E+01	TRANSISTOR:	NE021.?	C3= 140.E-13
R4= 260.E+00	CURRENT:	5mA	L4= 237.E-10
RL= 810.E+00	COMMENTS:	RAYTHEON	C5= 362.E-15
Cx= 386.E-12			L6= 776.E-10
CL= 296.E-13			C6= 362.E-14
Mg= 1.9			C7= 687.E-12
DF= 1396Hz			L8= 194.E-10
QL= 23761			QU= 25900
Gin= 5.7E-02			Bin= 1.1E-02

Figure 9.
1-Port SAWR Bridge
Oscillator at 354 MHz

41st Annual Frequency Control Symposium - 1987

CMOS GATE OSCILLATOR DESIGN

Thomas B. Mills

National Semiconductor Corporation

Santa Clara, CA. 95051

Introduction

The MOS gate oscillator, built with NMOS, PMOS, or CMOS processes, is probably the most widely - used oscillator circuit in the world: almost all microprocessors and computers, as well as many digitally tuned entertainment radios, use a gate oscillator. Some of these circuits have had problems in production, perhaps because they were designed by system engineers with little background in analog circuit design.

The design of gate oscillators in the frequency range of 1 to 50 MHz with fundamental and overtone crystals will be covered. Design techniques using scattering (S) parameters and more fundamental methods will be prsented.
A fundamental problem exists in the design of oscillators: they either work or they do not. It is difficult to tell how well they are working. In the design of an amplifier, specifications of gain, phase shift, time delay, pulse response, etc are made and can be designed to. These parameters can be measured and compared to specifications to evaluate the design. In the design of an oscillator, it usually oscillates or it does not. If it doesn't, component values are changed until it does. It is uncertain whether we are on the edge of oscillation and that a slight change in circuit parameters might result in no oscillation. Computer simulations are helpful in seeing how reliable the design is.

The Basic Gate Oscillator

The basic crystal gate oscillator and the familiar crystal equivalent circuit are shown in Fig 1. The configuration is the Pierce Oscillator with the crystal operating in the parallel mode. Fig 2 is a reactance plot of a quartz crystal showing the series resonant frequency, Fr, and the parallel or antiresonant frequency, Fa, of the crystal. Between Fr and Fa the impedance of the crystal is inductive and will resonate with capacitors Ca and Cb to produce the desired frequency of operation. At low frequencies, the gate, operating as an inverter, produces a phase shift of 180 degrees. The crystal and the network Ca, Cb, and Ro provide an additional 180 degrees of phase shift. As long as the power loss through the crystal and its associated network does not exceed the the power gain of the gate, oscillation is assured at a frequency where the overall phase shift is zero.

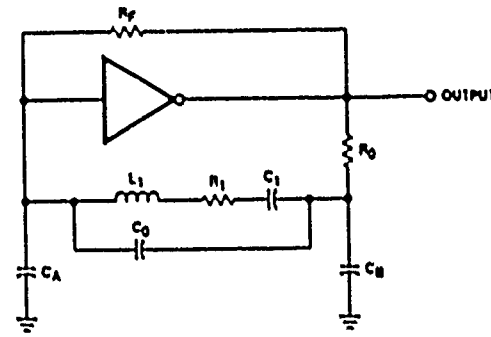


Figure 1. Gate Oscillator With Crystal Equivalent Circuit

Typical parameters for quartz crystals operating in various modes and at various frequencies are given in Table 1. These values are useful for starting design and computer simulation models.

Table 1
Typical Crystal Parameters

Frequency (MHz)		R1 (ohms)	L1 (mHy)	C1 (pF)	C0 (pf)
2		75	580	.011	3.6
10		15	14	.018	4.0
20	Fund.	20	3.1	.02	4.5
20	3rd overtone	25	40	.0015	3.1
40	3rd overton	25	9.3	.0017	3.8

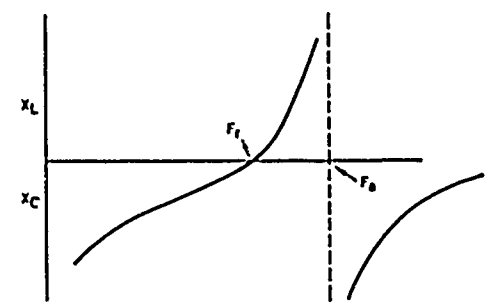


Figure 2. Reactance Plot Of Crystal Resonator

CH2427-3/87/0000-460 $1.00 C 1987 IEEE

CMOS Gate Construction

The simplest CMOS gate (inverter) consists of both a P channel and an N channel device connected as shown in Fig 3a [1]. This represents an unbuffered inverter. Today, most gates are built with two additional stages to form a three inverter gate: this is the configuration of the buffered gate shown in Fig 3b. The single inverter has less gain and slightly less delay time in the large signal mode; it was designed to be a better oscillator. On first inspection it would seem the buffered gate with its three inverters would have considerably more delay than the unbuffered gate used as a logic element, but internal stages are considerably faster than the output stage and the speed lost by buffering is relatively small. Due to the gain in the first two stages, the output FETs of the buffered gates are driven on and off somewhat faster than unbuffered gates and tend to consume less power than unbuffered gates operating at the same frequency. As a result, the buffered gate makes a better logic element than the unbuffered gate. However, the small signal characteristics of the buffered gate are considerably different noteably in their tremendous gain (S21) that tend to make them unstables as oscillators.

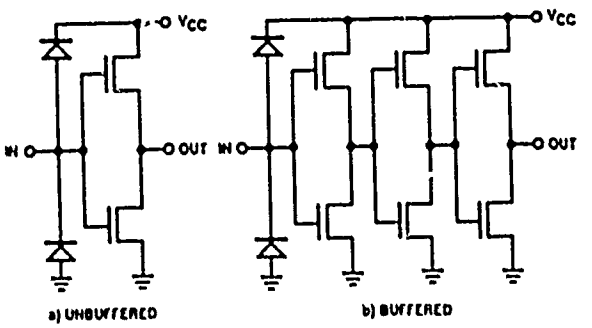


Figure 3. CMOS Gate Construction

S Parameter Analysis

Recently S parameter analysis has become the standard design technique for microwave circuits. S (scattering) parameters are so named because they refer to the waves flowing into and reflected from the ports of the network under measurement. S parameters are relatively easy to measure over wide frequency ranges, and provide good insight into device behavior. Traditional circuit analysis has used H, Y, or Z parameters to calculate circuit performance. These parameters are difficult to measure at high frequencies, requiring accurate opens and shorts, and are almost impossible to measure over a wide band of frequencies. From a practical standpoint, many devices oscillate when terminated with opens or shorts, further complicating the measurements.

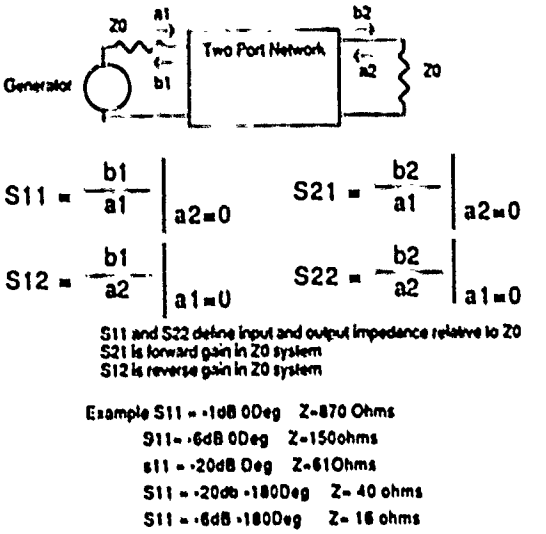


Figure 4. S Parameter Definitions

Fig 4 defines the four S parameters of a network. S11 and S22 are the input and output voltage reflection coefficients and S21 and S12 are the forward and reverse voltage transfer coefficients. All parameters are measured in a constant impedance network, usually 50 ohms.

S11 and S22 can be used to find the input and output impedance of the circuit using the following relation:

$$Zin = Zo \frac{1 + S11}{1 - S11}$$

The S parameters of a typical unbuffered HCMOS gate are shown in Table 2,while table 3 and 4 are the S parameters for a new high performance microcontroller [2] and a single chip 8 bit cmos microprocessor[3]. An explanation of the parameters gives some insight as to the device's performance. An S11 of .9 at -4 degrees indicates most of the energy is reflected from the input of the gate. The phase angle of -4 degrees indicates the impedance is high and capacitive. Solving the above equation yields an equivalent circuit of Rp = 950 ohms, Xp = -1435 ohms The same calculation can for the the output impedance. S21 of 2 at 178 degrees indicates a voltage gain of two when the device is driven from and terminated into a 50 ohm impedance. The phase shift shows it is an inverter with slightly less than 180 degrees of phase shift.

Circuit analysis with S parameters is almost impossible without the aid of a computer due to the large number of calculations with complex numbers. S parameter analysis programs for computers have been available for more than a decade, dating to Compact[4], which was originally written for the G E tymeshare mainframe computer systems in the late 1960s. With the advent of the IBM PC, many

Table 2
Typical S Parameters: Unbuffered HCMOS Gate (74HCU04)

Frequency MHz	S11 dB	S11 Ang	S21 dB	S21 Ang	S12 dB	S12 Ang	S22 dB	S22 Ang
1	-.01	-.6	8.31	179	-58.1	78	-1.58	-.8
5	-.03	-2.8	8.3	173	-43.2	81	-1.6	-3.2
10	-.1	-4.7	8.	168	-38.5	81	-1.6	-6
20	-.3	-8.9	7.9	157	-33.2	75	-1.8	-11
40	-1.0	-16.4	6.0	134	-27.7	63	-2.7	-22

Table 3
Typical S Parameters: High Performance Microcontroller

1	- .01	-.4	-1.4	178	-54.8	82	-.91	-.3
5	-.0	-2.7	-1.5	171	-39.0	85	-.94	-2.7
10	-.08	-4.9	-1.4	164	-34.9	84	-1.01	-4.7
20	-.24	-9.0	-1.5	149	-29.2	82	-1.21	-8.7
40	-.75	-15.3	-1.0	130	-25.3	79	-1.79	-15.2
60	-1.4	-22.4	-0.95	95	-20.2	94	-3.02	-21.1

Table 4
Typical S Parameters: 8 Bit CMOS Microprocessor

1	-.02	- .7	-8.7	-176	-60.8	63	-.29	.8
5	-.01	-1.0	-8.2	177	-47.2	82	-.31	-.8
10	-.03	-3.2	-8.2	173	-41.6	84	-.33	-2.3
20	-.02	-7.5	-7.9	167	-35.4	84	-.43	-5.6

low cost analysis programs have become available, and one has been reported in the public domain[5]. The advantage of these programs is their very fast execution speed when compared to more elegant software such as Spice[6], resulting in a more interactive mode. Some programs feature optimization routines that will vary component values for desired results.

S parameter analysis programs solve a network of elements and give the resulting S parameters of the overall circuit. The characteristic impedance in which the network is analyzed may be varied to suit the magnitude of the desired answers. One network suitable for gate oscillator S parameter analysis is shown in Fig 5. S11 parameters of gate 1 are used as S11 of gate 2 to present a proper load to the output of the resonator network at Ca. The remaining parameters of gate 2 are idealized to that of a perfect buffer (S21=2, 0 degrees, S22 = .99, -180 degrees). This allows the analysis of the oscillator with the loop opened as shown by the dashed line in Fig 5.

The analysis shows clearly the gain - phase characteristics of the overall network. Remembering the condition for oscillation to be phase shift of zero degrees and gain greater than unity, the operating conditions of the circuit can be visualized. More important, stability of oscillation can be estimated by noting how fast the phase characteristic goes through zero and whether maximum gain and zero phase shift occur at approximately the same frequency. Component values may be varied to see how much better or worse the oscillator starting condidtion will be.

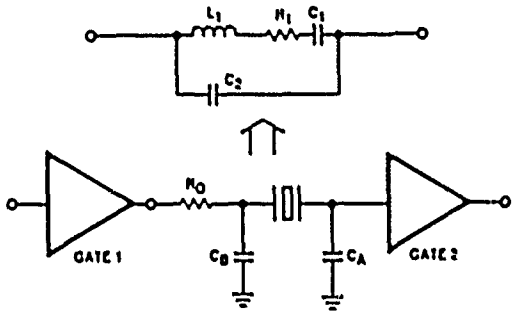


Figure 5. Network Used To Simulate Open Loop Characteristics of CMOS Gate Oscillator

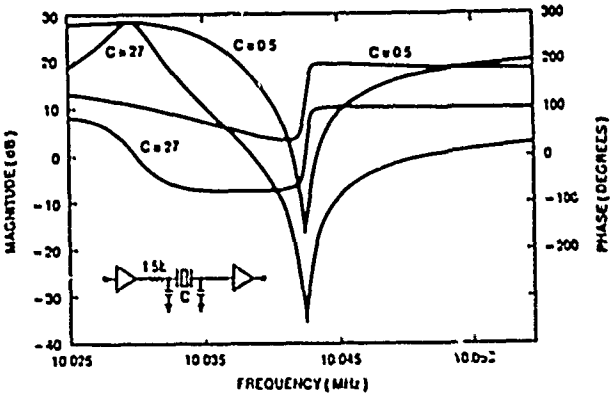


Figure 6. Load Capacitance Effects

The One Port Reflective Amplifier

The model just described gives a clear picture of a complete gate crystal oscillator but does not provide a feeling for the contribution of the various elements that make for a good or bad oscillator. Recent approaches to oscillator design [7] have used the one port reflection coefficient method. This technique analyzes the oscillator circuit as a one port network and requires that the input reflection coefficient of the one port network be greater than the loss associated with the network termination - in this case the crystal. This must occur at a frequency at which the phase of the termination is equal and opposite that of the active network. Oscillation will occur at this frequency.

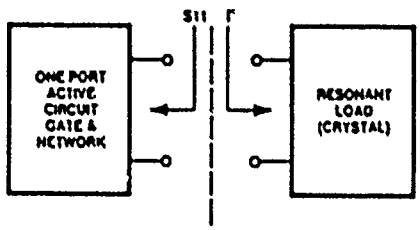


Figure 7. The Oscillator as a 1 Port Reflective Amplifier

Fig 8 shows how a network may be transformed mathematically in Star[8], an S parameter analysis program. The X2G statement transforms the output port with the common (ground) terminal. Rs has been added to the network to isolate from the termination impedance. Now the input scattering parameter S11 presented to the crystal by the network may be calculated. Similarly, the same technique may be used to determine S11 of the crystal from its equivalent circuit. In order to obtain reasonable numbers, the results are presented in S parameters relative to a 1000 ohm system, a characteristic impedance closer to the impedances associated with crystals and MOS gates.

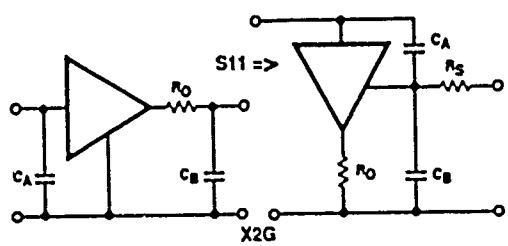


Figure 8. Transformation of Two Port (Feedback) Oscillator to a 1 Port(Negative Resistance) Oscillator

Fig 9 shows S11 for a 10 MHz crystal using the parameters given earlier. For maximum stability, it is desireable to operate at a frequency where the phase is changing rapidly with frequency -in the range of +50 to +200 degrees. In this region, the crystal has the hifgest loss, approaching 0.5 dB at +150 degrees, the frequency of maximum phase slope.

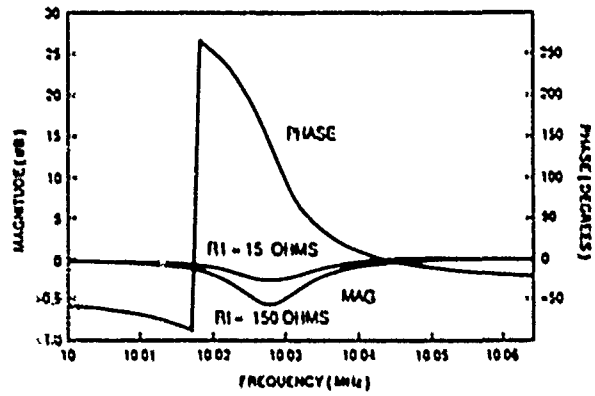


Figure 9. S11 Crystal

In the interest of optimizing the feedback network many simulations were run to provide a feeling of oscillator performance for the basic gate oscillator as the load capacitors were varied. Fig. 10 shows the results of changing CA for several values of CB on S11 for both magnitude and phase for an unbuffered gate at 10 MHz. It can be seen that to approach -100 degrees of phase, fairly large capacitors ,greater than 50pF, must be used. Interestingly, these values also produce the largest amount of input reflection, the criterion for maximum oscillator activity.

Simulations run on the high performance controller and the 8 bit CMOS microprocessor are shown in Fig. 11 and 12. It is interesting to note that the devices used in these processors are

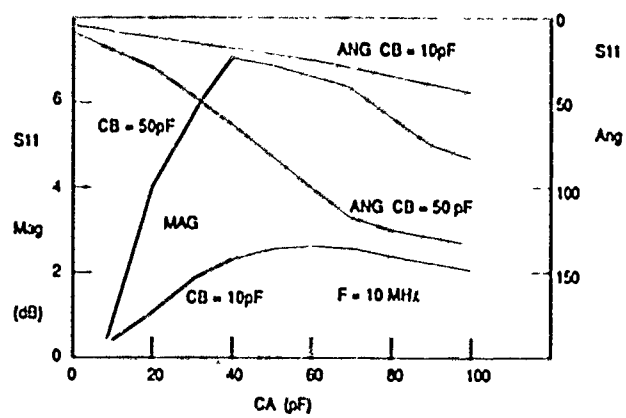


Figure 10. S11 Characteristics 74HCU04 Gate

smaller and result in less gain and higher impedances than the devices used in the 74HCU04 gate. Optimum network capacitance is less with these microprocessors.

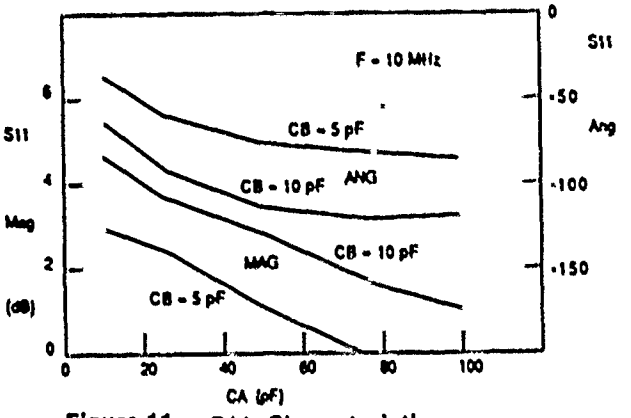


Figure 11 S11 Characteristics 8 Bit CMOS Microprocessor

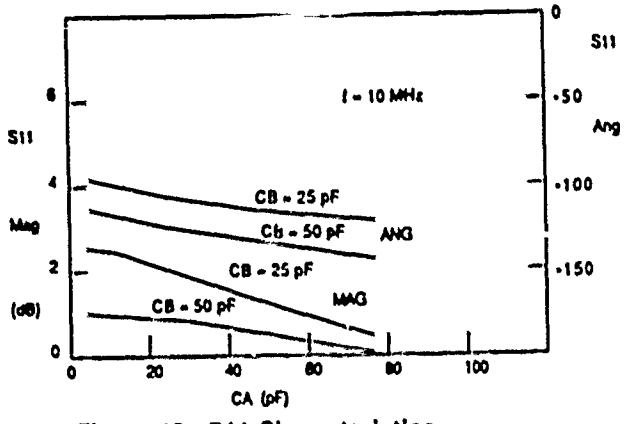


Figure 12. S11 Characteristics High Performance Controller

The results of simulations on an unbuffered gate over a frequency range of 5 to 30 MHz are shown in Fig. 13 for two values of capacitors in the feedback network. The magnitude and phase curves to the left show S11 for a network with 27 pF capacitors - fairly standard design values, and S11 for a network with 6 pf capacitors. The lower value of capacitors has appeal for monolithic IC designs that would try to incorporate "on-chip" capacitors, since capacitors eat up large ammounts of chip area and are therefore expensive. At the frequency of oscillation, the lagging phase angle of S11 for the 27 pF gate network matches the leading phase angle of S11 for the crystal, and S11 for the gate and capacitor network shows more gain (greater than unity) than the loss of S11 (less than unity) for the crystal. Also it should be noted that the slope of the phase is changing fairly rapidly in the vicinity of oscillation, indicating a stable oscillator.

However, for the 6 pf capacitor network, the magnitude of S11 is reduced

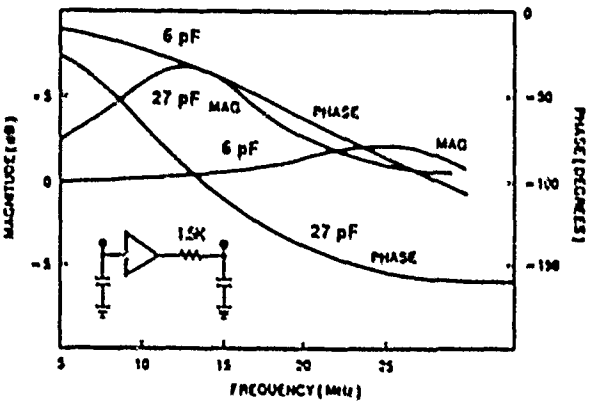


Figure 13. Load Capacitance Effects

considerably below 20 MHz while the phase shift in this area is less than 50 to 60 degrees. In this region, the crystal's phase shift is flattening out and approaching zero - a condition not desirable for good oscillator stability.

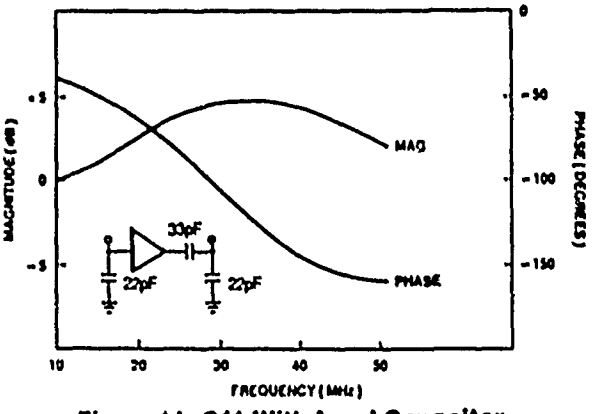


Figure 14. S11 With Lead Capacitor

Higher Frequency Oscillators

At higher frequencies where excess phase shift and lower gain in the gates reduce their performance, it is desirable to replace the output resistor with a capacitor to provide leading phase.The simulation results with this circuit, Fig. 14, show the improvement in magnitude and phase above 20 MHz.

Overtone crystals are often used above 30 MHz, but since the fundamental response is stronger than the overtone, it is necessary to reduce the magnitude of S11 to prevent operation in this region. The use of an L-C network will accomplish this. Simulation results in Fig 15 illustrate how a gate L - C oscillator can be optimized in the 30 to 50 MHz range to work with overtone crystals and reject the fundamental mode (10 to 17 MHz)

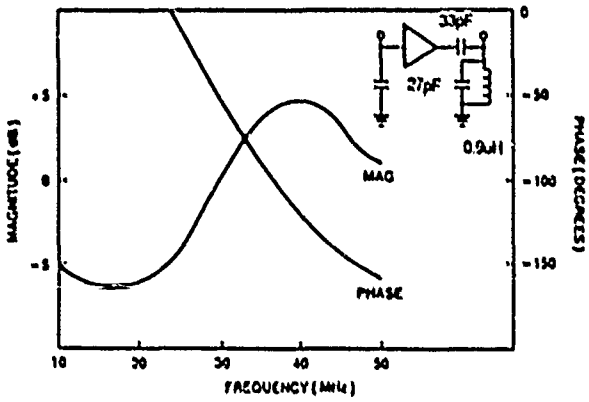


Figure 15. S11 LC Overtone Oscillator

Conclusions

The use of the reflective amplifier approach provides an insight to oscillator performance. Easily measured S parameters and low cost analysis software allows for rapid optimization of feedback networks to assure starting conditions with gate oscillators. Intrinsic characteristics of different gate designs require unique external crystal network design.

Acknowledgement

The author would like to thank Jim Moyer of the National Semiconductor technical staff for many helpful suggestions and criticisms of the analysis techniques and also for his help in proofreading the material for this paper.

[1] L. Wakeman,"D.C. Electrical Characteristics of MM54HC/74HC High Speed CMOS Logic, "AN-313, National Semiconductor Logic Databook, Vol 1.pp. 2-102. 1984

[2] "HPC16049... High-Performance Microcontrollers" datasheet. National Semiconductor Microcontroller Databook, pp 5-4, 1987

[3] "COP820C/COP821C/COP822C Single Chip microCMOS Microcontrollers" datasheet. National Semiconductor Microcontroller Databook, pp 2-4, 1987

[4] L. Besser,"Combine S-Parameter With Time Sharing", Electroic Design, Vol 16, August 1968, pp.62-68

[5] J.N. MacDonald, "Public Domain Library Gets Underway", R F Design,pp. 6, August , 1986

[6] L. W. Nagel, SPICE2: A Computer Program To Simulate Semiconductor Circuits," Univ. California, Berkeley, ERL Memo. ERL-M520, May 1975.

[7] J. W. Boyles, "The Oscillator As A Reflection Amplifier: An Intuitive Approach To Oscillator Design". Microwave Journal, pp. 83, June 1986

[8] R.W Rhea " Star 1.0 Manual",Lilburn, Ga, Circuit Busters

REVIEW OF THE REVISED MILITARY SPECIFICATION FOR QUARTZ CRYSTAL OSCILLATORS

S. Schodowski and V. Rosati

US Army Electronics Technology & Devices Laboratory (LABCOM)
Fort Monmouth, NJ 07703-5000

Summary

Revision B of the military specification for quartz crystal oscillators, MIL-O-55310B, is expected to be released during 1987.This document will supercede MIL-O-55310A. It will permit the specification and acquisition of more-reliable and higher-performance crystal oscillators required by modern military equipment.

Major modifications and improvements were made to requirements specifying design and construction, performance and test, quality assurance, and qualification. Definitions were expanded and two new oscillator types, the MCXO (Microcomputer-compensated-crystal oscillator) and the RbXO (Rubidium-crystal oscillator) were added.

Some of the changes discussed are modifications of frequency-stability and aging requirements, and the introduction of several new requirements such as "Initial frequency-aging," "Initial frequency-temperature accuracy including hysteresis and trim effect," and "Phase noise, vibration." Design and construction requirements changes, and new quality assurance provisions that include 100% environmental stress screening, and certification of hybrid facilities and lines, are reviewed.

Introduction

Revision B of the general specification for crystal oscillators, MIL-O-55310B, is expected to be released during 1987. The revised government-industry coordinated specification will supercede MIL-O-55310A, which was issued on 24 Nov 1976. The primary reason for updating this standardization document is to enable the specification and acquisition of devices that use the latest crystal oscillator technology and methodology. Changes were necessary in order to meet the more stringent frequency control and timekeeping requirements of new military systems. Defects existing in Revision A were corrected. It is also intended that Revision B provide for improved levels of quality and reliability. Revision goals and preliminary coordination efforts were described at the 1983 Frequency Control Symposium.[1]

Changes have been made in each major category of the specification. Two new oscillator types, primarily intended for low power timekeeping, have been added. Numerous modifications and additions have been made to requirements specifying performance, test, design and construction, and quality and reliability assurance. Definitions have been expanded. This paper describes some of the more significant changes.

New Device Types

The MCXO

The MCXO is a microprocessor-based system, employing external means of temperature compensation. Unlike a conventional TCXO, there is no pulling of the crystal frequency. Instead, time correction techniques such as the pulse deletion method shown in Figure 1 are used.[2]

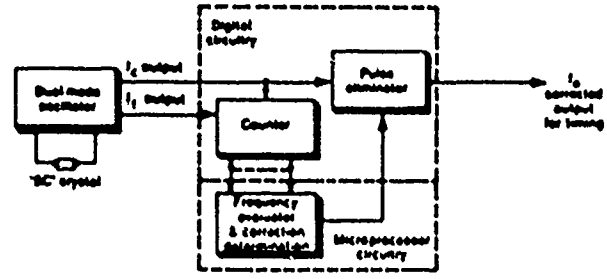


Figure 1. MCXO basic block diagram.

Simply described, the microcomputer reads the SC-cut crystal temperature, f_T, determines the necessary correction to the uncompensated crystal frequency, f_C, and then subtracts the correct number of pulses. The output can be described as a time-corrected pulse train whose average frequency is the compensated frequency. Because of the pulse deletion process, the output signal is exceptionally noisy and does not lend itself to conventional specification and inspection for frequency accuracy. The approach taken by Revision B is to specify MCXO timekeeping performance in terms of clock accuracy, and to specify a test for measuring accumulated time error under specified operating conditions.

A number of output variations are possible. The time-corrected signal can be divided to provide a 1 pulse per second time reference, and clock circuitry providing digital time of day output can be incorporated.

The RbXO

The second new device type, the RbXO, is an oven-controlled crystal oscillator combined with a rubidium reference source.[3] It is intended to make precise time and frequency available to military systems that lack the power required for sustained operation of atomic frequency standards. A basic block diagram of the RbXO is shown in Figure 2.

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

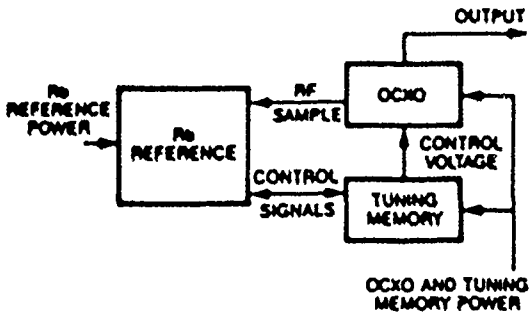


Figure 2. RbXO basic block diagram.

In operation, the rubidium reference is turned on, intermittently, for a period of about five minutes, to frequency adjust (syntonize) the voltage-controlled OCXO. The OCXO, designed for low power consumption, e.g., the Tactical Miniature Crystal Oscillator (THXO),[4] is on continuously. A digital tuning memory retains the frequency-control voltage until the next syntonization. Time between syntonizations is adjustable, depending on system accuracy requirements and operating conditions. This method provides near-rubidium standard accuracy at a total power consumption not much more than the OCXO itself.

The RbXO has introduced two new requirements. One is "Syntonization energy" which is computed from the time integral of rubidium reference supply power. The other is "Magnetic susceptibility," which specifies RbXO output frequency as a function of a DC magnetic field.

Frequency-Temperature Performance and Test

Initial frequency-temperature accuracy

A major defect in Revision A was a lack of standardization in specifying frequency-temperature performance. It was found that among QPL oscillator manufacturers and users, "Frequency-temperature stability" took on different meanings. As illustrated in Figure 3, some interpreted it as frequency-temperature deviation limits, (f_{max} and f_{min}, referenced to nominal frequency), some referenced the deviation limits to frequency measured at standard room ambient (T_{ref}), while still others used no reference, (i.e., they considered only relative peak-to-peak deviation). Each interpretation can provide different apparent performance and lead to confusion and conflict between user and manufacturer. This lack of standardization exists throughout the industry today, and is evident in published product specification sheets.

The solution was to establish "Initial frequency-temperature accuracy" as the preferred method for specifying frequency-temperature performance. It is defined as the "initial maximum permissible deviation of the oscillator frequency from the assigned nominal value due to operation over the specified temperature range," at nominal supply voltage and load conditions, other conditions remaining constant. For inspection, initial frequency-temperature accuracy is determined by:

$$\text{f-T accuracy} = \pm \text{ MAX } (\delta f_{max}, \delta f_{min})$$

where MAX ( ) is the maximum value of the fractional frequency deviations, and δf_{max} and δf_{min} are computed as follows:

$$\delta f_{max} = |(f_{max} - f_{nom})/f_{nom}|$$
$$\delta f_{min} = |(f_{min} - f_{nom})/f_{nom}|$$

For non-frequency-adjustable (manufacturer calibrated) oscillators, the initial frequency-temperature accuracy applies at the time of manufacture and for a specified period following shipment. For frequency-adjustable (manufacture-/user-calibrated) oscillators, the initial frequency-temperature accuracy applies at the time immediately following calibration by the manufacturer or user.

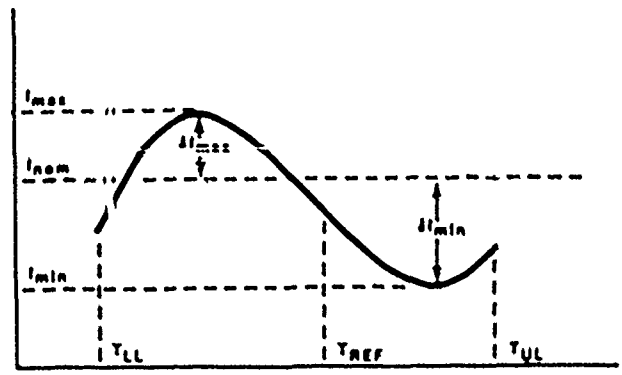


Figure 3. Representation of frequency-temperature performance. Over the temperature range (T_{UL} - T_{LL}), f_{max} and f_{min} have been arbitrarily referenced to either a.) f_{nom}, b.) frequency at T_{ref} or c.) relative to each other (no reference implied).

"Frequency-temperature stability," is retained in Revision B as a special requirement for use in cases where a relative frequency-temperature change is required to be specified. It is now explicitly defined, with no reference implied, as follows:

$$\text{f-T Stability} = \pm(f_{max} - f_{min})/(f_{max} + f_{min})$$

Initial f-T accuracy including hysteresis and trim effect

Figure 4 illustrates inspection procedures for a TCXO specifying "Initial f-T accuracy including hysteresis and trim effect." Thermally-induced hysteresis manifests itself as non-repeatability of the frequency-temperature characteristic over a complete quasi-static temperature cycle. It is indicated by the frequency displacement between curves obtained for each temperature test cycle. Trim effect is the skewing (distortion) of the f-T characteristic that results from frequency adjustment. For this test, the TCXO is subjected to three complete temperature cycles, following frequency adjustment to its marked offset from nominal (condition 1), and then to specified trim range setpoints simulating worst case recalibration (conditions 2 and 3). The oscillator's performance is acceptable if its "max" or "min" deviation from nominal for all 3 condi-

tions remains within specified accuracy limits.

This method effectively combines calibration and temperature-related errors into a total error requirement and minimizes problems related to specification and test.

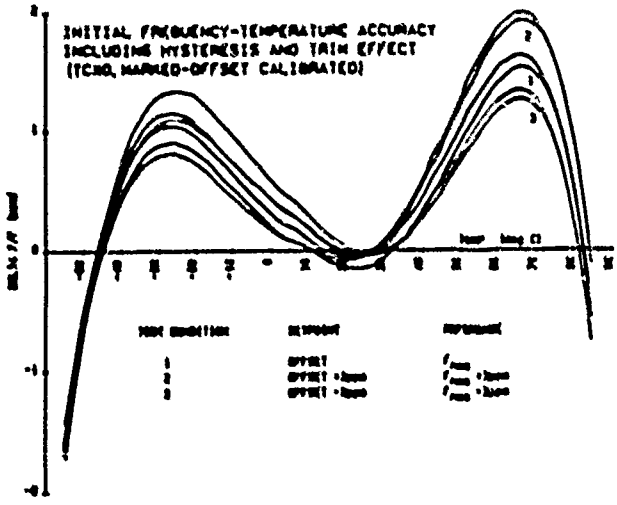


Figure 4. Illustration of inspection for "Initial frequency-temperature accuracy including hysteresis and trim effect" to specified limits of ±2 ppm (-55°C to +85°C).

Frequency Aging

Initial aging

Initial frequency aging is a critical requirement for many OCXO applications. Military equipment frequently requires stable performance for short-duration missions. Commencing with oscillator warmup, rapid frequency excursions, such as shown in Figure 5, may occur within 48 hours after thermal stabilization. This effect is believed to be due to contamination transfer, stress relief, or thermistor drift. It generally tends to be more pronounced after cold temperature storage. As a standard condition, Revision B specifies storage at -40°C for 24 hours preceding turn-on, and specifies that data be taken for 48 hours. The first data point is taken at the specified warmup time, 10 minutes in this example. Since the frequency change may be nonmonotonic, a maximum allowable frequency change over the test duration is specified. A maximum rate of frequency change can be specified if critical to the application. The test may be performed as an extension of the frequency warm-up test, provided that the storage conditions are met.

Long Term Aging

Specification of long-term frequency aging for OCXO and TCXO is based upon a method previously described for precision quartz crystal units[5]. Performance is now required to be specified by at least 2 parameters, i.e., total frequency change over a 30-day test period and projected frequency change for one year. A maximum aging rate per day at day 30 can also be specified. One can also specify total change for extended periods of time, e.g., 5 or 10 years.

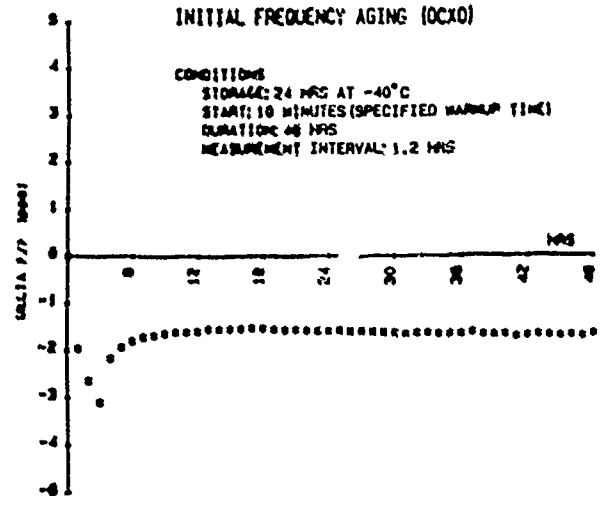


Figure 5. Initial frequency aging inspection for a 10 MHz OCXO.

Inspection consists of measuring frequency over a 30 day period and performing a least-squares fit of the data to the function:

$$f(t) = A(\ln(Bt + 1)) + f_o$$

where f(t) is the frequency of the crystal oscillator, t days after the start of the aging cycle, and A, B, and f_o are constants to be determined from the least-squares fit.

If analysis of the data indicates that the aging trend is not monotonic, i.e., exhibits a reversal, an extension of the 30 day test period is required.

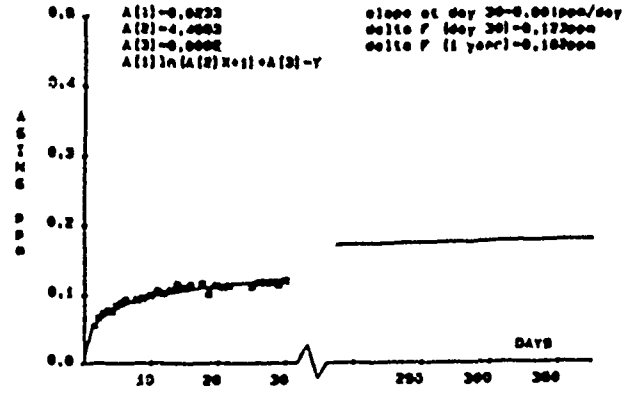


Figure 6. Long term frequency aging inspection data for a 3.2 MHz TCXO.

Figure 6 provides an example of TCXO aging inspection data at a specified temperature of 60°C. Frequency measurements are required to be made a minimum of 4 times per week, for four weeks , following a 2-day stabilization period. These are the points plotted. The solid line is a plot of the log equation extrapolated to one year using the

computed coefficients displayed in the upper left. Computed parameter values (upper right) are then compared to the specified values. The procedure requires that the r.m.s. of residuals of the data from the function be held to less than 5% of the specified total change. Five percent represents a limit within which a valid fit is assumed.

Phase noise, vibration

Phase noise, induced by vibration, is perhaps the most significant new requirement in Revision B. It is intended to meet the improved spectral purity requirements being demanded by new military systems. Because of the crystal's acceleration sensitivity[6], a crystal oscillator operating under the dynamic conditions encountered in a tactical environment will have phase noise considerably greater than at steady state conditions. The prescribed test requires measurement of the single-sideband noise to carrier ratio, (f), under the specified vibration type, frequencies, and levels. Figure 7 displays test results for a 3.2 MHz TCXO subjected to random vibration along one axis in accordance with its specified spectral density envelope. Away from the peaks, which result from mechanical resonances within the oscillator, the phase noise is 20 dB to 30 dB greater than that measured under quiescent conditions. At the major resonance near 850 kHz, g-amplification results in an additional 30 dB of phase noise.

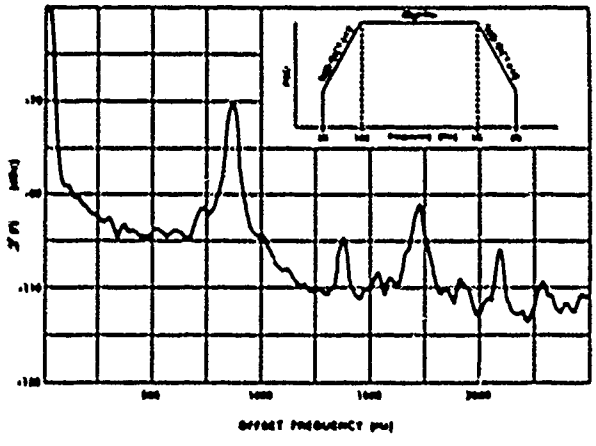


Figure 7. Random-vibration-induced phase noise for a 3.2 MHz TCXO.

Phase noise, vibration, can be specified as either a Group A inspection, i.e., the test is performed for each production lot, or a Group C (environmental) inspection requiring only periodic inspection. The latter case, applicable to oscillators where phase noise performance is less critical, is especially useful in identifying oscillator mechanical designs having problem resonances and associated high probability for mechanical failure.

Other New Performance Requirements

The majority of new performance requirements included in Revision B fall in the category of special requirements, i.e., they are applicable when specifically required by the oscillator specification sheet. "Initial frequency-temperature accuracy including hysteresis and trim effect", "Initial frequency aging", and "Phase noise, vibration", are included in this group. A partial listing of other new special requirements is as follows:

- Initial accuracy at reference temperature
- Frequency-temperature slope at reference temperature
- Frequency-temperature thermal transient stability
- Phase noise, acoustic
- Re-entrant isolation
- Output suppression
- Start-up time
- Modulation distortion
- Deviation frequency response
- Built-in-Test (BIT)
- Power consumption aging

Inspection methods for each of the new performance requirements have also been included.

Design and Construction

Design and construction requirements changes now require specification of the construction technology to be employed for each oscillator device. Three categories of construction technology are included, i.e., discrete, custom hybrid microcircuit, and mixed. Special criteria are applied to each category. Packaged devices used in discrete construction are required to comply with requirements for either established reliability or JANTX parts, and custom hybrid microcircuit construction is required to be in accordance with applicable requirements of Appendix G, MIL-M-38510[7]. Mixed construction, which combines discrete and hybrid microcircuit assemblies, requires that the hybrid portion be screened in accordance with MIL-STD-883[8]. Requirements have also been added for oscillators using uncased quartz resonators. These requirements are intended to result in improved reliability by eliminating inferior resonator mounting, sealing, and frequency adjustment methods.

Quality/Reliability Assurance

Two product assurance levels, Class B and Class S, have been established. Oscillators are now required to be 100% screened at stress levels appropriate to the device class and the construction technology used.

Within 24 months from issuance of Revision B, manufacturers of hybrid microcircuit oscillators will be required to show compliance with MIL-STD-1772[9]. That document currently imposes a uniform set of standards for line certification and process quality on manufacturers of general military hybrid microcircuits covered by MIL-M-38510. It is intended that hybrid crystal oscillator manufacturers provide at least comparable quality and reliability assurance.

Conclusion

Revision B has resulted in sweeping changes to MIL-O-55310. Modification and addition of requirements have been included that should be effective in supporting the

specification and acquisition of military crystal oscillators for at least the next decade. Because of the newness of some performance requirements and inspection procedures, some fine-tuning will inevitably be required as the specification matures. Minor changes will be incorporated by amendment as the need arises.

The inclusion of a class S quality assurance level should result in greater utilization of MIL-O-55310 in high reliability applications such as aerospace systems. These systems should benefit from oscillators manufactured with consistent criteria and uniform levels of quality and reliability.

It is hoped that the updated approaches, procedures, and definitions will become established and adopted by the frequency control community.

Acknowledgements

The authors would like to thank Dr. John Vig, and Dr. Raymond Filler, of LABCOM who contributed substantially to the revised specification; Ms. Jacqueline Messina, and Mr. Jeffrey Carver, also of LABCOM and Mr. Clarence Brezius and Ms. Janice Schere of DESC for their valuable support.

Special thanks go to all those in government and industry who participated in the coordination of this revision, for their valuable comments and helpful discussions.

References

1. V. Rosati and S. Schodowski, "Revision of the Military Specification for Quartz Crystal Oscillators (MIL-O-55310)," Proc. 37th AFCS, p. 524, 1983.

2. M. Meirs, "Lateral Field Resonator Development - Phase II," Frequency Electronics Inc., Mitchel Field, NY and US Army LABCOM, Ft. Monmouth, NJ, Research and Development Technical Report No. SLCET-TR-85-0401-1, July 1987, (to be published)

3. J.R. Vig and V. Rosati, "The Rubidium Crystal Oscillator Hybrid Development Program," Proc. 16th PTTI Applications and Planning Meeting, pp 157-165, 1984.

4. D. Brown, E. Laszlo, R. McGill, and P. Stoerner, "Manufacturing Methods and Technology for Tactical Miniature Crystal Oscillator," Proc. 38th AFCS, pp. 350-356, 1984.

5. R.L. Filler, "Aging Specification Measurement, and Analysis," Proc. 7th Quartz Crystal Conference, pp. 93-104, 1985.

6. R.L. Filler, "The Acceleration Sensitivity of Quartz Crystal Oscillators: A Review," These Proceedings.

7. Military Specification: "Microcircuits, General Specification for," MIL-H-38510F (1988).

8. Military Standard: "Test Methods and Procedures for Microelectronics," MIL-STD-883C (1988).

9. Military Standard: "Certification Requirements for Hybrid Microcircuit Facilities and Lines," MIL-STD-1772A (1987).

41st Annual Frequency Control Symposium - 1987

NOISE IN OSCILLATORS EMPLOYING SUBMICRON FIELD-EFFECT TRANSISTORS

(Invited Paper)

Madhu S. Gupta
Department of Electrical Engineering and Computer Science
University of Illinois at Chicago
Chicago, IL 60680

Abstract

The various sources of noise in a microwave field-effect transistor are surveyed, and a noise equivalent circuit model for the device is developed. Two alternative methods for the calculation of the oscillator noise spectra are then described, and the factors influencing the FM noise are identified. These factors serve as the basis for classifying the various methods of oscillator FM noise reduction.

I. Introduction

Advances in the technology of submicron device fabrication have made available a number of new active electron devices to circuit designers in the last few years, including the submicron-gate GaAs MESFET, the high electron mobility transistor (HEMT, or MODFET, or TEGFET), the HBT (heterojunction bipolar transistor), the PBT (permeable base transistor), the ballistic transistor, the quantum-well device, and several others exploiting the properties of heterojunctions, superlattices, and submicron dimensions. Of these, the two devices of the field-effect family, the submicron-gate MESFET most commonly fabricated in GaAs, and the high electron mobility transistor (HEMT) most commonly fabricated with AlGaAs/GaAs heterojunction, are well established as low noise devices in linear applications at microwave and millimeter wave frequencies. As preamplifiers, they have the lowest noise figure among all presently known three-terminal active devices, and under cryogenic conditions, their noise figure closely approaches that of masers. In addition, these devices have the advantages of wide frequency range of applicability, high DC-to-RF conversion efficiency, high input--output isolation, suitability for monolithic integration, versatility of circuit applications, simple DC biasing needs, low sensitivity to environmental parameters, and sufficient power output for many applications without post-amplification.

Oscillators designed with MESFETs and HEMTs have not exhibited such outstanding low-noise performance at all [1]. The reason for this disappointing performance is well known : These devices generate high levels of low frequency noise, and this baseband noise modulates the oscillator output due to device nonlinearities, producing large noise sidebands around the carrier frequency. The purpose of this paper is to review our present state of understanding of this subject, and to summarize the available information on the physical origin of the low frequency noise, the influence of the various device and operating parameters on the magnitude of this noise, the mechanism whereby this noise is upconverted to the neighborhood of the carrier frequency, the design variables that govern the magnitude of FM noise spectra, and the state-of-the art reached in low noise oscillator performance.

Three limitations to the scope of the following discussions should be pointed out in advance :

(i) Although most of the following considerations apply to all FET oscillators, the discussions tacitly assume that the oscillators are operated at a microwave frequency and employ the currently available state-of-the-art devices having gate lengths of the order of 0.25 μm.

(ii) There is much more information available in the literature concerning the MESFET oscillators than concerning HEMT oscillators which are of more recent vintage. As a result, the subsequent discussions refer only to MESFETs, although there is no reason to expect any different behavior from HEMTs except in the numerical values of some parameters.

(iii) The present discussion will be confined to the FM noise of oscillators, both due to the focus of this symposium, and due to the fact that the AM noise of oscillators employing field-effect transistors is small and has not been a handicap in any significant system applications of these oscillators. The FM noise of the oscillator can be expressed either as the ratio of the FM sideband noise power, in a unit bandwidth at a specified offset frequency f_m from the carrier frequency, to the carrier power, or as the rms value of the corresponding frequency deviation.

II. Physical Sources of Noise in MESFETs

The purpose of this section is to describe, for each of the various sources of noise in a field-effect transistor, the physical mechanism of noise generation, the region of the device in which the generation of noise takes place, the power spectral density of the noise current component contributed by the noise source, and the factors influencing the magnitude of noise. Since devices with a variety of different structures have been fabricated, the generic version of the MESFET structure, schematically shown in Fig. 1, will be used as the basis for the present discussion. The following are the principal sources of noise in this device :

CH2427-3/87/0000-471 $1.00 © 1987 IEEE

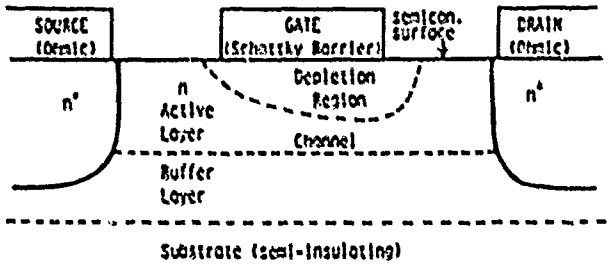


Fig. 1. Schematic Representation of the MESFET Structure.

1. Diffusion (or Velocity Fluctuation) Noise

Diffusion noise arises in the MESFET due to the fact that the majority carriers drifting in the channel in a direction parallel to the electric field have a randomly distributed velocity component in the longitudinal direction due to a finite carrier temperature. The source of this noise is therefore spatially distributed throughout the conducting part of the channel, and its net effect at the device terminals must be found by adding the contributions from all parts of the channel.

The highly conducting source and drain contact regions of the channel, which are outside the region of influence of the gate, contribute essentially a parasitic resistance to the channel; since the electric field in these regions is small and therefore the carriers are in near thermal equilibrium, the diffusion noise in this part of the channel is the same as the thermal noise in the parasitic resistances, and is easily calculated by Nyquist's thermal noise formula. By contrast, the channel region under the gate is a region where the electric field is high, the carrier drift velocity is saturated over some part of the region, and the carriers are far from equilibrium; the power spectral density of the total drain short-circuit noise current at the device terminals due to the diffusion noise arising in this active region alone can be written in a form which mimics the thermal noise formula :

$$S_i(f) = 4\,k\,T\,g_m\,P \qquad (1)$$

The similarity of this expression to Nyquist formula for thermal noise is a little contrived; in this expression, g_m is the transconductance of the device, rather than the conductance of a dissipative two-terminal element, and T is the lattice temperature in the channel region rather than the temperature of the carriers having the velocity fluctuations among them. Finally, P is a correction factor that must be introduced in order to make this expression valid; i.e., Eqn. (1) is only the definition of P, and is not a "result". What make Eqn. (1) useful are the results of actual diffusion noise calculations in MESFET channels from first principles, which show that the correction factor P is of the order of unity, and is only a slowly varying function of the DC bias of the MESFET.

The noise power spectrum in Eqn. (1) is frequency-independent; this is an approximation which is valid for all frequencies small compared to the inverse of the correlation time of the carrier velocity fluctuations. Therefore, diffusion noise may be treated as white noise for all frequencies upto about 10^{11} Hz, above which the noise power spectrum must roll-off. This roll-off is theoretically essential, but has not been experimentally explored.

The most important parameter influencing the magnitude of diffusion noise is obviously the electric field profile in the channel region, since the electric field determines the extent of carrier heating. The diffusion noise is therefore determined by DC bias, channel dimensions, doping, and the rate of carrier scattering (since scattering allows the carriers to loose energy by interaction with the lattice, and thus cools them down).

2. Avalanche Noise

When the electric field produced in the channel is high, the impact ionization due to field-accelerated carriers drifting in the channel causes the generation of secondary carriers. This avalanche generation process, being random, is accompanied by a form of shot noise called avalanche noise. The avalanche generated majority carriers drift along the channel, and contribute an additional component to the drain current. The minority carriers on the other hand can either be collected by the gate terminal, leading to an increase in the gate current, or can drift along the channel, so that the avalanching is caused by carriers of both species. The region of the device where this avalanching is most likely to occur is one where the electric field is the highest; for depletion-mode devices in which the gate voltage is negative, the drain end of the channel region is therefore the principal source of avalanche noise.

The magnitude and the spectrum of the output noise current due to avalanching will depend on whether avalanching is caused by one or both species of carriers. For the special case when both species ionize, the current multiplication factor M is not small, steady-state ionization occurs, and the ionization rate is spatially uniform, the noise power spectrum of short-circuit drain noise current contributed by avalanching can be written as :

$$S_i(f) = \frac{2\,q\,I_p\,M^3}{1 + (\omega M \tau_a / 2)^2} \qquad (2)$$

where I_p is the primary current causing avalanching, and τ_a the avalanche response time. Other expressions must be used in situations where any of the several restrictive conditions for the validity of Eqn. (2) do not hold.

The power spectrum of avalanche-generated noise is white for frequencies small compared to the inverse of the avalanche-response time. For typical device dimensions, the spectrum is therefore white over the microwave frequency range.

Clearly, the factor having the largest influence on the magnitude of the avalanche noise is the electric field reached in the channel, and this, in turn, is influenced by parameters like DC bias, channel length, and doping density.

3. Shot Noise

Shot noise accompanies any current composed of carriers crossing a potential barrier. As the current flow in the channel

in due to conduction, there is no shot noise associated with drain current. But the gate current consists of carriers crossing the Schottky-barrier between the gate and the channel; therefore the gate current can be expected to be accompanied with full shot noise.

The power spectrum of the gate noise current can be written with the help of Schottky's formula as :

$$S_i(f) = 2\,q\,I_G \qquad (3)$$

where q is the magnitude of the charge of an electron, and I_G is the gate current. It is apparent from Schottky's formula that the noise power spectrum is independent of frequency, a consequence of the assumption that the transit-time of the carrier through the barrier is negligible. The shot noise can therefore be taken as white noise for frequencies small compared to the inverse of this transit time, and therefore for all frequencies upto the millimeter wave range.

It also appearent from Schottky's formula that the only parameter determining the magnitude of shot noise is the gate current; therefore the shot noise will be influenced by such factors as the temperature and the doping density in the MESFET active layer. For well-designed MESFETs, and under most reasonable operating conditions, the gate current is small, so that the gate shot noise is very small and can be ignored in many applications. However, there are applications, such as transimpedance amplifiers following optical receivers, in which the shot noise in the gate current is a significant contributor to the device noise performance.

4. 1/f Noise.

Although there is considerable debate concerning the physical origin of 1/f noise, it is generally believed that this noise arises from resistance fluctuations, which are in turn caused either by the fluctuations in the number of carriers or by the fluctuations in mobility. Experimental studies to determine the region of origin of this noise within the device are not entirely unanimous, but 1/f noise is believed to originate primarily at the channel--substrate interface, the semiconductor surfaces, the contacts, and in the active layer itself.

The power spectrum of the drain noise current component due to 1/f noise can be expressed by the empirical formula :

$$S_i(f) = \frac{I_D^2\,\alpha_H}{N\,f^a} \qquad (4)$$

where α_H is an empirical parameter, called Hooge's parameter, N is the number of carriers participating in the flow of the drain curent I_D, and therefore depends both on the carrier density and the volume of the active region of the device under the gate, and the exponent a is also an empirical constant, approximately equal to 1.

The factors most influencial in determining the magnitude of 1/f noise are the quality of the buffer layer between the substrate and the active layer, the surface treatment or passivation, and the density of defects in the active region. In addition, empirical studies have shown that many other factors such as the gate leakage current and the external circuit impedance may also influence the magnitude of 1/f noise, but these results cannot be viewed as definitive because there are also reports of other devices where such effects are not observed.

5. Generation-Recombination (g-r) Noise.

Generation-recombination noise arises in MESFETs due to the trapping and release of carriers by the generation-recombination centers (so called traps) which causes random fluctuations in the free carrier density, and thus contributes a noise component to the current. Since the traps are most effective in carrier trapping and release when the Fermi level is close to the trap energy level, those regions of the device where Fermi level is varying are most likely to contribute to g-r noise. As a result, the g-r noise arises primarily in the space-charge regions in the device, including those (i) at the interface of the channel and the substrate layer, (ii) under the gate in the depletion region, and (iii) at the semiconductor surfaces.

The power spectrum of the noise current contribution due to each type of the traps can be written as :

$$S_i(f) = \frac{C}{1 + \omega^2\,\tau_l^2} \qquad (5)$$

where τ_l is the lifetime of the carriers in the trap. The constant C is directly proportional to the trap density , and also depends on the trap energy level, and the applied voltages which influence the Fermi level. The power spectrum is therefore of the form of the frequency response of a single-pole low-pass filter, often called the "Lorentzian" spectrum. It is therefore white at frequencies small compared to $1/\tau_l$, and decreases with frequency at higher frequencies at the rate of 20 dB/decade.

Typically, several types of traps will be present simultaneously, and therefore the total g-r noise spectrum will be a superposition of several spectra of the form of Eqn. (5), each having its own magnitude and corner frequency. It is apparent that the total g-r noise spectrum will not have the simple frequency dependence expressed in Eqn. (5). Indeed, if the trap lifetimes are distributed over a wide range of values, extensive and careful measurements are required to separate g-r noise from 1/f noise, and to distinguish the contribution of individual types of traps.

III. Noise Model of the Device

The individual components of the power spectral density of the drain noise current due to the above mentioned noise sources are shown together in Fig. 2. The figure is only illustrative, since the relative magnitudes and the corner frequencies of the various noise sources will be different in different types of devices, and for a given device, with vary with DC bias. Shot noise is not included in this figure, since it is small compared to the the noise contributions shown.

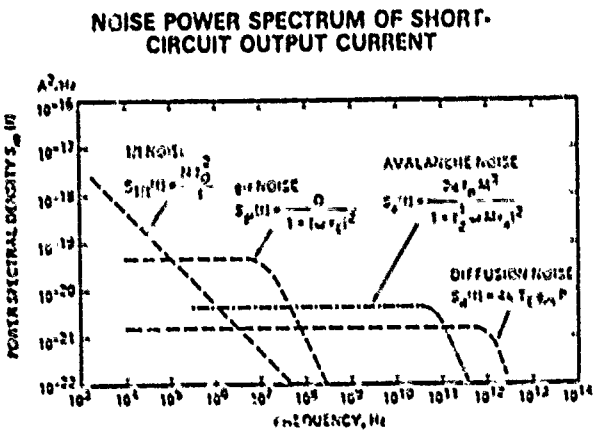


Fig. 2. Power Spectral Density of the Various Components of the Drain Noise Current.

Two conclusions can be drawn from this figure. First, the most significant noise sources at low frequencies (1/f noise and g-r noise) are different from those at high frequencies (diffusion and avalanche noise). Second, the noise produced in the device can be significantly larger at low freqeuncies, typically below 1 MHz, than that at higher frequencies. In nonlinear applications of the device, such as in oscillators, the low frequency noise can modulate signals and generate noise at higher frequencies, thereby deteriorating the noise performance. These two observations explain why the MESFETs are a low-noise device in linear amplifier applications, and yet have a poor noise performance in oscillator circuits.

In linear circuit applications, a MESFET can be represented by a linear (small-signal) circuit model at the desired DC bias conditions, and a noisy linear twoport can always be represented by a noiseless version of the same twoport along with two noise sources, which are partially correlated in general. This is the so called Rothe-Dahlke representation of linear noisy twoports. For nonlinear circuits, there are no such rigorous equivalent-circuit representations of guaranteed validity, and a noise model must be found by inductive methods.

Since quasi-linear models of MESFETs have had some success in describing the operation of MESFET oscillators, one possible model to consider is a Rothe-Dahlke type model. Such a model is shown in Fig. 3 (a), in which the MESFET is represented by a very simple signal model, consisting of just four noiseless circuit elements : the total input resistance R_T, the gate-to-source capacitance C_{gs}, the output resistance R_o, and the voltage-controlled current source $g_m v$, controlled by the voltage v across C_{gs}, where g_m is the transconductance of the MESFET. To this are added the two noise sources e_n and i_o at the input and the output respectively, which will be partially correlated in general.

When the equivalent circuit elements are nonlinear, the foregoing simplified noise equivalent-circuit model is still too complex, due to the presence of two separate noise sources. All MESFET oscillator noise analyses to date have therefore introduced further simplification in the noise model by transforming all noise sources inherent in the device to only one of the two noise sources e_n and i_o, and setting the other source to zero. The circuit models of Fig. 3 (b) and (c) show these two possibilities. The first of these models, in Fig. 3(b), is useful in characterization of the baseband noise generation in the device by experimental measurements, and allows the device noise measurements to be carried out at the output (drain-source) port of the device in the form of output current fluctuations. The second model, in Fig. 3 (c) is more convenient for calculating the oscillator noise spectra, as will be seen shortly.

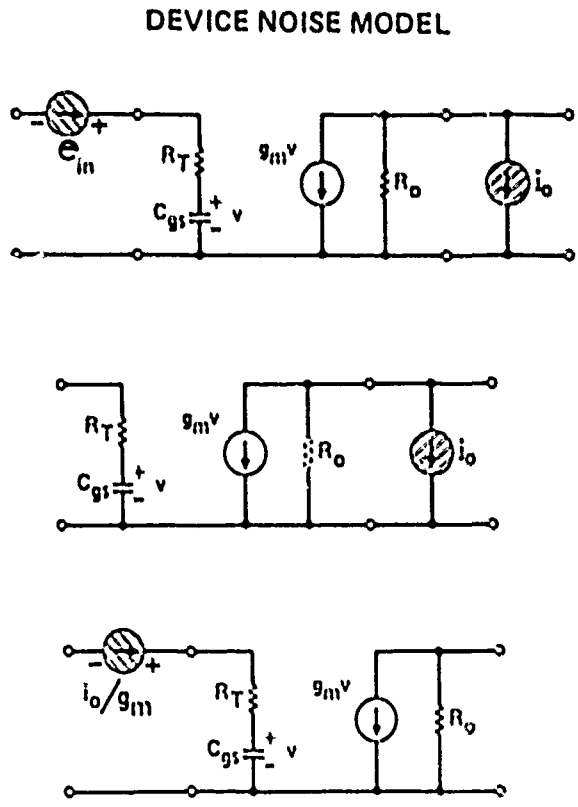


Fig. 3. Noise Equivalent Circuit Models for a MESFET.

The power spectral density of the noise source e_n can b measured in the laboratory as follows. With the device D biased at the desired operating point, the input (gate-to-source port is short-circuited at the frequency of measurement, and th output noise power is measured by a narrow-band low-nois receiver. The power spectral density $S_i(f)$ of the output noise current source is thus found. This can be transformed to th input port as a voltage through the intervening circuit model of the device, treated as a linear circuit. At low frequencies, whe $\omega C_{gs} R_T << 1$, the transformation involves only the devic transconductance, so that the spectral density of the noise vo tage source e_n is given by :

$$S_e(f) = S_i(f) / g_m^2 \qquad (1$$

IV. Noise Model of the Oscillator

Many different kinds of oscillator circuits, based on different device configurations, feedback arrangements, resonator connections, and load coupling, have been designed. All such oscillators employing only one MESFET device as the active device can be represented by the circuit model of Fig. 4 (a). The noise spectrum of such a noisy oscillator can be analyzed in one of two different ways :

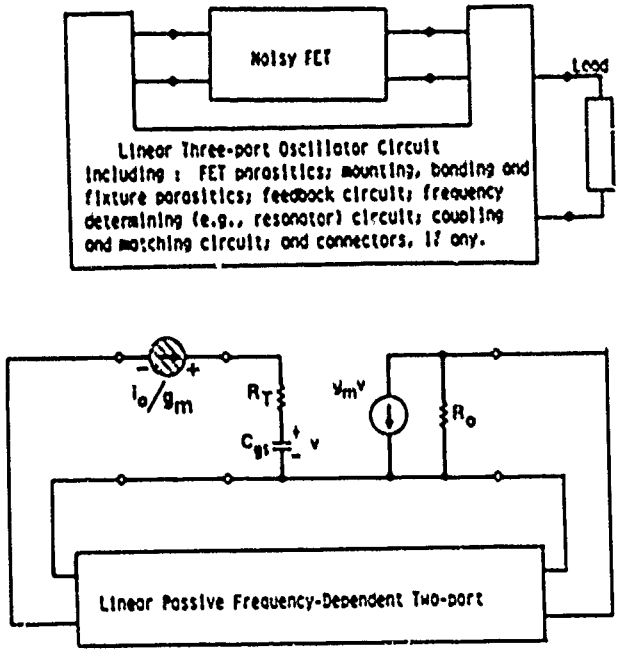


Fig. 4. Circuit Representations of a Noisy FET Oscillator.

1. As an active one-port (negative-resistance) device oscillator

As seen from the load port and looking back towards the oscillator, the MESFET along with its feedback circuit must appear as an active one-port device. More generally, since a field-effect transistor is a two-port active device, the oscillator circuits must use feedback to induce oscillations, so that one port of the device is used for introducing feedback, and the other port serves to deliver the output oscillations to the remainder of the circuit and load. Therefore, the transistor together with its feedback is reduced to a one-port active network, which contains all of the nonlinearity in the oscillator circuit, while the remainder of the circuit is reduced to a linear, passive one-port "load". The active oneport including the device can be represented as a frequency and signal-amplitude dependent large-signal impedance $Z(\omega,V_{RF})$ having a negative real part, which is connected across the load (including any other linear circuit elements), having a frequency dependent circuit impedance $Z_c(\omega)$.

The calculation of the noise spectra of such negative-resistance oscillators has been well-established for a long time, and has been summarized by Kurokawa [2]. In order to use the results of Kurokawa, it is only necessary to transform the noise source e_n present in the device to the terminals of the impedance $Z(\omega,V_{RF})$. To a first approximation, the MESFET may be treated as a linear circuit in carrying out this transformation, and the power spectrum of the new noise source e_t is thus determined. For the low frequency noise sources of interest here, it may be possible to simplify the transformation considerably. The oscillator noise spectra are thus found in terms of the impedance of the one-port device and circuit, their sensitivities to frequency and signal amplitude, and the power spectrum of the noise source connected between Z and Z_c.

The one-port model can be used to represent many different types of oscillator circuits, merely by a suitable choice of a reference plane dividing the oscillator circuit into two parts : an active device and a load. For example, Joshi and Debney [3] designed a common-source FET oscillator, and used the drain--ground port as the dividing plane, while Sechi and Brown [4] designed a common-drain FET oscillator and used the gate--drain port as the dividing plane. If the port used as dividing plane is externally accessible, so that the "device" and "load" impedances can be experimentally measured at this port looking on the two sides, an optimization of the oscillator design can be carried out for minimization of noise spectra. Sechi and Brown [4] show the results of such measurements.

2. As a feedback oscillator using a twoport active device.

Alternatively, the oscillator may be represented as in Fig. 4 (b), wherein the oscillator continues to be treated as a feedback circuit around a twoport active device, which is perturbed by the presence of the noise source e_n. It is clear that if the MESFET model was strictly linear, the frequency at which the condition of oscillations is satisfied in this oscillator circuit would not be influenced by the presence of the noise voltage e_n. The nonlinearity of the MESFET allows the noise voltage e_n to influence the oscillation frequency f_o and thereby produce FM noise sidebands.

Perhaps the simplest method of accounting for the nonlinearity of the device is to allow some of the elements in the equivalent circuit of the device to be nonlinear. The principal nonlinear elements in the MESFET are the three elements C_{gs}, g_m, and R_o, and each of these is influenced by e_n in general. The oscillator noise spectra can therefore be calculated in terms of the oscillator circuit elements, the nonlinearity of the above three elements, and the power spectrum of the noise source e_n in the circuit. Such an analysis is presented by Siweris and Schiek [5].

For the present purposes, the FM noise spectrum of the oscillator will be expressed in less detail, in terms of some composite parameters, rather than in terms of the individual circuit elements of a particular implementation of the oscillator circuit. One of the conclusions from the detailed analysis of Siweris and Scheik [5] is that, of the three equivalent circuit elements, only C_{gs} influences f_o to a first order. Therefore, the calculation of FM noise spectra is reduced to the determination of the sensitivity of f_o to e_n via C_{gs}. In the following, this observation allows the identification of the factors that influence the FM noise spectrum.

V. Factors Influencing the FM Noise Spectrum

The small change in the frequency of oscillations due to the noise voltage e_n can be expressed as follows :

$$\Delta f = \frac{\partial f_o}{\partial e_n} e_n \tag{7}$$

Therefore the FM noise spectrum, expressed in rms frequency deviation at an offset frequency f_m away from the carrier frequency f_o, can be written as :

$$\left[\Delta f_{rms}(f)\right]^2 = \left|\frac{\partial f_o}{\partial e_n}\right|^2 S_e(f_m) \tag{8}$$

The frequency sensitivity to noise voltage can, in turn, be expressed in terms of the factors that influence it :

$$\left|\frac{\partial f_o}{e_n}\right| = \left|\frac{\partial f_o}{\partial C_{gs}}\right| \cdot \left|\frac{\partial C_{gs}}{\partial e_n}\right| \tag{9}$$

Since the gate-to-source capacitance C_{gs} is the capacitance of a Schottky-barrier metal-semiconductor junction, its dependence on the gate-to-source voltage is known from the following equation applicable to abrupt junctions :

$$C_{gs} = \frac{C_o}{\sqrt{1 - v_{gs}/\phi_{bi}}} \tag{10}$$

Therefore, Eqn. (9) can be written as

$$\left|\frac{\partial f_o}{e_n}\right| = \left|\frac{\partial f_o}{\partial C_{gs}}\right| \cdot \left|\frac{\partial C_{gs}}{\partial v_{gs}}\right| \cdot \left|\frac{\partial v_{gs}}{\partial e_n}\right| \tag{11}$$

The set of equations (8) and (11) places into evidence all of the factors that influence the FM noise spectrum of a MESFET oscillator. Furthermore, all of the methods of FM noise reduction that have been proposed can now be placed in perspective.

There are four groups of parameters that are instrumental in determining the FM noise of the MESFET oscillator :

(1) The power spectrum $S_e(f_m)$ of the baseband noise generated in the device, and referred to the gate voltage, at the offset frequency f_m. One obvious method of FM noise reduction is therefore to employ a MESFET with a lower level of 1/f noise or g-r noise, whichever is dominant at the frequency equal to the offset frequency of interest.

(2) The sensitivity of the oscillation frequency to the MESFET gate-to source capacitance, $|\partial f_o/\partial C_{gs}|$. This is dependent on the frequency-determining RF circuit of the oscillator. Thus the choice of the FET configuration, feedback circuit, and resonator coupling should be made to minimize this sensitivity in order to reduce FM noise.

(3) The nonlinearity of the device input capacitance, represented by $|\partial C_{gs}/\partial v_{gs}|$. This is an intrinsic property of the gate junction, and its intrinsic value is determined by the nature of the doping profile. Its effective value can be reduced by diluting the effect of C_{gs} (item 2 above), or diluting the effect of v_{gs} (item 4 below), by introducing other reactive and voltage-dividing circuit elements respectively.

(4) The sensitivity $|\partial v_{gs}/\partial e_n|$. This can be influenced by the low-frequency impedances in the DC gate bias circuit.

The optimization of the FM noise spectrum of the oscillator can be carried out experimentally if each of the above parameters can be measured, and if the effect of the design variables on the values of these parameters can be estimated. It is also possible to measure the two parameters appearing in Eqn. (8) directly. The baseband noise generated in the device at low frquencies can be measured by a low-noise amplifier and a wave analyzer. The oscillator frequency sensitivity to gate noise voltage can be measured by introducing a signal at the gate and measuring the change in oscillator frequency [6]. Thus the device and the circuit can each be selected for minimum FM noise.

VI. Device Selection

The device specifications in the manufacturer's data sheet for a MESFET or HEMT typically include small-signal S-parameters, broadband linear equivalent circuit parameters, DC characteristics, and some large-signal parameters, such as 1 dB gain compression point. The various procedures found in the literature for designing oscillators with these devices are therefore also based on these and other parameters in terms of which the device is characterized. By contrast, the device noise specifications supplied in the data sheets usually consist of the minimum noise figure of the device as a linear amplifier at some microwave frequencies. This information is of no use to the circuit designer wishing to design a low-noise oscillator; given two devices, the one with the lower noise figure may well lead to an oscillator with a higher noise spectrum close to the carrier frequency. The analysis carried out above has identified a small set of parameters which can be conveniently measured for a given device, serve as a specification of the device noise capability in oscillator applications, are useful for comparing two devices in respect of their noise performance as oscillator, can be used to estimate the magnitude of the expected oscillator noise, and thus assist in optimizing the oscillator design for low noise performance.

A number of circuit techniques [7-11] have been proposed in the literature in recent years for lowering the noise in FET oscillators. Each of these techniques, when viewed in isolation, appears as an independent, clever idea. A general analysis of the FM noise in MESFET oscillators, given above in a form which is independent of the specific circuit implementation, has the further advantage that it provides a systematic classification of the various circuit techniques of oscillator noise reduction based on the variable through which the reduction is achieved, so as to put the various methods in proper perspective and understand the relationship between them, and to estimate the level of noise reduction achievable by each technique.

REFERENCES

[1] R. A. Pucel, "The GaAs FET Oscillator -- Its Signal and Noise Performance," *Proc. 40th Annual Frequency Control Symposium*, pp. 385-391, 1986.

[2] K. Kurokawa, "Injection Locking of Microwave Solid-State Oscillators," *Proc. IEEE*, vol. 61, no. 10, October 1973.

[3] B. T. Debney and J. S. Joshi, "A Theory of Noise in GaAs FET Microwave Oscillators and Its Experimental Verification," *IEEE Trans. Electron Devices*, vol. ED-30, no. 7, pp. 769-776, July 1983.

[4] F. N. Sechi and J. E. Brown, "Ku-Band FET Oscillator," *International Solid State Circuits Conference Digest*, pp. 124-125, 267, February 1980.

[5] H. J. Siweris and B. Schick, "Analysis of Noise Upconversion in Microwave FET Oscillators," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-33, no. 3, pp. 233-242, March 1985.

[6] H. Rohdin, C.-Y. Su, and C. Stolte, "A Study of the Relation Between Device Low-Frequency Noise and Oscillator Phase Noise for GaAs MESFETs," *1984 IEEE-MTTS International Microwave Symposium Digest*, pp. 267-269, San Francisco, Calif., May-June 1984.

[7] A. N. Riddle and R. J. Trew, "A New Method of Reducing Phase Noise in GaAs FET Oscillators," *1984 IEEE -- SMTT International Microwave Symposium Digest*, San Francisco, Calif., May - June 1984, pp. 274-276.

[8] Z. Galani, M. J. Bianchini, R. C. Waterman, Jr., R. Dibiase, R. W. Laton, and J. B. Cole, "Analysis and Design of a Single-Resonator GaAs FET Oscillator with Noise Degeneration," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-32, no. 12, pp. 1556-1565, December, 1984.

[9] A. P. S. Khanna, J. Obregon, and Y. Garault, "Efficient Low-Noise Three-Port X-Band FET Oscillator Using Two Dielectric Resonators," *1982 IEEE -- SMTT International Microwave Symposium Digest*, Dallas, Tex., June 1982, pp. 277-279.

[10] G. Lan, D. Kalokitis, E. Mykietyn, E. Hoffman, and F. Sechi, "Highly Stabilized, Ultra-Low Noise FET Oscillator with Dielectric Resonator," *1986 IEEE -- SMTT International Microwave Symposium Digest*, Baltimore, MD, June 1986, pp. 83-86.

[11] M. Prigent and J. Obregon, "Phase Noise Reduction in FET Oscillators by Low-Frequency Loading and Feedback Circuitry Optimization," *IEEE Trans. Microwave Theory and Techniques*, vol. MTT-35, no. 3, pp. 349-352, March 1987.

41st Annual Frequency Control Symposium - 1987

REVIEW OF DIELECTRIC RESONATOR OSCILLATOR TECHNOLOGY

A. P. S. Khanna
Modular Components & Subassemblies Division
Avantek Inc.
3175 Bowers Avenue
Santa Clara, CA 95054

Oscillators represent the basic microwave energy source for all microwave systems such as radars, communications, navigation or electronic warfare. A typical microwave oscillator consists of an active device (a diode or a transistor) and a passive frequency-determining resonant element. With the rapid advancement of technology, there has been an increasing need for better performance. The emphasis has been on low noise, small size, low cost, high efficiency, high temperature stability and reliability. Dielectric resonator oscillators offer the system designer a viable alternative in an effort to meet these challenges[1]. This paper will introduce the fundamentals of the transistor dielectric resonator oscillators, discuss various oscillator configurations, their performance, special circuits, limitations and future trends.

The quartz crystal oscillator is a highly stable source, however its operation is limited to less than 100 MHz. Stable microwave sources have commonly been realized using frequency multiplication of a the output of a quartz crystal oscillator. This method increases the FM noise power by N, where N is the multiplication factor, and has very low efficiency in addition to being very complex and expensive. Stable signals have also been generated in the past using metallic high Q cavities in passive cavity stabilization systems or in complex and bulky frequency discriminator systems.

The origin of solid state microwave oscillators using Gunn and Impatt diodes dates back to the late 1960s, before which microwave frequency generation depended upon klystron or magnetron tubes requiring massive power supplies. In less than two decades, solid state oscillators have evolved significantly. The extension of the bipolar transistor oscillator to microwave frequencies and the development of GaAs MESFET devices in the early 1970s has made available highly cost-effective, miniature, reliable and low-noise sources for use in the microwave and millimeter frequency ranges.

The basic active elements which can be used for microwave solid state oscillators are Gunn diodes, Impatt diodes and transistors. While Gunn oscillators have the advantage of low FM noise compared to Impatt oscillators, the latter has a higher efficiency and is a higher power device compared to the Gunn. Transistor oscillators on the other hand are low noise as well as high efficiency sources. Compared to Gunn oscillators, the transistor oscillators do not have the problems of threshold current, the necessity for heat sinking and the tendency to lock at spurious frequencies.

Gunn and Impatt diodes are negative-resistance devices, requiring only the application of a D.C. bias. Using them, the design of the oscillator is limited to the design of the output matching circuit in order to deliver the desired power output. Application of D.C. bias to the bipolar or the field-effect transistor, on the other hand, is not a sufficient condition for oscillation. Suitable series/parallel feedback is required to induce the negative resistance. The frequency range over which the negative resistance is present in the diodes is determined by the physical mechanisms in the device, while in the case of transistors this frequency range is also influenced by the chosen circuit topology. The only disadvantage to the transistor at present is a limit on the maximum oscillation frequency. While a Gunn or Impatt oscillator is capable of delivering 100 mW up to 100 GHz, the transistor oscillators are presently limited to approximately 10 mW at 40 GHz.

Transistor oscillators can be realized using either bipolar or GaAs FET devices. The maximum oscillation frequency for bipolar transistor oscillators is lower than that of GaAs FET oscillators. GaAs FET oscillators have been reported up to 100 GHz while oscillators using bipolar transistors have reached 20 GHz. However, the bipolar transistor offers lower phase noise: typically a bipolar oscillator has 6 to 10 dB less FM noise, (very close to the carrier) compared to a GaAs FET oscillator operating at the same frequency.

Dielectric resonators, due to their high Q, small size and excellent integrability in MIC circuits, can be directly used as the frequency determining element for realizing a MIC transistor oscillator. With the recent advent of temperature stable dielectric material, the transistor DRO is fast-becoming the most desirable choice in a vast number of applications.

The primary characteristics of the ceramic material to be used for dielectric reconators are:

- The Q factor, which is approximately equal to the inverse of the loss tangent.
- The temperature coefficient of the resonant frequency, τ_f, which includes the combined effects of the temperature coefficient of the dielectric constant and the thermal expansion of the dielectric resonator and the shielding package.
- The dielectric constant, ϵ_r.

The Q, τ_f and ϵ_r values required for various applications differ and, in general, satisfactory oscillator operation under most conditions can be achieved by choosing an appropriate material composition. Until several years ago, the lack of suitable materials (i.e., possessing acceptable combinations of Q, τ_f and ϵ_r) severely limited dielectric resonator applications. Materials such as TiO_2 (rutile phase), which has an unloaded Q of about 10000 at 4 GHz, and ϵ_r of 100 were most often used. However, TiO_2 has a T_f value of 400 ppm/°C which makes it impractical for most applications.

The development of temperature-stable dielectric resonators dates back less than a decade. A number of material compositions have been explored in attempts to develop suitable dielectric materials, including ceramic mixtures containing TiO_2, various Titanates and Zirconates, glass, ceramic systems and alumina-based ceramics. At present, several

CH2427-3/87/0000-478 $1.00 C 1987 IEEE

ceramic compositions have been developed offering excellent dielectric properties. Complex perouskite compounds with the general formula A $(B'_{1/3} B''_{2/3}) O_3$ (where A = Ba, Sr; B' = Zn, Mg, Co, Ni ; B" = Ta, Nb) have proved to possess acceptable properties for dielectric resonators. These compounds have dielectric constants between 20 and 40, a high quality factor, some greater than 10000 @ 10GHz, and a temperature coefficient which is variable through modification of the composition. Table I compares the important properties of different materials developed commercially.

TABLE I

DIELECTRIC RESONATOR MATERIALS

COMPOSITION	DIELECTRIC CONSTANT	Q	TEMPERATURE COEFFICIENT	FREQ. RANGE	MANUFACTURER
$Ba_2 Ti_9 O_{20}$	40	10,000 @ 4 GHz	+2	1 To 100 GHz	Bell Labs
$(Z_r - S_n) Ti O_4$	38	10,000 @ 4 GHz	-4 To 10 Adj	1 To 100 Ghz	Trans Tech Thomson, Murata
Ba (Zn 1/3 Ta 2/3) O_2	30	10,000 @ 10 GHz	0 to 10	4 To 100 Ghz	Murata
Ba (Mg 1/3 Ta 2/3) O_2	25	25,000 @ 10 GHz	4	4 To 100 GHz	Sumimoto
Ba O - PbO - $Nd_2 O_3$ - Ti O_2	88	5,000 @ 1 GHz	0 to 6	< 4 GHz	Murata/Trans Tech
$Al_2 O_3$	11	50,000 @ 10 GHz	0 to 6	> 18 GHz?	NTK/Trans Tech

Whether any of the dielectric compositions shown in Table I have overall superiority is not clear, since other factors, such as ease of ceramic processing and ability to hold tolerances on the dielectric properties must also be considered. The performance limitations, if any, of the lower dielectric constant materials, remain to be determined, since most component work reported thus far has used dielectric resonators with ϵ_r in the range of 37-100. The lower dielectric constant material is likely to be more sensitive to shielding, due to the increase in fields outside the resonator.

Resonant Frequency

The resonant frequency of a dielectric resonator is determined both by its dimensions and its surroundings. Although the geometrical form of a dielectric resonator is extremely simple, an exact solution of the Maxwell equations is considerably more difficult than for the hollow metallic cavity. For this reason, the exact resonant frequency of a certain resonant mode, such as $TE_{01\delta}$, can only be computed by rigorous numerical procedures. A number of theories on the subject, which can predict resonant frequency to an accuracy of $\pm$ 1% for the given configuration, appear in the literature. Unfortunately, these methods call for the use of high-powered computers. Kajfez[1] has presented an approximate solution of the involved equations both for the isolated case and for the more commonly-used MIC configuration. This method is typically accurate to $\pm$2%.

Currently-practical dielectric resonators cover a frequency range of 1 to 100 GHz. The lower frequency limit is imposed by the resulting large size of the resonator, while the upper frequency is limited by the reduced Q of small resonators, as well as by dimensions that become too small to effectively couple with the transmission line.

To a first approximation, a dielectric resonator is the dual of a metallic cavity. The radiation losses of the dielectric resonators with the commonly used permittivities, however, are generally much greater than the energy losses in the metallic cavities, which makes proper shielding of the dielectric resonator a necessity. The dimensions of a dielectric resonator are also considerably smaller than those of an empty metallic cavity resonant at the same frequency by a factor of approximately $1/\epsilon_r$. If ϵ_r is high, the electric and magnetic fields are confined in region near the resonator, which results in small radiation losses. The unloaded quality factor Q_u is thus limited by the losses in the dielectric resonator.

The shape of a dielectric resonator is usually a solid cylinder, but tubular, spherical and parallelopiped shapes are also used. The commonly-used resonant mode in cylindrical resonators is denoted by $TE_{01\delta}$. In the $TE_{01\delta}$ mode, magnetic field lines are contained in the meridian plane while the electric field lines are concentric circles around z-axis as shown in Fig. 1. For a distant observer, this mode appears as a magnetic dipole, and for this reason it is sometimes referred-to as the "magnetic dipole mode." When the relative dielectric constant is around 40, more than 95% of the stored electric energy, and more than 60% of the stored magnetic energy, is located within the cylinder. The remaining energy is distributed in the air around the resonator, decaying rapidly with distance away from the resonator surface.

To effectively use dielectric resonators in microwave circuits, it is necessary to have an accurate knowledge of the coupling between the resonator and different transmission lines. The $TE_{01\delta}$ mode of the cylindrical resonator can be easily coupled to a microstrip line, fin line, magnetic loop, or to a metallic or dielectric waveguide[1]. Figure 2 shows the magnetic coupling between a dielectric resonator and microstrip line. The resonator is placed on top of the microstrip substrate. The lateral distance between the resonator and the microstrip conductor primarily determines the amount of coupling between the resonator and the microstrip transmission line[2]. This has a direct effect upon output power, frequency stability, and harmonic content, as well as resonant frequency. Proper metallic shielding, required to minimize the radiation losses (hence to increase Q) also affects the resonant frequency of the $TE_{01\delta}$ mode. Figure 3 shows the equivalent circuit of the dielectric resonator coupled to the line.

DRO Circuits and Comparisons

There are two means of incorporating a dielectric resonator in a microstrip circuit: as a passive stabilization element (stabilized DRO), or as a circuit element in a frequency determining network (stable DRO).

A stabilized DRO is a device which utilizes a dielectric resonator in the output plane of a DRO circuitry to stabilize an otherwise free running oscillator (Fig. 4). This approach has several disadvantages including a tendency toward mode jumping, frequency hysteresis problems, higher insertion loss due to the resonator being coupled to the output circuitry and increased output power variation.

The stable DRO configuration, which uses the dielectric resonator as a feedback/frequency determining element, is the most commonly used, having greater efficiency, simpler construction and more resistance to mode jumping and hysteresis effects than the stabilized DRO.

To realize a stable DRO, the resonator may be used as either a series or parallel feedback element in the frequency determining circuit. Figure 5 shows two common configurations of each type[1]. An advantage of the series feedback design is the relative ease of coupling to a single line, compared to the parallel circuit's requirement for double line coupling. In addition, the two coupling coefficients in the parallel case are not independent, increasing the difficulty of alignment. With the parallel circuit, however, the use of a high gain amplifier can allow significant decoupling of the resonator from the microstrip lines, resulting in a higher loaded Q factor with associated reduction in phase noise.

Electrical Performance Parameters

Center Frequency and Power Output: As shown in Figure 6, Transistor DRO's are available spanning the frequency range of 3 GHz to 40 GHz with power outputs ranging to greater than +25 dBm. As noted earlier, the oscillator can be either Bipolar or GaAsFET, each with associated tradeoofsin performance, and the oscillator can be followed by one or mroe buffer amplifier stages as required to meet power output specifications. The center frequency is usually specified in MHz, to the MHz, with an associated stability (temperature dependent) or accuracy (temperature, pulling and pushing dependant) specification.

Frequency Stability is the measure of change in frequency over the specified operating temperature range. Typically, this has been expressed in parts per million per degree centigrade (ppm/°C). As the actual frequency change versus temperature is not a linear function about a reference of +25°C, other methods of specifying stability are preferred. One worthwhile method is to specify a percent change in frequency over a temperature range, another is to specify parts per million over the temperature range, referred to room temperature.

The principal cause of the DRO frequency drift with temperature the phase deviation between the resonant circuit and the active circuit, including device, feedback and output circuitry. Using the oscillation condition in the reflection coefficient form, it can be easily proved that the temperature coefficient of DRO frequency is a function of the following parameters:

- The temperature coefficient, τ_f, of the dielectric resonator placed in a shielded MIC configuration.
- The unloaded Q of the dielectric resonator.
- The coupling coefficient of the D.R. with the microstrip line.
- The temperature coefficient (τ_{ph}) of the device (transistor) input reflection coefficient phase that is known to decrease linearly with temperature.

In order to achieve a temperature-compensated DRO, a resonator with a temperature coefficient of +1 to +4 ppm/°C is generally required to offset the negative temperature coefficient of the device phase temperature coefficient. Figure 7 shows some of the typical frequency drift curves ove temperature for free-running DROs.

With present technology it is now possible to repeatably produce DROs with frequency drift of less than ± 100 ppm over the -55 to +85 °C military temperature range, at frequencies up to 40 GHz. However, as shown in Fig. 7, results reported in R & D environment are significantly better (± 10 ppm over temperature)[3].

Certain system applications require highe temperature stabilities than the ones achievable using free running DROs. A number of techniques used to improve the temperature stability are:

The Digitally Compensated DRO (DC-DRO), in which a temperature sensor is mounted in the oscillator to detect the temperature changes. Using EPROMs, pre-programmed with temperature characteristics of the DRO and a look-up table in the ROM, A/D & D/A convertors provide the necessary correction signal, which is applied to the varactor bias of a electronically-tuned DRO. Using this technique frequency stability of ± 15 ppm can be obtained over temperature[4].

The Analog-Compensated DRO (AC-DRO), which uses less-complex circuitry, consisting of an analog compensator circuit in conjunction with temperature sensor, to achieve up to ±20 ppm frequency stability. Each oscillator is characterized by generating a custom tuning voltage-vs.-temperature curve required to maintain a constant frequency. The compensation circuit is aligned to fit the tuning voltage vs. temperature curve of the specific oscillator. The correction voltage from the output circuit is applied to the tuning varactor of the ET-DRO, thus maintaining a constant frequency[4].

Ovenization is also used to enhance the temperature stability of a DRO. To achieve temperature stability, the oscillator package is inserted in a temperature-stabilized oven. Using a heater element, a quick-response thermistor and associated control circuitry, the package temperature can be

maintained within ±5 oC at 5 to 10 degrees above the maximum ambient temperature. A total frequency stability of better than ±10 ppm can be obtained using this approach.

Ovenized DROs offer better phase noise compared to the AC-DRO and DC-DRO because the DRO does not need to incorporate electrical tuning circuitry. Analog- and digitally-compensated DROs, however, are smaller in size and do not need the substantial amount of heater power required by the ovenized DRO.

Phase Locked DROs (PL-DRO) and Injection Locked DROs (IL-DRO) are used when the requisite frequency stability and phase noise cannot be achieved using the previously-described approaches. A PL-DRO or IL-DRO approach also becomes necessary when multiple oscillators are required to be phase or frequency coherentor both. For locked systems, a highly stable crystal-controlled signal source operating at HF or VHF is used as a reference oscillator.

In injection locking, a VHF power amplifier driving a step-recovery diode is used to generate a wideband harmonic comb, which includes the required locking frequency. A bandpass filter is used to select the frequency of interest, and a free-running DRO is locked to the harmonic of the reference source through the injection locking circuit shown in Fig. 8. The main requirement in this case is to make sure that the DRO frequency drift under all operating conditions is less than the injection locking bandwidth, Δf. This bandwidth is a function of the injection power, oscillator output power, and external Q.

Figure 9 shows a typical phase-locked DRO circuit. A DC-coupled sampler/phase detector is used to mix the nearest harmonic of the amplified crystal oscillator with the incoming frequency from the DRO. If the difference frequency is small enough, the loop will be driven towards a point where the difference frequency out of the sampler becomes zero. The loop then drives the DRO towards a zero phase error condition. A search mechanism is generally included in the system so that the loop will be forced to tune through a stable lock point if the initial difference frequency is too large for capture to occur. An AC-coupled phase-locked DRO circuit is used when the output frequency of the DRO is not harmonically related to the reference oscillator frequency.

Injection locking is simpler and less expensive approach compared to phase locking, but the RF output is more likely to contain spurious signals at the harmonics of the reference oscillator frequency. Injection locking, in reality, is frequency locking as opposed to phase locking.

Single Sideband Phase Noise

Phase noise can be defined as short-term frequency stability, characterized by frequency variations in the output frequency which appear as FM energy around the carrier frequency. Phase noise is specified in dBc/Hz measured at specified offests from the carrier frequency: typical offsets are 10 and 100 kHz. In a DRO, phase noise is primarily dependent on four factors:

1. The low frequency noise inherent in the active device.
2. F_c, defined as the upconversion factor, a measure of the efficiency in the conversion of the low frequency noise to the phase noise of the microwave oscillator.
3. The loaded Q factor of the dielectric resonator.
4. The output power (or external Q) of the oscillator.

Optimization of phase noise performance calls for special design considerations, as well as the use of a high-Q dielectric resonator and proper device selection. It has been proven that the low-frequency noise in a GaAs FET is inversely proportional to the gate length and width of the device. Figure 10 compares the phase noise performance of a number of different oscillators.

The following techniques can also be used to further reduce phase noise:

Low frequency feedback, using a parallel feedback circuit designed at low frequencies (up to 1 MHz) in order to reduce upconversion of the low frequency noise. Phase noise improvement of up to 20 dB has been reported using this technique[5].

Noise degeneration, using the same dielectric resonator both as the frequency determining element of the oscillator and the dispersive element of the frequency discriminator. The DC output of the discriminator is applied to the frequency control point of the DRO. This technique has been reported to achieve phase noise as low as -120 dBc @ 10 KHz at 10 GHz from the carrier[6].

Mechanical Tuning

The frequency of oscillation of the dielectric resonator is dependant on a number of factors, not the least of which is its proximity to the ground plane. To take advantage of this, a tuning screw can be installed in the top cover directly above the resonator which, by reducing the distance between the resonator and the apparent ground plane, will provide for a certain amount of change in the resonant frequency (Fig. 11). The reason for such behavior can be found in the cavity perturbation theory. Namely, when a metal wall of a resonant cavity is moved inward, the resonant frequency will decrease if the stored energy is predominantly in the electric field. Otherwise, when the stored energy close to the magnetic walls is mostly magnetic, as is the case for the shielded TE_{01} dielectric resonator considered here, the resonant frequency will increase when the wall moves inward. Current designs allow for up to 5 percent tuning range without degradation of other performance parameters. A properly-designed mechanical tuning option will provide a maximum of tuning range while still maintaining hermeticity and reliability, and will not appreciably affect the resonator Q factor (apparent as a degradation in noise and power performance) or temperature stability.

Electronic Tuning

The YIG-tuned oscillator (YTO) and varactor-tuned

oscillator (VCO) are the commonly used electronically-tunable oscillators. YTO's are selected for high tuning linearity, large tuning bandwidth (>2 octaves) and low phase noise. Their disadvantages are low tuning speed, cumbersome mechanical size, considerable weight and significant tuning power consumption. VCO's, on the other hand, offer small size and weight, high tuning speed, low tuning power, but have relatively poor tuning linearity, inferior phase noise and generally lower bandwidth (< one octave). For some applications, however, such as FMCW radar sources, narrowband-modulated communication systems or PLL systems, bandwidth needs are on the order of 0.1 to 1% ; and these applications require sources with low phase noise, high tuning speed and low tuning power. Electronically tunable DRO's (ET-DRO) can now meet the requirements for many of these applications.

ET-DRO's are also commonly used for analog/digital temperature compensation of the oscillator. This application requires that the frequency tuning of the DRO should exceed the frequency drift of the oscillator under *any* combination of operating conditions (temperature, load and bias variations). Various means are used to electrically tune the DRO[1], including ferrite tuning, optical tuning, and the more popular varactor and bias tuning.

Varactor tuning (a typical scheme shown in Fig. 12) can provide up to 1% frequency adjustment. The dielectric resonator is coupled to another microstrip line connected to a varactor, resulting in mutually-coupled resonant circuits. The bias-voltage-dependant capacitance of the varactor varies the resonant frequency of the low-Q resonant circuit with the tuning voltage. The amount of frequency tuning range can be controlled by varying the coupling between the low-Q microstripline/varactor circuit and the dielectric resonator circuit. Tighter coupling permits greater tuning range, however the attendant degradation in the Q factor manifests itself primarily as an increase in phase noise. Varactor tuning is by far the most common means of incorporating electronic tuning.

Bias voltage tuning takes advantage of the sensitivity to changes in the supply voltage of the device used as an oscillator. By not using an internal voltage regulator, the oscillator can be designed to provide the necessary tuning range by varying the supply voltage, typically within .1% of center frequency. As output power is often a function of supply voltage, care must be exercised to maintain suitable output power variation characteristics.

Figure 13 shows the phase noise characteristics for an electronically-tuned DRO using a varactor. It may be noted that increase in electrical tuning results in increased phase noise. Figure 14 reflects the frequency tuning and modulation sensitivity characteristics with tuning voltage and also shows the output power fluctuation over the tuning range. The electrically tunable DROs are presently available up to 26.5 GHz.

In an optically-tunable dielectric resonator configuration (Fig. 15), photosensitive material, such as high-resistivity silicon, is placed directly on the dielectric resonator. Light from a laser or light-emitting diode (LED) is brought through an optical fiber to illuminate the photosensitive material, changing its conductivity and perturbing the electromagnetic fields in and around the resonator. This perturbation results in a shift in the center frequency of the DRO. Using this technique tuning bandwidths of >0.1% can be obtained.

Special DRO Circuits

Excellent integrability and high performance of the dielectric resonators has generated a large number of interesting DRO configurations for various system applications. Some of those special circuits will be described in this section:

A dual-resonator oscillator, shown in Fig. 16, presents a highly stable DRO circuit using identical resonators in a series feedback configuration in both the source and gate circuit of a FET[7]. This oscillator has three output ports, and its microstrip circuitry is wideband due to the simple, minimally-tuned 50-ohm lines on all three ports of the transistor. This oscillator is not particularly susceptible to spurious oscillations, and can provide high temperature stabilities by selecting appropriate dielectric mixes for the resonators.

A push-push DRO (Fig. 17) can use a common dielecric resonator for two transistors. In this approach the fundamental frequency of the oscillators is cancelled and the second harmonics are added at the output plane of the oscillator. This circuit helps to generate low-noise oscillations at frequencies much higher than otherwise possible. For example, using an 18 GHz resonator, a GaAs FET DRO was reported to have a phase noise of -100 dBc @ 100KHz[8]. Figure 18 shows a comparison of phase noise between an Avantek 18 GHz push-push bipolar DRO and fundamental GaAs FET oscillator.

A selectable multi-frequency oscillator development appears in Fig. 19 The diagram shows a fast-settling tri-frequency selectable DRO. A single GaAs FET is used in conjunction with a single single-pole three-throw (SP3T) switch to select the dielectric resonator corresponding to the desired output RF frequency. Compared to the old approach of using several continuously-operating DROs and a high isolation, matched SP3T switch, this approach is free of the spurious signals at the unselected frequencies. This approach is also less expensive, uses less number of components and is more reliable compared to the old approach. The output frequency settles within $\pm$10 ppm in less than 2 microseconds[9].

A silicon monolithic self-oscillating mixer using a Darlington pair is shown in Fig. 20. The dielectric resonator is used in parallel feedback between the input and output of the device. RF input is mixed with the LO in the first transistor Q1 and the second transistor offers gain at the IF frequency. Using this configuration an RF signal at 3.7 to 4.2 GHz was down converted to 1 to 1.5 GHz IF with 9 dB conversion gain [10].

Limitations of Transistor DROs

Free-running DROs presently do not have the low phase noise and temperature stabilities required for certain high-performance applications. This

limitation necessitates the use of phase locking, injection locking, oversizing or analog or digital compensation circuits when necessary. Another important limitation of the DRO is the effect of vibration on the performance of the DRO. Under vibration the variation of the distance between the resonator and the outer shielding is the major cause of the deterioration of the phase noise. A random vibration density of 0.2 G^2/Hz can increase the phase noise by more than 20 dB up to the highest frequency of vibration. Phase locking or injection locking are typically used to minimize these effects.

Future Trends

The dielectric resonator oscillator technology is developing exceedingly rapidly. Emphasis will continue to develop higher performing DROs. The following aspects are likely to be dealt with in the near future:

- Lower phase noise and higher temperature stability oscillators.
- Extension of both lower and upper frequency coverage.
- Higher power at millimeter frequencies
- Reduction in cost and size
- Wider mechanical and electrical tuning bandwidths
- Improvement in DRO performance under vibration
- Optical tuning & injection locking
- Development of new materials for dielectric resonators for linear temperature coefficient and higher quality factor.
- Use of higher-order modes.

REFERENCES:

1. D. Kajfez and P. Guillon, Dielectric Resonators, Artech House 1986.
2. A. P. S. Khanna,"Q measurements of microstrip coupled dielectric resonators," Microwaves and RF, vol. 23, pp. 81-85, Jan. 1984.
3. C. Tsironis and V.Pauker,"Temperature Stabilization of GaAs MESFET Oscillators Using Dielectric Resonators," IEEE Trans. MTT., Vol. MTT-31, pp. 312-314, March 1983.
4. J. Lee et al., "Digital and analog frequency-temperature compensation of dielectric resonator oscillators," IEEE MTT-S Int. Microwave Symposium Digest., pp.277-279, San Fransisco, 1984.
5. M. Prigent & J. Obregon, "Phase Noise Reduction in FET Oscillators by Low-Frequency Loading and Feedback Circuitry Optimization," IEEE Trans. MTT., Vol.MTT-5, No.3, pp. 349-352, March 1987.
6. Z. Galani et al., "Analysis and design of a single-resonator GaAs FET oscillator with noise degeneration," IEEE Trans. MTT., vol. MTT-32, pp. 1556-1565, Dec. 1984.
7. A.P.S. Khanna et al, "Efficient, low noise three port X-Band FET oscillator using two dielectric resonators" IEEE MTT-S Int. Microwave Symposium Digest., pp.277-279, Dallas, 1982.
8. A.M. Pavio & M.A. Smith,"Push-Push Dielectric Resonator Oscillator," IEEE MTT-S Int. Microwave Symposium Digest., pp. 266-269, St. Louis, 1985.
9. A.P.S. Khanna & R. Soohoo,"Fast Switching X and Ku Band Multi Frequency Dielectric Resonator Oscillator Using A Single GaAs FET," IEEE MTT-S Int. Microwave Symposium Digest., pp. 189-191, Las Vegas, 1987.
10. I. Kipnis & A.P.S. Khanna,"10 GHz Frequency Convertor Silicon Bipolar MMIC," Electronics Letters 6th Nov. 1986 Vol.22 No.23 pp. 1270-1271.

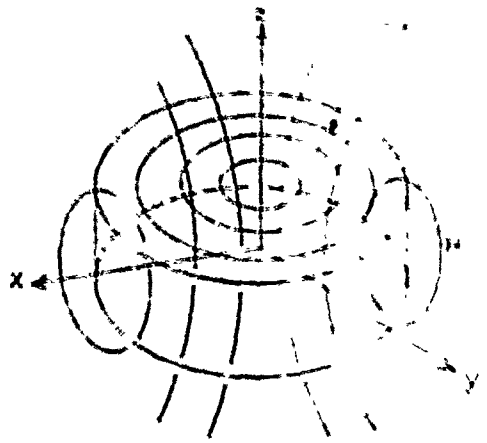


Fig. 1 Field distribution of $TE_{01\delta}$ mode in a dielectric resonator.

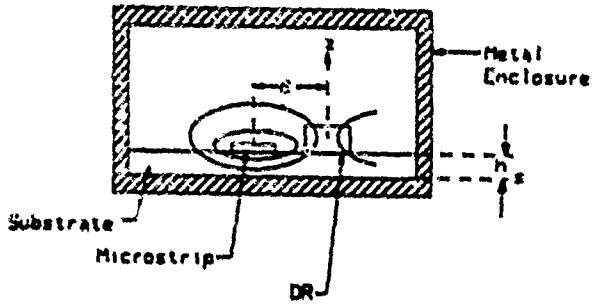


Fig. 2 Dielectric Resonator Coupled to a Microstrip line.

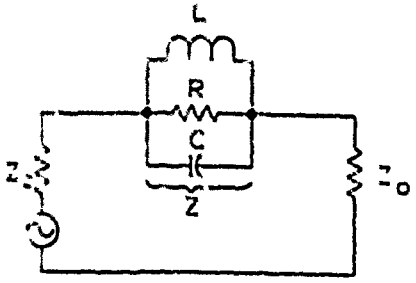


Fig. 3 Eqv. Ckt. of D.R. coupled to a microstrip line.

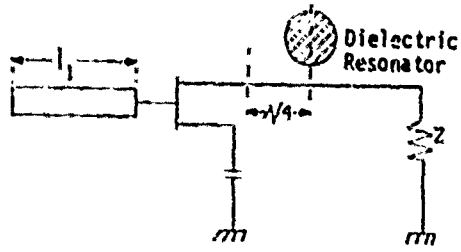
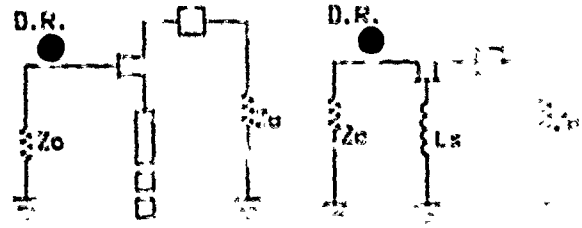
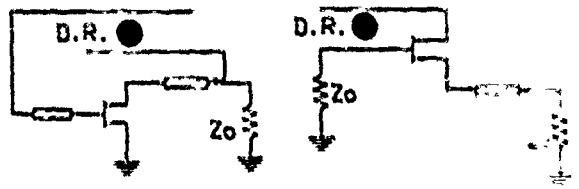


Fig. 4 Stabilized GaAs FET DRO.



SERIES FEEDBACK



PARALLEL FEEDBACK

Fig. 5 Stable Transistor DRO Configurations.

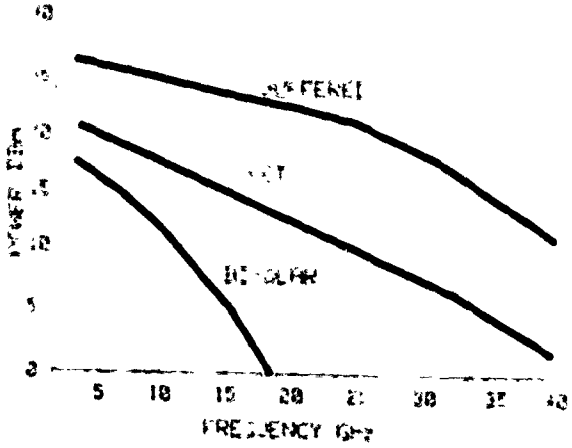


Fig. 6 Power vs. Frequency for Transistor DROs.

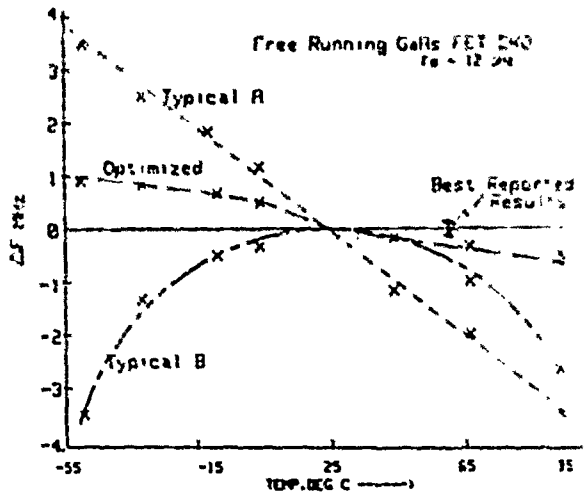


Fig. 7 Frequency Drift vs. Temperature for typical DROs.

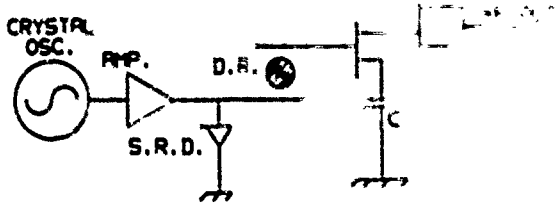


Fig. 8 Injection Locked DRO.

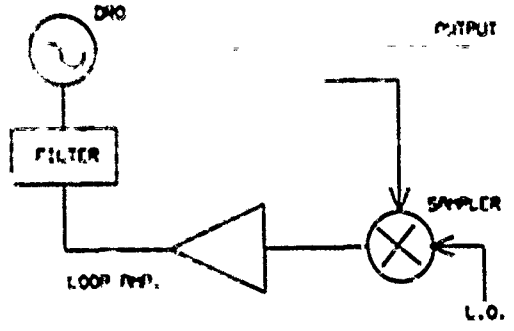


Fig. 9 Phase Locked DRO.

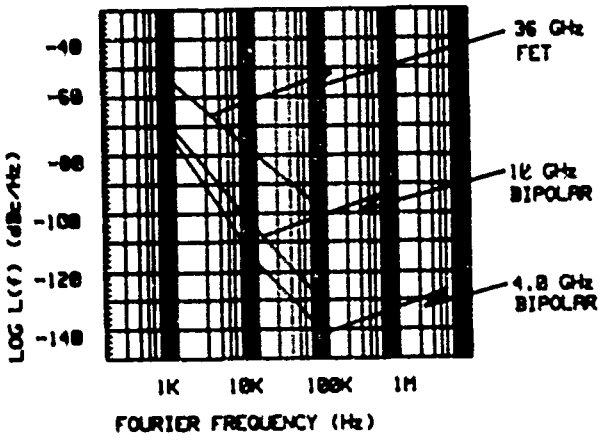


Fig. 10 Phase Noise of Transistor DROs.

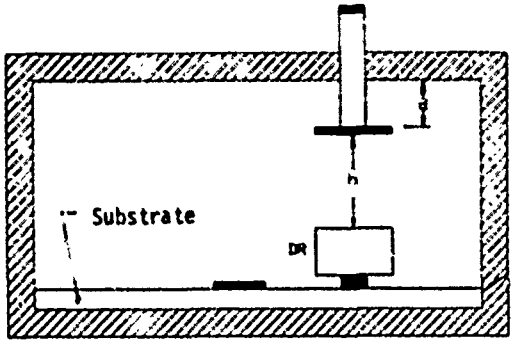


Fig. 11 Mechanically Tuned DRO Configuration.

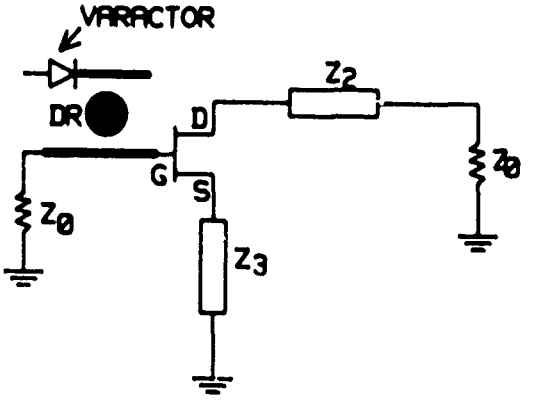


Fig. 12 Varactor Tuned DRO.

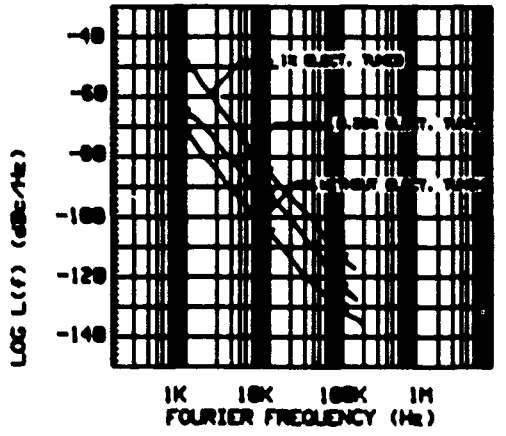


Fig. 13 Phase Noise Deterioration with Electronic Tuning.

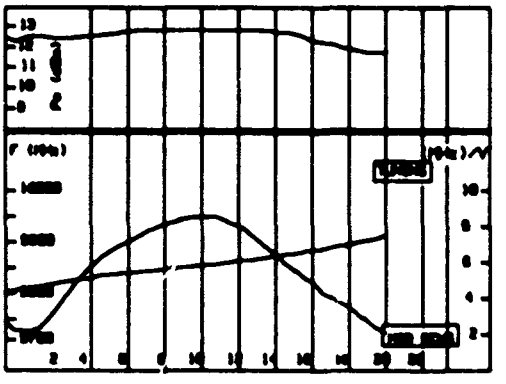


Fig. 14. Freq. tuning, modulation sensitivity & output power vs. tuning voltage for a VT-DRO.

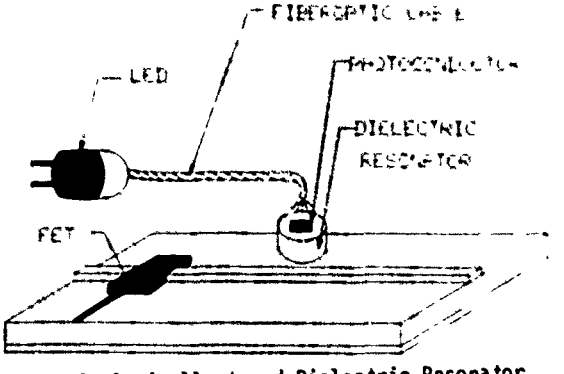


Fig. 15. Optically tuned Dielectric Resonator.

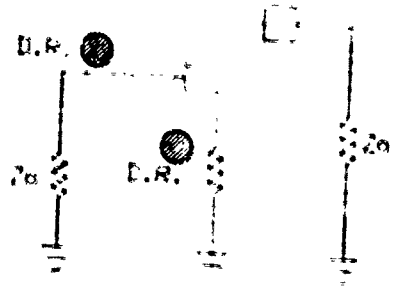


Fig. 16 Dual Dielectric Resonator Oscillator.

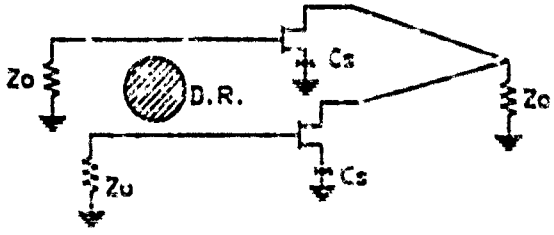


Fig. 17 Push Push Transistor DRO.

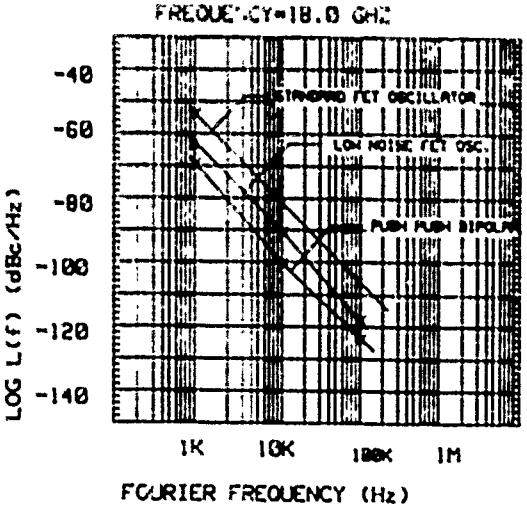


Fig. 18 Phase Noise Comparison between Push-Push Bipolar DRO & Fundamental GaAs FET DROs.

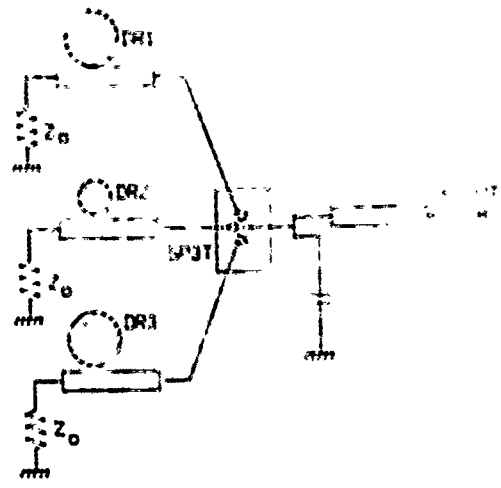


Fig. 19 Selectable Multi-frequency D.R.O.

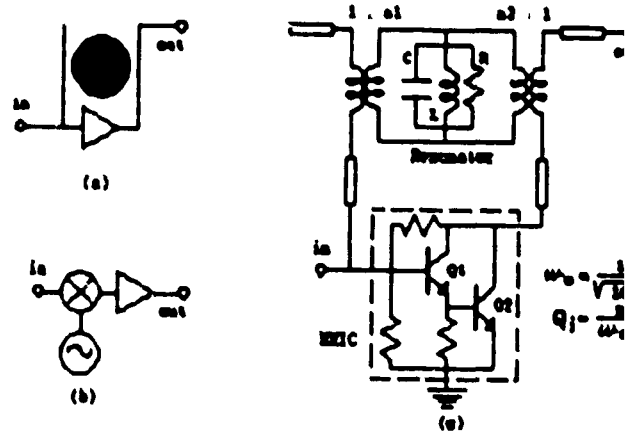


Fig. 20 D.R. Self Oscillating Mixer a)Schematic b)Functional Block Diagram and c)Equivalent Circuit.

41st Annual Frequency Control Symposium - 1987

Measurements and Analysis of Cryogenic Sapphire Dielectric Resonators and DRO's

G. J. Dick and D. H. Strayer
California Institute of Technology
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, Ca. 91109
(818) 354-6393

ABSTRACT

Experimental and computational results of a study of a new kind of dielectric resonator are presented. It consists of a cooled, cylindrically symmetric sapphire resonator surrounded by a metallic shield and is capable of higher Q's than any other dielectric resonator. Isolation of fields to the sapphire by the special nature of the mode allows the very low loss of the sapphire itself to be expressed. Calculations show that the plethora of modes in such resonators can be effectively reduced by the use of a ring resonator with appropriate dimensions. Experimental results show Q's ranging from $3x10^7$ at 77 Kelvin to 10^9 at 4.2 Kelvin. Performance is estimated for several types of DRO's incorporating these resonators. Phase noise reductions in X-band sources are indicated to values substantially lower than presently available.

INTRODUCTION

A new kind of dielectric resonator promises to enable an important advance in the capability of dielectric resonator oscillators (DRO's). This resonator consists of a cooled sapphire ring or cylinder surrounded by a metallic shield; and is capable of higher Q's than any other dielectric resonator, equalling those of quartz crystals at temperatures which can be reached by means of thermoelectric cooling.[1,2] At 10 to 20 Kelvins it rivals the performance of superconducting resonators requiring temperatures 10 times lower. We report on results of tests on such a sapphire resonator at X-band which show Q's ranging from $3x10^7$ at 77 Kelvin to 10^9 at 4.2 Kelvin.

The high Q's of these resonators depend not only on a reduction of losses internal to the sapphire but also on isolation of the resonant energy from losses in the surrounding metallic shield. With a dielectric constant (≈10) only a fraction of that of other dielectric resonator materials, sapphire resonators are at a substantial disadvantage in this regard. This is overcome in the resonators of the present study by a process similar to the optical phenomenon of total internal reflection.

We present the results of both experiment and calculation which show that effective isolation can be obtained in modes with 5 to 10 full waves around the perimeter of the resonator. New computations for mode Q's and frequencies for high mode numbers are presented based on previously published[4] solutions to the wave equation for an isolated isotropic dielectric sphere, the only finite geometry for which, to the authors' knowledge, closed form solutions have been developed. An approximate method is developed to allow calculations for right cylinders and for rings with rectangular cross section. This method is based on (also approximate) solutions for a rectangular dielectric waveguide[5]. The ring is assumed to be just such a waveguide bent around on itself. Losses in the metallic shield are explicitly considered.

The plethora of modes in the cylinder and sphere have led us to consider the ring resonator for further analysis and study. We find that an appropriate choice of ring dimensions can greatly increase the mode spacing without sacrificing the isolating properties of the mode.

Analysis of several different types of oscillator applications are presented. Possible applications include low noise microwave oscillators using only thermo-electric cooling, and oscillators with both extremely low noise and high stability at temperatures of 77K and below.

BACKGROUND

Cryogenic sapphire resonators have been studied experimentally by Blair[1] in Australia and by V. B. Braginsky[2] and coworkers in the USSR, with the aim to develop ultra-stable microwave oscillators and discriminators. Previous work has included measurements of mode frequencies and evanescent field decay lengths, measurement and calculation of temperature and frequency dependence of the Q's, measurement of the fractional thermal coefficient of the resonant frequency, and development and study of stabilized oscillator performance.

In these experimental studies sapphire losses are found to drop dramatically as the temperature is reduced below ambient, showing an approximately T^5 dependence for temperatures down to about 50 Kelvin, where a Q of approximately 10^8 is attained (for X-band). The loss mechanism responsible for this behavior has been identified by Gurevitch[6] as phonon generation due to lattice anharmonicity. The T^5 dependence of the losses as well as a linear dependence on frequency are predicted by this theory. Both are born out in experimental data, indicating that this source of loss is inherent in the sapphire, and probably cannot be improved upon by improved sample preparation. It seems appropriate, then, to use the presently observed high temperature behavior as a basis to engineer filters and DRO's.

The temperature dependence of the frequency of sapphire dielectric resonators has also been studied by both groups.[1,2] The fractional frequency variation with temperature dF/dT/F is found to saturate at about $6x10^{-5}$/Kelvin at high temperatures (>300K), dropping as the coefficient of expansion "freezes out" at lower temperatures.[1] It reduces to $3x10^{-6}$ at 77K and falls as T^3 at lower temperatures to a value estimated to be 10^{-12} /Kelvin at a temperature of 1 Kelvin.[2] The values found at 77K and below could allow very impressive oscillator stability, with a stability equal to that of quartz crystals reached by 40K. At 10 Kelvin, the readily available temperature variation of 10 microdegrees would cause a fractional frequency variation of only 10^{-14}.

CH2427-3/87/0000-487$1.00 C 1987 IEEE

The stability demonstrated by oscillators using sapphire and sapphire-filled resonators shows the efficacy of this reduction in expansion coefficient. Frequency stability of 10^{-13} was demonstrated by the Russian group using a GUNN-excited oscillator,[1] and stability better than 10^{-13} has been reported by the Australian group[2] using a frequency-locked GUNN oscillator at room temperature. The authors, using a sapphire resonator coated with superconducting lead, have demonstrated stability better than 10^{-14} at 100 seconds. In this last case, the higher stability is not attributable to the superconducting coating, but rather to the use of a ruby MASER as the source of excitation.[3] The lead coating is rather seen as a distinct disadvantage in this regard, its greater thermal sensitivity forcing operation at temperatures below 1 Kelvin to achieve the observed stability.

While all of the oscillators just mentioned operate at temperatures below 2 Kelvin, the prospect of both high stability (due to the low expansion coefficient) and extremely low phase noise (due to high Q) in the temperature range from 10 K to 77K is perhaps the most exciting aspect of their performance. Of great significance here are the relatively small and inexpensive cryocoolers available in this temperature range. In addition, comparison to conventional DRO's and cavity-stabilized microwave oscillators also indicates a dramatic reduction in phase noise using a sapphire resonator at approximately 170 Kelvin, a temperature achievable using thermoelectric cooling. Here the Q of 2×10^8 compares with values of $1\text{-}3\times10^4$ available from other microwave resonators, indicating a correxponding reduction of phase noise of 36 to 46 db.

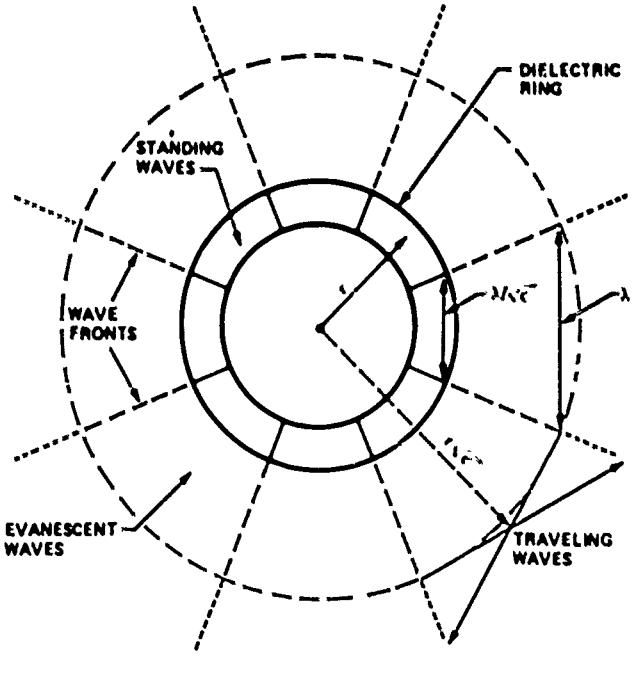


Figure 1 Diagram showing the character of the electromagnetic field in the vicinity of a dielectric ring for an 8-fold cylindrically symmetric mode.

ISOLATED MODES IN DIELECTRIC RESONATORS

Isolated modes in dielectric resonators achieve we coupling to the surrounding space not primarily by impedance mismatch due to the large dielectri constant, but rather by isolating properties of t mode itself. These modes can be understood from Fi 1 as consisting of a wave trapped and slowed by circular dielectric waveguide. The wave equation:

$$(k_r)^2 + (k_\phi)^2 + (k_z)^2 = \epsilon(2\pi/\lambda)^2$$

allows a large value of k_ϕ inside the dielectric the thickness and width of the ring are large enou to allow only small values of k_ϕ and k_z, respectivel Outside the dielectric, however, the dielectr constant ϵ is 1, and this large value of k_ϕ, ari required by the symmetry of the mode for some distan outside the dielectric, requires an imaginary part one of the other components (found in k_r) to sati this same wave equation. This region of evanescen decaying fields forms a buffer between the waves the dielectric and allowed travelling waves farth out. These modes have been misnamed "whisperi gallery" modes[2] but are more properly seen as analo ous to the phenomenon of total internal reflection optics.

To the authors' knowledge, solution in closed form f the modes of cylindrically symmetric dielectr resonators are available only for the isotropic sphe and the infinite cylinder. Of these, the spher being a finite structure is appropriate for consider tion here. Following solutions published by Cast'n et al[4] for the modes $TE_{n,1}$, we have calculat frequencies and Q's for m=1, r=1,2 and for n rangi up to relatively large values. These values a plotted in Fig. 2 and show an exponential increase Q as n and the frequency are increased. A shortcomi in this calculation is the inability to account f the effect of a metallic shield, necessary to allow reasonably small overall size and desirable increase the radiation-limited Q's as shown in Fi 2. It seems apparent that replacing the (complete absorbing) space surrounding the sphere by on slightly absorbing copper should improve the Q, but

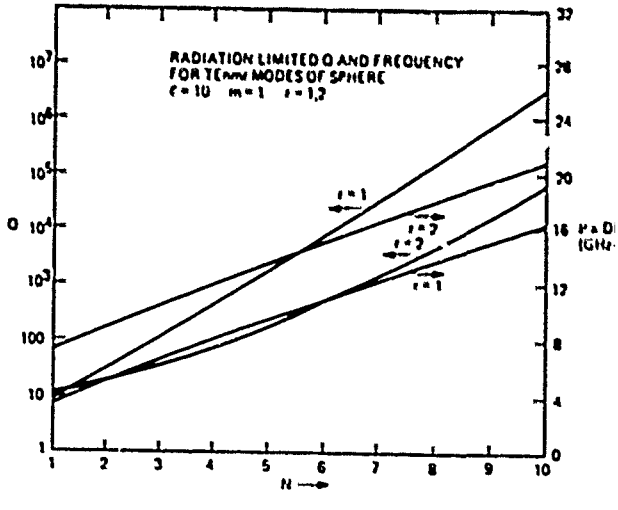


Figure 2 Radiation-limited Q and frequency for $TE_{n,r}$ modes of an isolated sphere with $\epsilon = 10$ m=1, r=1,2.

how much? An upper limit would seem to be the product of the Q's, e.g., for n=7, r=1 a radiation limited Q of 3×10^4 (from Fig. 2) combined with a copper can Q of 10^4 would indicate that Q's up to 3×10^8 might be possible, an attractive prospect. It also seems clear from Fig. 1 that the containing can must be in the evanescent region, and that there would be some trade-off between isolation from can losses and position of the can (overall size).

In order to test these ideas we mounted an uncoated sapphire cylinder with length and diameter both approximately 5 cm inside a copper can large enough to provide a 1cm gap at the outside and on the ends. At liquid nitrogen temperature and below, we found modes with high Q ($Q > 10^7$) for frequencies above about 7.5 GHz. This frequency corresponds to n = 8 or 9 from Fig. 2 with a corresponding free-space radiation-limited Q of 10^5 to 10^6. Since the measured Q is higher than these values, some enhancement of the Q results from the low-loss properties of the shielding can.

However the plethora of modes which we found gave us no hope of successfully identifying the modes on the basis of the spherical solutions. Furthermore, the prospect of oscillator design is daunting, given the existence of strongly coupled low-Q modes very near in frequency to weakly coupled high Q modes.

A simple application of the wave equation to the geometry of Fig. 1, forcing a correspondence of k_r and k_z to half-wave solutions in the r and z directions, respectively, indicated that the number of modes might be reduced without great penalty by a resonator with geometry shown in Fig. 3. As a next step, and in order to obtain a more complete picture of the modes, we have constructed a mode picture based on solutions for the modes of a rectangular dielectric waveguide derived by Marcatili[4]. He identifies modes E_{pq}^x and E_{pq}^y with electric polarization in the x and y directions, respectively and with p and q half waves in the x and y directions. Identifying the x, y, and z coordinates of these solutions with the r, z, and ϕ directions indicated in Fig. 3; mode indices p and q with the mode multiplicity in the r and z directions, and introducing a mode number n corresponding to the number of full waves around the perimeter of the ring, we identify modes $E_{n,pq}^x$ and E_{npq}^z for the ring.

Following Marcatili we find E_{pq}^y modes for the rectangular dielectric waveguide as the solutions of

$$p^2X + q^2Y = 1$$

where $X = (\pi/a)^2 (1 + 2\Delta/\pi a)^{-2} (k_1^2 - k_z^2)^{-1}$,

and $Y = (\pi/b)^2 (1 + 2\Delta/\pi^2 b)^{-2} (k_1^2 - k_z^2)^{-1}$,

where $\Delta = \lambda/2 \quad (\epsilon-1)^{-0.5}$,

$k_1 = 2\pi n/\lambda$,

$n = \sqrt{\epsilon}$, $\lambda = c/f$,

and a and b are the height and width of the ring cross section, c is the speed of light and f the microwave frequency.

Explicitly accounting for the ring geometry by constraining the solution to exhibit n full waves around an effective ring perimeter r_{eff}, we define $k_z = 2\pi n/r_{eff}$, where r_{eff} is defined in terms of the inner and outer ring radii as $r_{eff}^2 = (r_i^2 + r_o^2)/2$.

Decaying fields (imaginary components to the wave vector k) are required in the space just outside the dielectric by the wave equation due to the large value of k_z allowed by the dielectric. A lower limit to the decay rate is obtained by identifying the decay length l_d as

$$l_d = (k_z^2 - 2\pi/\lambda)^{-0.5}$$

Assuming that the gap is much smaller than the radius, we identify the Q enhancement factor as the square of the field decay to the metallic wall a distance l_{gap} away,

$$Q \text{ ratio} = \exp(2\pi l_{gap}/l_d).$$

We have calculated modes for a solid cylinder 5 cm in diameter and 5 cm long, identifying parameters

$r_i = 0$, $r_o = 2.5$ cm, $r_{gap} = 1$ cm, a = 2 cm, and $b = r_o - r_i = 2.5$ cm,

and for the ring in Fig. 3 with parameters given by, $r_i = 1.5$cm, $r_o = 2.5$ cm, $r_{gap} = 1$ cm, a = 2 cm, and $b = r_o - r_i = 1$ cm.

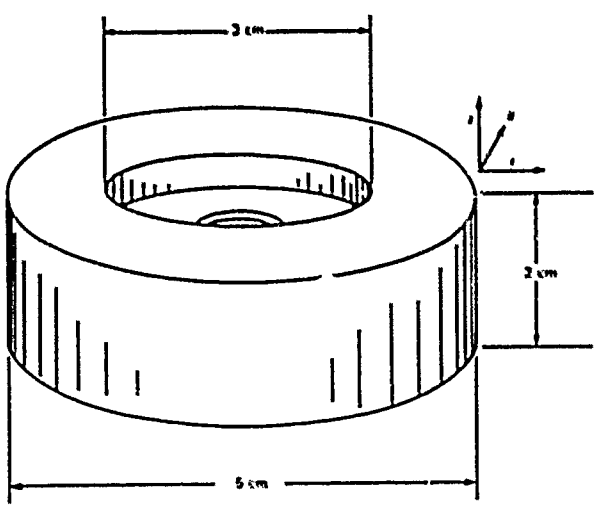


Figure 3 Sapphire ring construction showing directional axis identification at ring perimeter.

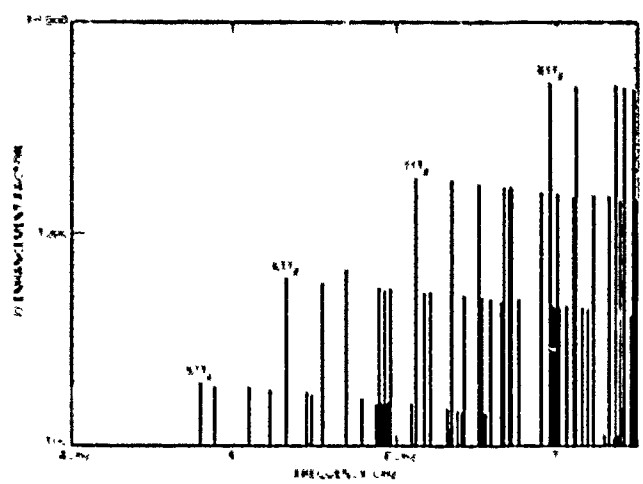


Figure 4 Calculated mode frequencies and Q enhancement factor (see text) for a dielectric sapphire cylinder 5 cm diameter, 5 cm high surrounded by a lossy shield 1 cm from the surface

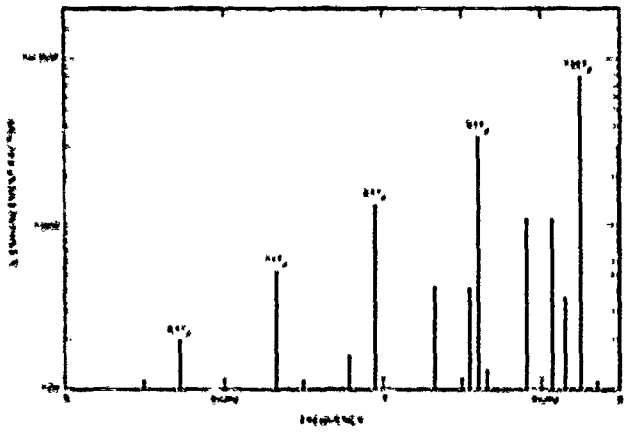


Figure 5 Calculated mode frequencies and Q enhancement factor for the ring shown in Fig. 3 surrounded by a lossy shield 1 cm from the surface.

The results of these calculations are shown in Figures 4 and 5. The predictions shown in Fig. 4 are in excellent qualitative agreement with the results of our measurements on the cylinder, confirming the validity of our approach. The efficacy of the ring geometry in reducing mode density is dramatically shown in a comparison of the two figures. The actual number of modes is larger than the number shown because modes with poor or no isolation are not shown. The calculations found 398 modes below 9 GHz for the cylinder and 60 for the ring. These modes are all doubly degenerate, a fact which was noted for many of them during the measurement process. Typical splitting of the modes was observed to be 10^{-5} to 10^{-4} fractional frequency deviation.

An inherent problem in the use of these resonators in active oscillators, and an important reason for choosing the ring for further study is that the coupling of any mode to the external electronics will tend to scale in direct proportion to the coupling to the wall. This means that, even though two modes may have very different Q's, if they are near to each other in frequency, mode selection may very well be a difficult problem. For example, if one of the modes is critically coupled to the active electronic elements, the other is likely to be nearly critically coupled as well.

Q MEASUREMENTS

Figure 6 shows results of measurements of the Q of two of the modes of the 5 cm by 5 cm cylinder for temperatures below 77 Kelvin. Also plotted are higher temperature results reported by Braginsky et al[2]. Good agreement is found with the results of these higher temperature data, confirming that these losses are inherent in the sapphire itself, and not due to impurities, surface treatment, etc. The leveling off of the loss reduction at about 10^{-8} is characteristic of the results reported by others and is probably due to impurities. The further Q improvement at the lowest temperatures is also typical, with the lowest point being marginally better than any others reported to date.

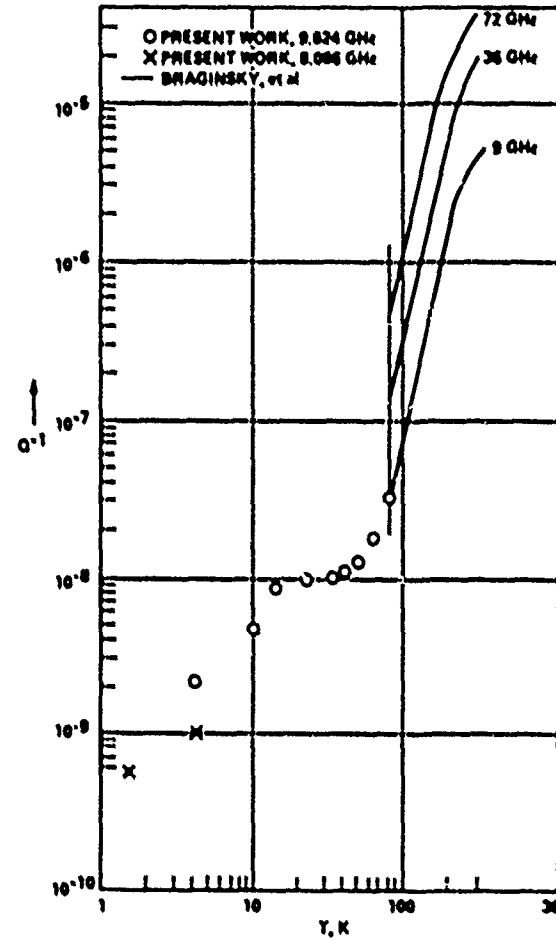


Figure 6 Q measurements for a sapphire cylinder 5 c diameter, 5 cm high contained in a lead-plat shielding can approximately 1 cm away. Also shown a higher-temperature data by Braginsky, et al. (refe ence 2).

A consideration for resonator design is the requir ment for surface finish and dimensional uniformity f the shaped dielectric cavity. Braginsky, et al. (re 2, p. 87) have used methods developed for optic fibers to estimate the losses caused by scatteri from surface roughness. They find that, for a reson tor of centimeter dimensions, roughness of 1 micr characteristic height will cause losses of order 10^{-10}. Although this small value is smaller than a losses measured so far, the resonator used in t measurements reported here was fabricated with optical quality polish on all surfaces to assure loss contributions from this source. Precaution such as acid etch, and purified alcohol rinse, we taken to assure that no foreign material adhered the surface.

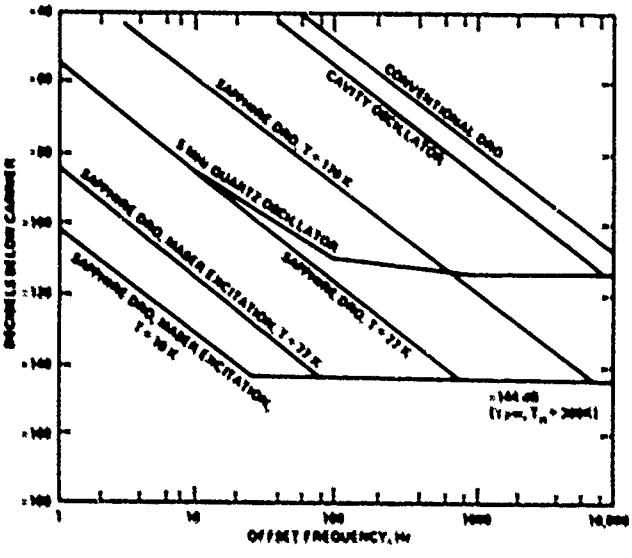


Figure 7 Phase noise for various X-band sources including conventional DRO and cavity oscillators, a state-of-the-art quartz crystal oscillator referenced to 10GHz, and predictions for several sapphire DRO's.

PREDICTIONS OF OSCILLATOR NOISE PERFORMANCE

The reduction in phase noise over that in conventional DRO and cavity oscillators which would result from the high Q of a cooled sapphire resonator are shown in Fig. 7. Q's of 10,000 and 30,000 are assumed for the conventional oscillators, respectively, and values from Figure 6 for the sapphire DRO's. Also shown is a further reduction which would result from the application of ruby maser technology to such oscillators.

Multiplicative 1/f noise $S_\phi(f)$ in the active device is assumed to be -100dbc/f (/Hz),[7] -130dbc/f(/Hz) for the curves indicating maser excitation. This latter value corresponds an upper limit obtained in tests of a low-Q S-band ruby MASER oscillator,[8] a value substantially quieter than that reported for any other active microwave device. It has been well documented that multiplicative 1/f noise in semiconducting devices can be reduced by operating devices in parallel or, similarly, by large gate dimensions. Thus it seems likely that the low 1/f noise in the ruby is due to its very large volume (≈1 cm^3). Ruby MASERS have been operated at temperatures as high as 90 Kelvin and at frequencies up to 42 GHz[9]. Their application to low noise oscillators could open a new window in low noise oscillator capability.

ACKNOWLEDGEMENTS

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

REFERENCES

1. D. G. Blair and S. K. Jones, IEEE Trans. Magnetics MAG-21, 142, 1985. D. G. Blair and I. R. Evans, J. Phys. D: Appl. Phys., Vol. 15, 1651-1656 (1982).

2. V. B. Braginsky, V.P. Mitrofanov and V.I. Panov, Systems with Small Dissipation (Univ. of Chicago Press, Chicago, 1985), 85-89; Frequency Stabilization of oscillators with high-Q leucosapphire dielectric resonators, V. I. Panov and P. R. Strankov, Radiotekhnika i Electronika 31, 213, 1986.(In Russian).

3. D. M. Strayer, G. J. Dick, J. E. Mercereau, IEEE Trans Magnetics MAG 22, (1986 to be published). Also S. Thakoor, D. M. Strayer, G. J. Dick and J. E Mercereau, J. Appl. Phys. 59, 854, 1986. G. J. Dick and D. M. Strayer, Proc. 38th Ann. Freq. Control Symp. (Elec. Industries Assoc., 1984), 414.

4. M. Gastine, L. Courtois, J. L. Dormann, IEEE Trans. Microwave Theory and Techniques, Vol. MTT-15, 694 (1967).

5. E. A. J. Marcatili, Bell System Technical Journal 48, 2071(1969)

6. V. L. Gurevich, Kinetics of Phonon Systems, Moscow (1980).

7. This compares to the best X-band GAASFET multiplicative noise (-109dbc/f (/Hz)) so far discovered by the authors: F. Walls, private communication.

8. H. Reitbock and A. Redhardt, Z. Naturforsch Vol 17a, 187 (1962). C. Moore and D. Neff, IEEE Trans. MTT 30, 2013 (1982)

9. J. A. Barnes, R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.A. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M.R. Winkler, IEEE Trans. Instr. and Meas. IM-20, 105, 1971.

41st Annual Frequency Control Symposium - 1987

COMPACT AND SIMPLE x 3 (9 to 27 GHz) PLL FREQUENCY MULTIPLIER USING HARMONIC PHASE DETECTION

Jordi Berenguer, Javier Bará and Adolfo Comerón.

Departamento de Electrofísica. E.T.S. Ingenieros de Telecomunicación de Barcelona
Universidad Politécnica de Cataluña
Apartado 30002, 08080-Barcelona, Spain.

SUMMARY

Frequency multiplication in the microwave and millimeter-wave range can be performed in compact form using harmonic phase detectors, in which the VCO phase is directly compared to a multiple of the reference frequency internally generated in the same device. In this paper we present a x3 P.L.L. frequency multiplier following this technique that phase-locks a 27 GHz Gunn diode VCO to a 9 GHz reference, in a simple and compact arrangement. The Gunn diode frequency is controlled via the voltage applied to it. The phase detector used is a balanced mixer, implemented with a 90° microstrip hybrid ring and Schottky diodes, and designed at the 9 GHz reference frequency. This system can be employed through the microwave and millimeter-wave ranges to obtain low order odd-index frequency multiplications.

INTRODUCTION

The expansion of satellite communication systems demands for receivers with good sensitivity and dynamic range. This often requires designs involving coherent detection and the use of local oscillators with high spectral purity /1/.

The design of low phase noise oscillators in the microwave and millimeter-wave range is usually implemented by phase-locking a Voltage Controlled Oscillator (VCO) to a spectrally pure reference signal (indirect synthesis by phase locked loops) /2/ /3/. This technique often calls for frequency multipliers, where the locked oscillator frequency is a certain multiple of a more stable frequency, and the phase detection is made at lower frequency after down-converting the VCO's high frequency by means of a harmonic mixer or regenerative frequency dividers /4/.

Indirect frequency multiplication by small odd multiples can be performed in a very compact manner using harmonic phase comparators, in which the phase of the VCO is directly compared to the multiple of the reference frequency generated internally in the same device /5/ (Figure 1).

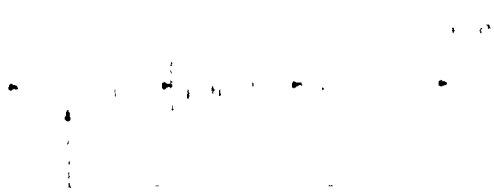


Fig. 1 - Block diagram of small odd multiples P.L.L. frequency multiplier using harmonic phase detector.

In this paper we present a x3 P.L.L. frequency multiplier following this technique which allows to phase-lock a 27 GHz VCO to a 9 GHz reference, in a simple and compact arrangement.

THEORY OF THE HARMONIC PHASE DETECTOR USING A 180° HYBRID RING

The phase detector function can be performed by a balanced mixer with DC coupled IF port. In the microwave range this is usually implemented with microstrip 90° or 180° hybrid and Schottky diodes (Figure 2).

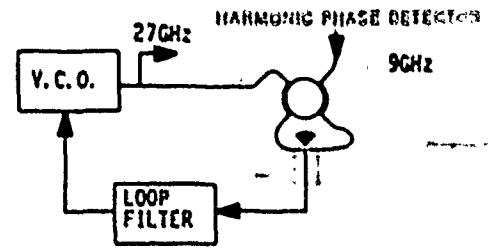


Fig. 2 - Block diagram of x3 (9 to 27 GHz) PLL frequency multiplier, with a microstrip hybrid structure as phase detector.

This structure has the property of behaving as a hybrid at odd multiples of the freouency at which the lengths of the lines connecting the ports are λ/4. Although a bandwidth reduction and a certain degradation of the hybrid are involved, they do not affect the harmonic phase detection ability.

As a result, a phase detector realized with a balanced structure using a 180° hybrid ring (or equivalent) can be used to perform a frequency multiplication by an odd factor in a set-up like that sketched in Figure 1.

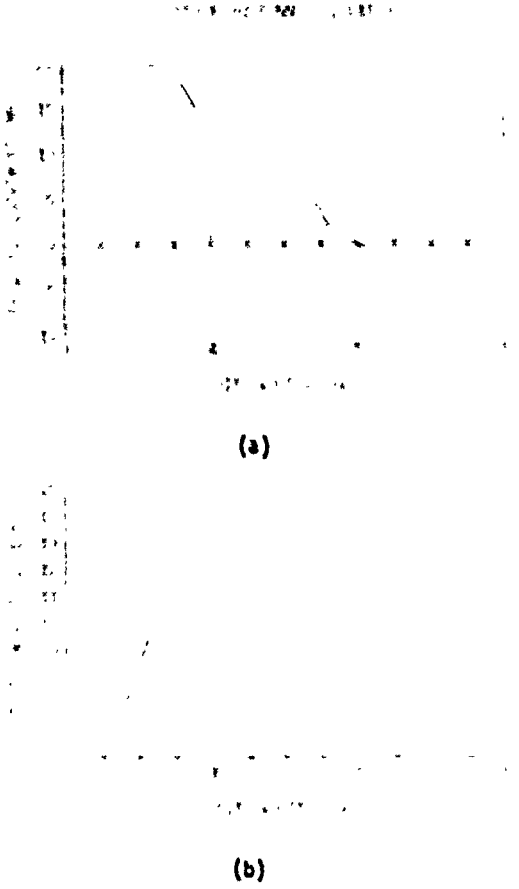
EXPERIMENTAL RESULTS

A x3 frequency multiplier between 9 and 27 GHz using a harmonic phase detector following the set-up of Figure 2 has been built. It is part of a coherent receiving system designed for the propagation experiments on the 12, 20 and 30 GHz beacons on board of the European Space Agency Olympus Satellite, to be launched by 1988. The 27 GHz output is employed as local oscillator for the first mixer in the 30 GHz channel chain.

CH2427-3/87/0000-492 $1.00 © 1987 IEEE

The phase detector uses a microstrip hybrid structure like that shown in Figure 2 made in 0.01" Cu-clad 217 and designed at the 9 GHz fundamental frequency. A pair of AsGa Schottky diodes in a T-package provide for the nonlinear function.

The 27 GHz VCO consists of a cavity mounted Gunn diode. The frequency is controlled by varying the diode bias voltage. Figure 3a shows the VCO frequency increment referred to 27.03 GHz versus the d.c. bias voltage and Figure 3b shows the ripple in the VCO output power level.



(a)

(b)

Fig. 3 - 27.03 GHz V.C.O. a) Frequency increment versus bias voltage. b) Output power level versusbias voltage.

The set-up employed for the characterization of the harmonic phase detector is sketched in Figure 4. The DC output voltage of the harmonic phase detector was recorded versus the phase shift introduced on the 9 GHz port (Figure 5). For a 9 GHz power of -2 dBm and a 27 GHz power of 0 dBm the harmonic phase detector slope was 2.65 mV/°.

The return loss at the phase detector 9 GHz port when it was pumped with a +4 dBm 27 GHz signal has also been measured (Figure 6). The isolation obtained between both ports is shown in Figure 7.

Figure 8 shows the improvement of the 27 GHz oscillator spectral purity when locked using this technique to a more stable 9 GHz frequency. The P.L.L. noise bandwidth is 21 KHz, with a damping coefficient of 1.14 and a 5 KHz loop filter resonance frequency. The hold-in range is 6 MHz.

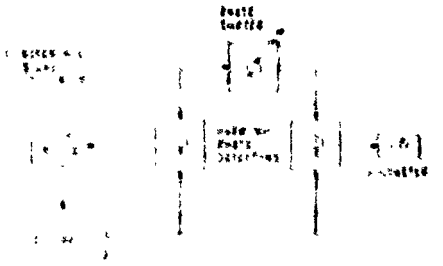


Fig. 4 - Harmonic phase detector characterization set-up.

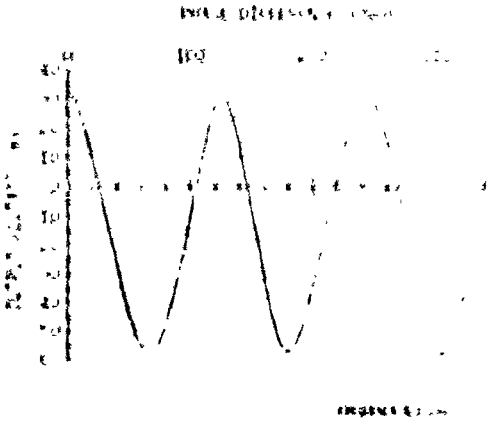


Fig. 5 - Harmonic phase detector D.C. output voltage, versus 9 GHz port phase shift.

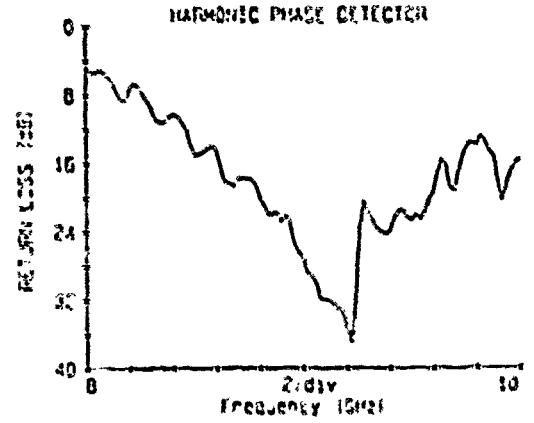


Fig. 6 - Harmonic phase detector return loss measured at the 9 GHz port.

CONCLUSIONS

A very compact x3 frequency multiplier from 9 to 27 GHz has been presented. It is based on P.L.L. techniques with harmonic phase detection that can be employed through the microwave and millimeter-wave ranges to obtain low order odd-index frequency multiplications.

In the coherent receiver final configuration, the 9 GHz frequency will be generated by a FET VCO, itself ultimately locked to a 5 MHz quartz standard, the quartz oscillator stability being transferred in this way to the 27 GHz VCO.

REFERENCES

[1].- W. P. Robins, "Phase Noise in Signal Sources". Peter Peregrinus, 1982.

[2].- R.G. Davis, M.J. Lazarus, "Phase Locking of mm-Wave Two-Port Gunn Oscillator By Bias Tuning". Microwave Journal, June 1986, pp. 103-107.

[3].- V. Manassewitsch, "Frequency Synthesizers Theory and Design". John Wiley & Sons, 1980.

[4].- K. Honjo, M. Madihian, "Novel Design Approach for X-Band GaAs Monolithic Analog 1/4 Frequency Divider". IEEE Trans on MTT-34, no. 4, April 1986, pp. 436-441.

[5].- F.M. Gardner, "Phaselock Techniques". John Wiley & Sons, 1979.

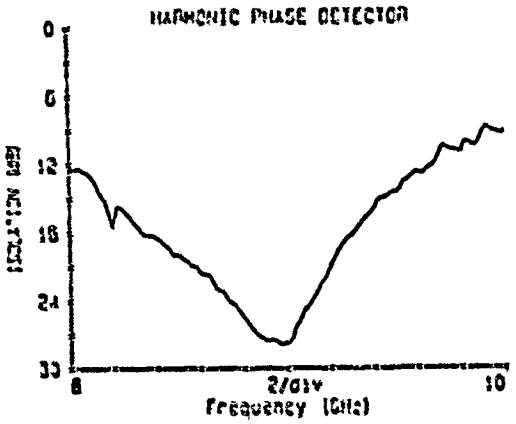
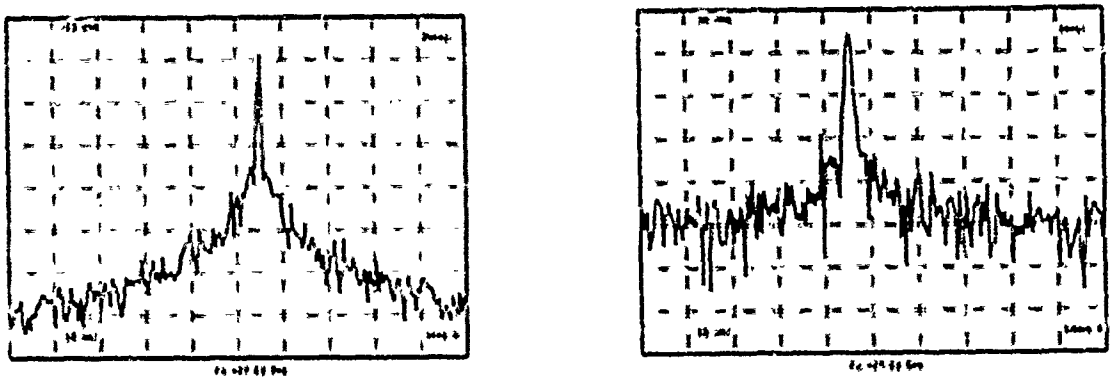


Fig. 7 - Harmonic phase detector isolation between ports, measured at 9 GHz.



(a) (b)

Fig. 8 .- 27 GHz V.C.O. spectrum. a) Free running. b) Locked. (Note the difference in frequency span and resolution bandwidth between a) and b)).

An Analysis of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase-Accumulator Truncation

Henry T. Nicholas, III and Henry Samueli
Integrated Circuits and Systems Laboratory
Electrical Engineering Department
University of California, Los Angeles
Los Angeles, CA 90024

ABSTRACT

An algorithm is presented for the calculation of the spectrum of direct digital frequency synthesizers (DDFS's) as a result of phase accumulator truncation. This algorithm, which is derived using number theoretic methods, includes a closed form expression relating the magnitude, number, and position of the spurious noise lines in the output spectrum of a DDFS to the read-only memory (ROM) look-up table size, the amount of phase accumulator truncation and the input frequency control command. The combined finite word length effects of the ROM and the Digital-to-Analog converter (DAC) nonlinearities are also examined in the light of these new results and new design guidelines are developed. The spectrums predicted by these closed form expressions are compared against spectrums generated by a discrete Fourier transform (DFT) and are shown to have comparable accuracy.

As a result of obtaining an expression for the magnitude of the spurious noise frequencies, a relationship between the greatest common divisor of the input frequency command word and the ROM table size, the phase accumulator word lengths, and the magnitude of the worst case spur is obtained. This relationship is used as the basis for a novel modification to the conventional phase accumulator structure which results in a 3.922dB reduction in the magnitude of the worst case spurious response. This hardware modification is also shown to average out the error effects of DAC nonlinearities and roundoff in the stored sine ROM samples.

I. INTRODUCTION

Modern communication systems are placing increasing demands on the resolution and bandwidth of frequency synthesizer subsystems. Spread spectrum applications impose an extremely stringent set of specifications on a frequency synthesizer which include the achievement of extremely fast switching times over a wide bandwidth in a design that is compact, reliable and easily reproducible[1]. In order to meet these demanding requirements many designers are exploring the use of the direct digital approach to frequency synthesis as an effective alternative to conventional analog synthesizers. Direct digital frequency synthesizers (DDFS's) exhibit fast switching speed, excellent temperature and ageing stability, and the ability to switch frequencies while maintaining constant phase, which are all properties difficult to achieve with analog techniques.

The most popular technique for direct digital frequency synthesis is the sine look-up table method first introduced by Tierney, Rader, and Gold[2]. This method, which is well described in the literature, synthesizes a sinewave by successively scanning through a look-up table stored in a read-only-memory (ROM) and converting the recalled sine samples to an analog waveform via a digital-to-analog converter (DAC). As shown in Figure 1, one sample of the sinewave is recalled from the ROM every period of the stable frequency reference. Each of these recalled samples differs from the previous one by a constant phase increment; thus different frequencies may be synthesized by changing the phase difference between the recalled samples via a digital frequency control command.

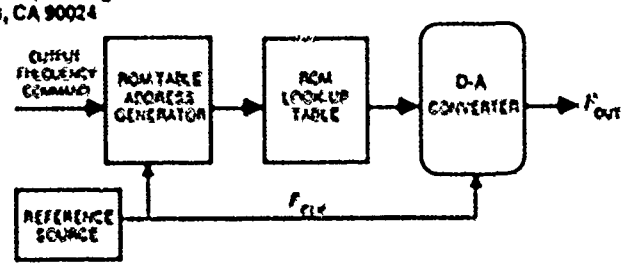


Figure 1. Simplified block diagram of a DDFS.

The generation of the constantly increasing phase argument to the ROM is normally implemented by accumulating a binary frequency control word using a standard two's complement adder and register as shown in Figure 2.

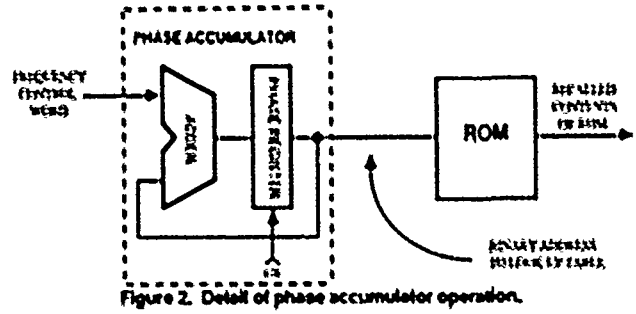


Figure 2. Detail of phase accumulator operation.

Using this architecture, both the frequency and phase resolution of the synthesizer are determined by the wordlength of the phase accumulator. For most applications, a phase accumulator wordlength in excess of 22 bits is required to achieve acceptable resolution. The use of all 22 bits of the phase accumulator wordlength to address 2^{22} stored sine samples would require a prohibitively large ROM, even if the quarter wave symmetry and course-line trigonometric reductions are applied to reduce the storage. Therefore, in actual practice, all sine DDFS's requiring fine phase resolution truncate part of the phase accumulator output when addressing the ROM.

The effective implementation look-up table based DDFS's has been seriously hampered by the lack of a precise theoretical characterization of the finite wordlength effects due to phase accumulator truncation. To illustrate this problem, three different output spectrums are depicted in Figures 3 through 5 showing the spurious response due to phase accumulator truncation for several different output frequencies. The three cases shown are the results of a hardware simulation of a DDFS with a phase accumulator word length of 12 bits and a ROM address length of 5 bits corresponding to the phase truncation of the 7 least significant bits of the phase register. These spectrums were generated by a discrete Fourier transform (DFT) of the numeric output of a hardware simulation of the digital logic, and thus show only the response due to algorithmic nonlinearities inherent in the system. They therefore represent the best case attainable given an "ideal" DAC.

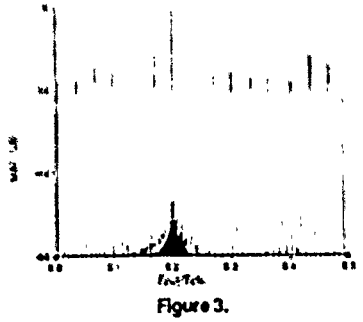


Figure 3.

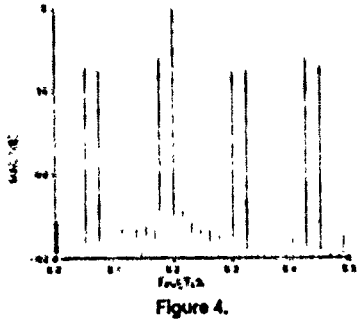


Figure 4.

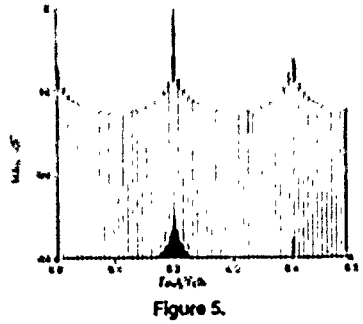


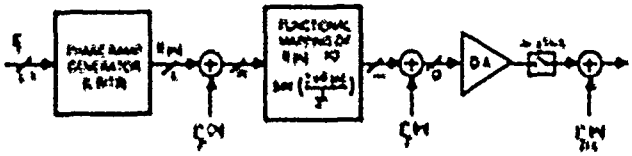
Figure 5.

CH2427-3/87/0000-495 $1.00 © 1987 IEEE

Figure 3 shows the output spectrum when the synthesizer is tuned to a frequency of .1000 times the clock frequency, F_{clk}. In contrast, Figure 4 shows the output spectrum for the next higher frequency resolvable by the synthesizer, .1992 F_{clk}. It is evident from comparing the spectrums in Figure 3 to Figure 4 that a change to the synthesizer's output frequency of only .12%, the smallest frequency change resolvable by this design, results in a completely different spurious response. The third plot in Figure 5 shows a change of only 1.1% (4 times the smallest resolvable frequency difference) and yet displays another response totally different from the first two. The dramatic change in the character of the response for all three cases demonstrates the importance of being able to explicitly describe the response of a DDFS for any combination of output frequency, ROM storage capacity, and phase accumulator length.

II. ANALYSIS

The operation of all direct digital frequency synthesizers that use some variation of the look-up table technique can be analyzed mathematically using the model shown in Figure 6. Although many techniques exist for performing the functional mapping from phase to sine, (i.e. Taylor series, ROM look-up, course/fine segmentation into multiple ROMs) the following model is valid for all methods not possessing feedback or memory elements other than pipeline registers.



W = Word length of phase accumulator output used to address ROM.
L = Length in bits of phase accumulator register word.
D = Word length of sine values stored in look-up table.
F_t = Frequency control word.

Figure 6. Sources of noise in DDFS's

Contained within this model are three sources of distortion. The first source, ε_{DAC}, is due to the DAC conversion noise at the output. The second, ε_T, is due to the finite wordlength of the sine samples stored in the ROM, and the third, ε_P, is due to the truncation of phase accumulator bits addressing the ROM. The most important of these finite wordlength effects is the truncation of phase accumulator bits. There are several reasons for this. First, in order to minimize ROM storage requirements, one would like to truncate as many of the phase bits as possible without incurring too much spurious distortion. In addition, the spurious response due to ε_T and ε_{DAC} can be controlled by the designer with less of a hardware penalty, since the hardware complexity is a linear function of the wordlength of both the ROM output and DAC output, while it is an exponential function of the ROM address length. Finally, the characterization of the spurious response due to phase truncation is important because it involves a detailed analysis of the interaction of the phase accumulator output sequence and its properties under quantization, which once understood, yields insight into the spurious response due to all finite wordlength effects in the DDFS.

Efforts have been made to characterize the effects of phase accumulator truncation using statistical methods. These types of analyses have the drawback that results are only obtained for the average spur energy rather than for the magnitude of the worst case spur. Furthermore, a stochastic analysis gives no information about the relative distribution of large spurs in the output spectrum. This information is extremely important because some combinations of ROM table size, phase accumulator word length, and frequency control register contents result in worst case spurs occurring very close to the desired output frequency.

The development of the algorithm for the generation of the spurious spectrum of a DDFS will be divided into three sections. First, some observations are made on the behavior of the phase accumulator. Then, an algorithm is developed for the determination of the spectrum of the phase error sequence. Finally, the analysis techniques and results gained from generating the spectrum of the phase error are applied to develop an algorithm for the generation of the spectrum of the DDFS due to phase truncation.

PHASE ACCUMULATOR OPERATION

The behavior of the spurious spectrum a DDFS for different values of the frequency control word, F_t, is intrinsically linked to the numerical properties of the phase accumulator. For this reason, we will begin the analysis of the DDFS spectral response with a description of phase accumulator operation.

The phase accumulator operates on the principle of overflow arithmetic using the modulo 2^L property of an L bit periodically overflowing accumulator register to simulate the modulo 2π property of the sine function by exploiting the relationship

$$\sin\left(2\pi\frac{\langle\theta(n)\rangle_{2^L}}{2^L}\right) = \sin\left(2\pi\frac{\theta(n)}{2^L}\right) \qquad (1)$$

where $\langle\cdot\rangle_{2^L}$ represents taking the integer residue of a number modulo 2^L. The phase accumulator thus acts as a digital integrator followed by a modulo 2^L operator. The frequency of the output of the DDFS is related to the binary frequency control word by the relationship

$$F_o = \frac{\omega_o}{2\pi} = \frac{\Delta\theta(n)}{\Delta t}\frac{1}{2^L} = \frac{F_t F_{clk}}{2^L} \qquad (2)$$

Due to the discrete-time nature of the system, values of F_t larger than 2^L will result in an aliased output frequency equal to $F_{clk} - F_o$. Thus there are $2^L/2$ discrete output frequencies that can be generated, each corresponding to a unique value of F_t less than 2^{L-1}.

The phase sequence $\theta(n)$ generated by the phase accumulator can be represented as the samples of an idealized sawtooth waveform, where, after normalizing with $F_{clk}=1$, the slope of the sawtooth is given by

$$\frac{\Delta\theta(n)}{\Delta t} = F_t \qquad (3)$$

One concept that is key to the theoretical understanding of phase accumulator operation is the relationship between the period of the idealized continuous time waveform and the period of the discrete time phase accumulator output sequence. As shown in Figure 7, the phase accumulator outputs the samples of a hypothetical continuous time sawtooth waveform of amplitude 2^L and period $2^L/F_t$. Since the sine function is computed from the phase sequence using combinational logic without feedback, the period of the synthesized analog sinewave at the output of the DAC is also $2^L/F_t$. The numerical period of the phase accumulator output sequence is defined as the minimum value of N for which $\theta(n) = \theta(n+N)$ for all n. In general, the numeric period of the phase accumulator sequence is given by $2^L/(F_t,2^L)$, where $(F_t,2^L)$ represents the greatest common divisor of F_t and 2^L. Thus the period of the sampled sawtooth will equal the period of the idealized continuous time sawtooth only for the special case when $F_t = (F_t,2^L)$, i.e., when F_t is purely a power of two. This behavior is illustrated for the case of L=6 and F_t=17 in Figure 7. In this figure it is evident that the normalized output frequency of the continuous-time sawtooth wave is $2^L/F_t = 17/64 = .26563$, resulting in a period of $64/17 = 3.7647$. In contrast, however, the period of the numeric sequence of samples of this waveform is $64/(17,64) = 64$. The most important consequence of this phenomenon is that the numerical period of the sequence of samples recalled from the sine ROM will have the same value as the numerical period of the sequence generated by the phase accumulator. Therefore, the spectrum of the output waveform of the DDFS prior to digital-to-analog conversion is characterized by a discrete spectrum consisting of $2^L/(F_t,2^L)$ points.

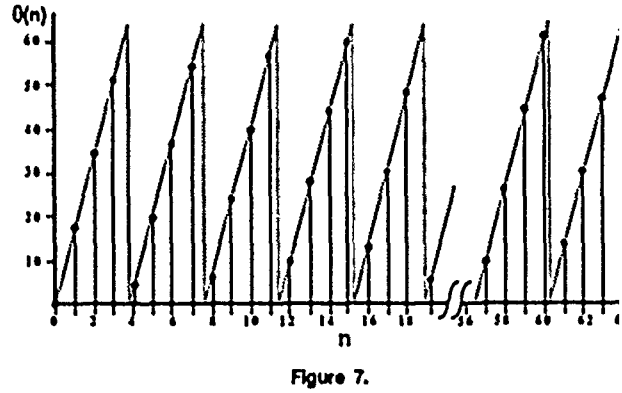


Figure 7.

SPECTRUM OF THE PHASE ERROR SEQUENCE

In the ideal case of infinite precision in the look-up table word length and no phase quantization, the output sequence of a DDFS is given by

$$S(n) = \sin\left(2\pi\frac{F_t}{2^L}n\right)$$

which yields the sampled values of a sinusoid of frequency $F_f/2^B$. We now consider the operation of a DDFS in which the output of the phase accumulator is quantized to W bits by truncation. Under this condition, where we define B to be the number of bits truncated such that L - W = B, the output sequence of the DDFS is given by

$$S_f(n) = \sin(2\pi\frac{2^B}{2^L}\left[\frac{F_f}{2^B}n\right]) \qquad (5)$$

where the operator [ ] represents truncation to integer values. Equation (5) can alternatively be expressed as

$$S_f(n) = \sin(\frac{2\pi}{2^L}(F_f n - \mathcal{E}_f(n))) \qquad (6)$$

where $\mathcal{E}_f(n)$ represents the phase error sequence. This phase error sequence can be modeled as the sampled values of a continuous-time sawtooth waveform which will be referred to as $\mathcal{E}_f(t)$. The amplitude of the sawtooth waveform, $\mathcal{E}_f(t)$, is 2^B, and the frequency is $F_f/2^B$ as exemplified in Figure 8.

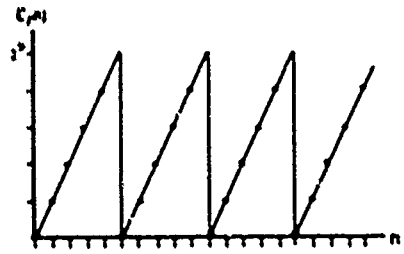


Figure 8. Phase accumulator error sequence

This sawtooth waveform is identical to the waveform that would be generated by a phase accumulator of word length B with a input frequency control command of $(F_f)_{2B}$.

In order to characterize the spectral properties of the error sequence, $\mathcal{E}_f(n)$, the Fourier series must be obtained for the idealized continuous-time representation of the error waveform $\mathcal{E}_f(t)$. Unfortunately the sawtooth waveform defined above is specified to have a value of zero at the discontinuity which is in violation of the required Dirichlet conditions for a Fourier representation. Dirichlet conditions require that a function take on its average value at a point of discontinuity, which in this case is $2^B/2$.

Therefore, in order to represent the phase error as a Fourier series, the error waveform must be reexpressed as the superposition of an ideal "Dirichlet" sawtooth and a correcting waveform, such that the both waveforms satisfy Dirichlet conditions and such that the sampled values of the superposition of the two waveforms equals the actual error sequence, $\mathcal{E}_f(n)$.

SAWTOOTH: B=2 F_f=3 frequency=3/4

PULSE TRAIN: B=2 F_f=3 frequency=3/4 pulse width=1/3

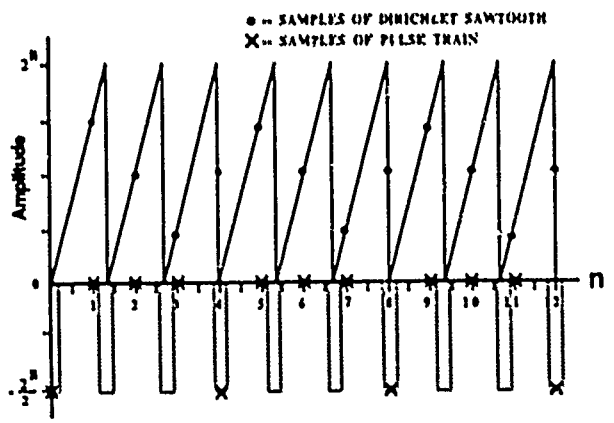


Figure 9.

The correcting waveform that is added to the Dirichlet sawtooth to make the sampled values agree with the actual phase error samples is a pulse train of frequency $F_f/2^B$ with each pulse having a width of $(2^B,F_f)/(2^B)$ as shown in Figures 9 and 10. This choice of frequency and pulse-width gives the correcting waveform two important properties. first, that pulses are only sampled at the midpoint of the pulse, and second, that the only pulses sampled are those whose centers lie exactly on the discontinuity of the sawtooth.

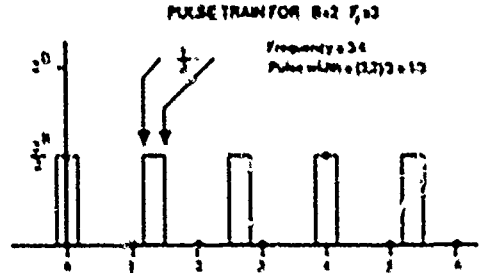


Figure 10.

The Fourier series for $\mathcal{E}_f(t)$ can now be expressed as the sum of the individual Fourier series for the "Dirichlet sawtooth" and the pulse train waveforms that were depicted in Figure 9. This representation is given in equation (7) where the sawtooth waveform is represented as a pure sine series and the pulse train as a pure cosine series (ignoring the D.C. term). In the term containing the cosine series the substitution $\Lambda = 2^B/2(2^B,F_f)$ is introduced, where Λ is equal to one half the reciprocal of the pulse train duty cycle.

$$\mathcal{E}_f(t) = \sum_{K=1}^{\infty}\frac{2^B}{\pi K}\sin(2\pi K\frac{F_f}{2^B}t) - \frac{2^B}{2\Lambda}\sum_{K=1}^{\infty}\frac{4\Lambda}{2\pi K}\sin(\frac{2\pi K}{4\Lambda})\cos(2\pi K\frac{F_f}{2^B}t) \qquad (7)$$

This Fourier series may be sampled at integer time points to obtain

$$\mathcal{E}_f(n) = \sum_{K_o=1}^{\infty}\frac{2^B}{\pi K_o}\sin(2\pi K_o\frac{F_f n}{2^B}) - \frac{2^B}{2\Lambda}\sum_{K_o=1}^{\infty}\frac{4\Lambda}{2\pi K_o}\sin(\frac{2\pi K_o}{4\Lambda})\cos(2\pi K_o\frac{F_f n}{2^B}) \qquad (8)$$

Because of the discrete nature of the time index, the error sequence, $\mathcal{E}_f(n)$, will be periodic with a period of at most 2^B. If F_f contains common divisors with 2^B, the sequence will repeat every $2^B/(F_f,2^B)$ samples. This periodicity in the time domain will result in a sampled frequency domain representation whose magnitude contains only $2^B/2(F_f,2^B)$ unique points because the magnitude spectrum of a real sequence is symmetric about the $F_{out} = 0$ axis. Therefore, $\Lambda = 2^B/2(F_f,2^B)$ as defined in (7) can be interpreted as the number of unique spectral noise lines in the discrete magnitude spectrum of $\mathcal{E}_f(n)$. By exploiting the fact that the sine and cosine functions in (8) are periodic in K_o with period $2^B/2(F_f,2^B)$, the sine and cosine series can be rewritten as double summations, where the inner summation enumerates the coefficients of the sine and cosine functions in (8) over one period of K_o.

If we define $\mathcal{E}_{sf}(n)$ as the sine series summation in the left term of (8) we obtain

$$\mathcal{E}_{sf}(n) = \frac{2^B}{\pi}\sum_{g=1}^{\infty}\left(\sum_{K=1}^{\Lambda-1}\frac{1}{(K+2\Lambda g)}\sin(2\pi(K+2\Lambda g)\frac{F_f}{2^B}n) + \sum_{K=1}^{\Lambda-1}\frac{1}{(-K+2\Lambda g)}\sin(2\pi(-K+2\Lambda g)\frac{F_f}{2^B}n)\right) + \frac{2^B}{\pi}\sum_{K=1}^{\Lambda-1}\frac{1}{K}\sin(2\pi K\frac{F_f}{2^B}n) + \frac{2^B}{\pi}\sum_{g=1}^{\infty}\frac{1}{g\Lambda}\sin(2\pi g\Lambda\frac{F_f}{2^B}n) \qquad (9)$$

Using trigonometric identities and noting that $F_f/(F_f,2^B)$ is always an integer, and by interchanging the summations for g and K, equation (9) can be simplified to

$$\mathcal{E}_{sf}(n) = \frac{2^B}{2\Lambda}\sum_{K=1}^{\Lambda-1}\left(\left(\frac{2\Lambda}{K\pi} - 2\left(\frac{K\pi}{2\Lambda}\right)\sum_{g=1}^{\infty}\left((g\pi)^2 - \left(\frac{K\pi}{2\Lambda}\right)^2\right)^{-1}\right)\sin(2\pi K\frac{F_f}{2^B}n)\right) \qquad (10)$$

Using Jolley (3) (Eqn 769) and noting that the argument of the summation is zero at K = Λ, equation (10) can be rewritten without loss of generality as

$$\varepsilon_{sp}(n) = \frac{2^B}{2\Lambda}\sum_{K=1}^{\Lambda}\left(\cot\left(\frac{K\pi}{2\Lambda}\right)\sin\left(2\pi K\frac{F_f}{2^B}n\right)\right) \qquad (11)$$

Thus the sine portion of the phase error defined in (8) is expressed as a numerical summation of Λ different harmonics of the sampled sinewave $\sin(2\pi F_f n/2^B)$, with each amplitude determined by the coefficient $\cot(K\pi/2\Lambda)$.

Returning to the analysis of the Fourier series for $\varepsilon_s(n)$ in (8), let $\varepsilon_{cp}(n)$ be defined as the cosine part of the phase error sequence. This sequence is now rewritten in a manner similar to that in (9) as a double summation over K and g as follows.

$$\varepsilon_{cp}(n) = \frac{-2^B}{2\Lambda}\sum_{g=1}^{\infty}\left(\sum_{K=1}^{\Lambda-1}\operatorname{sinc}\left(\frac{\pi K+2\pi\Lambda g}{2\Lambda}\right)\cos\left(2\pi K\frac{F_f}{2^B}n\right) + \sum_{K=1}^{\Lambda-1}\operatorname{sinc}\left(\frac{-\pi K+2\pi\Lambda g}{2\Lambda}\right)\cos\left(2\pi K\frac{F_f}{2^B}n\right)\right)$$
$$- \frac{2^B}{2\Lambda}\sum_{K=1}^{\Lambda-1}\operatorname{sinc}\left(\frac{\pi K}{2\Lambda}\right)\cos\left(2\pi K\frac{F_f}{2^B}n\right) - \frac{2^B}{\pi}\sum_{g=1}^{\infty}\operatorname{sinc}\left(\frac{\pi g}{2}\right)\cos\left(2\pi g\Lambda\frac{F_f}{2^B}n\right) \qquad (12)$$

After trigonometric and algebraic manipulation, and using Jolley (eqn 770, pg. 146), we obtain

$$\varepsilon_{cp}(n) = \frac{-2^B}{2\Lambda}\sum_{K=1}^{\Lambda-1}\left(\cos\left(2\pi K\frac{F_f}{2^B}n\right) + \frac{1}{2}\cos\left(2\pi\Lambda\frac{F_f}{2^B}n\right)\right) \qquad (13)$$

In order to obtain a more concise representation of $\varepsilon_{cp}(n)$ that is in the same form as $\varepsilon_{sp}(n)$ in (11), the $\cos(2\pi\Lambda F_f n/2^B)$ term of the summation in (13) is multiplied by a "correction factor" of two. This operation is justified later by showing that the K = Λ coefficient is always doubled due to aliasing. Using this correction factor, $\varepsilon_{cp}(n)$ is rewritten as

$$\varepsilon_{cp}(n) = \frac{-2^B}{2\Lambda}\sum_{K=1}^{\Lambda}\cos\left(2\pi K\frac{F_f}{2^B}n\right) \qquad (14)$$

Upon substituting back into (8) and using (11) we obtain

$$\varepsilon_p(n) = \frac{-2^B}{2\Lambda}\sum_{K=1}^{\Lambda}\left(\cot\left(\frac{K\pi}{2\Lambda}\right)\sin\left(2\pi K\frac{F_f}{2^B}n\right) - \cos\left(2\pi K\frac{F_f}{2^B}n\right)\right) \qquad (15)$$

The error sequence is now expressed as a sum of Λ distinct frequencies with unique amplitudes corresponding to the discrete spectrum of the phase error. By combining the terms from the sine series and cosine series into a complex Fourier series representation we obtain

$$\varepsilon_p(n) = \sum_{K=1}^{\Lambda}\zeta_K\, e^{j\left(2\pi K\frac{F_f}{2^B}n\right)}\, e^{j\Psi(K,\Lambda)} \qquad (16)$$

where the magnitude of the complex phasor, ζ_K, and the phase angle, $\Psi(K,\Lambda)$, are defined as

$$\zeta_K \triangleq \frac{2^B}{2\Lambda}\operatorname{cosec}\left(\frac{K\pi}{2\Lambda}\right) \qquad \Psi(K,\Lambda) \triangleq -\cot\left(\frac{K\pi}{2\Lambda}\right) \qquad (17)$$

The variable $\Gamma = F_f/(F_{clk}2^B)$ is now defined so that the frequency $F_f/2^B$ can be expressed as a rational fraction of two mutually prime integers. Using this definition, each of the spur magnitudes, ζ_K, corresponds to a unique frequency $K\Gamma(F_{clk}2^B)/2^B$. This, however, is only a partial solution, since an explicit result is required that will define the exact location of the spur frequencies F_{sp} in the range $0 < F_{sp} < F_{clk}/2$. In general $K\Gamma(F_{clk}2^B)/2^B$ will range over several decades of the frequency range of interest. Because aliasing, the position of each spur in the spectral range $0 < (F_{sp}/F_{clk}) < 1$ is equal to the residue of the frequency $K\Gamma$ modulo $2^B/(F_{clk}2^B)$.

We now introduce the normalized integer frequency, F_n, which is defined such that all frequencies in the discrete spectrum (as would be generated by a DFT) are represented by integers. From the definition of the DFT, we know that the discrete spectrum of a sequence contains the same number of points that are guaranteed to be non-zero as the numerical period of the sequence. Using this and realizing that the period of the sequence $\varepsilon_p(n)$, is 2Λ, we define the relationship between the un-normalized output frequency and F_n as $F_n = \frac{f_{sp}2\Lambda}{f_{clk}}$. Where this definition guarantees that all possible spur locations in the spectrum are represented by integer values.

Using this definition for frequency axis of the discrete spectrum we can determine the frequency of the K-th spur from

$$\left\langle K\Gamma\right\rangle_{2\Lambda} = F_n \qquad (18)$$

Although (18) calculates the frequency that each spur, ζ_K, maps to in the range $0 < (F_{sp}/F_{clk}) < 1$, a much more powerful description of the spectrum would yield the amplitude (derived from K) as a function of the frequency, F_f. Furthermore, when using (18), it must be ensured that more than one value of K does not yield the same F_n. Thus an assurance must be gained that each of the frequencies represented by the summation in (16) correspond to a unique frequency in the range $0 < (F_{sp}/F_{clk}) < 1$, and, given this assurance, a solution to the congruence in (18) for K must be obtained given any F_{sp} in that range. The congruence in (18) is actually a representation of the integral Diophantine equation,

$$K\Gamma + bY = F_n \qquad (19)$$

where Y is an integer, $b = 2\Lambda$ is an integer, and all solutions are defined to be an equivalence class modulo b. Because it is a Diophantine equation, integral values of K that solve (19) can be found if and only if $(F_n)_{(\Gamma,b)}=0$. From number theory, by the coprimality of Γ and $2^B/(F_{clk}2^B)$, it can be shown that each set of K's less than $2^B/(F_{clk}2^B)$ will map into a unique set of F_n's less than $2^B/(F_{clk}2^B)$. Thus, each frequency will map uniquely into the range $0 < (F_{sp}/F_{clk}) < 1$. Therefore, exploiting the fact that $(\Gamma,2^B/(F_{clk}2^B)) = 1$, the solution to (18) can be expressed without loss of generality by either Euclid's algorithm or by Euler's Theorem. From Euler's Theorem, $(F_f^{\Phi(x)})_x=1$ for all F_f and X coprime, where $\Phi(x)$ is defined as the Euler Totient function. Therefore by multiplying both sides of (18) by $F_f^{\Phi(x)}$ with $X=2^B/(F_{clk}2^B)$ and further factoring out F_f one obtains:

$$\left\langle F_n\Gamma^{\Phi(2\Lambda)-1}\right\rangle_{2\Lambda} = K \qquad (20)$$

where $\Phi(n)$ is the Euler totient function which is defined as the number of integers smaller than n which are relatively prime to n. There is no explicit function that can be used to calculate $\Phi(n)$, therefore a table must normally be used to obtain its values. Fortunately, the Totient function can be easily obtained for the special case of n strictly a power of 2. Thus for $n = 2^x$ where x is an integer; all even numbers less than n will not be mutually coprime to n and all odd numbers less than n will be coprime to n. Therefore the Totient function for $n=2^x$ is simply the number of odd integers less than n, which is $n/2$ or 2^{x-1}. In the above analysis, $n=2^B/(F_{clk}2^B)$, which is by definition always a pure power of 2. Therefore we obtain

$$\Phi\left(\frac{2^B}{(F_{clk}2^B)}\right) = \frac{2^B}{2(F_{clk}2^B)} = \Lambda \qquad (21)$$

and thus,

$$K = \left\langle F_n\Gamma^{\Lambda-1}\right\rangle_{2\Lambda} \qquad (22)$$

Therefore, by using (22) to get the value of K to substitute into (18), the amplitude of the spurious response can be calculated at any point in the digital spectrum. A problem, however, arises when one considers that (22) was derived from (18), which maps each spur from a spectrum in which each ζ_K lies on a frequency $F_{sp} = K\Gamma(F_{clk}2^B)/2^B$ to one in which all of the spurs lie in the frequency range $0 < F_{sp}/F_{clk} < 1$. This "post mapped" frequency domain can be thought of as consisting of 2Λ possible spur locations. In the "post mapped" spectrum the spur location denoted by $F_n = 2\Lambda$ corresponds to the frequency F_{clk}, and the location $F_n = \Lambda$ corresponds to $F_{clk}/2$. This mapping is shown graphically in Figure 11 for the case of 3 bits of truncation and $F_f = 11$.

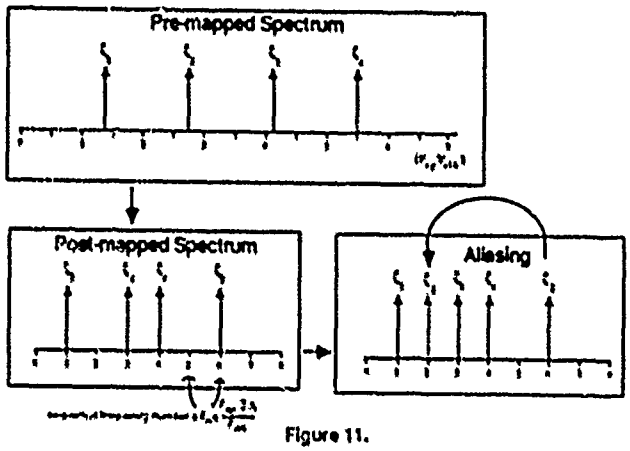


Figure 11.

In some cases one or more of the Λ different spurs will map onto the frequency numbers between $F_n=\Lambda$ and $F_n=2\Lambda$. This is illustrated in Figure 11 by the spur which was mapped to $F_n = 6$, as predicted by (18). One might question then, how a value for K could be extracted using (22) for all values of F_n in the range $0 < F_n < \Lambda$ (corresponding to $0<F_{sp}<F_{clk}/2$) when the mapping defined by (18), from which (22) was derived, leaves some of the spur locations in that range empty. This is exemplified by the empty $F_n=2$ location for the example in Figure 11. The answer to this question comes from sampling theory. Because the spur location $F_n=\Lambda$ actually corresponds to the frequency $F_{clk}/2$, the spur lying on $F_n=6$ in the "post mapped" spectrum in Figure 11 is aliased back to the frequency $F_n=2$, as shown in the "Aliasing" box in Figure 11.

Ideally, it would be desirable to associate a spur number directly with each spur location in the range $0 < F_{sp} < F_{clk}/2$, rather than calculating the entire spectrum and then aliasing back the spurs in the range $F_{clk}/2<F_{sp}<F_{clk}$. To calculate the value of K for the spurs lying on these "empty" spur locations, an aliasing phenomenon is used similar to that encountered in sampling theory. Since the summation expressing the number of spurs is defined to have Λ terms, and since the modulus operator in (22) is defined modulo 2Λ, calculations using (22) can yield a value of K for an empty F_n that is larger than Λ (which is impossible since there are only Λ different ζ_K's). This is similar to the aliasing of sampling theory because the frequency associated with the spur, $K\Gamma(F_f,2^B)/2^B$ can be associated with $(2\Lambda-K)\Gamma(F_f,2^B)/2^B$ for $\Lambda > K > 2\Lambda$. Therefore, if (22) yields a value for K that is in the range $\Lambda > K > 2\Lambda$, the true K value can be computed by subtracting the old K from 2Λ. This eliminates the need to determine whether a frequency location was empty or not when applying (22). In order to find the spur magnitude from the frequency number, simply use (22) and subtract K from Λ for K's greater than Λ. Fortunately, the expression for the spur amplitudes given in (17) is symmetric about the $K = \Lambda$ axis, so the aliased values of K generated by (22) produce the correct amplitudes when substituted into (17) without calculating $K - 2\Lambda$ for values of K in the range $\Lambda > K > 2\Lambda$.

The procedure for obtaining the discrete spectrum for the phase error function, $\varepsilon_f(n)$, can now be summarized as follows.

1. Find the number of spurs from $\Lambda = 2^{B-1}/(F_f,2^B)$ where:
   B = number of bits truncated
   F_f = contents of the frequency control register
2. Divide the frequency range of 0 to $F_{clk}\cdot 2(F_f,2^B)$ into Λ equally spaced spur locations.
3. Find the spur number K for ζ_K from the sequential spur number F_n using the equation:
$$K = \left\langle F_n\Gamma^{\Lambda-1}\right\rangle_{2\Lambda}$$
4. Calculate the amplitude of the spur using the value of K derived in step 3 from:
$$\zeta_K = \frac{2^B}{2\Lambda}\operatorname{cosec}\left(\frac{K\pi}{2\Lambda}\right)$$

DDFS Output Spectrum Due to Phase Truncation

Now that an algorithm has been developed for computing the discrete spectrum of the phase error, $\varepsilon_f(n)$, we can determine the effect of this phase error sequence on the output spectrum of the DDFS. Applying simple trigonometric identities, the ROM output sequence, $S_t(n)$, of (6) can be written as

$$S_t(n) = \sin\left(2\pi\frac{F_f n}{2^L}\right)\cos\left(2\pi 2^{B-L}\frac{\varepsilon_f(n)}{2^B}\right) - \cos\left(2\pi\frac{F_f n}{2^L}\right)\sin\left(2\pi 2^{B-L}\frac{\varepsilon_f(n)}{2^B}\right) \quad (23)$$

Because $\left|\frac{\varepsilon_f(n)}{2^B}\right| \le 1$ for all $\varepsilon_f(n)$, the supremum of the arguments to the sine and cosine terms containing $\varepsilon_f(n)$ in (23) is $2\pi 2^{B-L}$. Since it is reasonable to assume that $2^{B-L} << 1$ for all realistic DDFS implementations, (23) can be approximated with very little error by

$$S_t(n) = \sin\left(2\pi\frac{F_f n}{2^L}\right) - 2\pi\frac{\varepsilon_f(n)}{2^L}\cos\left(2\pi\frac{F_f n}{2^L}\right) \quad (24)$$

Therefore, to a reasonable approximation, the output spectrum of the DDFS will be composed of a sinewave at the desired output frequency corrupted by the cosine modulated harmonics of the sawtooth waveform $\varepsilon_f(n)$. Substituting the summation for the phase error spectrum obtained in (16) for $\varepsilon_f(n)$ yields the result

$$S_t(n) = \sin\left(2\pi\frac{F_f n}{2^L}\right) - \frac{2\pi}{2^L}\left(\sum_{K=1}^{\Lambda}\zeta_K e^{j(2\pi K\frac{F_f}{2^B}n)} e^{j\Psi(K,\Lambda)}\right)\cos\left(2\pi\frac{F_f n}{2^L}\right) \quad (25)$$

which can be rewritten as

$$S_t(n) = \sin\left(2\pi\frac{F_f n}{2^L}\right) - \frac{2\pi}{2^L}\sum_{K=1}^{\Lambda}\zeta_K\left(e^{j2\pi\left(\frac{F_f}{2^L}+K\frac{F_f}{2^B}\right)n} + e^{-j2\pi\left(\frac{F_f}{2^L}-K\frac{F_f}{2^B}\right)n}\right)e^{j\Psi(K,\Lambda)} \quad (26)$$

The spurious response now consists of two series of spectral lines whose amplitudes are proportional to the same set of ζ_K's that were derived for the the spectrum of the sawtooth. However, the discrete spectrum will consist of $2^L(F_f,2^L)$ points onto which the two sets of Λ different ζ_K's will be aliased. Because 2Λ is in general less than $2^L(F_f,2^L)$ for $L > B$, there will always be points in the discrete spectrum containing no aliased spurs. Additionally it is possible for a single frequency to contain a noise line that results from the superposition of two different aliased ζ_K's. The variable $\zeta_{K\pm}$ is now defined to be the magnitude of the individual spurs in the output spectrum of a DDFS. This is expressed in (27) as a function of ζ_K, the magnitude of the spurs in the spectrum of the phase error sequence $\varepsilon_f(n)$.

$$\zeta_{K\pm} = \frac{\pi}{2^L}\zeta_K = \frac{\pi 2^{B-L}}{2\Lambda}\operatorname{cosec}\left(\frac{K\pi}{2\Lambda}\right) \quad (27)$$

As part of the definition in (27) we define notation ζ_{K+} to represent the spurs mapped from the first term, $e^{j2\pi(F_f/2^L + KF_f/2^B)}$, in (26), and ζ_{K-} to be the spurs mapped from the $e^{-j2\pi(F_f/2^L - KF_f/2^B)}$ term. The location of these spurs can be calculated using the same techniques derived earlier for the sawtooth spectrum. However the procedure for calculating the spectrum is now complicated by the fact that a noise spur does not lie on each frequency location. Therefore, in addition to the calculation of spur amplitudes for a given frequency, a criteria for determining if a spur exists at all at a given F_n must be included in the algorithm. To accomplish this, the frequency range $0<F_{sp}<F_{clk}$ must be divided into $2^L(F_f,2^L)$ frequency locations. As a result, F_n is now redefined to be the integer frequency number of the frequency location $0<F_n<2^L(F_f,2^L)$ where $F_n = 2^L F_{sp}/F_{clk}(F_f,2^L)$. If equation (26) is rewritten to express the frequency $F_f/2^L$ as the ratio of two mutually prime integers using $\Gamma = F_f/(F_f,2^B)$ we obtain

$$S_t(n) = \sin\left(2\pi\frac{F_f}{2^L}n\right) - \sum_{K=1}^{\Lambda}\zeta_{K\pm}\left(e^{j2\pi\frac{(F_f,2^B)}{2^L}(\Gamma+K\Gamma 2^{L-B})n} + e^{-j2\pi\frac{(F_f,2^B)}{2^L}(\Gamma-K\Gamma 2^{L-B})n}\right)e^{j\Psi(K,\Lambda)} \quad (28)$$

From (28) it is evident that each $\zeta_{K\pm}$ is mapped onto two different frequencies with frequency numbers given by

$$F_n = \left\langle\Gamma + 2^{L-B}K\Gamma\right\rangle_{2^L} \quad \text{or} \quad F_n = \left\langle\Gamma - 2^{L-B}K\Gamma\right\rangle_{2^L} \quad (29)$$
$$(F_f,2^{L-B}) \qquad\qquad (F_f,2^{L-B})$$

Thus using (29) it can be determined, for a given K, onto what frequency locations ζ_{K+} and ζ_{K-} are mapped. This is helpful in determining where in

the frequency spectrum the largest spurs lie. However, normally when evaluating the spectral response, one is more interested in determining the magnitude of the response for a given frequency. Thus, to completely mathematically characterize the spectrum of the DDFS one would like to determine for a given frequency whether any spur exists there, and if so, whether that spur is due to the superposition of one or more $\zeta_{K\pm}$'s and the value of K for each one. From (29) it is evident that spurs exist only at frequency numbers for which the fraction $(F_n - \Gamma)/2^{L-B}$ or $(F_n + \Gamma)/2^{L-B}$ is an integer, or equivalently, for which the following congruences are satisfied

$$\langle F_n \rangle_{2^{L-B}} = \langle \Gamma \rangle_{2^{L-B}} \quad \text{or} \quad \langle F_n \rangle_{2^{L-B}} = -\langle \Gamma \rangle_{2^{L-B}} \tag{30}$$

If a spur does exist at a given frequency it is possible that this spur is due to the superposition of two $\zeta_{K\pm}$'s. There are in general two ways in which $\zeta_{K\pm}$'s can overlap. The first is when the spurs from one family are aliased on top of spurs from another family (i.e. when a ζ_{K+} spur lies on top of a ζ_{K-} spur). The second way is when spurs from the same family are aliased on top of each other (i.e. ζ_{K-} on top of $\zeta_{K'-}$ or ζ_{K+} on top of $\zeta_{K'+}$). In general ζ_{K+} and ζ_{K-} will overlap only when one of the following equivalence relations are satisfied:

$$\langle F_f - 2^{L-B}K_1F_f \rangle_{2^L} = \langle F_f + 2^{L-B}K_2F_f \rangle_{2^L} \tag{31}$$

or

$$\langle F_f - 2^{L-B}K_1F_f \rangle_{2^L} = \langle F_f + 2^{L-B}K_2F_f \rangle_{2^L} \tag{32}$$

Equation (31) can be rearranged and simplified to obtain

$$\left\langle \langle 2^{L-B-1} \rangle_{2^{L-1}} \langle K_1 - K_2 \rangle_{\frac{2^{L-1}}{(F_n,2^{L-1})}} \right\rangle_{\frac{2^{L-1}}{(F_n,2^{L-1})}} = 1 \tag{33}$$

It can be seen that (33) can only be satisfied when 2^{L-B-1} is an odd number. Since 2^{L-B-1} is odd only for $L - B - 1 = 0$, the overlap will occur when L and B satisfy $L - B = 1$. A second condition for equivalence in (33) is that $K_1 - K_2 = 1$. Therefore for the case of $L - B = 1$ (corresponding physically to a DDFS with all but the most significant bit truncated) the ζ_{K+} spurs are always aliased on top of the ζ_{K-} spurs. Thus the only spurs that do not fall on top of other spurs for this case are ζ_1 and ζ_Λ. Fortunately the special case of $L - B = 1$ is a condition that would never be realized in a practical DDFS.

The second condition for ζ_{K+} and ζ_{K-} overlap as expressed in (32) can be rearranged and simplified to obtain

$$\langle K_1 + K_2 \rangle_{2\Lambda} = 0 \tag{34}$$

Since $MAX(K_1) = MAX(K_2) = \Lambda$, (34) is satisfied only for

$$K_1 + K_2 = 2\Lambda \tag{35}$$

In other words overlap always occurs for the case of $K_1=\Lambda$ and $K_2=\Lambda$. Therefore independent of the conditions on L and B, $\zeta_{\Lambda+}$ is always aliased on top of $\zeta_{\Lambda-}$. This means that when calculating the magnitude of the spectral response at a frequency where it is determined that $\zeta_{\Lambda+}$ resides, the calculated magnitude of $\zeta_{\Lambda+}$ must be doubled since $\zeta_{\Lambda-}$ will always be aliased on top of it. This, it turns out, is the justification for the factor of 2 "correction factor" that multiplied the $(1/2)\cos(2\pi\Lambda F_f n/2^B)$ term in (15) to get the summation in (16). Therefore no provision needs to be made for the doubling of the $\zeta_{\Lambda\pm}$ terms when (27) is used to calculate the spur amplitudes.

As mentioned earlier, spurs can also overlap when spurs from the same family are aliased on top of one another (ζ_{K+} on top of $\zeta_{K'+}$ or ζ_{K-} on top of $\zeta_{K'-}$). The conditions for this type of overlap are expressed as

$$\langle F_f + 2^{L-B}KF_f \rangle_{2^L} = \langle F_f + 2^{L-B}(K+n)F_f \rangle_{2^L} \qquad n = \pm 1, 2, 3, \ldots, \Lambda \tag{36}$$

or

$$\langle F_f + 2^{L-B}KF_f \rangle_{2^L} = 2^L - \langle F_f + 2^{L-B}(K+n)F_f \rangle_{2^L} \qquad n = \pm 1, 2, 3, \ldots, \Lambda \tag{37}$$

where spurs of the same family overlap if and only if either (35) or (37) is satisfied. Equation (31) can be rewritten as

$$\langle K \rangle_{\frac{2^B}{(F_n,2^B)}} = \langle K + n \rangle_{\frac{2^B}{(F_n,2^B)}} \tag{38}$$

Therefore the equivalence in (37) is satisfied only for n a multiple of $2^B/(F_n,2^B)$. Since n is bounded by Λ because there are only Λ spurs, (31) yields no self overlap conditions.

The other condition for self overlap, given by (37), can be rearranged and simplified to obtain

$$\langle 2^{L-B-1}(2K+n) \rangle_{\frac{2^L}{(F_n,2^L)}} = \langle n \rangle_{\frac{2^L}{(F_n,2^L)}} \tag{39}$$

Since -1 is an odd number, equivalence (39) can only be satisfied if 2^{L-B-1} is an odd number. 2^{L-B-1} is odd only for $L-B-1=0$ or if $L=B+1$. This yields the result that self overlap only exists for the trivial case of truncation of all bits addressing the ROM except the phase accumulator MSB.

Therefore, all of the conditions for spur overlap have now been exhaustively determined. Summarizing the results, it has been determined that there is always overlap among spurs of the same family for the condition $L-B=1$, a condition that is not realized for any practical DDFS design. Furthermore, other than for the case $L-B=1$, there is no overlap between spurs of the same family. For the case of overlap among spurs of different families it has been shown that the spurs $\zeta_{\Lambda+}$ and $\zeta_{\Lambda-}$ overlap under all conditions and that for the special case of $L-B=1$ all of the spurs are aliased on top of one another. In conclusion, for any practical DDFS design, it has also been proven that the only spur overlap in the spectrum will be between the $\zeta_{\Lambda+}$ and the $\zeta_{\Lambda-}$ Spurs. It has also been shown that when calculating the spectrum of a DDFS no adjustment has to be made to add these two spurs together since their values are "pre-doubled" when their amplitude is calculated using (27).

Now that the conditions for spur overlap and the criteria for spur existence at a frequency number have been determined, the congruences given in (29) can be solved for K. This will enable the amplitude of the spurious response to be calculated from the frequency number in the digital spectrum. Equation (29), which describes the frequency location in terms of K, can be rewritten as

$$F_n = \langle \Gamma \pm 2^{L-B}K\Gamma \rangle_{\frac{2^L}{(F_n,2^L)}} \tag{40}$$

By subtracting and applying the modulus operator to both sides and simplifying we obtain

$$\left\langle \frac{F_n - \Gamma}{2^{L-B}} \right\rangle_{\frac{2^L}{(F_n,2^L)}} = \langle \pm K\Gamma \rangle_{\frac{2^L}{(F_n,2^L)}} \tag{41}$$

By applying Euler's Theorem using the same justification as in (18) and exploiting the fact that $2^B/(F_n,2^L) = 2^B/(F_n,2^B)$ in all cases where spurs due to phase truncation exist

$$\left\langle \frac{F_n - \Gamma}{2^{L-B}} \Gamma^{\phi(2\Lambda)} \right\rangle_{2\Lambda} = \pm K \tag{42}$$

As shown earlier $\phi(2\Lambda) = \Lambda$ for Λ a power of 2 therefore

$$\left\langle \frac{F_n - \Gamma}{2^{L-B}} \Gamma^{\Lambda-1} \right\rangle_{2\Lambda} = \pm K \tag{43}$$

This yields an expression for the K value of a spur (from which its amplitude can be calculated) from its frequency location in the digital spectrum. This result however presumes that one is interested in the spectrum modulo $2^L/(F_n,2^L)$. In reality the spectrum from 0 to $2^{L-1}/(F_n,2^L)$ is desired, with all of the spurs whose frequency numbers lie between $2^{L-1}/(F_n,2^L)$ and $2^L/(F_n,2^L)$ aliased back into the range 0 to $2^{L-1}/(F_n,2^L)$. If the location of a spur were calculated using (40), and if the resulting spur number, F_n, was larger than $2^L/(F_n,2^L)$, then the true aliased position of the spur could be calculated by $F_n = -F_n + 2^L/(F_n,2^L)$. Modifying (40) to account for the aliasing of the spurs in the range $2^{L-1}/(F_n,2^L)$ to $2^L/(F_n,2^L)$ we obtain

$$\frac{2^L}{(F_n,2^L)} - F_n = \langle \Gamma \pm 2^{L-B}K\Gamma \rangle_{\frac{2^L}{(F_n,2^L)}} \tag{44}$$

which simplifies to

$$\left\langle \frac{-F_n-\Gamma}{2^{L-B}} \right\rangle_{2\Lambda} = 2K \qquad (45)$$

Combining the result from (45) with that obtained in (43) gives an explicit expression for the spur amplitude for the two conditions given in (30) that determine if a spur exists at a frequency number. Therefore, if the fraction $F_n - \Lambda$ is an integer, then (43) can be used to calculate the spur amplitude, and if $F_n + \Lambda$ is an integer, then (45) can be used.

Therefore, the complete algorithm for determining the digital spectrum of a DDFS in the presence of phase truncation can be summarized as follows

1) Divide the frequency range from 0 to $F_{clk}/2$ into a sequence of $2^{L-1}/(F_f,2^L)$ evenly spaced potential spur locations, each sequentially numbered. The sequential frequency number of a spur location is related to the actual analog spur frequency, F_{sp}, by

$$F_n = \frac{F_{sp}\,2^L}{(F_f,2^L)F_{clk}}$$

2) Find the spur number, K, from the frequency number, F_n using the following rules:

   a) If both $F_n - \Lambda$ and $-F_n - \Lambda$ are not divisible by 2 then the magnitude of the spur at F_n is 0.

   b) If 2 divides $F_n - \Lambda$ then

$$\left\langle \frac{F_n-\Gamma}{2^{L-B}} \right\rangle_{2\Lambda} = K$$

   If 2 divides $-F_n - \Lambda$ then

$$\left\langle \frac{-F_n-\Gamma}{2^{L-B}} \right\rangle_{2\Lambda} = K$$

   where we define $\Lambda = \frac{2^{B-1}}{(F_f,2^B)}$

3) The magnitude of the spurious noise line at F_n is

$$\zeta_{Kx} = \frac{\pi 2^{B-L}}{2\Lambda}\operatorname{cosec}\left(\frac{K\pi}{2\Lambda}\right)$$

III. IMPLICATION OF RESULTS

The spectrums generated using this algorithm were compared to those generated using a DFT and were found to exactly predict the spectrum of the DDFS to within the roundoff error involved in the FFT algorithm. All spur magnitudes larger than -120dB agreed with the DFT results to within 1%.

Now that an explicit relationship has been derived for the output spectrum due to phase truncation, several observations can be made about the interaction of the finite word length effects and the frequency control command, F_f. From (27) it is evident that the spectrums generated from values of F_f that have the same value of $(F_f,2^B)$ have a one to one correspondence between the magnitudes of the spurs. In other words, in the two different spectrums the magnitudes of all the spurs will be the same, and only their position in the spectrum will change. This behavior provides a clue as to the number theoretic operation of the phase accumulator. For simplicity we will consider only odd values of F_f such that $(F_f,2^B) = 1$. If one considers the output of one numerical period of the phase accumulator to be a vector, $\underline{V}$, of length 2^L where the basic vector, $\underline{V}_1$, is provided by $F_f = 1$ and is composed of the consecutive integers from 0 to 2^L-1. Under this model the phase accumulator can be considered as a permutation generator, where each value of F_f provides a different permutation of the values from 0 to 2^L-1 as given by

$$\langle F_f\, n \rangle_{2^L} = \underline{V}_{F_f}[n]$$

Where $\underline{V}_{F_f}[n]$ is defined as the n^{th} element of the vector $\underline{V}_{F_f}$.

This property provides considerable insight into the nature of the spectrum of the DDFS. It can be shown that permutation operators of this class are commutative with the DFT operator. Therefore, a permutation of this type to the input vector to a DFT will result in an identical permutation of the transformed output vector. Therefore, the spectrum of a DDFS due to any distortions in the system which are the result of a memoryless phenomenon can be generated by using one DFT for each class of $(F_f,2^B)$, and permuting the results for the desired F_f.

Returning to the analysis of the spurs due to phase truncation only, a relationship for the magnitude of the largest spur in the spectrum may be obtained from (27). Because ζ_K is a monotonically decreasing function of K, the magnitude of the largest spur may be obtained by substituting K = 1. Using this fact the magnitude of the largest spur in the spectrum is given by

$$\zeta_{worst} = \zeta_{1x} = 2^{B-L}\frac{\frac{\pi(F_f,2^B)}{2^B}}{\sin\left(\frac{\pi(F_f,2^B)}{2^B}\right)} \qquad (46)$$

From (47) we see an important relationship between the magnitude of the worst case spur and $(F_f,2^B)$. For values of F_f such that $(F_f,2^B) = 2^B$, spurs due to phase truncation are nonexistent. However for values of F_f such that $(F_f,2^B) < 2^B$, the magnitude of the largest spurs in the spectrum is a decreasing function of $(F_f,2^B)$ with the maximum value provided by $(F_f,2^B) = 2^{B-1}$. This relationship is shown graphically in Figure 12

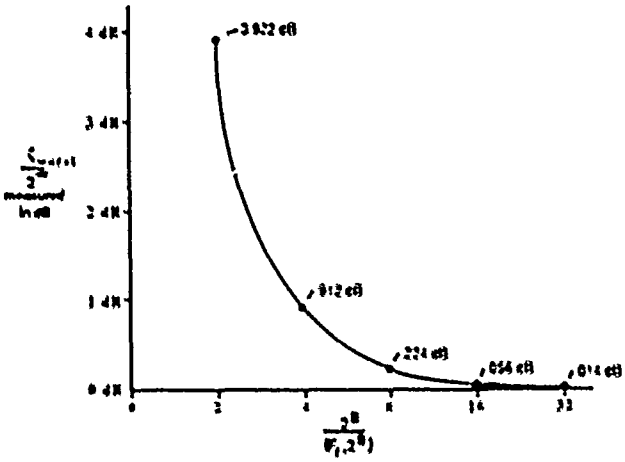


Figure 12. Worst Case Spur Amplitude vs. $\frac{2^B}{(F_f,2^B)}$

Note from Figure 12 that if $(F_f,2^B) = 1$, then as B gets larger, the magnitude of the worst case spur asymptotically approaches a value of $2^{B-L} = 2^{-W}$. This is 3.922dB less than the largest worst case spur which would result from the same B, but unrestricted values of F_f. The magnitude approaches this limit very quickly, so that for B larger than 4 bits and with $(F_f,2^B) = 1$, the magnitude of the worst case spur can be considered to be 2^{-W}.

PHASE ACCUMULATOR MODIFICATION

This phenomenon suggests a simple method for improving the worst case spur performance. If a scheme could be developed to prevent values of F_f that would result in $(F_f,2^B) = 2^{B-1}$ from being input to the phase accumulator, then the performance of the DDFS could be significantly improved. One brute force approach would be to only allow odd values of F_f to be used. This would provide a performance improvement of 3.922 dB predicted above for B > 4 at the expense of halving the frequency resolution of the synthesizer. However, a much more elegant solution is possible involving only a small amount of additional hardware that provides the 3.922 dB improvement, without reducing the frequency resolution. This solution is obtained by observing the behavior of the least significant bit of the phase accumulator when the input frequency control command, F_f, is odd. This condition, as mentioned previously, corresponds to the minimum worst case spur condition. Under these circumstances, since the carry-in to the adder in the least significant bit of the phase accumulator is zero, the least significant bit of the phase accumulator register will oscillate between one and zero every clock cycle. Therefore, the goal of the hardware addition is to modify the existing L bit phase accumulator structure to emulate the operation of a phase accumulator with a word length of L+1 bits under the assumption that the least significant bit of the frequency control word is always a one. Such a hardware addition is shown in Figure 13. As seen in the figure, the additional flip-flop and inverter synthesize the same carry-in sequence to the L bit adder that the least significant bit of an L+1 bit adder with an odd F_f input would propagate to its L most significant bits. Thus, this circuit simulates the doubling of the frequency resolution of the phase accumulator by adding an extra bit to its word length, while simultaneously halving the resolution by restricting the input F_f to be an odd number. The net result is a phase accumulator with the same frequency resolution but with a 3.922 dB performance improvement and no restriction on the input value of F_f. Fortunately this spur performance gain is implemented without adding any hardware that would increase the critical path delay through the adder therefore the throughput of the DDFS is not degraded by the modification.

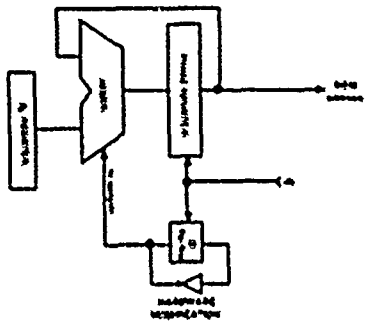


Figure 13. Hardware Modification to Force $(F_f,2^B) = 1$

INTERACTION OF ALL FINITE WORD LENGTH EFFECTS

This modification also has the benefit of simplifying the analysis of the DDFS since the magnitude of the worst case spur is now independent of F_f. Using previously derived relations for the spurious response due to phase truncation we may now graph the worst case response due to all finite word length effects. In this graph it is assumed that all error effects are additive and that the errors due to finite word length effects in the ROM and DAC both contribute a potential spur of magnitude 1/2 LSB.

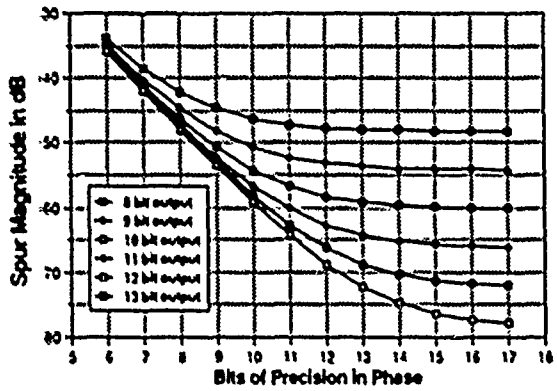


Figure 14. Worst Case Spurious Response of DDFS for all Finite Word Length Effects.

IV. CONCLUSION

In this paper, a simple algorithm is presented for generating the output spectrum of DDFS's resulting from phase accumulator truncation. This algorithm is fast and requires very little memory since the spectrum is computed sequentially. The computational burden of this algorithm is small enough that all the significant spurs in spectrums containing 2^{12} points have been generated on a hand held calculator (HP-41C). These results have been compared to those obtained by a DFT and have been shown to have comparable accuracy.

In the process of deriving the algorithm, several important insights were gained into the relationship between the amount of phase truncation, the choice of frequency control word, and the resulting output spectrum. One of these was the realization that input values of F_f with the same $(F_f,2^B)$ will generate DDFS spectrums that differ only in the positions of the spurs. The relative amplitudes of these spurs and the total number of spurs in these spectrums will be the same. Another realization was that all phase accumulator output sequences with the same value of $(F_f,2^B)$ are related to each other by a number theoretic permutation which is commutative with the DFT operator. This property thereby allows the spectrum of all DDFS output sequences with the same $(F_f,2^B)$ to be generated from a single basic spectrum.

The derivation of Equation (27) for the DDFS spur amplitudes due to phase truncation yielded an important relationship between the value of $(F_f,2^B)$ and the amplitude of the worst case spur. This relationship was used to propose a modification to the conventional phase accumulator structure to reduce the magnitude of the worst case spur. This hardware modification also reduced the detrimental effects of roundoff error and DAC nonlinearities on the spectral purity of the output sinewave. The modification also allows the output spectrum for any value of F_f to generated from a permutation of a singe spectrum without worrying about each different class of $(F_f,2^B)$.

The use of the new phase accumulator structure also allowed the generation a design table which generalizes the worst case spurious behavior due to the additive effects of all finite wordlength phenomena. The table gives the worst case spurious response as a function of phase accumulator precision and the wordlength of the stored ROM samples.

In conclusion, the relationships derived in this paper showing the interaction of phase accumulator truncation, ROM table length, and output spurious noise level should greatly facilitate the design of DDFS's with the optimum hardware configuration for any requirements. Also, the implementation of the novel phase accumulator structure should greatly improve DDFS performance by reducing the worst case spur levels and by increasing the tolerance to roundoff errors and DAC nonlinearities.

V. ACKNOWLEDGMENTS

This research was supported in part by the University of California MICRO Program in cooperation with TRW Inc. under Grant 86-110.

The authors would like to thank Dr. Douglas N. Green of TRW for originally introducing them to the problems associated with the analysis of direct digital frequency synthesizers and also Mr. Gerald R. Fischer, Mr. Peter J. Hadinger, and Mr. William F. McDonnell of TRW for their valuable discussions and comments on this work.

VI. REFERENCES

[1] Albert L. Bramble, "Direct Digital Frequency Synthesis", Proc. 35th annual Frequency Control Symposium, USERACOM, Ft. Monmouth, N.J., May 1981, pp.406-414.

[2] J. Tierney, C.M. Rader, and B. Gold, "A digital frequency synthesizer," IEEE Trans. Audio Electroacoust., vol. AU-19, pp. 48-57, 1971.

[3] L. Jolley, Summation of Series, Chapman and Hall, London, 1925.

[4] H.T. Nicholas, "The Determination of the Output Spectrum of Direct Digital Frequency Synthesizers in the Presence of Phase Accumulator Truncation", Thesis, Masters of Science in Electrical Engineering, UCLA 985.

[5] C.E. Wheatley III and D.E. Phillips, "Spurious suppression in direct digital synthesizers," Proc. 35th annual Frequency Control Symposium USERACOM, Ft. Monmouth, N.J., May 1981, pp.428-435.

[6] D.A. Sutherland, R.A. Strauch, S.S. Wharfield, H.T. Peterson, and C.R. Cole, "CMOS/SOS Frequency Synthesizer LSI Circuit for Spread Spectrum Communications" IEEE Journal of Solid State Circuits, vol. SC-19, pp.497-505, Aug. 1984.

[7] J. Gorski-Popiel, Frequency Synthesis: Techniques and Applications, New York: IEEE press, 1975.

[8] V. Manassewitsch, Frequency Synthesizers: Theory and Design, New York: John Wiley, 2nd Ed., 1979.

41st Annual Frequency Control Symposium - 1987

THE IMPACT OF DIGITAL SIGNAL PROCESSING ON CRYSTAL FILTER REQUIREMENTS

Marvin E. Frerking
Collins Defense Communications
Rockwell International Corporation
Cedar Rapids, Iowa

ABSTRACT

The availability of low-cost digital signal processing (DSP) hardware and A/D converters has made it possible to perform many of the filtering functions, traditionally accomplished using crystal filters, by digital means. Finite impulse response (FIR) digital filters are often used, and they are designed to have an absolute linear phase response. The basic configuration of these filters is discussed along with typical specifications and frequency response curves.

The use of DSP often requires a crystal filter at one of the IF frequencies to limit the bandwidth. If this is not done, aliasing occurs, which cannot be removed by subsequent digital processing. The reasons for these requirements as well as typical crystal filter specifications are discussed.

The primary advantages of digital filters are that: (1) a linear phase can be obtained, (2) once designed, each filter has exactly the same response in production, and (3) additional bandwidths can be provided without significantly increasing the cost. Once the decision is made to use DSP, many other functions such as demodulation and AGC can also be performed.

In the near future, DSP is expected to replace the most exacting analog filters, particularly those requiring equalizers. These receivers will continue to require crystal filters, but with somewhat wider bandwidths. Many receivers, particularly those requiring only one bandwidth or a relatively wide bandwidth, will likely remain analog for some time, although DSP has the potential to replace large portions of these receivers. As digital hardware costs continue to decline more and more of the analog functions will be accomplished digitally.

INTRODUCTION

Digital signal processing techniques have been known for many years; however, until recently they have attracted little attention among communications equipment manufacturers. This is primarily the result of the cost of digital signal processing hardware which has been prohibitively high. During the early 1980's, however, a new class of components was introduced by the semiconductor industry. These devices integrate a microcomputer with a hardware multiplier on the same chip. The execution speed to perform a multiplication plus an addition was less than 250 nsec. Since their introduction, the state of the art has been rapidly improving and most of the devices being introduced at the present time have execution speeds in the order of 100 nsec. In addition, dedicated chips are available which perform multiplication and addition as rapidly as 5 nsec. When these devices are combined with high-speed memory, even wide-band systems can be designed using digital filters.

The state of the art has also improved in the manufacture of A/D converters and 12- to 14-bit devices sampling at 100 kHz can be readily obtained. In addition, flash A/D converters with 8-bit accuracy can be obtained with sample rates well in excess of 100 MHz.

The costs have fallen rapidly on the signal processors and several of the early chips, which are quite capable of performing practical digital filters, are available for less than $10.

It is not surprising that the availability of this technology has resulted in a large amount of activity in the communications design area. At this time at least one digital receiver, the R-5099/U, has been manufactured by Rockwell International in large quantities. In addition, many trade studies are currently being made to evaluate the potential of digital signal processing (DSP) for new designs. There are many advantages which can be realized with digital filters. Among them are:

1. Better filter shape factors can be obtained when necessary.

2. A linear phase response can be achieved.

3. Receivers can be produced with matched delay from unit to unit.

4. Multiple bandwidths can be provided by changing the filter coefficients.

5. Filters have no production variation.

6. Filters can be made field programmable.

DIGITAL FILTER STRUCTURE

Later in this paper, a typical digital receiver architecture will be discussed along with the requirements of the crystal filter which precedes the A/D converter. The factors which determine the crystal filter requirements are also established. We shall pause briefly, however, to examine a digital filter and provide an explanation of how it works. Two types of digital filters are used: (1) infinite impulse response (IIR) filters, and (2) Finite impulse response (FIR) filters. IIR filters have not been used to a great extent in communications equipment to date, and because of time limitations we will not discuss them in this paper other than to indicate that IIR filters have characteristics much like the familiar analog filter types, e.g., Butterworth, Chebyshev, and Elliptic filters. IIR filters, unfortunately, also have the nonlinear phase delay characteristics of their analog counterparts as well as some unique problems due to finite arithmetic truncation. FIR filters, on the other hand, are easier to implement digitally and generally provide better delay characteristics. There are additional advantages in cases where the sample rate is to be reduced after the filter.

The block diagram of an FIR filter is shown in figure 1. In this diagram the signal X(nT) is the digitized input voltage at time t=nT, where T is the time between samples and n is the sample number n=0, 1, 2, ... The first delay element holds the previous input value from t=(n-1)T. The second delay holds the voltage from (n-2)T, etc. Each delay block requires a latch with b bits, where b is the number of bits of precision. When a new input value at t=(n+1)T becomes available, the previous values are shifted one place to the right and the value in the last delay block is lost. An output value y (nT) is calculated by taking the sum of all the products as shown. Thus,

$$y(nT) = h_0x(nT) + h_1x[(n-1)T] + \ldots + h_{N-1}x[(n-N+1)T] \qquad (1)$$

where N = the total number of taps.

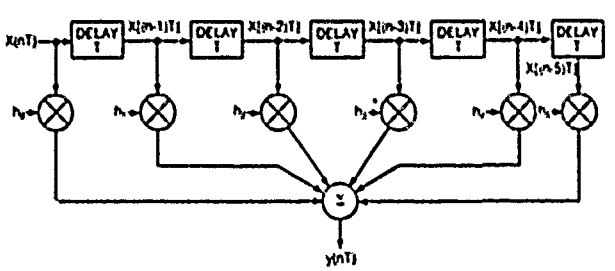


Figure 1. Block Diagram of FIR Digital Filter.

The output can be more concisely written as

$$y(nT) = \sum_{k=0}^{N-1} h_kx(n-k)T \qquad (2)$$

Obviously, for an N-tap filter a total of N multiplications and additions are required each sample time to calculate the outout value. For an FIR

CH2427-3/87/0000-503 $1.00 C 1987 IEEE

filter the coefficient values, h_k, also represent the impulse response which has additional advantages in some cases.

The frequency response of the filter is determined by the values of the coefficients, h's, and it can be shown that the transfer function is given by

$$H(\omega) = \sum_{k=0}^{N-1} h_k e^{jk\omega T} \qquad (3)$$

It can also be shown that the filter has a linear phase response (constant time delay) if the coefficients are symmetrical about the center of the filter. The filter of figure 1 is, therefore, linear in phase if $h_0 = h_5$, $h_1 = h_4$, and $h_2 = h_3$. In general, it is required that

$h_k = h_{N-1-k}$ for $0 \le k \le N/2$ for N even and

$$0 \le k \le \frac{N-1}{2} \text{ for N odd.}$$

The absolute value of the delay through the filter is equal to NT/2.

The filter coefficients can easily be determined by a filter design program, several of which are available to run on personal computers. The program requires as inputs the sample rate, the bandwidth and ripple of the passband, the bandwidth of the stopband, and the stopband attenuation. In some cases the designer is then asked for the number of taps and the program does the best job possible given the available number of taps. The frequency response is then displayed. If the specifications are not met, the designer must increase the number of taps and repeat the design cycle. Most of the programs use the Chebyshev (minimax) approximation, based on the Remez algorithm to optimize the coefficients.[1] The designer must be aware that the use of finite arithmetic may alter the response, and the programs normally allow the designer to plot the response with a specified number of bits of accuracy in the coefficients. Scaling is also necessary to prevent overflow.

The frequency response characteristics which can be obtained are very impressive for a large number of taps. Figure 2 shows the calculated frequency response for a receiver using 80-tap FIR filters operating with a 16-ks/s rate. The response meets the requirements of MIL-STD-188C which requires a passband from 330 - 3050 Hz with 60 dB of attenuation below 300 Hz from the carrier frequency and also above 4000 Hz. The specified delay distortion is 1400 μsec; however, the digital filter has no delay distortion.

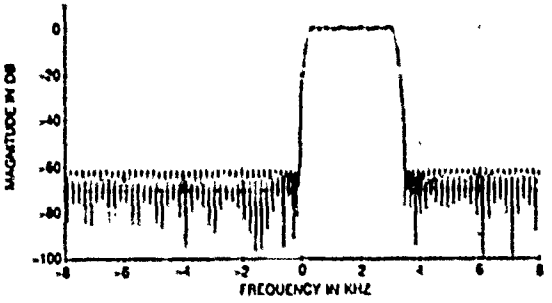


Figure 2. Sideband FIR Filter Response, 80 Tap.

It is interesting to calculate the computation load for a filter of this type. If the signal processor has a multiply-accumulate instruction with a 100-nsec execution time, then $80 \times (100 \times 10^{-9}) = 8 \times 10^{-6}$ sec are required to calculate an output value. For a 16-ks/s rate, 62.5 μsec are available and the chip would only be 13 percent loaded. The chip could, therefore, provide multiple simultaneous filter functions or perform other tasks in addition to filtering. Because of the architecture of a digitally implemented receiver, it has been possible to accomplish more with a digital signal processor than merely to provide filters implemented digitally to replace the analog filters. It is very desirable, however, to use a crystal filter in the analog portion of a digital receiver. These conclusions can be understood by examining the architecture of a digital receiver.

DIGITAL RECEIVER ARCHITECTURE

The block diagrams of a digital receiver are shown in figures 3 and 4. They do not represent a specific receiver, but are typical of the way in which a digital receiver might be organized. This receiver might tune from below 10 kHz to perhaps 30 MHz. Only the single-sideband mode is discussed, although independent sideband, AM, and CW can also be accommodated by changing the software. In the receiver shown in figure 3, the signal from the antenna is translated to a first IF which may be in the region of 100 MHz to avoid image responses. This IF signal is then translated to a second IF with a center frequency, perhaps in the 3- to 10-MHz region. It is at this frequency where the primary analog selectivity is obtained and where the crystal filter is located. The signal of this IF may now be digitized with a flash A/D converter or it can be translated to a third lower IF frequency. For the case shown, a 12-kHz last IF is used and a 12-bit A/D converter samples at a 48-ks/s rate. The digital words are then inputted to the signal processor shown in figure 4. Here the signal is first translated to baseband by multiplying the IF signal by $\cos \omega_0 nT$ and $\sin \omega_0 nT$. Note that since the sample rate in this case was chosen to be four times the IF, the sequence $\omega_0 nT = 0, \pi/2, \pi, \ldots$ etc, results in $\cos \omega_0 nT = 1, 0, -1, 0, \ldots$ and $\sin \omega_0 nT = 0, 1, 0, -1, \ldots$. If this is done, the need to calculate values for cos and sin is eliminated. If the signal is sampled at a high rate with a flash A/D converter, say at 12 MHz, then a 3-MHz IF would be used. In the latter case, a box-car filter would be used after the mixer which would accumulate 250 samples and then input the sums to the digital filters at a 48-kHz rate. After the signal is translated to baseband, and baseband in this case is the center of the audio band rather than the frequency of the carrier, low-pass filters rather than bandpass filters can be used. It is beyond the scope of this paper to discuss the properties of complex signals; however, we simply state that using both cos and sin multipliers results in a complex signal which may have distinct positive and negative frequency components. These frequencies are later translated to the normal audio band by the BFO.

We note from figure 4 that the filter in each signal path is broken into a decimation filter followed by a sample rate reduction before the main filter. The decimation filter narrows the bandwidth sufficiently so that a lower sample rate can be used. This has two benefits. First, more time is allowed to perform the multiplications of the main filter, and secondly, fewer taps are required to achieve the same transition bandwidth if the sample rate is lower. An AGC algorithm is also included in the signal processor to compute the AGC control voltage for the analog portion of the receiver as well as for the digital multiplier for fine adjustment to the gain. The purpose of the analog AGC is to prevent large signals which fall in the passband of the crystal filter from overloading the A/D converter. The block diagram of figure 4 may typically require from one to three processors depending on the speed of the processors, the bandwidths involved, and the characteristics of the analog crystal filter.

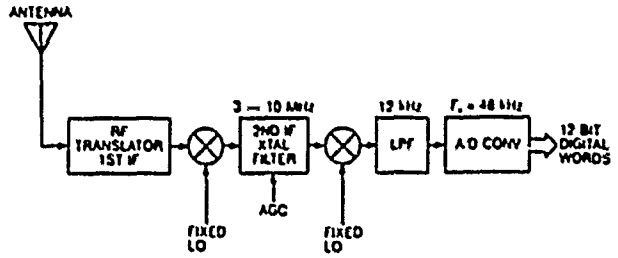


Figure 3. Analog Portion of Digital Receiver.

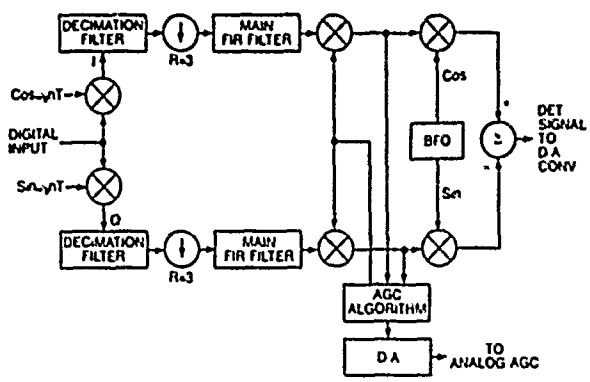


Figure 4. Digital Portion of Digital RCVR.

ANALOG CRYSTAL FILTER REQUIREMENTS

The constraints which determine the requirements of the analog crystal filter preceding the A/D converter will now be discussed. They can most easily be explained with the aid of figures 5, 6 and 7. Figure 5 shows the frequency response of the crystal filter which has a center frequency f_o equal to the second IF and may be in the 3- to 10-MHz region in many cases. Here the passband is assumed to be ±4 kHz to accommodate all the modes of the receiver. The stopband in this case must be no wider than ±20 kHz for a shape factor of 5:1. This can be seen from figure 6 which shows the frequency spectrum of the 12-kHz IF. We have used the Fourier frequency representation which includes both positive and negative values. Note that the image response of the last mixer must be considered, and an undesired signal 20 kHz below f_o results in a signal at 8 kHz just as a desired signal 4 kHz below f_o would. Since the ultimate selectivity of the digital filter removes signals beyond ±4 kHz, this is acceptable. However, if a signal further below in the range from -20 kHz to -28 kHz were allowed to reach the mixer, it would fall in the desired passband. The crystal filter is, therefore, required to remove the image frequencies. Obviously, the lower the final IF, the more difficult the requirement.

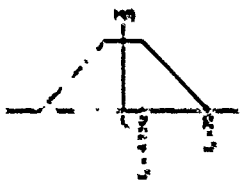


Figure 5. Crystal Filter Response.

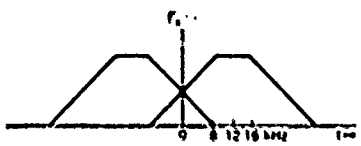


Figure 6. Frequency Spectrum of Last IF.

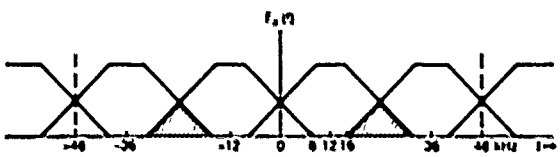


Figure 7. Frequency Spectrum of Digital Signal.

There is yet another requirement on the crystal filter, and this results from the sampling or digitizing process. When an analog signal is sampled at a rate, say f_s, the digital signal contains not only the frequency spectrum of the sampled signals, but the spectrum is repeated at all harmonics of f_s. This is shown in figure 7. The basic spectrum is centered around zero. The entire spectrum is reproduced around frequencies of ±48 kHz, ±96 kHz, etc. As shown, an overlapping region occurs between the upper portion of the signal centered at zero and the lower portion of the signal centered at -48 kHz. This is shown by the shaded area. Distortion of this type is referred to as aliasing. Aliasing cannot be allowed to fall into the desired passband, in this case within ±4 kHz of the 12-kHz IF. The region of overlap is obviously a function of the crystal filter stopband which in this case is the same as the image attenuation requirement. An examination of figure 7 shows that if the crystal filter stopband were increased, the shaded region would expand into the desired passband. The aliasing region could also be reduced by increasing the sampling rate of the A/D converter. For example, if a 96-kHz sampling frequency had been used, the harmonically generated spectra would be located at 96-kHz increments rather than at 48 kHz. Unfortunately, the higher sample rate would result both in a larger computation load for the processor and a more expensive A/D converter. Therefore, the design engineer is faced with a trade-off between the crystal filter cost and the digital hardware cost. In some cases, it may also be necessary to increase the passband of the crystal filter to avoid nonlinear phase distortion in the region of the final filter passband.

VERSATILITY OF DIGITAL SIGNAL PROCESSING

As indicated earlier, one of the significant advantages of using digital signal processing is the ability to change the program to provide multiple functions in a receiver or an exciter.

As an example, figure 8 illustrates the block diagram of a receiver which can simultaneously recover both the upper sideband and the lower sideband. In this case the lower FIR filter provides a 90-degree phase shift with respect to the upper section. Such a filter is said to include a Hilbert transformer. In this case two identical filters are used. Recall that the frequency of the carrier, if it were present, would be zero. Therefore, if the outputs of the filters are subtracted, the upper sideband is recovered. If they are added, the lower sideband is recovered. Obviously, both can be recovered simultaneously with only a small increase in computation load. Conversely, in an analog receiver, a second complete crystal filter would have to be added. The digital hardware could also be reprogrammed to receive AM or CW. It could also demodulate SSB signals.

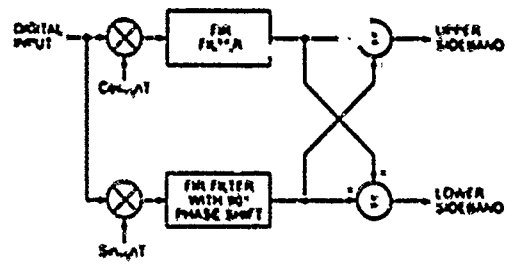


Figure 8. Block Diagram of Independent Sideband Receiver.

The use of DSP may also have an impact on the frequency synthesizer. We recall, in the discussion earlier, that the cosωnT and sinωnT computation could be eliminated if the sample rate is chosen to be four times the final IF frequency. If the processor is programmed to generate these functions, however, then the small frequency steps, for example, 1 Hz to 1 kHz can also be obtained in the signal processor. The frequency synthesizer could then step in 1-kHz increments, provided the crystal filter passband were increased somewhat.

It should be mentioned that another class of digital signal-processing systems has been built in which dedicated multiplier/accumulator chips are used with separate memories. Sample rates of several hundred ks/s to Ms/s can then be obtained. In addition, digital polyphase filters have been built to efficiently provide many simultaneous channels. Systems of this type can replace analog systems which contain banks of narrow-band receivers or even wide-band receivers employing SAW filters.

DSP chips designed primarily to perform filter functions have also appeared in the marketplace and doubtless these chips will find areas of application in communications systems.

The systems described above are certainly not exhaustive of the many techniques used in digital signal processing. They do, however, illustrate several of the main constraints on the analog filters required when DSP is used.

CONCLUSIONS

Because of the constraints and because of the functions DSP can provide, several conclusions can be drawn which may be reflected in future equipments. Among these are:

1. In the near future DSP will replace the most exacting analog filters, particularly those requiring equalizers.

2. Receivers using DSP will often provide multiple bandwidths by changing coefficients in the software, and only one analog crystal filter will be required.

3. Receivers in the near term will continue to use a significant analog section including a crystal filter; however, the filter will have a somewhat wider bandwidth.

4. DSP has the potential to replace larger and larger portions of the analog circuits and filters in communications hardware; however, the cost of high-speed A/D converters will limit wide-band digital signal processing to specialty systems for some time to come.

5. Many receivers, particularly those requiring only one bandwidth or relatively wide bandwidths will likely remain analog for some time.

6. Crystal filters used in receivers for direction-finding and adapti array applications will be required to have low differential time de even though the shape factor may be quite modest

REFERENCES

[McClellan, J.H., Parks T.W., and Rabiner, L.R., "A Computer Program Designing Optimum FIR Linear-phase Digital Filters," IEEE Trans tions on Audio and Electroacoustics, VOL. AU-21, No. 5, December 1

41st Annual Frequency Control Symposium - 1987

RANDOM NOISE IN DIGITAL GATES AND DIVIDERS

Donald C. Phillips

Collins Defense Communications, Rockwell International Corporation
Cedar Rapids, Iowa

Summary

Phase noise in digital gates and dividers is investigated for better theoretical understanding, with improved measurement techniques, and data measured under carefully specified conditions. The goal is to characterize digital hardware for more predictable design of low-noise phase-lock loops or similar applications. Measurements to date include CMOS and TTL gates and special squaring amplifiers, for three identifiable noise mechanisms which will be described. Measurements were first made at 1 MHz because of the importance of reference noise in PLL's (phase-locked loops).

Introduction

A major source of noise in high performance frequency synthesizers is the noise in digital frequency dividers and phase detectors. Any phase noise in the digital chain is carried through to the analog circuits, along with AM noise from the last digital (limiting) stage. The noise on a PLL reference frequency includes the frequency standard noise, the additional noise of the squaring amplifier, and any other noise in the digital system, including the phase detector and integrator. Because of the noise multiplication of most loops with frequency dividers, low noise design requirements place a limit to the amount of multiplication (loop division ratio) allowed in any one loop. Therefore, a quieter reference may result in a synthesizer with fewer loops.

This paper reports on an IR&D study program to characterize and analyze these digital noise sources. Although the work is not complete, the results so far are illuminating. The paper is divided into three sections.

The theoretical section includes the effects of various noise sources in phase-locked loops and the increased phase noise due to the impulse nature of digitized signals. Second, some special noise measurement techniques for low frequencies are discussed. Finally, measurements taken to date are interpreted with the goal of verifying principles that will be useful in analyzing and predicting noise performance of frequency synthesizers.

Uniform (white) noise sources are assumed for analysis purposes, at least within the designated bandwidths. Nonuniform noise (such as 1/f) can then be analyzed by altering the white noise results accordingly.

Effects on PLL's

Fig. 1 shows the frequency response of various noise sources, derived from basic PLL theory. The closed-loop response shown in (a) holds true for phase noise in the reference source, reference and feedback dividers, phase detector, and any noise voltage on the VCO control voltage path, except after the integrating amplifier in a type-2 loop which is shown in (b). VCO phase noise is reduced by the high-pass effect shown in (c) for type-1 loops and (d) for type-2 loops. These response curves help to determine the source of noise when applied to spectral plots of the VCO output. Note that these are not necessarily the resulting noise spectra, but they indicate the way the loop alters the spectrum of a noise source, which often is not uniform.

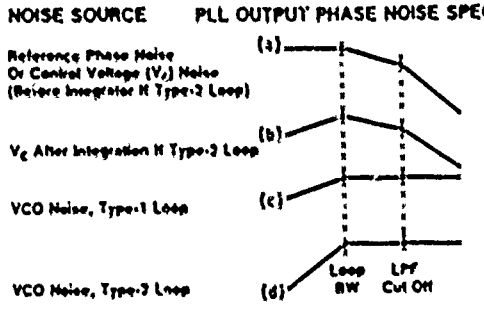


Fig. 1 PLL Output Transfer Functions from Various Sources

Theoretical Considerations

Due to the lack of verified noise data, some conjecture is warranted and even necessary to guide the measurement effort. It is hoped that the various noise mechanisms will become understood and used to clarify noise specifications and the reporting of noise measurements.

Individual Logic Gate Noise

The logic gate is the unit building block for most digital circuits. Its noise characteristics are needed to understand and predict the noise of dividers and phase detectors. Of concern is the phase jitter on one of the logic transitions or edges, which will eventually drive an edge-triggered device. These transitions are like the impulses in a sampling system and do not necessarily have the same spectrum as that seen on a spectrum analyzer for the entire wave. The one exception is the phase detector where the entire output, both phase and amplitude noise, is of concern.

The gate output phase noise is assumed to be detected in a "narrow" bandwidth, less than one-half the gate switching frequency. For a PLL reference frequency, this band contains the Fourier frequencies within the PLL bandwidth, or reference noise input to the loop, as in Fig. 1(a). Three major sources for this noise are:

Input Signal Phase Noise: Obviously, the narrow-band output phase noise includes the input narrow-band phase noise sidebands.

Input Voltage Noise: Input noise voltages include driving source noise as well as device input noise. The rms value of these voltages act as a noisy switching threshold for the driving signal, with a zero-crossing slope of Kp volts/radian.

CH2427-3/87/0000-507 $1.00 C 1987 IEEE

The sampling effect of a zero-crossing detector causes noise around each harmonic of the switching frequency to affect the output transition time and produce a narrowband demodulated output, as in a sampling scope. The noise output then is the sum of all the harmonic noise sideband powers, as shown in Fig. 2. Both external and internal parameters determine the gate input circuit bandwidth, or maximum significant harmonic H. H is not easily measured, but can be estimated from other measurements.

The equivalent input phase noise $\phi_{in}(v)$ from these noise voltages can be found by dividing by Kp, and multiplying by $\sqrt{H}$ (since power is proportional to the square of phase), and by $\sqrt{2}$ to obtain the equivalent sine-wave peak value since FM theory is based upon peak radians. Then the output noise

$$\phi_o(v) = \sqrt{H}\,\sqrt{2}\,v_n/K_p.$$

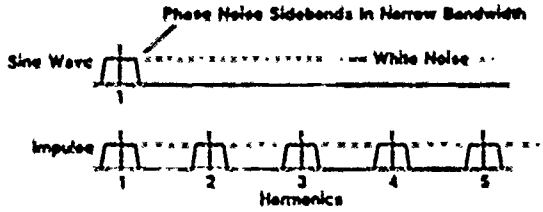


Fig. 2 Broadband Noise Effect on Gate Impulse Harmonics

This noise is higher in low-frequency squaring amplifiers because, for the same device type, there are more harmonics with a lower switching frequency. Low frequency noise is also more significant because the noise is divided down less, if at all. This noise can be reduced by band-limiting the input signal, thus reducing the input noise at baseband and harmonic frequencies. Even a low-Q tuned circuit may help significantly.

Propagation-Delay Noise: Variations in the gate propagation delay with time cause a time jitter t_j. If t_j has a uniform distribution with respect to frequency, the resulting output phase noise is the gate switching frequency f_{sig} times t_j. This noise is expected to be greater at high switching frequencies.

Total Gate Output Phase Noise: The total detected output phase noise power is the sum of the individual noise powers from these three mechanisms:

$$\phi_o^2 = (\phi_{sig})^2 + 2H(v_h/K_p)^2 + (f_{sig} * t_j)^2.$$

Note that (1) the output from the input signal modulation is independent of f_{sig}, (2) the effect of broadband voltage noise for a particular device type is inversely proportional to frequency, since the higher the frequency, the fewer the harmonics, and (3) the propagation delay noise is proportional to frequency.

Noise in Frequency Dividers

The three noise mechanisms of digital dividers are similar to those in individual gates, except for the division ratio N and its effect on sampling, and the number of gates in the signal path, operating at different frequencies.

Input Signal Phase Noise: The narrow-band input phase noise is divided by N at the output, according to basic FM theory.

$$\phi_o(sig) = \phi_{in}(sig)/N$$

Input Voltage Noise: Wideband input voltage noise, internal and external to the divider, has components which beat with multiples of the divider output. This is due to the sampling process (not leakage). Egan (Ref. 2, pp 75-81) has shown that interfering discrete signals near multiples of the divider output frequency cause difference frequencies in the output spectrum. This effect has been verified by measurement and is well known for the additional spurious signals in a divider following a mixer.

Extending this principle to noise, additional noise sidebands are formed as shown in Fig. 3. In this example a divider input frequency of 2 units is divided by 2 to an output frequency of 1 unit. Now there are twice as many harmonics and twice the noise power from a uniform noise spectrum. In general, the resulting close-in noise sidebands at the input are thus not only divided by N, but multiplied by the square-root of N, so that the input signal noise is reduced by the square root of N. This output phase noise is:

$$\phi_o(v) = \sqrt{H_o}\,\sqrt{2}\,v_n/(N\,K_p) = \phi_{in}/\sqrt{N}.$$

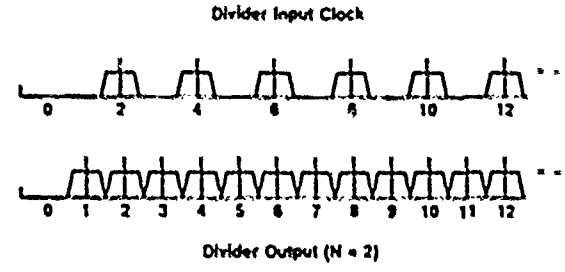


Fig. 3 Broadband Noise Effect on Multiples of Divider Output

This may explain the widely observed effect that when frequencies are divided by N and multiplied by M in either order, the resulting output to input phase noise ratio is higher than M/N. The higher close-in noise of phase-locked loops with lower reference frequencies and high division ratios is often said to be due to the "high multiplied-up noise". It may also be partly due to the poor division of noise.

Propagation-delay Noise: The propagation-delay noise ϕ_p in dividers is the rms value of this noise for M gates in the signal path, according to the switching frequency Fg and time jitter t_j for each gate.

Total Divider Output Phase Noise: The total divider noise is the rms value of the above three noise mechanisms. Note that a gate is a special case where N = M = 1.

$$\phi_o^2 = (\phi_{sig}/N)^2 + (v_n/K_p)^2(2H/N) + \phi_{p1}^2 + \phi_{p2}^2 \ldots \phi_{pm}^2$$

These mechanisms are plotted in Fig. 4 for different variables.

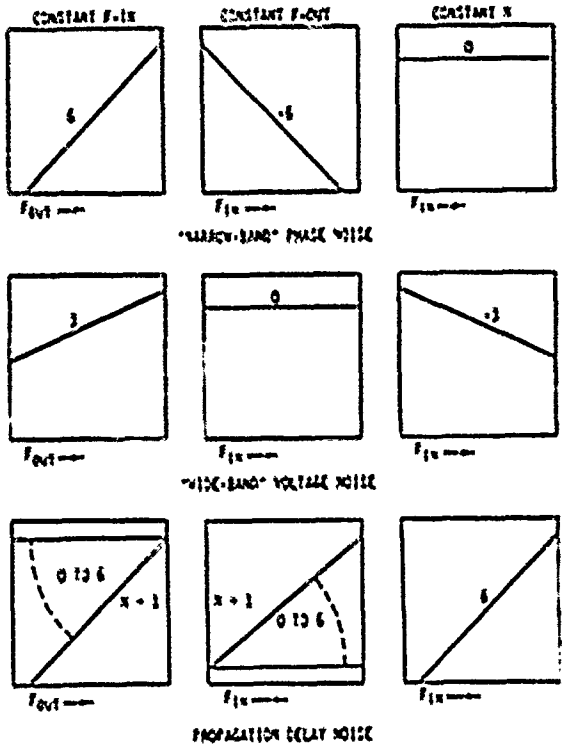


Fig. 4 Divider Output Noise as Affected by f_{in}, f_{out}, and N

Digital Phase Detector Noise

Because the phase detector interfaces with analog circuits, there are special sources of phase detector noise:

1. The phase noise of the two digital inputs will, of course, appear at the phase detector output. However, the rising edge of its output rectangular wave contains the noise of one edge of one input wave, and the falling edge contains the noise of one edge of the other input wave. Therefore, if a divider is to be followed by a phase detector, the divider should be analyzed for the phase noise on one edge only. The two edges do not necessarily have equal noise characteristics. For example, a saturating transistor has a noisier turn-off than turn-on.

2. Some types of phase detectors operate with the two input clocks very nearly at coincidence, where the output may switch over only part of the normal voltage swing. This quasi-linear mode may add to the phase noise output. This is related to the nonlinearity of the phase detector function (Kp, volts per radian) at the coincident region.

3. AM noise is found mostly on the high and low level states. For a succession of gates in a signal path, the limiting effect of each gate clips much of the AM noise of the preceding gate, until the last gate before the analog section of the phase lock loop. The equivalent phase noise is the noise voltage divided by the phase detector gain constant Kp (in volts/radian).

4. Both AM and phase noise are increased in digital devices by noise on the power supply, which often modulates at least one of the logic states directly. Power supply noise also modulates the device propagation delay, with resulting output phase noise. Intrinsic propagation delay noise is usually low at typical loop reference frequencies.

Measurement Techniques

An HP11740A Noise Measurement System is used because of its high sensitivity (-170 to -190 dBc/Hz) and automatic calibration. It is commonly used to measure dividers with high input frequencies. Two dividers with their outputs in quadrature drive a balanced diode mixer acting as a phase detector for the divider noise, with the signal generator noise greatly reduced. We wished to measure individual gate noise with carrier frequencies in the 1 to 10 MHz range of phase detector and divider output frequencies for high performance phase-locked loops. Sine waves from frequency standards at these frequencies are converted to digital signals in squaring amplifiers with a significant contribution to phase noise.

The test block diagram is shown in Fig. 5. An HP8642A signal generator was used because of its good waveform at high output levels. Its noise level of -140 dBc/Hz is too high for the measurements desired. The following sections show how the generator noise was reduced for 1 MHz tests.

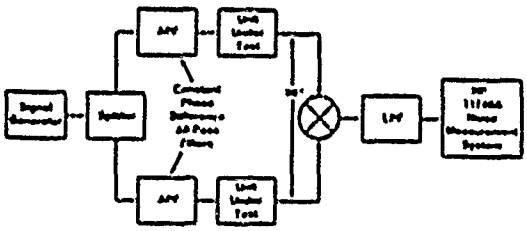


Fig. 5 Basic Test Block Diagram

The signal generator noise in the 1 MHz region was reduced by over 40 dB by adjusting the constant phase-difference circuits (Fig. 6) for nearly equal time delays, or nearly constant phase difference (90 degrees) over the desired noise detection bandwidth (900 to 1100 kHz). The coils are adjusted for minimum output with frequency modulation of the generator, and readjusted for each device under test. With 25 kHz modulation and 500 Hz peak deviation (-40 dBc sidebands) the output sidebands were below -80 dBc. Amplitude modulation at 2% (also -40 dBc sidebands) should be below -80 dBc if the mixer is close to exact quadrature.

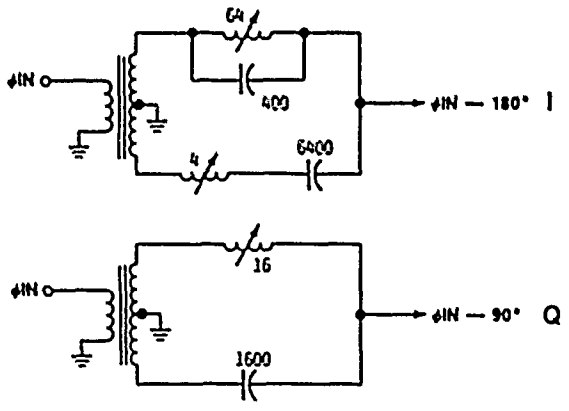


Fig. 6. Constant Phase-Difference All-pass Filters (Values for 1 MHz in uH and pf)

For some tests, it is also necessary to reduce the broadband generator noise at frequencies other than 1 MHz. A HPF and LPF combination are used as a BPF (Fig. 7), to reduce baseband noise below 100 kHz, and noise on harmonics from 2 MHz up. This results in a virtual "noiseless" (< -180 dBc/Hz) signal source.

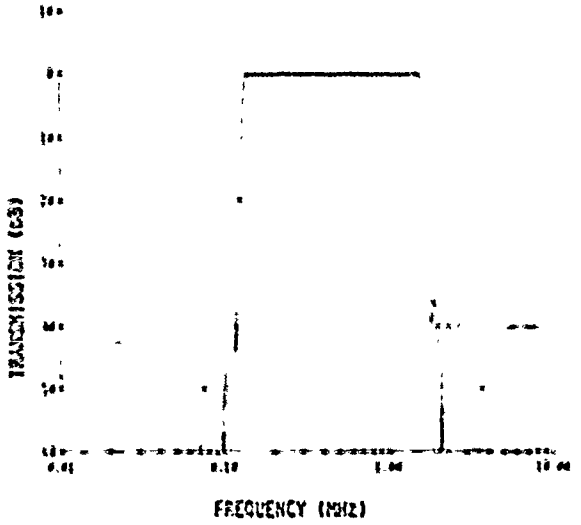


Fig. 7 HPF-LPF Response for 1 MHz Input

Measuring noise with diode mixer

Fig. 8 shows measured noise from a Minicircuits ZAD-1 mixer. Because the quadrature outputs require a precise load, they were connected to the mixer through 6 dB pads, with 7 dBm at each port. The upper trace shows the measured signal generator noise by itself, which is reduced to below -180 dBc by the quadrature circuit. The upper trace of noise from the mixer was made without the input LPF (low noise at 1 MHz, but higher noise at harmonics). The next lower trace shows the reduction of noise when the LPF was added, reducing the broadband noise above 1 MHz. Lower noise levels have been measured (Ref. 6, p. 11) with square waves from the output of TTL dividers. Apparently the effect of the input slope affects mixer noise as discussed above under gate noise. When used with digital gates it is difficult to apportion the noise between the gates and the mixer; therefore, a second approach is used for some tests.

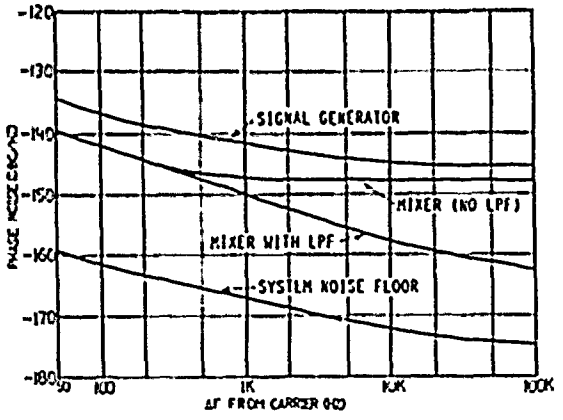


Fig. 8 Diode Mixer Noise Measurements at 1 MHz

Measuring noise with digital phase detector

As shown in Fig. 9, the diode mixer phase detector function is replaced by the digital circuit under test, followed by a super low-noise LT1028 gain-of-3 op-amp circuit used to drive the HP11740A noise amplifier 50-ohm input. This test circuit has a proven noise level well below the gate switching noise, as will be shown. Large low-impedance Ni Cad batteries with noise levels below that of the best voltage regulators are used to minimize noise from these sources.

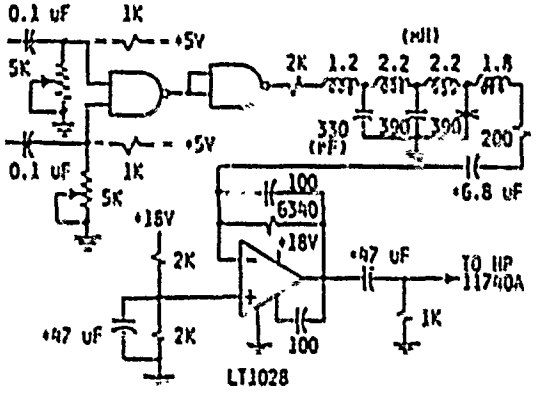


Fig. 9 NAND-gate Phase Detector and Preamplifier

Measured Data

A NAND-gate phase detector (Fig. 9) is used for measurement of individual gates with the LT1028 preamplifier to the HP11740A. After the APF's are adjusted for minimum generator FM, the bias resistors are adjusted for minimum generator AM (accurate zero-crossing detection). Over 40 dB attenuation is achieved on both AM and FM sidebands.

Measurements are shown of 'LS00 TTL and 'HC00 CMOS gates under conditions of 1 MHz switching conditions with the gate used as a squaring amplifier, compared with a special differential squaring amplifier (Fig. 10). The steady-state low and high logic levels were below the system noise with the preamplifier. Note that the levels shown are for two independent gate input noise sources. Noise from input sources for a single gate would be about 3 dB less.

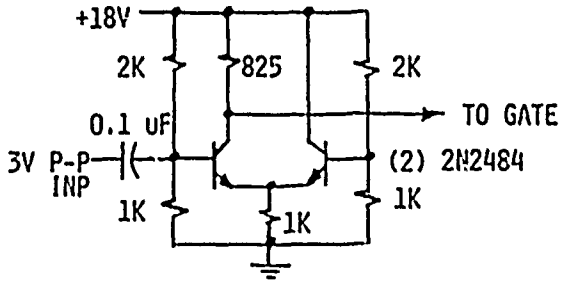


Fig. 10 Differential Squaring Amplifier

In the measurement shown in Fig. 11, the 'HC00 CMOS gate is a lower noise squaring amplifier than the 'LS00 TTL gate in this test with the "noiseless" driving signal. The differential squaring amplifier was noticeably better than either gate, and the switching noise is greater than the steady-state noise under the test conditions.

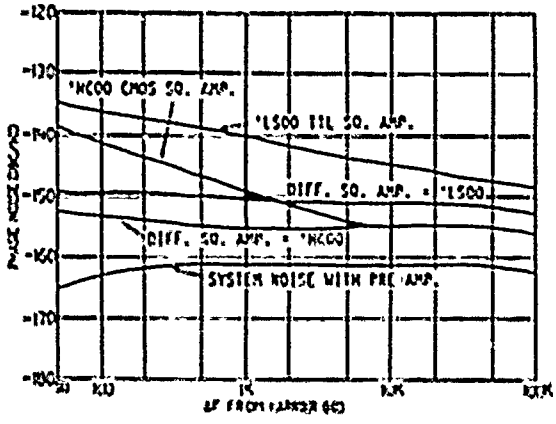


Fig. 11 Gate Noise Measurements at 1 MHz (with "Noiseless" Signal Input)

A noticeable difference occurred when the input LPF was removed (Fig. 12), allowing noise from 1.4 MHz up to be applied to the squaring amplifier. Both the TTL and CMOS gates were now considerably noisier, and the differential squaring amplifier was only a little better. This points out the need for bandpass filtering just before the squaring amplifier.

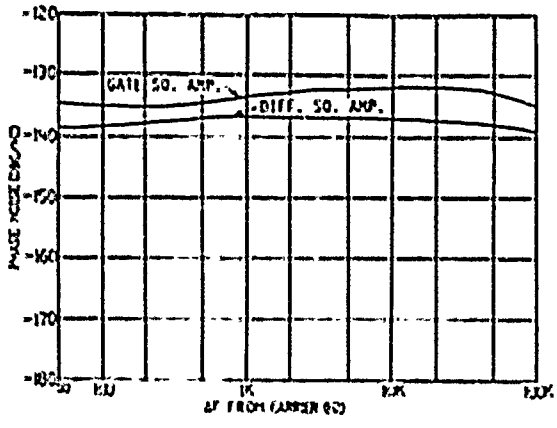


Fig. 12 Gate Noise Measurements at 1 MHz (with Broadband Noise Above 1.4 MHz at Signal Input)

Conclusions

Tentative theoretical analysis has shown three noise mechanisms in gates and dividers: narrow-band input noise, wide-band input noise, and propagation-delay noise. Techniques were described that allow measurement with an effective driving signal that can be essentially noiseless, or with narrow band and/or wide-band noise, allowing measurements of these three basic noise mechanisms in both gates and dividers. The terms "narrow band" and "wide band" have been given strict definitions to make measured data more meaningful.

Measurements at 1 MHz have shown differential nonsaturating squaring amplifiers to have less internally generated noise than TTL or CMOS gates. Noise both above and below the input frequency modulation band can modulate the output. Diode mixers behave somewhat like squaring amplifiers, in that the output noise level depends upon the input zero-crossing slopes.

Further measurements need to be made to verify these principles and provide better design guidelines. They include similar measurements at 10 MHz to determine the effect of carrier frequency, both for gates and squaring amplifiers. Then a measurement of a 1 MHz output of a divider with an input frequency of 10 or 100 MHz will test the effect of division on noise. As the input frequency becomes higher and produces less squaring-amplifier noise at the output, it may be desirable to revert to a diode mixer since it could then be driven with higher zero-crossing slopes. ECL and GaAs devices also need to be compared.

It should be evident from the examples given in this paper that measured noise levels of digital devices need to have an accurate detailed description of the test methods used. In addition, these measurements need to be interpreted in regard to the intended application, such as the effect on a PLL output spectrum.

References

1. Egan, William F, "The Effects of Small Contaminating Signals in Nonlinear Elements Used in Frequency Synthesis and Conversion," Proceedings of the IEEE, Vol. 69, No. 7, July 1981.

2. Egan, William F, "Frequency Synthesis by Phase Lock," John Wiley & Sons, 1981.

3. Hewlett-Packard, "Phase Noise--RF and Microwave Phase Noise Measurement Seminar."

4. Kroupa, V. F, "Noise Properties of PLL Systems," IEEE Transactions on Communications, October 1982, pp. 2244-2252.

5. Robins, W. P, "Phase Noise in Signal Sources," Peter Peregrinus Ltd., London, 1982.

6. Scherer, Dieter, "Design Principles and Test Methods for Low Phase Noise RF and Microwave Sources," RF and Microwave Measurement Symposium and Exhibition, Oct. 1978, Hewlett Packard.

LOW NOISE FREQUENCY SYNTHESIS

F. L. Walls
Time and Frequency Division
National Bureau of Standards
Boulder, Colorado 80303

C. M. Felton
Felton Electronics Design
P.O. Box 4013
Boulder, Colorado 80306

Abstract

This paper reviews the various definitions of phase noise and changes in the phase noise of a signal under noiseless multiplication, division, and translation. Next the phase noise in selected non-cryogenic rf and microwave oscillators is reviewed. Using a systems approach one can synthesize a microwave signal where the close in phase noise is controlled by a low frequency crystal oscillator while the high frequency phase noise is controlled by a microwave source. This approach yields a phase noise performance that is superior to that possible with a single source. Finally the phase noise of various amplifiers, multipliers, and dividers is compared. The phase noise of dividers while generally inferior to that of the best multipliers, is often sufficient for most applications. Special note should be made that the phase noise quoted in the literature for some dividers is perhaps pessimistic and that the phase noise of the system will often be seriously affected if placed under vibrational or thermal stress.

Introduction

This paper is to reviews the various aspects which govern the synthesis of signals with low phase noise. The synthesis of such signals from the rf region to the microwave region and beyond requires both low noise local oscillators and processing techniques. We therefore briefly review the model of a signal, definitions of phase noise, and the theoretical change of phase noise under noiseless multiplication, division, and translation. Next we review the phase noise of selected non-cryogenic rf and microwave sources. In the rf region the low phase noise sources are virtually all controlled by quartz crystal devices. At present the best available quartz devices have phase noise close to the carrier (typically Fourier frequencies from about 0.01 Hz to the half bandwidth of the crystal resonators) which varies approximately as $S_\phi(f) = K/(Q^4 f^3) = \nu_0^4/(K^3 f^3)$ where Q is the quality factor of the crystal, K is a constant for a given level of crystal technology, and ν_0 is the oscillation frequency. For the best quartz crystals currently available, $K = Q\nu_0$ is of order 1.2×10^{13}. The phase noise far from the carrier generally depends only on the oscillator technology and crystal drive.

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

The phase noise of microwave (and lumped circuit LC) oscillators is usually dominated by the noise in the sustaining circuit at a level determined by the quality factor of the resonator. The phase noise of selected microwave oscillators is compared to that obtained from multiplying various quartz controlled oscillators to the same frequency. It is shown that the lowest phase noise microwave reference signals have the phase noise within about 10-100 kHz of the carrier controlled by several selected quartz crystal devices while further out the noise is determined by a microwave source.

The phase noise of frequency multipliers and dividers is also explored. Both odd and even-order multipliers offer phase noise which is usually lower than that of available sources. Even-order multipliers, however, tend to be more sensitive to input amplitude and environmental effects than odd-order ones. The phase noise of several selected frequency dividers including microwave types is compared. In general the phase noise of presently available frequency dividers is inferior to that of good frequency multipliers although it is sufficient for most applications. There is a wide disparity between the phase noise quoted in the literature for emitter coupled logic families and that presented here. Special attention must be paid to the environment under which the system has to operate. Vibration, thermal stress, or electrical noise can easily lead to seriously degraded performance.

Signal Model

The signal can be expressed as

$$V(t) = (V_0 + \epsilon(t)) \sin[2\pi\nu_0 t + \phi(t)]. \qquad (1)$$

where V_0 is the nominal peak output voltage, and ν_0 is the nominal frequency of the signal. The time variations of amplitude have been incorporated into $\epsilon(t)$ and the time variations of the actual frequency, $\nu(t)$, have been incorporated into $\phi(t)$. (Complex waveforms such as a square wave can be expressed as the sum of several terms such as given in equation 1.) The actual frequency can now be expressed as

$$\nu(t) = \nu_0 + d\frac{[\phi(t)]}{2\pi dt}. \qquad (2)$$

The fractional frequency deviation is defined as

"US GOVERNMENT WORK IS NOT PROTECTED BY US COPYRIGHT"

$$y(t) = \frac{\nu(t)-\nu_0}{\nu_0} = \frac{d[\phi(t)]}{2\pi\nu_0 dt} . \quad (3)$$

Power spectral analysis of the output signal V(t) combines the power in the carrier ν_0 with the power in $\phi(t)$ and $\varepsilon(t)$ and therefore is not a good method to characterize $\varepsilon(t)$ or $\phi(t)$. Since in many precise signals, understanding the variations in $\phi(t)$ or $y(t)$ is of primary importance, we will confine the following discussion to frequency domain measures of $y(t)$, neglecting $\varepsilon(t)$ except in cases where it sets limits on the measurements of $y(t)$.

Definitions of Phase Noise

Spectral (Fourier) analysis of $y(t)$ is often expressed in terms of $S_\phi(f)$, the spectral density of phase fluctuations in units of rad^2/Hz of measurement bandwidth centered at Fourier frequency f from the carrier ν_0. Intuitively $S_\phi(f)$ can be understood as the mean squared phase deviation in a measurement bandwidth BW centered at Fourier frequency f from the carrier as shown in equation 4.

$$S_\phi(f) = \phi^2(f)/BW \quad rad^2/Hz \quad (4)$$

In practice the measurement bandwidth must be small compared to f especially when $S_\phi(f)$ is changing rapidly with f. $S_\phi(f)$ is uniquely related to the spectral density of fractional frequency fluctuations, $S_y(f)$, as

$$S_\phi(f) = \nu^2/f^2 \; S_y(f). \quad (5)$$

It should be noted that these are single-sided spectral density measures containing the phase or frequency fluctuations from both sides of the carrier. Other measures encountered are $\mathcal{L}(f)$,dBc/Hz, and $S_{\Delta\nu}(f)$. These are related by [1]:

$$S_{\Delta\nu} = \nu_0^2 \; S_y(f) \quad Hz^2/Hz$$

$$\mathcal{L}(f) = \tfrac{1}{2} S_\phi(f) \qquad f_1 < |f| < \infty \quad (6)$$

$$\text{for } \int_{f_1}^{\infty} S_\phi(f)\, df \ll 1 \text{ rad}^2$$

$$dBc/Hz = 10 \log \mathcal{L}(f).$$

$\mathcal{L}(f)$ and dBc/Hz are single sideband measures of phase noise which are not defined for large phase excursion and are therefore measurement system dependent. A more accurate specification of single sideband phase noise would be 1/2 $S_\phi(f)$, which is always well defined [1].

When the frequency of a signal is changed by a fractional amount N using any combination of perfect multipliers or dividers, $S_\phi(f)$ is changed by N^2. The multiplication or division process can be modeled as phase amplification or attenuation. Therefore scaling the frequency by N also scales the phase fluctuation by N as shown in equation 7, where $S_\phi(f,\nu_2)$ is the phase noise of the signal at carrier frequency ν_2, and $S_\phi(f,\nu_1)$ is the original phase noise at carrier frequency ν_1. This is in contrast to heterodyning the signal against a reference frequency (frequency translation), where the output signal contains the phase noise of both sources as shown in equation 8.

$$S_\phi(f,\nu_2) = \left(\frac{\nu_2}{\nu_1}\right)^2 S_\phi(f,\nu_1) + S_\phi^M(f) \quad \text{mult/div} \quad (7)$$

$$S_\phi(f,\nu_2) = N^2 S_\phi(f,\nu_1) + M^2 S_\phi(f,\nu_{ref}) + S_\phi^N \quad \text{translation}, \quad (8)$$

where $\nu_2 = N\nu_1 + M\nu_{ref}$, $S_\phi(f,\nu_{ref})$ is the phase noise of the reference signal, and $S_\phi^N(f)$ is the equivalent extra noise added by the synthesis electronics. The linewidth of the signal can change significantly due to the frequency the synthesis process. The linewidth of the signal, $2f_1$, can be roughly defined as[2]

$$\langle\phi^2\rangle = \int_{f_1}^{\infty} S_\phi(f)\,df = \frac{1}{f_2} \text{ rad}^2 . \quad (9)$$

The fractional power of the signal within a Fourier frequency f_c of the carrier is given by [2]

$$P_c = e^{-\phi_{rms}^2} . \quad (10)$$

An analysis of equations 9 and 10 shows that the linewidth of a signal grows relatively slowly as the frequency is multiplied until the phase modulation due the broadband noise (noise pedestal) approaches 1 rad^2, at which point the linewidth can increase many decades for small changes in N. The reverse is true for division [2]. The frequency at which this abrupt change in linewidth occurs is sometimes referred to the "collapse frequency."

Phase Noise of Available Sources

In order to assess the limits of achievable frequency synthesis it is helpful to review the phase noise of nominally available sources. The most common sources used for precision frequency synthesis are controlled by quartz acoustic devices. Figure 1 shows the

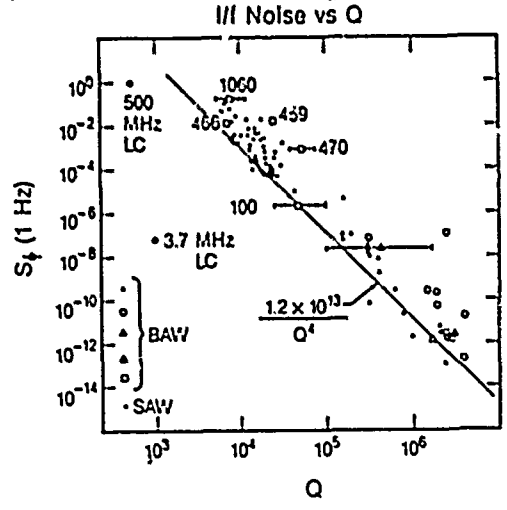


Figure 1. Flicker of frequency levels of a number of quartz resonator controlled oscillators as a function of unloaded Q factor. Adapted from [3-5].

nominally best achieved phase noise 1 Hz from the carrier for such devices measured at the oscillation frequency [3,4]. This figure includes both bulk wave devices (BAW) and surface acoustic wave devices (SAW). Measurements on many quartz devices show that $S_\phi(f) = K/(Q^4 f^3)$ where K depends on the acoustic losses within the quartz and not on the noise in the electronics. Q is the unloaded quality factor. The value of $S_\phi(1\ Hz)$ is characteristic of the flicker frequency level achieved in these oscillators and not the broadband phase noise which is nearly independent of frequency. The flicker performance typically dominates other effects for Fourier frequencies of approximately 0.01 Hz to the half bandwidth of the resonator, $\nu_0/2Q$. At lower Fourier frequencies, drift and temperature effects often dominate. For the best quartz resonators $Q\nu_0 = K = 1.2 \times 10^{13}$. The result of this is that close-in phase noise of the best quartz devices scales as $S_\phi(f) = \nu_0^2/(K^3 f^3)$. This is in contrast to LC oscillators where the phase noise is due to the phase noise in the electronics [3,5]. See Figure 1.

Figure 2 shows a more complete description of the phase noise in three different oscillators under

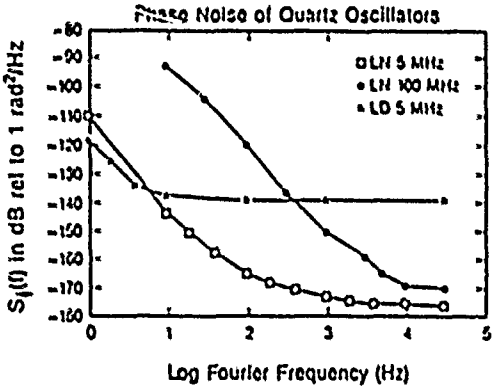


Figure 2. Phase noise of three selected BAW controlled oscillators as a function of Fourier frequency offset from the carrier.

benign laboratory conditions. Near the carrier the phase noise is usually dominated by flicker frequency at a level given by figure 1. At larger Fourier frequencies, the phase noise is determined by the added noise of the electronics and the signal level in the resonator. Figure 3 shows the phase noise of the oscillators of figure 2 multiplied to X-band, assuming a perfect multiplier chain. Figure 3 shows that a composite oscillator system would provide better phase noise over a wide region of Fourier frequencies than a single oscillator. This could be accomplished by phase locking the LN 100 MHz oscillator to the LN 5 MHz oscillator at a unity bandwidth of approximately 400 Hz which is in turn locked to the LD 5 MHz oscillator at a bandwidth of approximately 10 Hz [6]. The oscillators chosen for this example are only used to illustrate the system approach to providing a reference frequency and may not be optimum for all situations. Specifically, SAW oscillators at several hundred MHz have very low phase noise and could be used to reduce the wideband phase noise even further [3,7]. One could also consider using a quartz post filter to reduce the wideband phase noise.

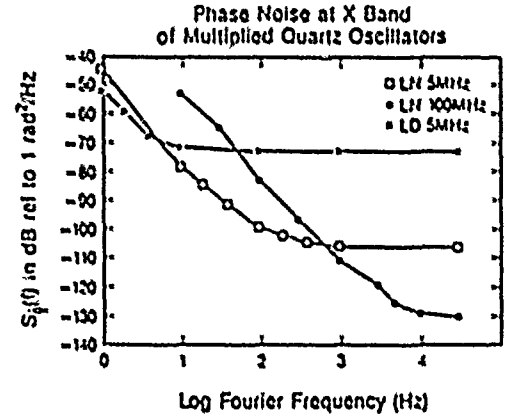


Figure 3. Phase noise of the oscillators displayed in Figure 2 multiplied to 10 GHz.

Under field conditions the performance is often compromised by orders of magnitude due to the vibration sensitivity which for quartz devices is typically of order $d\nu = 2 \times 10^{-9}/g$ acceleration. (For a review see references 8 and 9.) If the vibrations are severe enough, the power in the carrier can be lost as described above [2]. Microwave oscillators also have a relatively high vibration sensitivity which in some cases is worse than that of quartz oscillators. Recently, selected units showing something of an order of magnitude improvement have been tested. Also it has been possible to reduce this sensitivity for Fourier frequencies up to several hundred Hz using compensation techniques [8]. This adds considerable complexity to the oscillator. Temperature variations generally affect the long-term frequency of the oscillator and not the phase noise for frequencies above about 0.1 Hz.

Figure 4 shows a comparison of the phase noise from an optimum combination of the multiplied quartz oscillators shown in figures 2 and 3 along with several microwave oscillators at X-band [10]. Below Fourier frequencies of some tens of kHz, the multiplied quartz oscillators generally have the lowest phase noise and are often used to phase lock microwave sources. With modern solid state amplifiers it is possible to raise the power level of these sources to at least watts without serious compromise of the phase noise. Figure 5 shows the phase noise of several GaAs FET amplifiers along with that traditionally measured in Si bipolar amplifiers up to about 1 GHz [11]. Also included is the typical phase noise of a double balanced mixer using Schottky diodes.

The wideband phase noise of the available microwave oscillators is nearly universally dominated by the added phase noise of the sustaining circuit, $S_\phi(f,Amp)$ [3,12], and the loaded resonator quality factor, Q_L as shown in equation 11 . This relationship is easily derived from the phase shift around the loop required for oscillation, and the induced fractional change in oscillation frequency due to small phase fluctuations, $\Delta\varphi$, given in equation 12.

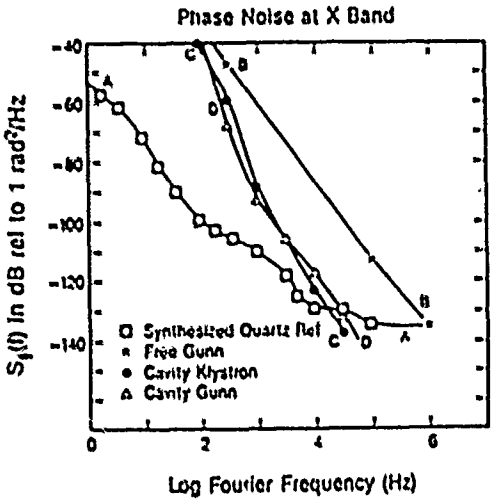


Figure 4. Curve A is the theoretical phase noise performance of a signal optimally synthesized from the multiplied quartz controlled oscillators from figure 3. Curve B is the phase noise of a free running Gunn oscillator [10], Curve C is phase noise of a klystron, injection locked to the reflected signal from a transmission cavity with an unloaded Q of 50,000 [10]. Curve D is the phase noise of a Gunn oscillator, injection locked to the reflected signal from a transmission cavity with an unloaded Q of 50,000 [10].

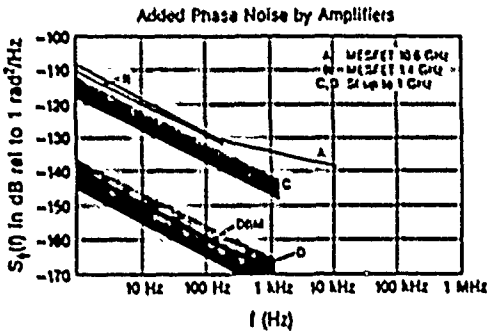


Figure 5. Curve A is the phase noise added to the signal from a MESFET amplifier operating at 10.6 GHz. Curve B is the phase noise of a MESFET amplifier operating at 1.4 GHz [11]. Curve C is the phase noise performance of a generic silicon bipolar amplifier for frequencies up to about 1 GHz with the emitter resistance capacitively bypassed. Curve D is the performance of generic silicon amplifiers with some unbypassed emitter resistance. The nominal value required is the reciprocal of the transconductance [14,16].

$$S_y = 1/(2Q_L)^2 \, S_\phi(f,\text{Amp}) \qquad (11)$$

$$\varphi = n2\pi \quad \Delta\varphi = 2Q_L dy$$

$$n = 0,1,2... \qquad (12)$$

Very close to the carrier the phase noise is often dominated by the frequency drift of the resonator. We know of no data which shows that the phase noise of quality L-C (including cavity) oscillators at all Fourier frequencies beyond a few Hz is not dominated by the sustaining circuit [3,5]. Figure 1 shows data from two different LC oscillators where the phase noise is from 25 to 86 dB less than that expected from a quartz oscillator of similar Q factor [5]. Therefore improvements in the phase noise of the sustaining circuits and or improvements in loaded quality factors would directly translate into better phase noise of the oscillators. This is in contrast to the crystal controlled oscillators where the phase noise within the half bandwidth of the resonators is usually dominated by noise in the quartz resonator [3,4].

Figure 6 shows one method of reducing the wideband phase noise of a source by using a reference cavity.

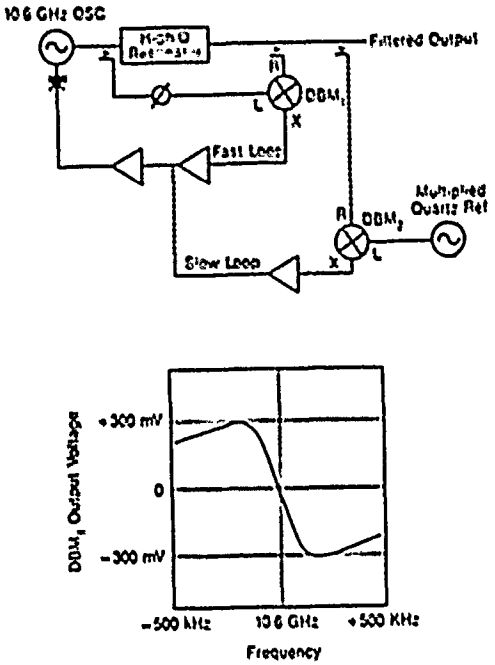


Figure 6. One possible scheme for locking an oscillator to the center of an external cavity with a fast loop in order to improve the phase noise out to Fourier frequencies of order several MHz. The low noise performance in primarily due to the availability of ultra low noise double balanced mixers. Also shown is a method of phase locking the system to a multiplied quartz reference signal. See also [13].

The key practical point here is that the phase noise of the locked system is limited by the expression in equation 11 with the phase noise now given by that of the double balanced mixer. Considerable improvement in phase noise is possible over that available from traditional microwave oscillators using a reference cavity of similar Q factor, since the phase noise of a good microwave double balanced mixer is typically some 30 dB better than that of presently available microwave amplifiers. Figure 7 shows the expected performance with a loop bandwidth of about 2 MHz. Truly extraordinary performance might be expected if the present revolution in high temperature superconductivity lead to high Q room temperature cavities [5,13].

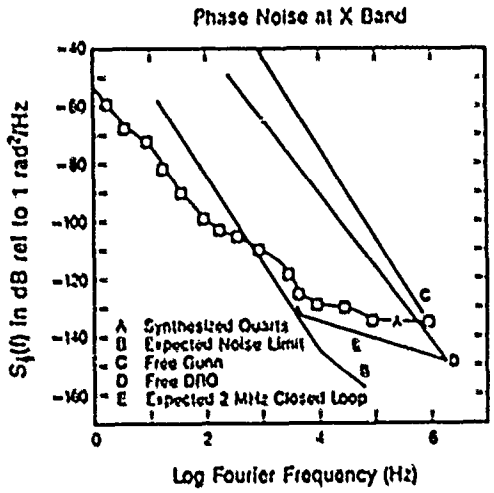


Figure 7. Curve A shows the theoretical phase noise performance of a signal optimally synthesized from the quartz oscillators of figure 3. Curve B shows the theoretical noise floor of the scheme shown in figure 6 A for a reference cavity Q factor of 20,000, a noise in DBM_1 of $S_\phi(f) = 10^{-14}/f$, and a phase sensitivity of 0.5 V/rad. Curve C shows the generic performance of a Gunn oscillator, and curve D shows the performance of a DRO oscillator. Curve E shows the approximate phase noise expected from the 10.6 GHz oscillator locked to an external reference cavity using the scheme shown in figure 6a with a loop bandwidth of 2 MHz and the assumptions used in calculating curve B.

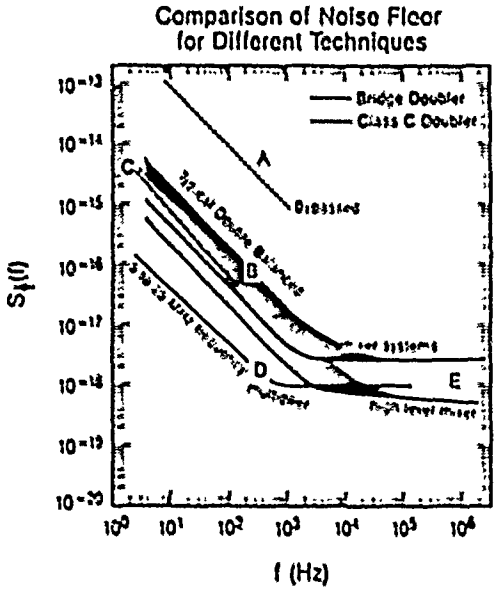


Figure 8. Curve A is the phase noise of a traditional class c doubler with capacitive bypassing of the emitter resister [14]. Curve B is phase noise of the same doubler with 34 ohms of unbypassed emitter resistance. Curve C is the phase noise of a full wave bridge doubler using Schottky diodes. Curve D is the phase noise of a 5 to 25 MHz multiplier using emitter coupled pairs [15]. For comparisons the phase noise in typical double balanced mixers is shown in curve E.

Phase Noise of Selected Multipliers, and Dividers.

The added phase noise due to frequency multiplication can be spectacularly low. Figure 8 shows the phase noise that was measured on a traditional class c doubler with and without emitter bypassing [14], a fullwave doubler using Schottky diodes, and a quintupler using emitter coupled pairs. The traditional even-order multipliers are particularly sensitive to phase variations with amplitude and environment since the zero crossings of the output are not directly associated with those of the input. In the odd-order multipliers such as the emitter coupled pair examined by Baugh [15], the zero crossings of the output are closely tied to that of the input and these devices show relatively low sensitivity to input power level and circuit parameters. In all these devices particular care must be paid to providing sufficient unbypassed emitter impedance in order to suppress the flicker phase modulation in the active junction [14,16].

The phase noise of frequency dividers and digital circuits in general is not very well documented. We have measured the noise in several families of emitter coupled logic (ECL) and have obtained numbers which are generally 6 to 20 dB better than that in the literature [17-20]. Figure 9 shows the phase noise that we measured for several families of divide by 20 circuits versus input frequency. In all cases the phase noise was referred to the output frequency. We found that special care was needed in order to cancel the phase noise of the driving source. Specifically in one divide by 10 circuit where the

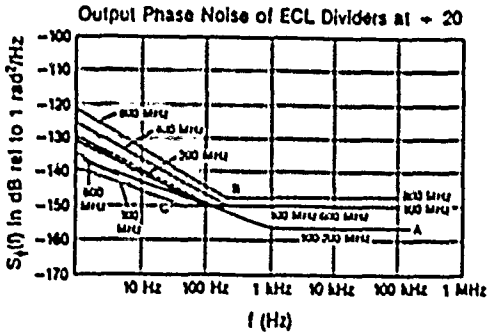


Figure 9. Shows the output phase noise which we measured for 3 different configurations of ECL logic. Curve A was obtained using the scheme shown in divider 1 for both channels of figure 10. Curve B was obtained using the scheme shown in divider 2 for both channels of figure 10. Curve C was obtained using the scheme shown in divider 3 for both channels of figure 10. In order to achieve these results it was necessary to trim the level of the low noise power supplies, to use a very low noise frequency synthesizer for the source, and to adjust the count of the dividers to obtain ~ 90° phase shift between the outputs. A low noise buffer amplifier to translate the logic levels to +10 to +13 dBm drive for the mixers was also necessary.

waveform was not symmetric, we found that the source noise dominated the measurements. Adding a divide-by-2 circuit improved the measured noise by 10 to 30 dB, which is much more than the 6 dB expected from the division process. Originally the measured phase noise was very sensitive to the bias voltage. After adding the divide-by-2 circuit the phase noise was independent of the bias voltage as long as it was within the manufacturer specifications. We used a symmetric push-pull buffer to translate the digital signal to a form appropriate to drive the double balanced mixer as shown in figure 10. These buffer amplifiers have a small signal gain of about 6 dB and typically drive a 50 ohm load at +13 dBm. The buffers are followed by 3 dB pads in order to reduce the standing wave ratio on the cables leading to the mixer. This also greatly reduces the variations in the dc output of the mixer due to changes in the signal amplitude or temperature variations. Typical mixer sensitivities were 0.3 to 1 V/rad at the zero crossing. Figure 11 compares the quoted phase noise for several different divider types available in the literature. We suspect some of the data in the literature were limited by the source noise and/or the output circuit used to drive the mixer. The phase noise of our buffers is so low as not to contribute to the noise.

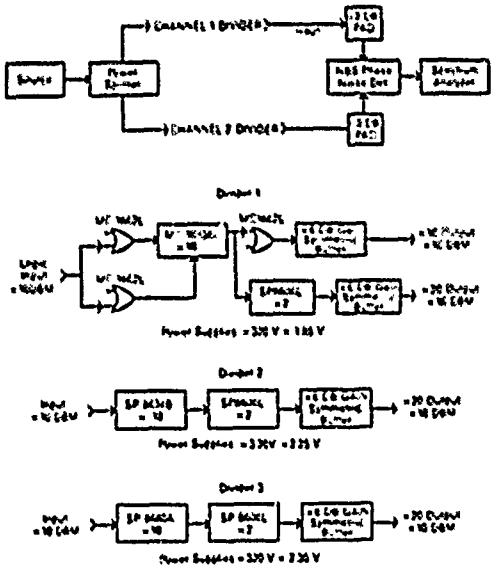


Figure 10. Shows the block diagram of the measurement set up for the phase noise measurements shown in Figure 9. (Certain commercial devices are identified in this paper in order to adequately specify the experimental procedure. Such identification does not imply recommendation or endorsement by the National Bureau of Standards, nor does it imply that the devices identified are necessarily the best available for the purpose).

Stone [21] has brought to our attention a type of regenerative divider which is self starting and should have very low noise, perhaps limited only by the available amplifiers. The general concept shown in figure 12 is similar to that used in divider circuits 20 years ago. We have not had the time to measure the phase noise in such circuits but have demonstrated that it can be used to divide by 2, 4, 6, 8, and 10 with a suitable emitter coupled pair multiplier in the feedback loop. Obviously other division factors are possible using even order multiplication within the loop. This design should operate to frequencies in excess of 40 GHz using present technology.

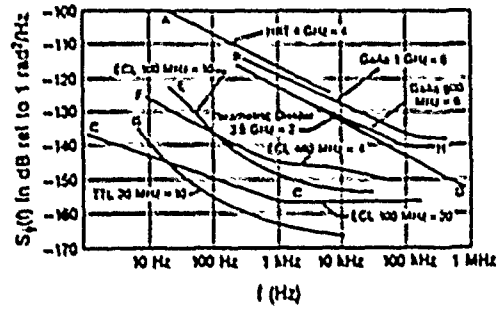


Figure 11. This figure shows the output phase noise of several types of dividers. Curve A shows the phase noise of a divider constructed using GaAs Hetrojunction bipolar transistors. Curves B and H show the output phase noise of a GaAs MESFET based divider for in input frequency of 1 GHz and 500 MHz [17]. Curve C shows the phase noise of the silicon ECL divider 1 of figure 10. Curve D shows the phase noise of a parametric divider that can operate as high as 18 GHz [20]. Curve E and F show the phase noise of two types of silicon ECL dividers from [17]. Curve G shows the phase noise in a TTL divider from [17].

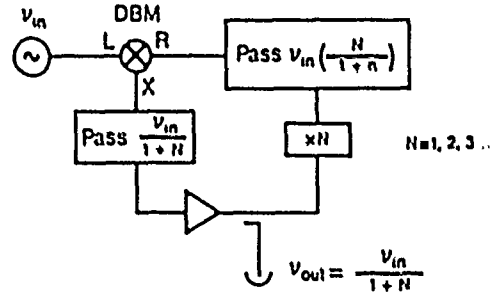


Figure 12. Block diagram of a self starting regenerative divider [21]. This scheme should operate with presently available technology to at least 40 GHz.

Conclusions

We have shown that in general a composite system is capable of synthesizing reference signals with better phase noise than that of any single source. In such a system a multiplied low frequency quartz controlled oscillator generally controls the phase noise close to the carrier while the wideband noise is determined by a higher frequency quartz oscillator and/or by a microwave oscillator. Frequency multipliers generally have lower phase noise than present dividers, although, the noise in several types of dividers is sufficient for many applications. The phase noise of quartz controlled oscillators is determined by the acoustic losses in the quartz and therefore is not likely to improve significantly. The phase noise of LC and cavity controlled oscillators is due to the phase noise of the sustaining amplifier and is likely to improve significantly. In some applications environmental effects can significantly degrade performance.

Acknowledgements

The authors are very grateful for the continued helpful comments and suggestions of their many colleagues especially Andrea De Marchi, Charles Stone, Charles H.Manney, Jean-Jacque Gagnepain, and Raymond Besson. This work has been supported in part by contracts from the Naval Research Laboratory and the CCG Microwave/Millimeter Wave Metrology Engineering and Electromagnetic Field Engineering Sub Groups.

References

[1] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen, Jr., W.L. Smith, R.L. Sydnor, R.F.C. Vessot, and G.M. Winkler, "Characterization of Frequency Stability," IEEE Trans. on I & M, IM-20, pp. 105-120, 1971.

[2] F.L. Walls and A. DeMarchi, "RF Spectrum of a Signal after Frequency Multiplication Measurement and Comparison with a Simple Calculation," IEEE Trans. on I & M. IM-24, pp. 210-217, 1975.

[3] T.E. Parker, "Characteristics and Sources of Phase Noise in Stabile Oscillators," Proc. of 41st Annual Symposium on Frequency Control, 1987.

[4] J.J. Gagnepain, J. Uebersfeld, G. Goujon, and P. Handel, "Relation Between 1/f Noise and Q Factor in Quartz Resonators at Room and Low Temperatures, First Theoretical Interpretation," in Proc of the 35th Annual Symposium on Frequency Control 1981, pp. 476-483; J.J. Gagnepain, M. Oliver and F.L. Walls, "Excess Noise in Quartz Crystal Resonators, in Proc. of the 37th Annual Symposium on Frequency Control, 1983, pp. 218-225.

[5] F.L. Walls, "Other Means for Precision Frequency Control," Chapter 14 in Precision Frequency Control, (Ed. E.A. Gerber and A. Ballato, Academic Press, 1985).

[6] S.R. Stein and F.L. Walls, "Composite Oscillator Systems for Meeting User Needs for Time and Frequency," Proc. of IEEE Trans. on I & M, IM-27, 249-252, 1978.

[7] E.A. Gerber, T.Lukaszek, and A. Ballator, "Advances in Microwave Acoustic Frequency Sources," IEEE Trans on Microwave Theory and Techniques, MTT-34, 1986.

[8] V.J. Rosati and R.L. Filler, "Suppression of Vibration-Induced Phase Noise in Crystal Oscillators: An Update," Proc. of 41st Annual Symposium on Frequency Control, 1987.

[9] F. L. Filler, "The Effect of Vibration on Frequency Standards and Clocks," Proc. of the 38th Annual Frequency Symposium, 31-40 1981.

[10] A. S. Risley, J. H. Shoaf and J. R. Ashley, "Frequency Stabilization of X-Band Sources for Use in Frequency Synthesis into the Infared," IEEE Trans. on I & M, IM-23, pp. 187-195, 1974.

[11] M. D. Hurimann and W. N. Hardy, "Measurement of Flicker Phase Noise of a 1.4 GHz MESFET Auplifer at Temperatures From 300K to 1.26K," Electron Lett. 23, pp. 283-284, 1987.

[12] G.S. Curtis, "The Relationship Between Resonator and Oscillator Noise, and Resonator Noise Measurement Techniques," Proc of 41st Annual Symposium on Frequency Control, 1987.

[13] S.R. Stein and J.P. Turneaure, "Superconducting-cavity Stabilized Oscillators with Improved Frequency Stability," Proc. IEEE 63, 992-993, 1975.

[14] S.G. Andresen and J.K. Nesheim, "Phase Noise of Various Frequency Doublers," IEEE Trans. on I & M, IM-22, 339-342, 1973.

[15] R.A. Baugh, "Low Noise Frequency Multiplication," Proc. of the 26th Annual Symposium on Frequency Control, 334-342 1972.

[16] D. Halford, A.E. Wainwright, and J.A. Barnes, "Flicker Noise of Phase in RF Amplifiers and Frequency Multipliers: Characterization, Cause, and Cure," Proc. of the 22nd Annual Symposium on Frequency Control, 326-331 1968.

[17] U.L. Rohde, Digital PLL Frequency Synthesizers, (Englewood Cliffs, NJ: Prentice Hall, Inc., 1983).

[18] P. Bailey, "Eliminate source noise from phase noise test," Microwaves & Rf, 25, pp. 87-91, 1986.

[19] K. K. Agarwal, W. J. Thompson, and P. M. Asbeck, "Dividers Implemented with GaAs Heterojunction Bipolar Transistors," Electron. Lett., 21, pp. 1005-1006, 1985.

[20] Private Communication, J. Bednarz, Telemus Electronic Systems Inc., 310 Moolin, Ottawa, Ontario Canada K2H6G3.

[21] Private Communication, C. Stone, Brightline corp., P.O. Box 1016, Cedar Park, Texas

LOW NOISE AIRBORNE SYNTHESIZER FOR FREQUENCY AGILE RADAR

A. Vulcan and M. Bloch
Frequency Electronics, Inc.
Mitchel Field, NY 11553

Summary

A low noise, KU-Band synthesizer developed for use in a high performance fighter aircraft radar is described. The synthesizer incorporates both direct and indirect techniques to generate exceptionally pure signals with low phase jitter and spurious content. Fast switching speed and phase noise of -125 dBc/Hz from 10 KHz to 40 MHz is obtained The radar operates in an environment which has a high level of random vibration and static G loading, and the mechanical design incorporates vibration isolation techniques. The selection of synthesizer electrical parameters also assures optimum performance over the full range of mechanical stresses. The master oscillator from which the various fixed and steppable frequency outputs are derived uses an SC cut crystal with a G sensitivity of less than $3 \times 10^{-10}/G$.

In order to fit within an existing package size, the control and test functions, as well as several RF subassemblies are hybridized in hermetically sealed thick-film assemblies. Two fixed UHF and L-Band sinewaves and 3 sets of digital clock signals are produced. Electrical and physical isolation of individual functional elements insures that cross-talk does not perturb the KU-Band output frequency.

System Description

The development of the KU-Band synthesizer emphasizes low phase noise performance over severe environmental conditions while maintaining electrical characteristics which are compatible with high resolution airborne multi-mode radar requirements. The major features are:

- Fast Warm-Up
  - SC-Cut Crystal in Master Oscillator
  - Thick Film Hybridized Oscillator
  - Booster Oven
  - Low Thermal Inertia
- Low Phase Noise During Vibration
  - Mechanical Vibration Isolation
  - Low G-Sensitivity Crystal
  - Encapsulants and Adhesives
- Low Spurious Signals
  - Multipole Crystal, Helical, and Cavity Filters
  - Generation of KU-Band Output by Dual Conversion Technique Minimizes Spurious Signals
  - High Isolation Between Functional Groups
- Small Size - Light Weight
  - Thick and thin Film Hybrids
  - Surface Mounted Miniature Components
  - Integrated Subassemblies
  - High Dielectric Constant Substrates

The first two features are implemented in the system by proper design of the master oscillator which supplies a VHF reference signal to the synthesizer. The reference source is a dual oven, Butler oscillator, using an SC cut crystal. The oscillator/buffer amplifier and oven controllers are hybridized in order to minimize size and weight. The components within the master oscillator are staked to the supporting structure and the overall assembly is encapsulated with a low density foam material. The master oscillator is suspended in a mechanical isolation structure which attenuates low frequency vibration inputs while encountering 4 G static loads in any axis.

Direct synthesis is used to enhance speed and minimize system phase noise. A high degree of filtering is used within the synthesizer to keep the spurious signals at extremely low levels. A crystal filter at 150 MHz, as well as helical filters at UHF frequencies, and cavity devices at KU-Band reduce the spurious signals to levels below -60 dBc.

The size and weight of the synthesizer is minimized through the use of thick and thin film hybrid components and the use of integrated subassemblies to a maximum extent. The various functions are grouped and packaged in separate mechanical chassis to facilitate maintenance and reduce manufacturing and test time.

Performance Characteristics

Digital clock signals as well as RF sinewaves at various microwave frequencies are synthesized from the master oscillator. Dual KU-Band outputs step over a 1 GHz range. The specifications for the KU-Band Synthesizer are summarized below.

- Output Frequencies
  - 50 MHz
  - 25 MHz
  - 12 MHz
  - 165 MHz
  - L-Band selectable
  - KU-Band
- Frequency Stability
  - +/-2 PPM Accuracy for Ten Years
  - Stabilize to +/-2 PPM of Nominal 2 Minutes after Turn-On
- Spurious Signals
  - -60 dBc
- Phase Noise
  - -46 dBc/Hz at 35 Hz for 0.7 G^2/Hz Input
  - -121 dBc/Hz Floor at 10 KHz

CH2427-3/87/0000-519 $1.00 C 1987 IEEE

- Size
  - 6.5 Inches Wide x 4.0 Inches High x 8.0 Inches Long

- Weight
  - 12 pounds

The frequency characteristics of the synthesizer outputs are determined by the long and short term stability of the VHF master oscillator. The crystals are selected for low aging since the radar mission dictates that the frequency shall be within +/-2 PPM for a 10 year period without adjustment. Stabilization of the frequency to +/-2 PPM occurs within 2 minutes of turn-on since an oven current boost circuit provides immediate warm-up. A hybridized oven controller is designed to minimize overshoot and ensure that the frequency vs. temperature characteristics remain within a 0.1 PPM window for ambient temperatures ranging from -55°C to +85°C.

Since the output signal is derived by direct synthesis, the Ku-Band phase noise is degraded by approximately 50 dB relative to the master oscillator. The contribution of the channel stepping oscillator/multipliers is minimized since the selected channel frequency is mixed with the multiplied signals and consequently, the phase noise requirements are less than the LO's derived by direct multiplication.

Electrical Design

Figure 1 is a block diagram of the KU-Band synthesizer. The master oscillator is the reference source for all output frequencies and is the only subassembly that is vibration isolated. The block diagram illustrates how the various output signals are generated through multiplying and mixing. The channel step synthesizer consists of crystal oscillators and hybrid X10 multipliers whose outputs are phase locked to the VHF master oscillator. The appropriate output of the oscillator/multiplier bank is selected by a high speed, high isolation pin diode switch. Since the loops are always locked to the reference signal, there are no loop dynamic considerations which influence the switching speed. Switchable LO's and switchable filter banks ensure that the system spurious signals are kept at a level consistent with the radar requirements.

The major criteria determining the loop bandwidth of the phase locked oscillators is the random vibration spectrum which the synthesizer is subjected to. The loop bandwidth of the channel select synthesizer must be sufficiently large to ensure that system phase noise is determined only by the vibration isolated master oscillator. Since the contribution of random noise above 2 KHz is negligible, and the crystal oscillator output is multiplied by a factor of 10, the loop bandwidth is kept at approximately 15 KHz. This ensures that the phase noise contribution of the channel step synthesizer is minimal and that the loops remain locked over full environmental extremes. Using this approach, the KU-Band phase noise is determined solely by the vibration induced noise in the reference oscillator multipled by 20 LOG N.

Single sideband phase noise is related to frequency deviation Δf and modulation frequency fm by;

$$L(f) = 20 \log\left(\frac{\Delta f}{2fm}\right) \text{dBc/Hz}$$

In the presence of random vibration $\gamma \frac{g^2}{Hz}$

$$\Delta f = \sqrt{2\gamma} \times fo \times S$$

Where fo is oscillator frequency and S is oscillator vibration sensitivity in parts per G.

$$\text{Thus, } L(f) = 20 \log\left(\frac{\sqrt{2\gamma} \times fo \times S}{2fm}\right) \text{dBc/Hz}$$

In the synthesizer, vibration induced oscillator phase noise is enhanced by 49 dB when translating from 55 MHz to 15 GHz.

The random vibration level which the synthesizer is subjected to in the aircraft is approximately 3G RMS. The spectrum peaks at 35 Hz and it is this energy which must be reduced by the vibration isolator to insure that the radar requirement is met. This limit is plotted in Figure 2 where the dynamic phase noise is plotted at offset frequencies from 10 Hz to 100 MHz.

Since the phase noise under vibration is directly proportional to the crystal G sensitivity, every crystal manufactured for use in the synthesizer is subjected to a sine vibration sweep in the 3 orthogonal axes to characterize the devices. In the operational environment, the worst vibration is encountered in the aircraft vertical axis and consequently, the best crystal axis from the standpoint of minimum vibration sensitivity is aligned with this axis. This is done within the oscillator assembly during the alignment procedure. Each completed oscillator is again characterized for both sine and random vibration spectra response after alignment is completed. This hard mounted test of the oscillator insures that no internal resonances exist and phase noise contributions due to factors besides the crystal are negligible.

Figure 3 compares the specified phase noise to the vibration induced phase noise of an oscillator having a crystal with a 3×10^{-10}/G sensitivity. It is shown that the hard mounted oscillator exceeds the specification limits by 25 dB at 35 Hz. Consequently, the vibration isolator must be designed to have attenuation at least equal to the difference between the two curves of Figure 3. It is imperative, of course, that all crystals have G sensitivity of less than 3×10^{-10}/G. In practice, values of 2.2×10^{-10}/G are routinely achieved and in some cases 1.5×10^{-10}/G is attained in the worst axis.

Performance Under Vibration

The anti-vibration mount design employs a system of looped, stainless steel, multi-strand cables to achieve a high level of isolation within a small space. Figure 4 illustrates the packaging concept. Each end of the oscillator package is supported by a pair of flexed cable loops and a pair of cantilevered cable posts. Isolation is achieved in three mutually perpendicular axes either by flexure of the cable loops (Axis 1) or by deflection of the cantiliver posts (Axes 2 and 3). The isolator acts as a second order low pass mechanical filter with a rolloff of 8 to 10 dB per octave. The design allows a sway space of 0.8 inches in each axis permitting 4.5 G's of static acceleration loading without bottoming.

Figure 5 shows the dynamic vibration attenuation characteristic of the mount. It is seen that over the band of 20 to 100 Hz, sufficient isolation exists to enable the oscillator to meet the phase noise requirement during vibration with a crystal having 3×10^{-10}/G or less vibration sensitivity.

Figure 6 is the KU-Band phase noise at the synthesizer output when the system is subjected to the random vibration spectrum. It is seen that the required performance is met over the full frequency range. Figure 7 is a block diagram of the test setup used to make the presented measurements. An identical synthesizer, which is not subjected to vibration, is phase locked to the unit under test and the Ku-Band outputs are mixed, amplified, and fed to a spectrum analyzer. Appropriate measures must be taken to calibrate the system and ensure that the mixer inputs are in phase quadrature.

Mechanical Design

The synthesizer, which consists of seven separate modules, is packaged in a carrier measuring 8.0 inches long x 6.5 inches wide x 4.0 inches high. The overall synthesizer weighs approximately 12 pounds. The carrier assembly, as well as the individual modules, are milled out of single aluminum blocks to ensure structural rigidity and minimize performance degrading mechanical resonances. The synthesizer mounting surface in the radar is the thick metal plate on which the master oscillator is mounted. Thus, the random vibration spectrum is transmitted directly to the oscillator without any intervening structural members having unpredictable dynamic characteristics.

Conclusion

A Ku-Band synthesizer having low vibration induced phase noise is described. Electrical and mechanical design parameters are selected to optimize performance when the synthesizer is subjected to simultaneous static and dynamic G forces.

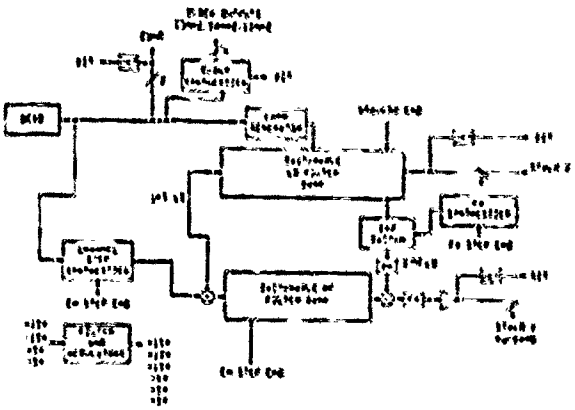
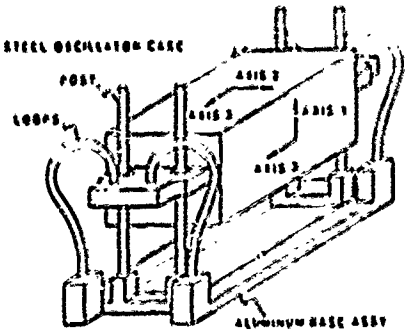


FIGURE 1
BLOCK DIAGRAM, KU-BAND SYNTHESIZER

FEI VIBRATION ISOLATOR

FIGURE 4



NATURAL FREQUENCY:	AXIS	IN
	1	5Hz
	2	6.5Hz
	3	7Hz
SUSPENDED MASS:	0.6#	
SWAY SPACE:	0.6"	

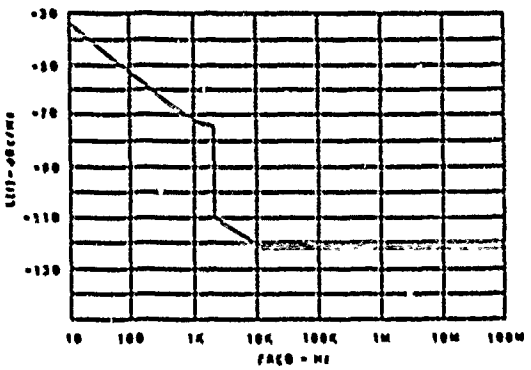
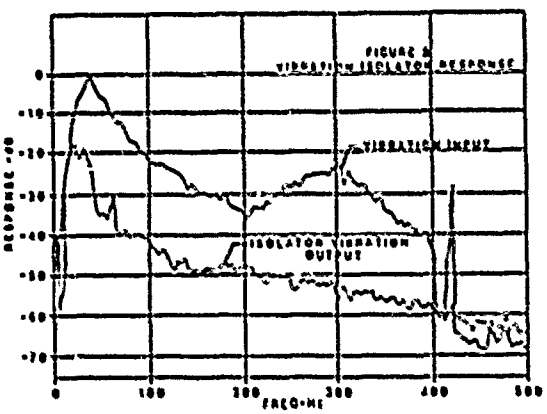
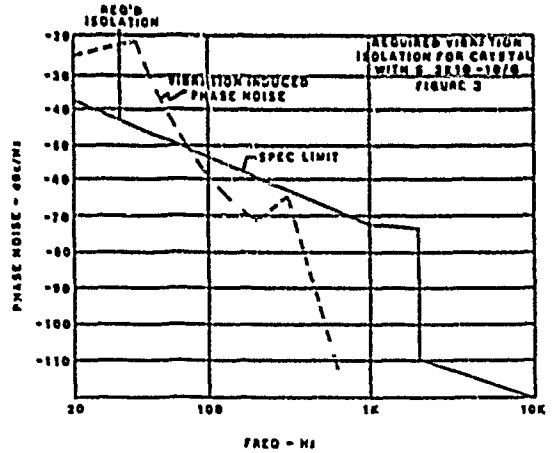
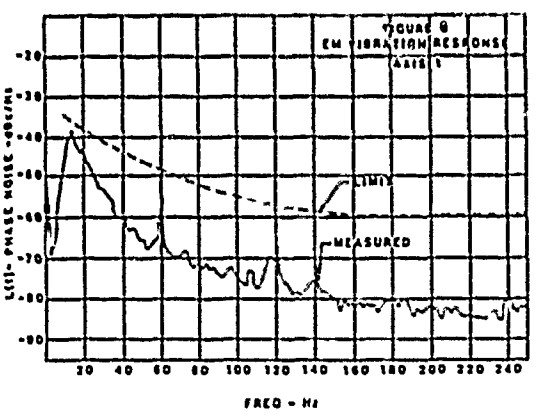


FIGURE 2
KU-BAND PHASE NOISE LIMIT







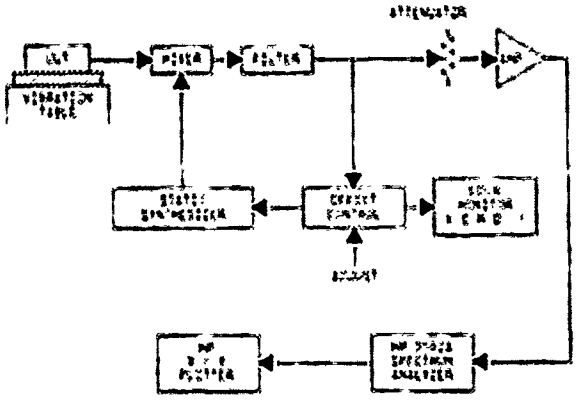


FIGURE 3
PHASE NOISE TEST SETUP

41st Annual Frequency Control Symposium - 1987

IMPROVEMENT OF THE PULL-IN RANGE AND ACQUISITION TIME OF A MICROWAVE P.L.L. SYSTEM BY INJECTION LOCKING THE V.C.O.

Jordi Berenguer, Javier Bará, Eduardo Artal, Ignasi Corbella and Adolfo Comerón.

Departamento de Electrofísica. E.T.S. Ingenieros de Telecomunicación de Barcelona.
Universidad Politécnica de Cataluña.
Apartado 30002, 08080-Barcelona, Spain.

SUMMARY

The requirements of large pull-in range and short acquisition time in microwave and millimeter wave phase locked oscillators are of great importance since the free running VCO frequency may be some tens of MHz off the reference frequency. A substantial improvement can be obtained by injection-locking the V.C.O. to the reference signal; as a result, the V.C.O behaviour changes to a phase controlled fixed frequency oscillator, giving a reduction by one in the system order. Then, frequency acquisition is performed by the injection-locking process, and phase tracking is made by the P.L.L. system. We present experimental comparisons between a microwave P.L.L. system operating at 11 GHz, with an without injection-locking synchronization.

INTRODUCTION

The synthesis of microwave signals often requires, in addition to spectral purity and frequency stability, a large pull-in range and short acquisition time. This is due to the fact that in the microwave range, the free-running V.C.O. frequency can be some tens of MHz off the reference frequency, thus making acquisition by pull-in very long, if possible at all; note that in second order P.L.L.'s with active loop filters, practical considerations may result in a reduction of the D.C. gain of operational amplifiers well below their nominal values of about 100 dB, therefore reducing the P.L.L. pull-in range.

This paper describes how to improve the pull-in range and the acquisition time by injection locking the V.C.O. to the reference signal.

Alternatively this system can be viewed as an injection-locking process in which a phase control between oscillators has been added. This results in some interesting possibilities such as power combination of oscillators, and design of electronic phase shifters and phase modulators.

PLL BEHAVIOUR WITH INJECTION LOCKING

The standard simplified equivalent circuit for a P.L.L. system is sketched in the Figure 1. Part (a) shows the physical model, and (b) the linearized block diagram, were Kd is the phase detector constant in V/rad, F(s) is the loop filter transfer function, and Kf is the V.C.O. constant in rad/s/V. The pull-in range is given by /1/ :

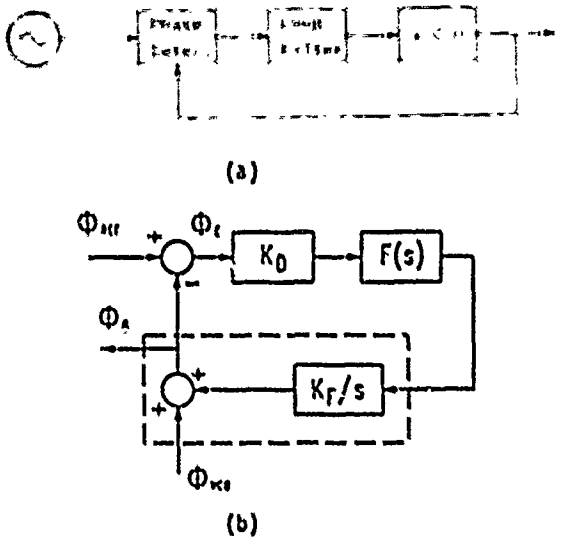


Figure 1.- Standard P.L.L. system. a) Physical model, b) linearized block diagram.

$$\Delta\omega_p = 2\sqrt{K_f K_d F(0) 2B \Big/ \left[1+\frac{1}{4\zeta^2}\right]}$$

and the approximated acquisition time for a second order P.L.L. with an imperfect integrator as loop filter is given by :

$$T_p = 2.4\,\zeta^2\left(1+\frac{1}{4\zeta^2}\right)^2\frac{\Delta f^2}{B^3}$$

in terms of the loop bandwidth B, the damping coefficient ζ, the loop filter dc gain F(0), and the initial open loop offset frequency Δf.

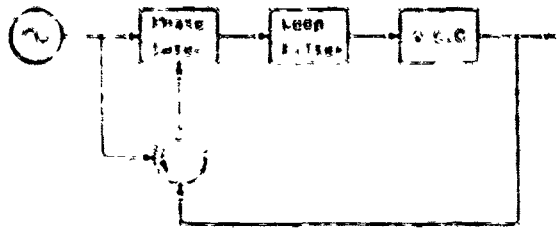
The performance of this P.L.L. can be improved by fundamental or subharmonic injection locking of the V.C.O. to the reference signal. Figure 2a shows synchronization at the fundamental; in this case, use of a circulator is required. On the other hand, if the injection locking is subharmonic, the circulator is avoided, but the synchronization bandwidth is reduced.

The equivalent functional model is given in Figure 2b, which differs from Figure 1b in the V.C.O. behaviour, now a phase controlled fixed frequency oscillator, characterized by a constant Kp in rad/V. As a consequence, the system order is reduced by one.

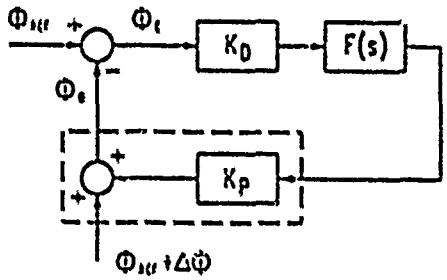
The synchronization bandwidth is now fixed by the injection locking process, and it is given by /2/ :

$$\Delta\omega_m = 2\,\frac{\omega_o}{Q_{ext}}\Big/\sqrt{\frac{P_i}{P_o}}\quad\frac{1}{\cos\upsilon}$$

CH2427-3/87/0000-524 $1.00 C 1987 IEEE



(a)



(b)

Figure 2.- P.L.L. system with a VCO injection locking synchronization at the fundamental. a) Physical model. b) Linearized block diagram.

in terms of the operating frequency ω_o, the external Q of the controlled oscillator, the injected power level Pi, the oscillator output power level Po, and the angle θ between the device line and the impedance locus of the controlled oscillator.

The acquisition time is also reduced, since it now depends only on the initial phase difference between the frequency synchronized signals. If an imperfect integrator with time constants $\tau 1, \tau 2$, is used as loop filter, the acquisition time is obtained from the step response of a first order system:

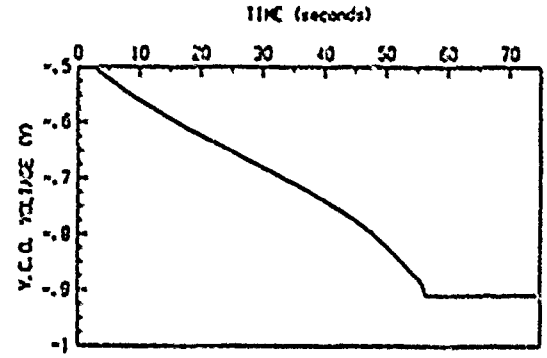
$$\phi_o(t) = \Delta\phi \frac{\tau_1}{\tau_1 + K_d K_p \tau_2} e^{-t/\tau} + \phi_{ref}(t)$$

$$\tau = \tau_2 + \frac{\tau_1}{K_d K_p}$$

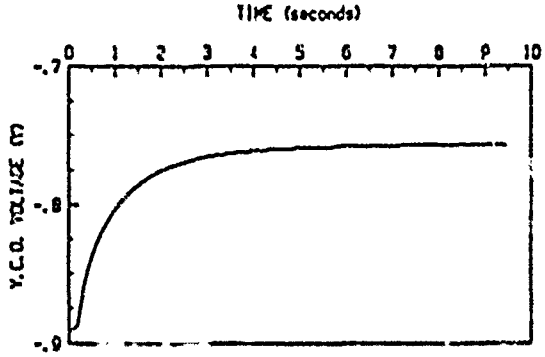
EXPERIMENTAL RESULTS

A comparison between a microwave P.L.L. system with and without injection locking was performed. The operating frequency was 11.43 GHz, the V.C.O. was a FET oscillator, its frequency controlled by the gate-source bias voltage with a value of Kf= 4.08E+6 rad/seg/V. As a phase detector a balanced microwave mixer using a microstrip hybrid ring was used, yielding a value of Kd= 0.381 V/rad.

In the P.L.L. circuit, a loop bandwidth of 10 KHz, and a damping factor ζ=1 were chosen. For practicals reasons, the loop filter dc gain was limited to 68 dB, giving a theoretical pull-in range of 21.8 MHz. The acquisition time for an initial offset of 20 MHz was 60 s (Figure 3a).



(a)

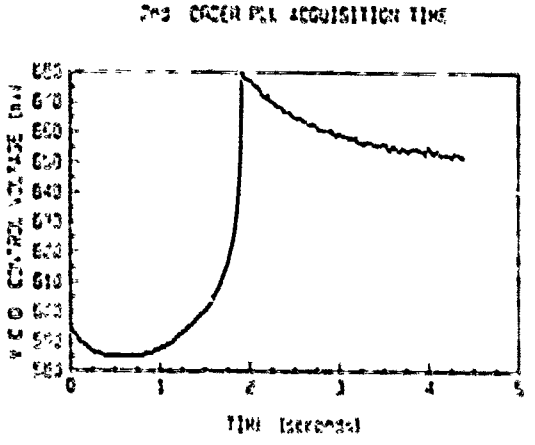


(b)

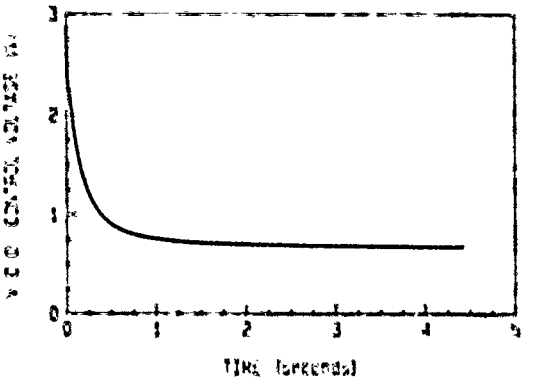
Figure 3.- Phase acquisition time. a) Second order P.L.L. b) with the injection locking synchronization.

If we include the injection locking process, with Pi= 3.5 dBm, Po= 14 dBm, Qext= 30, cos θ = 1 the synchronization bandwidth is 227 MHz. Figure 3b shows the measured values of the voltage applied to the V.C.O. with injection locking for an initial frequency offset of 20 MHz; the measured value of Kp is 20.7E+3 rad/V, and the loop filter parameters were τ1= 8.6E-3 and τ2= 1.5E-4. It is observed that locking is accomplished after 10 s, in good agreement with the theoretical time constant of τ = 1.08 s. The improvement is therefore a reduction by a factor 6.

The same comparison was made with another P.L.L. with loop bandwidth of 52 KHz, damping coefficient of 1.15, VCO constant 71.5E+6 rad/s/V, phase detector constant 0.1 V/rad and loop filter parameters τ1= 1.2E-3 and τ2=3E-5. The pull-in range obtained was 2 MHz using a D.C. loop filter gain of 38 dB. The acquisition time for a 300 KHz initial offset frequency was 2 s. (Figure 4a). If we include the injection locking process, with Pi=-12 dBm, Po=+14 dBm. Qext=14.5, cos θ = 1, the synchronization bandwidth measured is 78 MHz, with a Kp = 0.055 rad/V while the acquisition time is reduced around 0.5 s (Figure 4b).



(a)



(b)

Figure 4.- Phase acquisition time. a) Second order P.L.L. b) with injection locking synchronization.

CONCLUSIONS

A substantial improvement of the pull-in rang and the acquisition time in a second order microwav P.L.L. system can be obtained by the use of injectio locking synchronization of the V.C.O. As a result th V.C.O. changes its behaviour, and becomes a phas controlled fixed frequency oscillator, thus reducin the system order by one.

It is also possible to use this technique in a Injection Locking Amplifier, in order to keep a give phase difference between the oscillators. This featur is of special interest for power combination o oscillators (area in which results will be soo reported) and also for the design of a class of ± 90 electronic phase shifters or phase modulators.

REFERENCES

[1].- F.M. Gardner, "Phaselock Techniques". John Wile & Sons, 1979.

[2].- Kaneyuki Kurokawa, "Injection Locking o Microwave Solid-State Oscillator". Proc. of th IEEE, 61, 10, Oct. 1973, pp. 1386-1410.

41st ANNUAL FREQUENCY CONTROL SYMPOSIUM - 1987

PROGRESS REPORT ON THE EIA P-11 ROUND ROBIN CRYSTAL MEASUREMENTS EXPERIMENT

by

W. L. Smith
555 Union Boulevard
Allentown, PA 18103

ABSTRACT

In 1985 a new Standard Method for the Measurement of the equivalent Electrical Parameters of Quartz Crystal Units was adopted by the Electronics Industries Association. Designated EIA-512(1985), this document was later accepted by the American National Standards Institute as a National Standard. The method is based on the use of modern Network Analyzer equipment to determine the two-port scattering parameters of a crystal unit when inserted into a suitable test fixture. The test fixture and measuring system are "calibrated" with standardized reference impedance elements, permitting full correction of measured responses. In 1986 the P-11 Engineering Committee initiated a Round Robin experiment to estimate the precision with which the equivalent parameters of "representative" crystal units could be determined at several locations with different but similar equipment, using the procedures of EIA-512.

Crystal units ranging from 5 MHz fundamental mode to 146 MHz fifth overtone mode were made available for the experiment. Six locations have volunteered to participate, and the experiment has been in progress since about December 1st of last year. A brief description of the experiment is presented as well as a preliminary report of results from the tests completed to date.

CH2427-3/87/0000-527 $1.00 © 1987 IEEE

INTRODUCTION

The new measurement standard was accepted for the following reasons: the technique has been well proven through many years of wide application in microwave metrology; it has been adapted for crystal unit measurements as an internal reference method by several laboratories since it was first described over fifteen years ago; several published papers have reported very good precision of measurement at frequencies well above the capabilities of competing methods; it is one of the few methods of high frequency measurement which can readily be made traceable to National Standards laboratories; and probably more importantly the necessary automated equipment became available at greatly reduced costs during the past few years. However, concern was expressed within the P-11 group that very little data was available confirming the correlation of measurements between different locations. It is recognized that being able to repeat the measurements of devices to very high precision within one laboratory does not establish the practical tolerance which may reasonably be placed on devices which will be measured by manufacturers at one location, a quality assessment group at another, and a user's incoming inspections at still a third. It was decided at the Quartz Devices meeting held at Kansas City in August of 1986 that a Round Robin experiment including several locations, using crystals similar to the types in common manufacture, and providing sufficient data to have statistical significance would provide the needed information for guidance.

A subcommittee was formed to define and coordinate such an experiment on behalf of the P-11 Committee. Members are Clifford Mercer, Reeves-Hoffman; Thomas Schuyler, Colorado Crystal Corp.; Raymond Filler, Fort Monmouth; Tim Semones, Innovative Measurement Solutions; Robert Kinsman, Motorola; Wolfgang Pflaum, Hughes Aircraft; and the author. A brief meeting of this group was held at Kansas

City, and after some telephone discussion, an experiment was designed, as described in the following section. The Round Robin began last December, is now in progress and data has been submitted from half the participating locations. A preliminary analysis of the available data is presented here.

SCOPE OF THE EXPERIMENT

The experiment being conducted is based on the measurement of nine crystal units selected as "representative" devices. Frequencies cover the range from 5.38 MHz to 146.6 MHz, with fundamental, third and fifth overtone devices included. Series Resistance R1 values range from approximately 3 Ohms to 50 Ohms, and motional inductance L1 values from about 3 mH to 90 mH. The nominal parameters of the crystals are given in Table 1. The range of parameters and Q values is noted to encompass those commonly encountered in a large proportion of commercial manufacture. Admittedly, better precision and correlation of data would be obtained if the test units were all high precision overtone mode devices having Q's of two million and capacitance ratios of twenty five thousand; however, the results obtained would have much less significance to the Industry. The crystals selected were chosen to be somewhat special in having relatively small temperature coefficients of frequency in the vicinity of room temperature, as it was recognized that different locations would probably exhibit as much as several degrees deviation from each other, and small coefficients would probably be easier to correct. The units had all been manufactured at least a year before the experiment began, and had been measured several times to verify that little aging of frequency or of motional parameters would be expected during the several months required for completion of the tests.

It was agreed that the number of participants should be limited to about five in

order to shorten the duration of the experiment and keep the effort required for data analysis within reasonable bounds. Six locations were ultimately selected from those volunteering, and represent users as well as manufacturers. A list of the participating organizations is given in Table 2.

In the interest of keeping data within reasonable limits, it was agreed that only measurements in accordance with EIA-512 should be included in the experiment, and that the parameters recorded would consist of series resonance frequency Fs, series resistance R1, motional inductance L1 and shunt static capacitance Co. Each data entry however would be identified by the location code, the sequence number and measurement number within the sequence, the test fixture used, the device number, the temperature at the time the measurement was made, and the drive current used for the test.

As mentioned above, the experiment was drafted to provide enough data for meaningful statistical analysis. To this end a total of at least fifty measurements of each device are being made at each location. With automatic test equipment, such a scope becomes possible, although still a very significant contribution of time and effort on the part of each participant. These measurements are made in the following manner so that effects of "calibration" of the test equipment can be identified:

A total of five sets, or sequences, of measurements are made, each sequence consisting of a calibration of the system followed by a group of ten repeat measurements of each of the devices. A set of repeat measurements is made by placing the device into the test fixture, installing a draft shield, waiting for thermal stabilization, and then measuring the device ten times without disturbing either the device or the equipment.

As one source of systematic error which might be difficult to assess is that caused by differences in test fixtures, it was decided that one test fixture complete with required reference impedances for calibration would be circulated with the test kit, and each location would make five sequences of measurements with this 'reference' fixture. If possible, each participant is urged to carry out a second round of five sequences of tests with another fixture (which should be of the general type specified by the Standard, but may be of considerably different construction than the reference fixture).

Another source of systematic error is the effect of temperature. To allow at least an approximate correction for its effect, the temperature coefficient of frequency was measured for the test crystals beforehand and stored in the dedicated data base. Then, after each individual measurement, the temperature of the test fixture/draft shield in which the crystal is installed for measurement is determined and recorded. No attempt was made to assess temperature dependence of the other parameters beforehand (but their existence became evident from the data, as discussed later).

Because of the relatively large volume of data generated by an experiment of this scope, it was considered essential to be able to transfer the results from the measurement equipment into the dedicated data base without the need for manual transcription. Thanks to the efforts of Tim Semones, who made available a special copy of the IMS measurement program which provides for recording data on disc in a particular format, and to Tom Schuyler who implemented a means of transferring data from floppy discs to magnetic tape, it was possible for me to ultimately place all data in a dedicated Round Robin data base maintained at AT&T Bell Laboratories at Allentown.

To minimize the probability of catastrophic loss in transit between locations, the test kit is being hand carried from one location to the next. This imposes additional

burden on the participants, but seemed prudent in view of the large investment of time and effort represented by these units - loss of one unit would mean a significant reduction in information from the entire experiment.

Table 3 includes the specific instructions issued to each participant as a reminder of the procedures to be used.

RESULTS TO DATE

In October of 1986, a small group including Canon Bradley, Tom Schuyler, Tim Semones, Ed Miles, Tony Palmieri and the author, met at the AT&T Bell Laboratories in North Andover for a trial shakedown of the entire measurement and data transfer procedure. With some minor edits of software and final calibration of the test fixtures, several measurements of some of the crystal units were made according to the instructions provided, and data was saved to disc in the prescribed manner. The participant instructions were edited to reflect the findings of the test. The sample data disc was returned to Colorado Crystal, where the data was transcribed to magnetic tape and later stored on the data base at AT&T-Allentown without serious mishap.

The reference fixture and a special copy of the measurement program were made available by Tim Semones of Innovative Measurement Solutions, and together with the crystal units, the test kit was ready and the experiment officially got underway in December.

Scheduled measurements have been completed by three locations in time to be included in this report. This represents one half of the total experiment, providing results that should be at least indicative of the final assessment. Two of these locations were able to provide data with a user fixture, and all three provided a full complement of measurements with the reference fixture, giving a grand total of about 250 individual

tests of each of the nine devices, or a total of about 2250 separate measurements to be examined. Since four parameters are retained from each test, this amounts to about 9000 separate data elements. With such an abundance of data, it becomes difficult to decide just which of the possible groupings and dependences should be examined. The results presented in the following sections are one approach to understanding what the data is telling us.

RESULTS WITH REFERENCE FIXTURE

As the first step, the data obtained with the reference fixture were considered with only the measured Fs values being corrected to 25 °C using the temperature coefficients which had been measured beforehand. The mean values of each of the four parameters for each crystal unit and their standard deviation were calculated for the groups of 150 measurements completed to date. These are listed in Figure 1. Three of these results show standard deviations significantly greater than the rest: Fs measurements for the Unit #3, R1 data for Unit #1 and R1 data for Unit #8 are sufficiently different to warrant further examination before proceeding. Correlations between the groups of data and measurement temperature were then calculated, as it is not unusual to find that all motional parameters of crystal units are dependent on temperature to at least some degree. Figure 2 summarizes the results of these calculations, where it is noted that in many cases very significant correlation coefficients are obtained. Since even Fs is seen to be strongly correlated to temperature, we conclude that the coefficients measured beforehand were not as accurate as we would have liked.

One way of minimizing the effects of systematic deviations of data is to calculate the slope of a linear regression of the data on the suspected causative parameter. When a significant correlation exists, a near linear dependence is likely, and can be essentially

removed by using the slope to extrapolate all data to the mean value of the causative. These data were treated in this manner, and the mean and standard deviations once more calculated. Figure 3 lists the results after this "second order" temperature correction. Clearly, a significant improvement is achieved. (For example, the standard error of Fs measurements of Unit #9 has decreased from over 0.3 ppm to less than 0.1 ppm.). Examination of Figure 3 shows that in most cases the standard deviations of the measured values are reasonably distributed, but again three sets of data exhibit significantly larger variability the any other sets. Fs data for Unit #3 has nearly twice the standard error of its nearest competitor, and R1 values for Unit #1 and Unit #3 have considerably larger standard deviations than do all the other units.

As temperature is the only identifiable systematic casusative we can clearly identify at this point, we proceed with a more detailed analysis. The mean values and standard deviations of each sequence of ten repeat measurements were calculated, identifying a particular sequence as being all measurements made with the same calibration of the system. Since we are still examining only those measurements made with the reference fixture, there are fifteen such sequences in all, five made by each of the participants. These sequences were rather arbitrarily identified with numbers 1 thru 15. It is possible to trace a particular sequence number to a location and date, but that will not be done for purposes of this presentation.

Figure 4 lists the difference between the mean values of Fs of each sequence of data and the overall mean values listed in Figure 3. Examination of this list with particular interest in the results for Unit #3, we see that the wider divergence of these measurements is not due to a single large error in one sequence value, but that rather there seems to be a general instability of Fs here in comparison to the other units. No reason for this behavior has as yet been hypothesized. The only other sequence value

that appears to be markedly different from all others for the same unit is the value of Fs for Unit #1, sequence 1. While the actual offset is less than 0.2 ppm, this is significantly greater than any other offset of Unit #1 values.

Figure 5 lists the similar data for R1 measurements. Examination of this list indicates that sequences 6 and 8 have extreme errors for Unit #1, while sequences 6 and 12 for Unit #3 have excessive offsets. The reason for these abnormalities in R1 determination are not readily apparent at this time, although scatter plots of residuals indicate patterns which differ significantly from the same plots for the other units. It is interesting to note that omitting these four suspicious sets from the analysis, the standard error of R1 measurements for each unit range from 0.5℃ to 1.9℃ and if all residuals for all units are considered as a single group, the standard error becomes 1.1℃.

Figure 6, listing the sequence mean errors from reference L1 data, and Figure 7 for Co data, contain no extraordinary values. A standard error of all L1 data for all units of somewhat less than 0.4℃ is well within expected performance. It was not in fact expected that L1 data would have less error than R1 measurements; however, this may be only characteristic of these particular crystal units. The overall standard error in Co data of somewhat less than 2℃ is considered quite good, as this amounts to an actual error of only a few hundredths of a picofarad.

Figure 8 lists the standard deviations of the ten repeat measurements of Fs in each sequence. Here we note that the instability of Unit #3 is further substantiated by showing standard deviations of data within sequences more than twice those of the other units. Figure 9 shows the same measures of the R1 data, and here we note that the standard deviations of data within sequences is essentially the same for those sets which showed extreme offsets as for the others, thus, whatever caused the large offsets

in mean value for those sets of data remained stable during the particular sequence. Figure 10 and Figure 11 list the same standard deviations for L1 and Co data. Again there are no surprises.

In Figure 12, the correlations of the parameters to measurement temperature are once more calculated using the data with second order corrections, and we verify that indeed any linear dependencies have been removed, as we should expect.

RESULTS WITH OTHER FIXTURES

Similar analysis of data obtained with the two USER fixtures thus far has been carried out. The offsets are once more calculated as the differences from the mean values of all data for each unit obtained with the reference fixture, as before. The same regression slopes are used to make the second order temperature corrections as were used in the correction of the reference fixture data.

Figure 13 summarizes the Fs measurements with the USER fixtures. Here the differences from the reference values are consistent with the results of Figure 4, except that the apparent instability of Unit #3 is not evident in this case.

Figure 14 shows the offsets of R1 data, and here we note three instances of abnormally low values of R1 (negative offsets) in sequence 2 with fixture EIA2P. No good explanation is known at this time; we only offer the observation that this sequence of data was obtained at the same location as sequence 6 of Figure 5, on the next day. Unit #4 showed a large positive deviation in the same sequence. If sequence 2 with EIA2P is removed from consideration, the results are reasonable. Examination of the L1 data shown in Figure 15 show self consistent measurements with each USER fixture; however, the results with C-USER are consistently negative, while just the opposite is true with EIA2P. In both cases, there is a trend toward smaller values at higher

frequencies, which though not large are statistically significant. The reason is not clear, but might be related to the fact that different thru-line devices were used with the different fixtures, and a constant line length was assumed at all frequencies even though it was recognized that the bending of the line in fact makes its effective length vary with frequency. In the case of EIA2P, the two piece fixture is calibrated with no physical through line, the two ports being connected together with a special coupling.

Figure 16 indicates that the Co values determined with EIA2P are also significantly different from those obtained with either the reference fixture or with C-USER. This difference is not understood at this time, although it may be due to the same effects as mentioned above.

CONCLUSIONS

From the discussions above, we may form some tentative conclusions at this time, remembering that only half the total experiment is completed. First, it may be stated that if sufficient care in maintaining (or correcting for) temperature, and well-behaved devices are used, we can expect to correlate Fs measurements performed at different locations with a standard error only slightly greater than 0.1 ppm. Also, R1 data from different locations should achieve precision approaching 1°c with adequate temperature correction and with carefully characterized fixtures, while L1 values may agree to somewhat better than this. Some unexplained differences between results obtained with test fixtures of different construction have been noted, but as yet remain unexplained. We are hopeful that additional data may shed some light on the abnormalities thus far noted.

Based on the results to date, we also will conjecture that environmental effects will usually limit the precision with which parameter values may be expected to correlate

when measurements are made at different times or locations, unless great care is taken.

ACKNOWLEDGEMENTS

I would like to express my appreciation for the assistance of the subcommittee members in formulating and organizing this project. Special thanks are due to Colorado Crystal Corporation and to AT&T Bell Laboratories at North Andover for making available the crystal units needed and for the preliminary measurements of temperature coefficient; to Canon Bradly, Tom Schuyler, Ed Miles, Tim Semones, and Tony Palmieri for assistance with the 'shakedown' of the experimental procedure; to Tim Semones for making available the reference test fixture kit to be circulated, and for preparing the special version of the test program to permit machine handling of all the data; and especially to the volunteers who have (or will soon have) devoted so much time and effort to carrying out the measurements. Without the help of all of you, this experiment would not be possible.

41st ANNUAL FREQUENCY CONTROL SYMPOSIUM - 1987

HIGH FREQUENCY STABLE FREQUENCY SOURCES FOR ADVANCED SYSTEMS

Ann Harrison, John Bowsett
and David Sharpe

Quartz Crystal Division
STC Components, Harlow, Essex, ENGLAND

Summary

A novel method for the compensation of crystal oscillators has been developed. It is primarily intended for AT cut crystal oscillators and offers the advantages of higher stability, lower power consumption and smaller size when compared with more conventional methods of compensation.

Compensation is achieved by generating a power series representation of the control voltage for a voltage controlled crystal oscillator. A more detailed description of the method was given at this symposium in 1983 (ref. 1).

Initially an integrated circuit was developed which generated four voltages each varying as a function of temperature. One voltage is a constant whilst the others have a linear, quadratic and cubic variation with temperature. These voltages are combined in a summing amplifier to provide the compensation voltage required.

Oscillators have been manufactured using this IC over the last 3 years and ±1ppm over -40°C to +85°C is being achieved with very good yields.

A second IC is under development which generates additional voltages, with these ±0.5 ppm over -40°C and +85°C has been achieved.

It has also proved possible using the first integrated circuit to compensate third overtone crystals and tight tolerance devices up to 400 MHz are being developed.

Introduction

There has been a continuing demand over the years for oscillators having increased frequency stability, smaller size, lower power consumption, higher frequency and reduced phase noise.

The first of these parameters, the frequency stability, is usually the key parameter and very often determines the specification achievable for the other parameters.

Very tight frequency stability can be achieved by the use of oven control to keep the device at constant temperature effectively eliminating ambient temperature variations. There are however severe penalties to pay in terms of size, power consumption and cost with this approach.

Temperature compensation is an alternative which, whilst it cannot achieve the stability of an ovened device, provides a much smaller, cheaper solution. A temperature compensated oscillator is essentially an oscillator whose frequency can be varied by the application of a d.c. voltage combined with a circuit which generates a temperature varying voltage. The temperature varying voltage is fed into the crystal oscillator in such a way as to "compensate" for temperature variation in the oscillator frequency.

Conventional thermistor compensated oscillators were seen to have reached their practical production limit at around ±3ppm over -40°C to +85°C and a radical new approach was therefore considered necessary.

A bipolar integrated circuit was designed to generate a power series representation of the desired control voltage. The power series generated is ideally suited for the compensation of oscillators using AT cut quartz crystals.

It was also possible to incorporate, within the IC, much of the voltage controlled oscillator circuit and parts of the output circuit, thus creating a temperature controlled oscillator in which the majority of the circuit elements are contained on a single IC.

Compensation Voltages

The compensation voltage required can be expressed in the form

$$V(T) = a_0 + a_1 (T-T1) + a_2 (T-T1)^2$$

$$+a3 (T-T1)^3 + \ldots\ldots + a_n (T-T1)^n$$

where a ....a are the weighting coefficients.

T is the ambient temperature

T1 is the reference temperature

A good approximation can be obtained using only the first four terms of the series (a 3rd order curve), but for greater precision higher order terms can be added.

CH2427-3/87/0000-539 $1.00 C 1987 IEEE

An integrated circuit has been designed to generate four voltages. One voltage is constant while the others have a linear, quadratic and cubic variation with temperature respectively. To compensation voltage is obtained by summing these voltages in an operational amplifier using suitable coefficient-setting resistors.

In practice, the integrated circuit does not generate voltages of the form a (T-T1) since these terms are difficult to generate accurately and are subject to measurement errors. Instead it generates voltages corresponding to the Chebyshev polynomials. The equations of the first four Chebyshev polynomials are:

$T_0(X) = 1$
$T_1(X) = X$
$T_2(X) = 2X^2 - 1$
$T_3(X) = 4X^3 - 3X$

Figure 1 shows these polynomials plotted graphically.

Equation 1 can now be re-written as:

$$V = a_0 T_0(X) + a_1 T_1(X) + a_2 T_2(X) + a_3 T_3(X) + \ldots + a_n T_n(X)$$

Apart from To (X), the dominant terms for AT cut crystals are the linear and cubic terms, and T (X) alone is a good approximation to the desired compensation voltage.

Integrated Circuit Design

A block diagram of the elements of the compensation system is shown in figure 2. The constant voltage is generated using a voltage stabiliser and generates a temperature invariant voltage using a bandgap reference circuit. The linearly-varying voltage with temperature is generated using the temperature coefficient of the base-emitter junction voltage of a transistor, operated at constant emitter current.

Conventional analogue multipliers comprising four transistors with the collectors cross-coupled are used to generate voltages with quadratic and cubic variations with temperature. The summing amplifier is included and the integrated circuit also contains an oscillator circuit and two output stages to form the voltage controlled oscillator.

Compensation Methods

In order to compensate a crystal oscillator using this method, the values of five resistors have to be determined.

These resistors set the values of the coefficients and they are determined from data measured at a number of temperatures over the operating temperature range. At each of these temperatures the compensation voltage needed to keep the oscillator at the correct frequency is measured.

The output voltages from the linear, quadratic and cubic generators are also measured. A computer program is then used to calculate the coefficients. This program uses the exchange algorithm to derive the minimax linear approximation for the set of measurement points. The problem is similar to that of solving a set of simultaneous equations with no exact solution; the coefficients are adjusted until the errors between the synthesised compensation voltage and the required compensation voltage are of a specified form. This approximation minimises the maximum error between the synthesised compensation voltage and the desired compensation voltage. Once the resistors are fitted the performance of each oscillator is verified by measuring the frequency variation of the unit over the operating temperature range.

Production Results

The first IC has now been in production use for nearly three years and well over 5,000 oscillators have been manufactured. Oscillators ranging in frequency from 3MHz up to 40MHz are being produced against a variety of specifications. Typical plots of frequency against temperature are shown in fig. 3, 4 and 5. Figure 5 also shows the very excellent hysteresis performance obtained with these devices.

Phase noise is a parameter which is assuming greater and greater importance.

System designers are demanding ever improved phase noise from oscillators and Fig. 6 shows a typical phase noise response achieved with the integrated circuit oscillators.

Wherever possible surface mount components have been used in these devices to minimise size and weight. The smallest device produced to date with this technique is less than 4 cu cm in volume and weighs less than 3 gms.

Tighter Compensation

A second integrated circuit is under development which generates further voltages. Initial trials have given results of the order of ±0.5 ppm over -40°C to +85°C, a typical performance curve shown in Fig 7. This device has been compensated over -50°C to +90°C and is achieving better than ±0.5 ppm including hysteresis which is well under 0.1 ppm everywhere in the temperature band.

To achieve these tight tolerances it has proved necessary to introduce a further process stage to overcome the errors introduced by SOT resistor tolerances and changes in circuit load conditions caused by the adding of the SOT resistors. After the verification run the results from this run and the compensation run are compared and the resistor values trimmed to optimise the performance.

Higher Frequencies

Because of the greater pulling voltage range generated by the IC, compensation of more "rigid" crystals, that is overtone crystals, can be achieved to a much higher degree than with thermistor techniques. With traditional crystal manufacturing techniques, third overtone crystals are available up to around 80 MHz, but new developments in plasma etching carried out in conjunction with Standard Telephone Laboratories (STL) enable STC to produce fundamental crystals up to 150 MHz (third overtone to 450 MHz)

Compensated oscillators are now being produced using these crystals at 376 MHz. A typical frequency vs temperature performance is illustrated in Fig. 8 with better than ±1 ppm being achieved over 0°C to +70°C.

Phase noise and ageing are shown in Figs. 9 and 10 respectively.

The very excellent ageing figure indicates a stability of better than 1×10^{-8} per day after one months ageing.

Conclusions

Two integrated circuits have been developed to provide a novel compensation technique.

The first of these ICs has been successfully introduced into production. A wide range of oscillators has been developed using this technique and is being continually enhanced.

An oscillator is being produced which satisfies all the key criteria demanded by designers of advanced systems; high frequency stability, miniaturisation, low power consumption, high frequency and reduced phase noise.

A second integrated circuit is under development which will enhance still further the frequency stability available from this range of oscillators.

Coupled with these circuits the development of much higher frequency fundamental crystals is providing very stable high frequency devices.

References

i) James S Wilson
Precision Crystal Oscillator Design
Thesis for degree of Doctor of Philosophy
University of Aston, October 1983.

ii) An Improved Method of Temperature Compensation of Crystal Oscillators
J S Wilson
STC Components
37th Annual Symposium on Frequency Control, 1983.

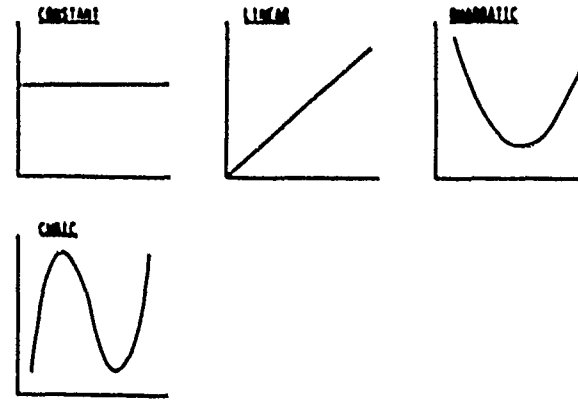


Fig. 1 - Cherbychev Polynominals

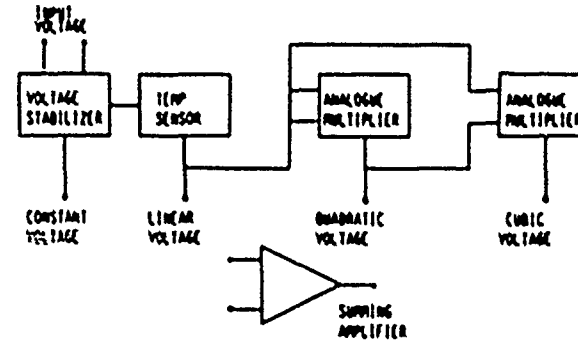


Fig. 2 - Integrated Circuit - Block Diagram

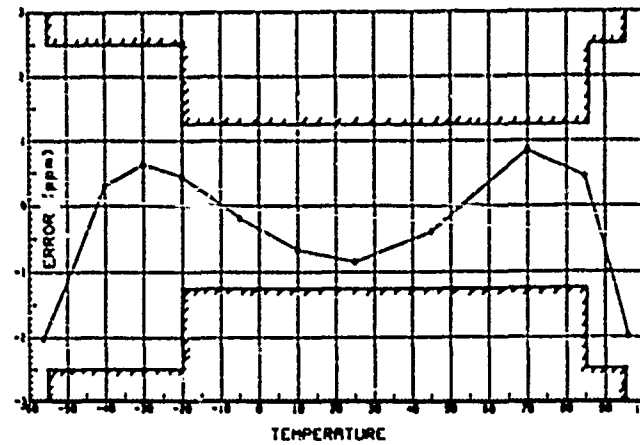


Fig. 3 - Third Order Compensation

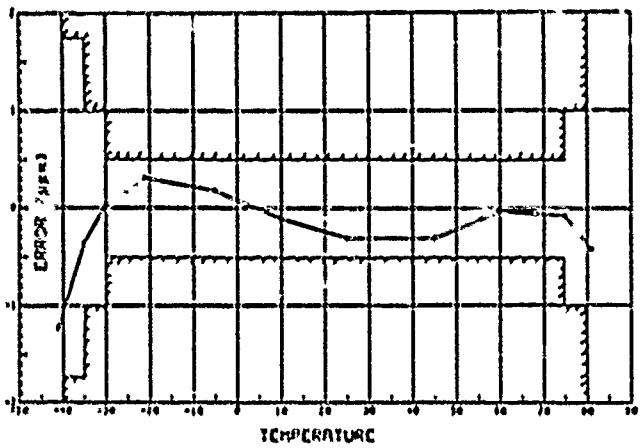


Fig. 4 - Third Order compensation

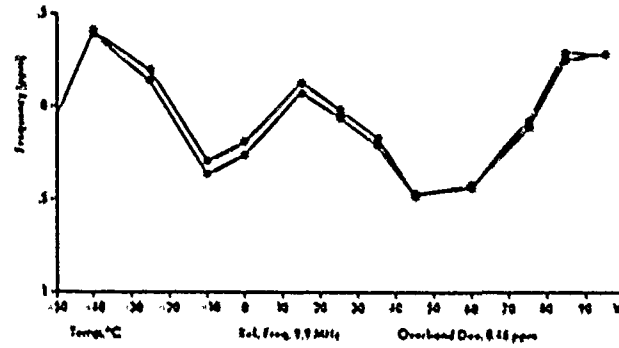


Fig. 7 - Fourth Order Compensation
- with Hysteresis

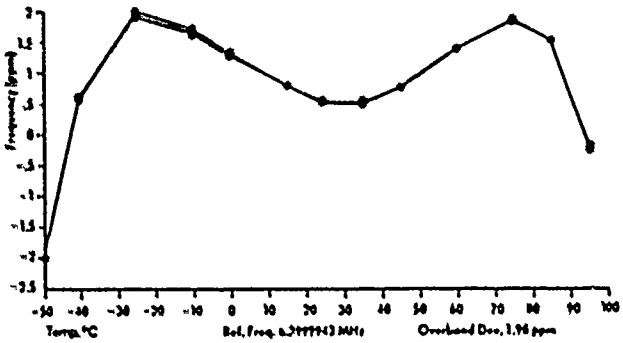


Fig. 5 - Third Order Compensation
- with Hysteresis

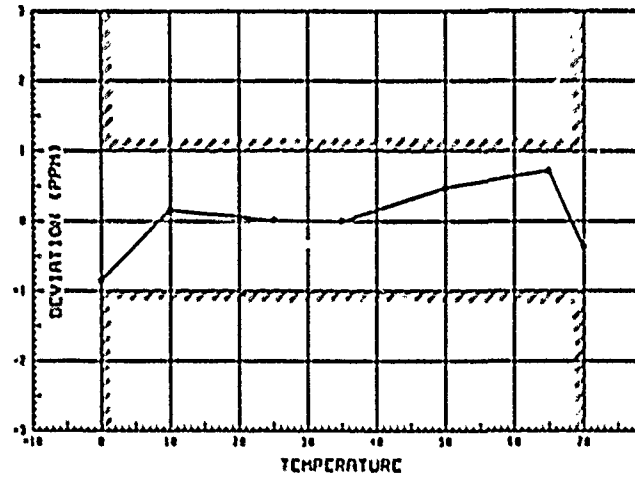


Fig. 8 - 376 MHz Third Order Compensation

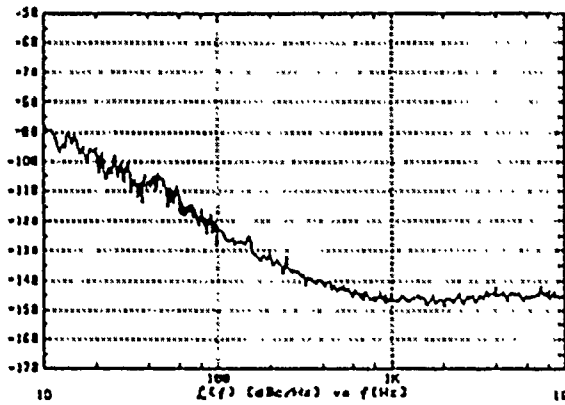


Fig. 6 - Third Order Oscillator - Phase Noise

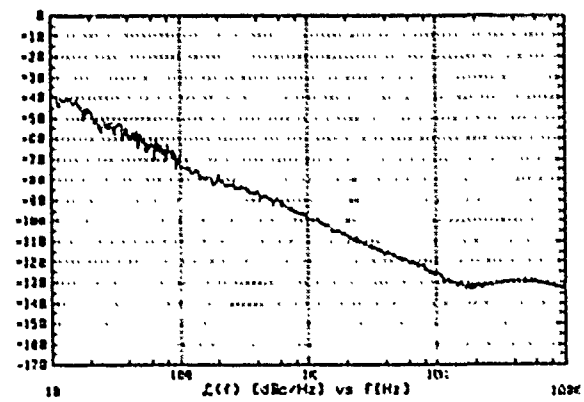


Fig. 9 - 376 MHz Oscillator - Phase Noise

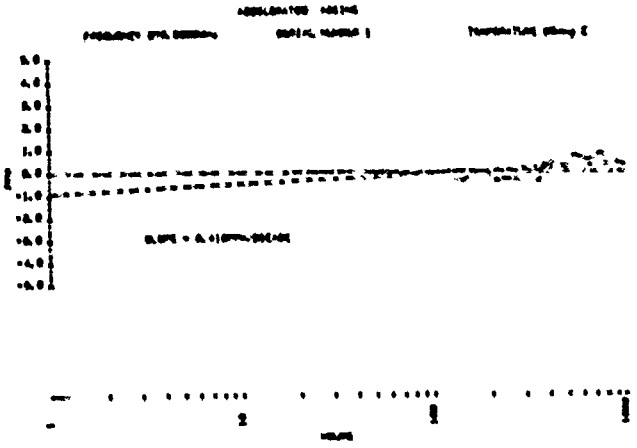


Fig. 10 - 376 MHz Oscillator - Ageing

41st ANNUAL FREQUENCY CONTROL SYMPOSIUM - 1987

A HIGH LINEARITY SAW ACCELEROMETER

David Hower, Mark Cracknell
and Ann Harrison

Quartz Crystal Division
STC Components, Harlow, Essex, ENGLAND

Summary

In this paper the design, implementation and performance of a SAW accelerometer aimed at the sub inertial applications area is described.

The device exploits the frequency shift of a SAW resonator when it is strained by acceleration. The use of a quartz substrate provides high stability with temperature and time. This stability is enhanced by using two SAW resonators, operating in a differential mode which provides an output directly compatible with digital electronics.

Developments in quartz cuts, mounting techniques and electronic circuitry are improving these devices still further.

Principles of Operation

The effects of strain on quartz bulk and surface wave devices have been well described in the literature. They are normally regarded as a nuisance to be minimised or eliminated.

In a sensor application we exploit the effects of acceleration induced strain and attempt to optimise the structure for maximum sensitivity.

The structure selected for the SAW accelerometer uses a quartz beam mounted as a cantilever. SAW resonator patterns are defined in an aluminium layer deposited on one surface of the beam. As the beam deforms, the resonance frequency of each of the SAW patterns shifts due to the changes in pattern dimensions and the piezoelectric constants of the quartz.

For the quartz cuts used to date, typically orientations between ST-X and AT-X, the pattern shift effect dominates.

Two resonators are used, one optimised for maximum acceleration sensitivity and one for minimum acceleration sensitivity.

Each of the resonators is used as the control element of an oscillator and the outputs of the two oscillators are mixed. The difference frequency output of the mixer is amplified and shaped for convenient measurement by a frequency counter. The circuit configuration is shown in Fig. 1.

The strain in the beam is proportional to the imposed acceleration and the shift in the output difference frequency is a very linear function of the strain.

The use of a differential output minimises temperature and ageing effects of the SAW devices, provided that the external circuitry for each SAW matches over temperature and time. The high phase slope of a resonator reduces the effect of external circuitry drifts and the low loss allows a simple, single-stage maintaining amplifier for each SAW device.

Mechanical Realisation

To optimise the performance of the devices several design features must be achieved in conjunction.

The two SAW resonators must have an acceleration sensitivity as different as possible.

The temperature coefficients and ageing characteristics of the two SAW resonators and their associated oscillators must be matched as closely as possible.

Several mechanical constructions have been tried, two of which are described.

The first of these, Fig. 2, uses two separate beams mounted in a single dual-in-line package. The beams are cantilever mounted on a quartz substrate which is then mounted on the dual-in-line header. Quartz of the same orientation is used as a substrate in all these assemblies to reduce to an absolute minimum any mounting strain effects which would distort temperature performance or increase ageing.

One of the beams was mass loaded with tungsten to increase its acceleration sensitivity.

The major disadvantage of this construction was the difficulty of obtaining a perfect match for temperature coefficient from the two resonators.

To overcome this problem the construction, as illustrated in Fig. 3, was tried. A single quartz beam was used and two resonator patterns deposited, one adjacent to the cantilever mount for maximum sensitivity and the second remote from the mount.

CH2427-3/87/0000-544 $1.00 C 1987 IEEE

A wedge-shaped beam was employed and mass loading added to increase the bending effect near the mount. The major problem encountered with this device was setting the gap between cantilever bam and quartz substrate.

Results

The resonators used in this device have typical SAW parabolic frequency versus temperature response. An example is shown in Fig 4 having a turn over at 46°C and a parabolic constant of 3×10^{-8}.

Calculations on two such curves predicts a combined overband performance as shown in Fig. 5, with a very linear portion within 1 ppm over 5°C to 65°C and a total spread of less than 4 ppm.

The overband performance actually achieved is shown in Fig. 6. The flat portion of the response is moved down in temperature, indicating that the oscillator circuits are not fully matched or that the mixer is distorting the results. It still, however, represents an order of magnitude improvement over the SAW resonator performance, a performance of better than 15 ppm over -40 deg. C to +80 deg. C being achieved.

Figs. 7 and 8 show the linearity of the accelerometer response in a centrifuge up to 30 g. The output shows a better than 5 ug/g^2 linearity in the +ve g direction and approximately 21 ug/g^2 in the -ve g direction. Both directions show a small hysteresis of around 10 mg.

Sensitivities in the range of 1000 - 1500 Hz/g have been achieved with these constructions and the sensitivities along the orthogonal axes are less than 1% of this value.

Performance Objectives and Achievements

Table 1 shows the initial performance objectives and the achievement against these to date.

The full scale range required for these devices has been fully met as have the linearity requirements but further development is required to meet the full sensitivity and stability requirements.

	Objectives	Achievements
Full scale range	±20g - ±200g	±20g - ±200g
Sensitivity	1000 Hz/g to 2000 Hz/g	400 Hz/g to 1500 Hz/g
Linearity	20 ug/g^2	20 ug/g^2
Zero drift - Time (48 hrs)	0.01% FSD	0.02% FSD
- Temperature (-40 - 80°C)	0.1 mg/°C to 0.5 mg/°C	10 mg/°C to 20 mg/°C

TABLE 1

Future Development

Development to improve the accelerometer performance still further is currently addressing three areas, alternative quartz cuts, improved oscillator and mixer circuitry and alternative physical construction.

Two alternative quartz cuts have been evaluated, LST and FST. Resonators constructed using LST quartz have shown improved temperature coefficients but have also been shown to be virtually insensitive to acceleration and therefore of little use in this application. The second cut FST, as well as showing small improvements in TC gives a greatly increased strain sensitivity of up to 4 KHz/g and this cut is now being used for further developments.

The electronic circuitry is being redesigned to minimise its effects on the very good TC match of the SAWs and the possibility of some form of temperature compensation being evaluated.

The new construction being evaluated is illustrated in Fig. 9. To increase the sensitivity of one resonator against the other the beam is thinned beneath the cavity region of the resonator nearer the hinge.

Initial trials indicate a sensitivity of around 2 KHz/g on ST-cut quartz. The groove in these initial samples was cut with a saw and because of the mechanical damage to the edges of the cut the mechanical strength of these devices was not good. It is expected to be able to overcome this problem by the use of plasma etching to cut the groove and using this method it will be possible to shape the profile of the groove to give optimum strain sensitivity and strength.

Conclusions

A SAW accelerometer using quartz as the sensing element has been developed giving a stable linear output and having many advantages over more conventional devices.

- The output is a pulse train which is easily interfaced to digital systems. Mass-balance and piezo accelerometers require extensive signal conditioning before further use.
- The structure is relatively simple and robust.
- The power requirements are low (9 mA at 5V)and independent of acceleration.
- The device is unaffected by external magnetic fields.
- The inherent stability of quartz gives high stability with time and temperature.

Acknowledgement

This work was supported by the Royal Aircraft Establishment, Farnborough, UK. We would particularly like to thank Dr D Miles of RAE for general advice on the accelerometer field and providing some of the performance data on the STC prototypes.

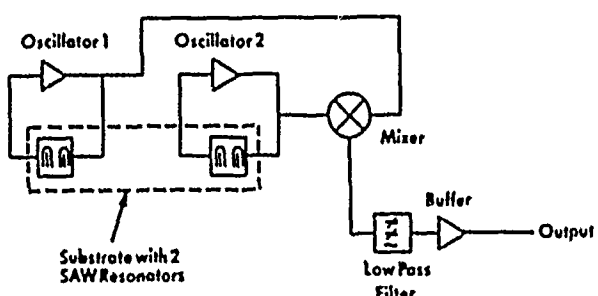


Fig 1. - Circuit Configuration

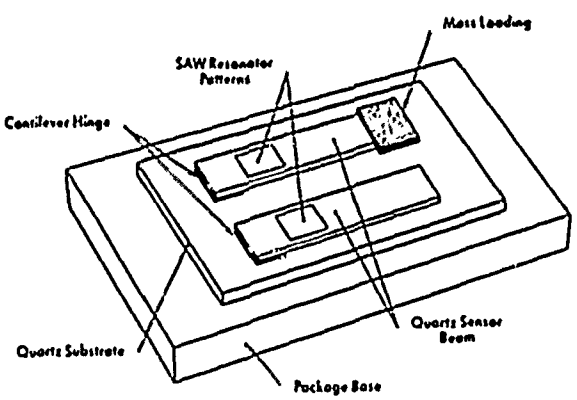


Fig. 2 - Accelerometer - Separate Substrates

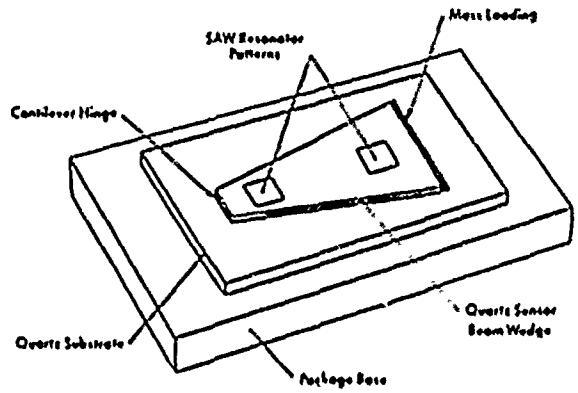


Fig. 3 - Accelerometer - Wedged Beam Substrate

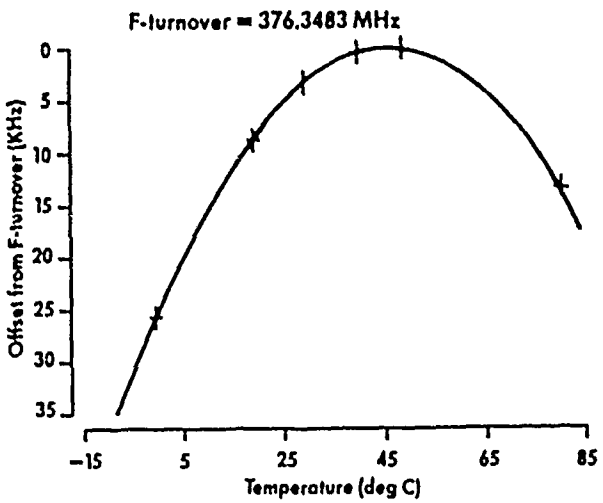


Fig. 4 - SAW Resonator - Typical Frequency vs Temperature Characteristic

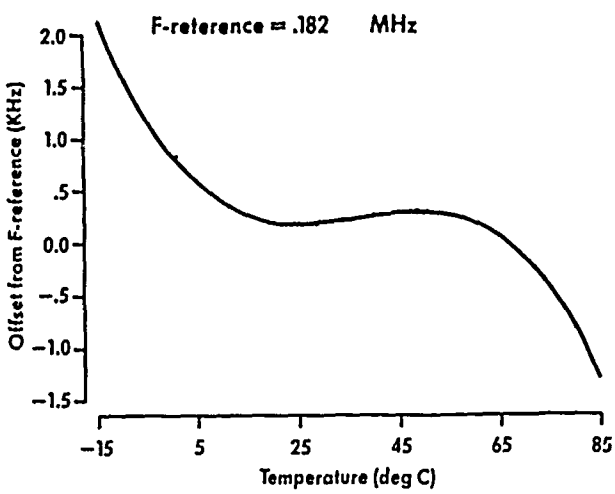


Fig. 5 - Computed Difference Frequency vs Temperature Characteristic

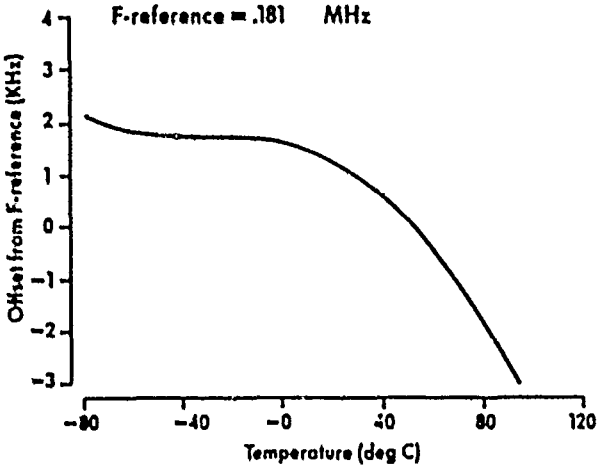


Fig. 6 - Actual Difference Frequency vs Temperature Characteristic

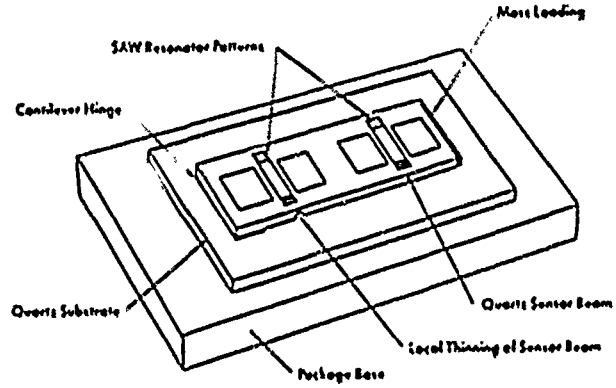


Fig. 9 - Accelerometer - Selectively Thinned Beam

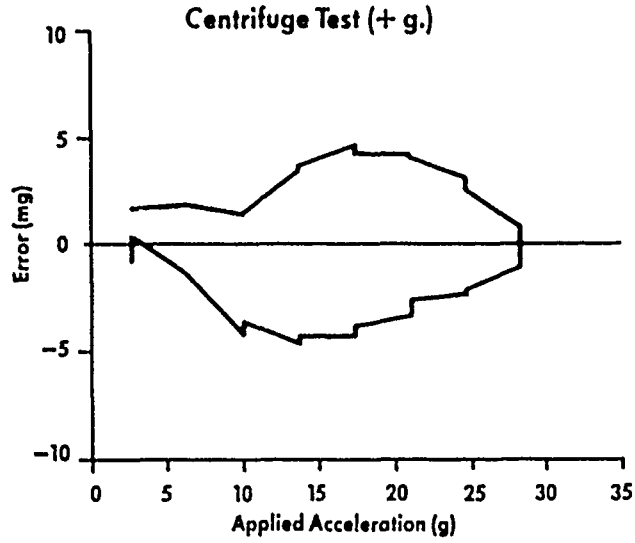


Fig. 7 - Centrifuge +ve g

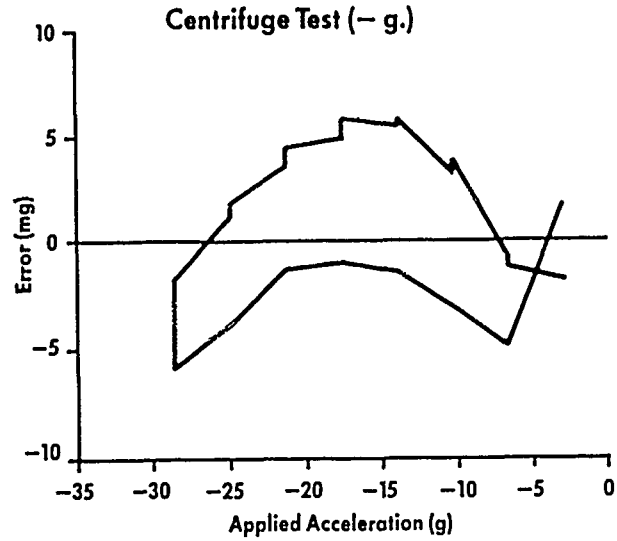


Fig. 8 - Centrifuge -ve g

41st Annual Frequency Control Symposium - 1987

USE OF ANNULAR SAW FOR CUTTING QUARTZ RESONATOR BLANKS AND COMPARISON WITH OTHER METHODS

J Dowsett, R N Spencer and Dr A F N Wood

STC Components
Edinburgh Way
Harlow
Essex CM20 2DE
England

Introduction

For many years the Quartz Crystal Industry has used as a standard method the multiblade slurry saw for the cutting of Quartz resonator plates with some use of the peripherally loaded diamond blade saw for larger or thicker slices. Both these methods have severe limitations on either angle accuracy or economy of cutting and in sub-surface damage, parallelism and thickness control.

The demands of the crystal user for continuing improvements in the temperature performance of, for example, the AT resonator has lead to a significant tightening of angle control both in ZZ and XX angles. As a result of this together with the market forces driving down resonator prices these earlier methods have become very uneconomical. This paper describes the introduction of the large annular saw with a thin stressed blade for the cutting of tight angle tolerance blanks with good parallelism and thickness uniformity.

Brief Description of Earlier Saws

The peripheral diamond saw, see Figure 1., was one of the earliest developed for cutting hard materials such as quartz and remains a very good method with respect to angle control it however, suffers from the need to maintain a very thick self supporting blade to hold this performance advantage, consequently it has poor economy in quartz utilisation and produces sub-surface damage which requires heavy primary lapping to remove.

The multiblade slurry saw, see Figure 2., introduced in the 1960s produced a revolution with its capability to produce large numbers of slices with good economy from cultured quartz blocks grown with a convenient orientation. The angle control produced at this time was adequate for the current specifications and the poor control of parallelism and thickness was out weighed by the economy of blank production.

However, when tighter angle control both in ZZ and XX became a requirement this saw could only produce a very small percentage of blanks within these new limits, results achieved on this type of saw are given later in the paper.

The Annular Saw

Recent developments in the Semiconductor industry have lead to the development of larger annular saws using a thin stressed blade with an inner edge diamond loaded, capable of cutting large diameter silicon bars. We have modified this machine to handle the cutting of large dimension quartz bars typically 43 x 24 mm in cross section.

Figures 3-6 shows views of the machine blade construction when the quartz block is mounted in the machine it can be angle corrected in the ZZ and XX directions. This can be carried out by obtaining corrections from the first cut slice which is sufficiently flat and parallel and of good finish to obtain correction to 0.25" by X ray orientation.

The machine is capable of indexing each feed step to maintain good thickness control. A blade distortion sensor with a continuous analogue plot gives warning of blade wear and the need to redress the diamond cutting edge.

The quartz block is lumbered in the normal way, a reference face prepared on the -X, this face is prepared with an angle tolerance of better than 5' to minimise machine resetting between block loads.

The blank is cemented to a glass plate to ensure a clear cut through the quartz block, see Figure 7.

The diamond loaded blade is shown in Figure 8., this blade has 9" internal diameter and a thickness of 175 uM, the diamond cutting edge produces a kerf of 300 uM and blank thickness of less than 200 uM have been achieved.

Results

A comparison of results achieved from this annular saw with those from a multiblade slurry saw are given in Figures 9-12. These figures show the angle distribution from the two saws after the sawing process and after the primary lapping process. It can be seen that the angle distribution on this annular saw after lapping is still held within tight limits while that of the slurry saw has more than doubled and is in the order of two to three times that achieved on the annular saw.

These results were obtained from two batches taken at random from production and represent a very typical achievement.

Figure 13 shows the "rocking curves" achieved on annular sawn slices compared to a carefully prepared and etched check piece. The width of this "rocking curve" is a good indication of sub-surface damage and it can be seen that the sawn blank is not substantially different to that of the lapped slice. The lapped slices had been reduced in thickness from 500 uM to 350 uM, more than sufficient to remove the expected sub-surface damage.

CH2427-3/87/0000-548$1.00 C 1987 IEEE

A summary of the results achieved in comparison to the multiblade saw is given in Tables 1 and 2. It can be seen that the blank angle control has been considerably improved with large increases in yield with respect to target angles and angle spreads. The control of parallelism and thickness shows improvements in the order of 6:1. These results are based on finished blanks through all stages of lapping.

The sawing economy in terms of sawing rate, blade life and quartz utilisation also showed significant improvements. The only disadvantage of the annular saw is the longer time to change blades but this is easily balanced by considerably longer blade life.

Conclusion

The advantages of the annular saw are:

(1) Ease of setting to target angle.

(2) Improved yield to ± 0.5 angle tolerance.

(3) Improved thickness and parallelism control.

(4) Significant improvement in operating economy.

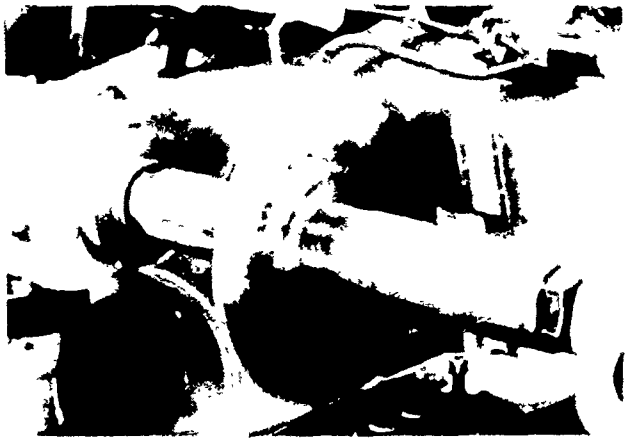


Fig. 1 - Peripheral Diamond Saw

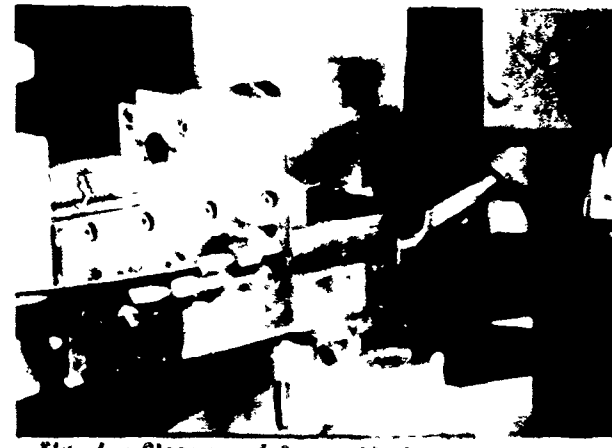


Fig. 4 - Close up of Quartz Blank during a cut

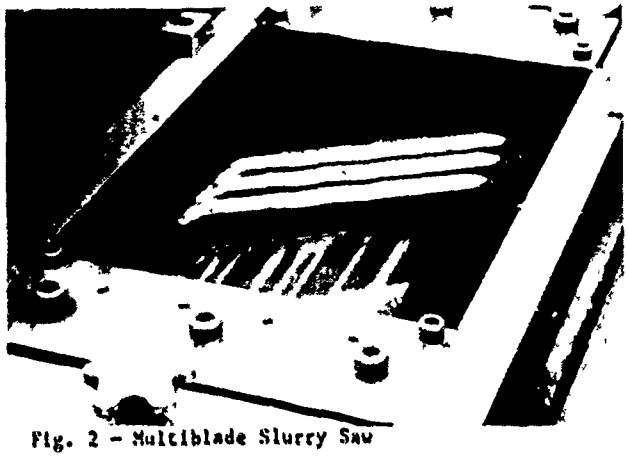


Fig. 2 - Multiblade Slurry Saw

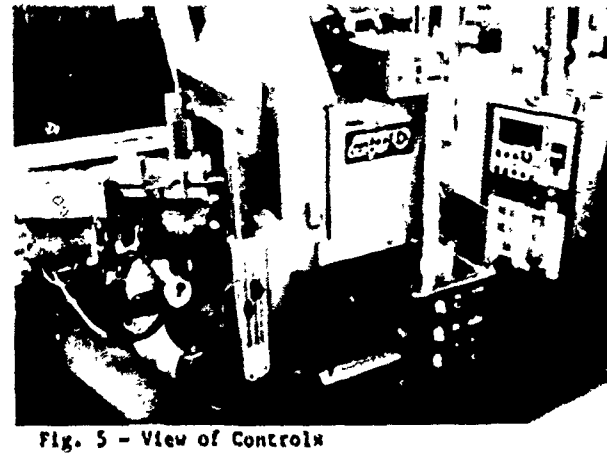


Fig. 5 - View of Controls

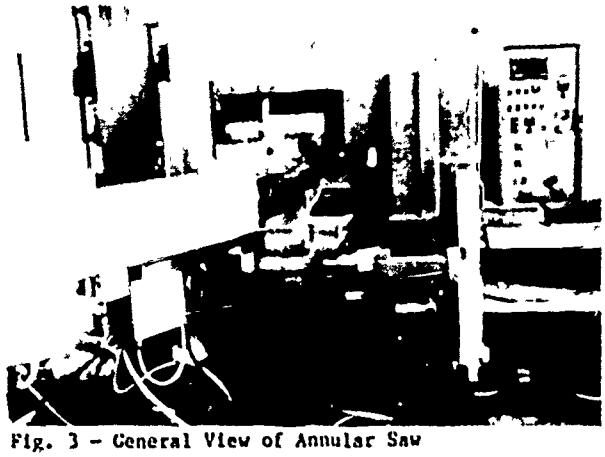


Fig. 3 - General View of Annular Saw



Fig. 6 - Slice Pick-up Mechanism



Fig. 7 - Mounting of Lumbered Quartz block



Fig. 8 - Saw Blade

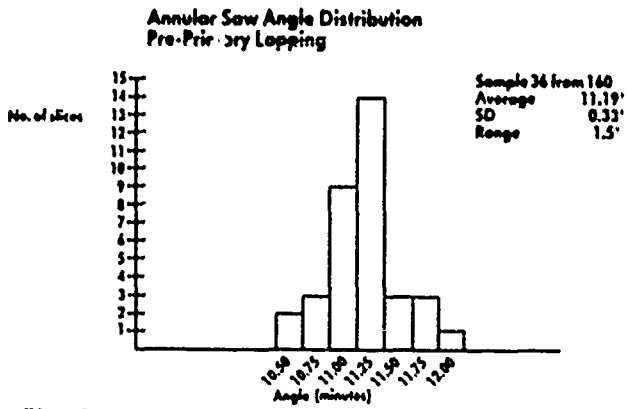


Fig. 9

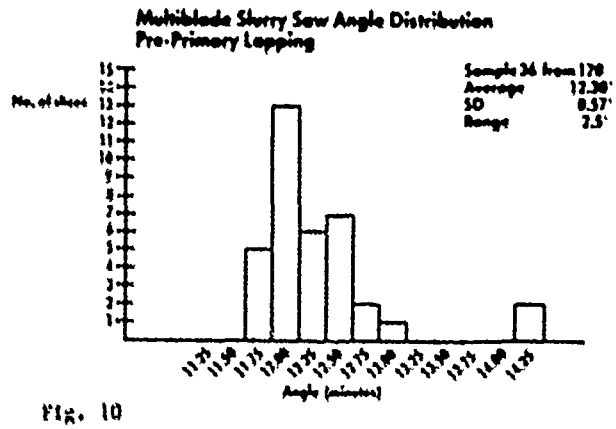


Fig. 10

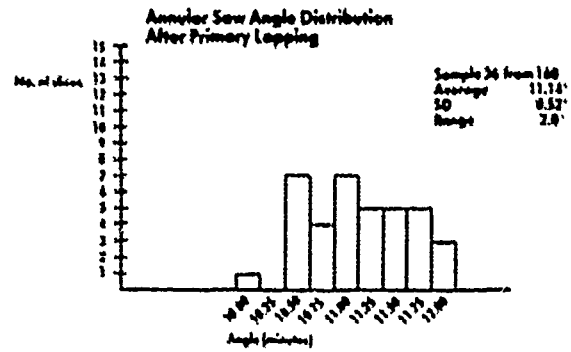


Fig. 11

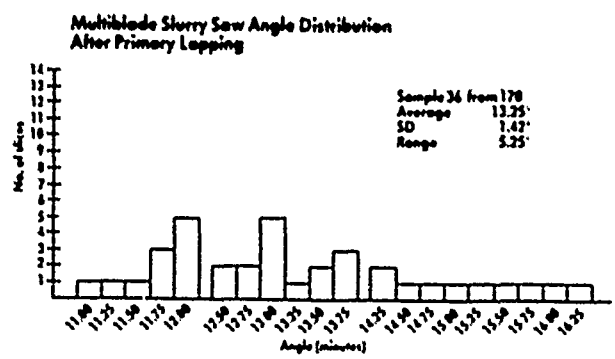


Fig. 12

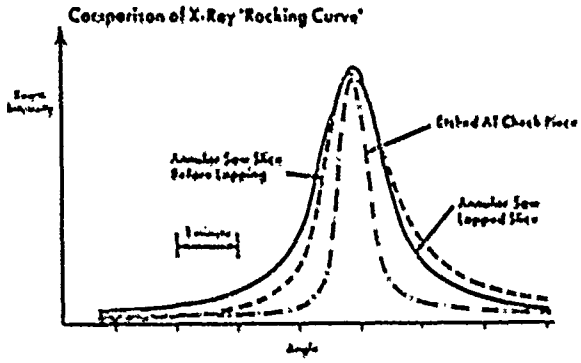


Fig. 13

Summary of Results (Blanks)

	Annular Saw	Multiblade Saw
Angle Control ZZ angle:		
Target Angle ±0.25'	95%	75%
Angle Spread	2'	3-4'
Blank Yield ±0.5'	30%	10%
Angle Control XX Angle: ±10'		
Yield	95%	10-15%
Blank Parallelism	5-10 μm	60 μm
Blank Thickness Control	15 μm	100 μm

Table 1

Summary of Results (Sawing Economy)

	Annular Saw	Multiblade Saw
– Sawing rate	3.5 cm^3/min	3.2 cm^3/min
– Blade life	18,920 cm^3	8,512 cm^3
– Quartz utilisation/block	175 slices	150 slices
– Blade change time	1.5 hrs	0.75 hrs

Table 2

AUTHOR INDEX

AUTHOR	PAGE
Adams, C.	249
Allan, D.	2,144
Armington, A.	213
Artal, E.	524
Bain, C.	95
Balascio, J.	213
Ballato, A.	325
Bara, J.	492,524
Benjaminson, A.	452
Beno, L.	75
Berenguer, J.	492,524
Bernacki, B.	42
Bernier, L.	116
Bhaskar, N.	47
Bloch, M.	519
Bourquin, R.	289
Bower, D.	544
Bradley, D.	249
Braymen, S.	371
Burkland, W.	371
Burns, S.	371,382
Camparo, J.	36
Capelle, B.	236
Caret, G.	413
Chaffee, J.	112
Chen, C.	216
Comeron, A.	492,524
Cook, R.	47
Corbella, I.	524
Cracknell, M.	544
Crampton, S.	91
Curtis, G.	420
Cutler, L.	12
Debaisieux, A.	413
De Marchi, A.	53
Detaint, J.	236,314
Dick, G.	20,87,487
Dowsett, J.	539,548
Driscoll, M.	429
Dulmet, B.	289
Dworsky, L.	295
EerNisse, E.	339,344
Ehrlich, D.	351
Euler, F.	216
Ezekiel, S.	42
Feinberg, A.	360
Felton, C.	512
Filler, R.	398,444
Frerking, M.	503
Frueholz, R.	36,47
Fujishima, S.	391
Furuta, H.	25
Gagnepain, J.	266
Gerard, E.	413
Giffard, R.	12
Gottlieb, M.	388
Greenhall, C.	126
Greer, J.	351
Gualtieri, J.	192
Gupta, M.	471
Halliburton, L.	216,223
Hanson, W.	228,429
Hara, T.	435
Harris, M.	213
Harrison, A.	539,544
Harrison, R.	122
Hashimoto, M.	25
Hemmer, P.	42
Horrigan, J.	213
Hou, J.	365
Howe, D.	149
Huang, H.	82
Hunt, J.	183
Irvine, R.	175
Jiang, G.	82
Johnson, G.	175
Joly, C.	314
Jones, K.	199
Kahan, A.	216
Kahla, C.	47
Kalivas, G.	122
Kanazawa, K.	350
Katsuta, Y.	435
Keilson, S.	223
Ketcham, R.	371
Khanna, A.	478
Klepczynski, W.	130
Kline, G.	371,382
Koelman, J.	91
Kosinski, J.	325
Kudo, T.	435
Kusters, J.	249
Lakin, K.	371,382
Lamela-Rivera, H.	42
Landin, A.	371
Lee, P.	277
Lin, C.	82
Lindenmuth, R.	444
Ling, S.	223
Lipson, H.	216
Lu, J.	82
Luiten, O.	91
Lukac, C.	130
Lukaszek, T.	325
Luo, W.	82
Lutes, G.	161
Malecki, L.	20
Martin, J.	167
Mattison, E.	95
Mc Avoy, B.	389
Mc Carron, K.	371
Mc Clelland, T.	59,66
Meirs, M.	59,65
Mercer, C.	258
Messina, J.	444
Mills, T.	460
Mizan, M.	325
Natoli, V.	42
Nicholas III, H.	495
Nowick, A.	223
Oakley, T.	75
Ogou, S.	435
Ohta, S.	391
Ohtsu, H.	25
Owings, B.	75
Parker, T.	99,351
Pascaru, I.	59,66
Peters, H.	75
Philippot, E.	314
Phillips, D.	507
Ping-ping, L.	144
Prestage, J.	20
Reed, C.	350
Robichon, G.	413
Rosati, V.	409,444,466
Rothschild, M.	351
Saita, H.	435
Salvo, Jr., H.	388
Samueli, H.	495
Sawyer, B.	241
Schodowski, S.	466
Schwartzel, J.	236,314
Shahriar, M.	42
Sharpe, G.	539
Sherman, Jr., J.	243
Shick, D.	282
Smith, S.	42
Smith, W.	527
Smythe, R.	311
Spencer, R.	548
Stoof, H.	91
Strayer, D.	487
Tang, M.	277
Tiersten, H.	282,311
Togawa, K.	391
Tran, H.	59
Tucker, T.	87
Uriya, S.	435
Verhaar, B.	91
Vessot, R.	95
Vig, J.	444
Vulcan, A.	519
Walls, F.	439,512
Ward, R.	344
Wasserman, S.	95
Wheeler, P.	12
White, R.	333
Whitesides, G.	95
Will, C.	111
Winkler, G.	12
Wood, A.	548
Yong, Y.	303
Zacharski, J.	59
Zarka, A.	236
Zhai, Z.	82
Zhuang, Q.	130

INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS

Order through: IEEE Service Center
445 Hoes Lane
Piscataway, NJ 08854
(201) 981-0060

176-1978 Piezoelectricity

177-1966 Piezoelectric Vibrators, Definitions and Methods of Measurements for (ANSI C83.17-1970)

180-1962 Ferroelectric Crystal Terms, Definitions of

319-1971 Piezomagnetic Nomenclature

ELECTRONIC INDUSTRIES ASSOCIATION

Order through: Electronic Industries Association
2001 Eye Street, NW
Washington, DC 20006
(202) 457-4900

(a) Holders and Sockets

RS-192-A, Holder Outlines and Pin Connections for Quartz Crystal Units (Standard Dimensions for Holder types).

RS-367, Dimensional and Electrical Characteristics Defining Receiver Type Sockets (Including crystal sockets).

RS-417, Crystal Outlines (Standard dimensions and pin connections for current quartz crystal units-1974).

(b) Production Tests

RS-186-E, (All Sections), Standard Test Methods for Electronic Component Parts

RS-512, Standard Methods for Measurement of Equivalent Electrical Parameters of Quartz Crystal Units, 1 kHz to 1 GHz, 1985.

(c) Application Information

Components Bulletin No. 6, Guide for the Use of Quartz Crystals for Frequency Control

(d) EIA-477 Cultured Quartz (Apr. 81)

EIA-477-1, Quartz Crystal Test Methods (May 1985)

INTERNATIONAL ELECTROTECHNICAL COMMISSION (IEC)*

Order through: American National Standards Inst. (ANSI)
1430 Broadway
New York, NY 10018

*ANSI can quote prices on specific IEC publications on a day to day basis only. All IEC and ISO standards have been removed from its Standards Catalog. Call ANSI, (212) 354-3300 for prices.

DEPARTMENT OF DEFENSE

Order through: Naval Publication & Form Center
5801 Tabor Avenue
Philadelphia, PA 19120

MIL-C-3098 Crystal Unit, Quartz, Gen Spec for
MIL-H-10056 Holders (End), Crystal, Gen Spec for
MIL-STD-683 Crystal Units, Quartz/Holders, Crystal
MIL-O-55310 Oscillators, Crystal, Gen Spec for
MIL-F-18327 Filters, High Pass, Low Pass, Band Pass Suppression and Dual Functioning, Gen Spec for
MIL-O-39021 Oven, Crystal, Gen Spec for
MIL-O-55240 Oscillators, Audio Frequency
MIL-F-28734 Frequency Standards, Cesium Beam, Gen Spec
MIL-F-288111 Frequency Standard, Cesium Beam Tube
MIL-C-24523 (SHIPS), Chronometer Quartz Crystal

IEC PUBLICATIONS

122: Quartz crystal [illegible] frequency control and selection.

122-1 (1976) Part [illegible]s and test conditions.

122-2 (1983) Part 2: [illegible]e of quartz crystal units for freque[illegible]d selection

122-3 (1977) Part 3: St[illegible]dines and pin connectors.

122-3A (1979) First supplement.

122-3B (1980) Second supplement

122-3C (1981) Third supplement.

283 (1968) Methods for the measurement of frequency and equivalent resistance of unwanted resonances of filter crystal units.

302 (1969) Standard definitions and methods of measurement for piezoelectric vibrators operating over the frequency range up to 30 MHz.

314 (1970) Temperature control devices for quartz crystal units. Amendment No. 1 (1979)

314A (1971) First supplement.

368: Piezoelectric filters.

368A (1973) First supplement
Amendment No. 1 (1977)
Amendment No. 2 (1982)

368B (1975) Second supplement.

368-1 (1982) Part 1: General information, standard values and test conditions.

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 (1973) Basic method for the measurement of resonance frequency and equivalent series resistance of quartz crystal units by zero phase technique in a π network.

Note: This publication, where revised, will be issued as a second edition of Publication 444-1

444-2 (1980) Part 2: Phase offset method for measurement of motional capacitance of quartz crystal units.

48[illegible] (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 and methods.

679-2 (1981) Part 2: Guide to the use of quartz crystal controlled oscillators.

689 (1980) Measurements and test methods for 32 kHz quartz crystal units for wrist watch[illegible] 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.

PROCEEDINGS

ANNUAL FREQUENCY CONTROL SYMPOSIA

NO.	YEAR	DOCUMENT NUMBER	OBTAIN FROM*	COST
10	1956	AD298322	NTIS	$41.50
11	1957	AD298323	*	44.50
12	1958	AD298324	*	46.00
13	1959	AD298325	*	49.00
14	1960	AD246500	*	32.50
15	1961	AD265455	*	28.00
16	1962	PB162343	*	35.50
17	1963	AD423381	*	43.00
18	1964	AD450341	*	43.00
19	1965	AD471229	*	47.50
20	1966	AD800523	*	47.50
21	1967	AD659792	*	41.50
22	1968	AD844911	*	44.50
23	1969	AD746209	*	25.00
24	1970	AD764210	*	28.00
25	1971	AD746211	*	28.00
26	1972	AD771043	*	26.50
27	1973	AD771042	*	34.00
28	1974	ADA011113	*	31.00
29	1975	ADA017466	*	34.00
30	1976	ADA046089	*	40.00
31	1977	ADA088221	*	44.50
32	1978		EIA	20.00
33	1979		*	20.00
34	1980		*	20.00
35	1981		*	20.00
36	1982	ADA130811	NTIS	41.50
37	1983	83CH1957-0	IEEE	59.00
38	1984	84CH2062-8	*	59.00
39	1985	85CH2186-5	*	59.00
40	1986	86CH2330-9	*	59.00
41	1987	87CH2427-3	*	59.00

* NTIS - National Technical Information Service
Sills Building
5285 Port Royal Road
Springfield, VA 22161

* EIA - Annual Frequency Control Symposium
c/o Electronic Industries Association
2001 Eye Street
Washington, DC 20006

*IEEE - Institute of Electrical & Electronics Engineers
445 Hoes Lane
Piscataway, NJ 08854

Remittance must be enclosed with all orders. Prices are subject to change without prior notice.

A subject and author index for the Proceedings of the 10th through the 38th Symposia appears as a supplement to the 38th Proceedings volume.